PROSPR II METRICS Data Dictionary [Version 3]

# Overall Instructions

The **M**ultil**E**vel *Op****T****imization of the* Ce**R**v**I**cal **C**ancer **S**creening Process *in* Diverse Settings & Populations (METRICS) PROSPR II *Research* *Center* (PRC) was designed to elucidate multilevel factors that hamper or facilitate the cervical cancer screening process and reduce disparities. **METRICS has three data-contributing sites from diverse healthcare settings –** a mixed model healthcare system,a public safety-net system, and a primary care provider network**.** Sites collected **11 years** of high-quality, comprehensive cervical cancer screening process **data at the patient, provider, clinic/facility, and system levels** using a rich array of electronic clinical information systems and other data sources. The estimated size of the combined cohort is ~1.1 million screened and unscreened females.

This data dictionary provides information and guidance for using the PROSPR II METRICS dataset (calendar years 2010-2020) for analyses. The METRICS PRC harmonized and collected data from the contributing sites to create a limited consolidated data subset (LCDSS) for dissemination to the National Cancer Institute contractor managing dissemination to the PROSPR Coordinating Center (PCC) and approved external collaborators.

The METRICS PRC sent a data request packet (DRP) with instructions to the contributing sites to guide data extraction for each common data element (CDE). Each site noted when data extraction deviated from given instructions. This data dictionary provides the information presented in the DRP to guide data extraction as well as site-specific deviations in the extraction process for each CDE, the rationale for the deviation, and other site-specific nuances or contextual information.

This data dictionary is divided into the following three sections:

1) Section 1: Overview. A summary of the METRICS data contributing sites, cohort definitions, data file list, and data flow schematic;

2) Section 2: File Description. A description of each data file submitted by the contributing sites and the analytic Cervical Timeline and Screening Episode files, with site-specific notes and exceptions; and

3) Section 3: Appendices. A list of abbreviations used throughout the document, decision making guidance for select submitted data elements, and an orientation to the development of the derived data elements in the analytic Cervical Timelineand Screening Episode files.

# Section 1: Overview – Description of the METRICS Data Contributing Sites & Cohort Definitions

## METRICS Data Contributing Sites and Cohort Definitions

| **Definition** | **Site A (Enrollment-Based)** | **Site B (Utilization-Based)** | **Site C (Utilization-Based)** |
| --- | --- | --- | --- |
| Type of Healthcare system | Mixed-model healthcare system providing coverage to members for health care received at system owned and operated medical offices and other outpatient facilities, as well as for health care received from contracted providers, medical groups and hospitals to provide services to members. | Integrated county tax-supported public safety-net system comprising primary care clinics, women’s health clinics, HIV clinics, and other specialty clinics in areas of high need throughout the community. | Not-for-profit integrated delivery system composed of two networks that include primary care practices, community health centers, hospital-based clinics, and community-based clinics. |
| Data Sources | Epic electronic health record system (EHR) 2010-2019  Internal data warehouse 2010-2019  Central cancer registry 2010-2019  State immunization registry 2010-2019 | Epic electronic health record system (EHR) 2010-2019  State and hospital cancer registry 2010-2019  Internal data warehouse | Homegrown electronic medical record (EMR) 2010-2015  Epic electronic health record system (EHR) 2015-2019  State and hospital cancer registry 2010-2019  State vaccination registry 2010-2019  State immunization registry 2010-2017  Internal data warehouse 2015-2019 |
| Cohort Entry | Enrolled healthcare system members entered the study cohort from January 1, 2010 through December 30, 2020 when all of the following criteria were met: female sex; age 18-89 years; elected, assigned, or attributed to a healthcare system primary care provider; resident of cancer registry catchment area. | Female county residents aged 18-89 years at first visit to a qualifying primary care clinic from January 1, 2010 through December 30, 2020. | Females aged 18-89 years at first visit to a qualifying primary care clinic from January 1, 2010 through December 31, 2020. |
| Cohort Exit | Cohort members exited the cohort at the earliest date as of any of the following: age 90 years; death; healthcare system disenrollment (gaps of no more than 90 days were allowed); no longer have an elected, assigned or attributed healthcare system primary care provider (gaps of no more than 90 days were allowed); residential relocation outside the cancer registry catchment area (gaps of no more than 90 days were allowed); or end of the cohort period (i.e., December 31, 2020). | Cohort members exited the cohort at the earliest date as of any of the following: date as of age 90; death; move out of county for longer than 6 months; or administrative cut-off due to the end of the cohort period end (i.e., December 31, 2020). Death date was estimated as day after a primary care visit if vital status is known to be deceased but exact death date is unknown. | Cohort members exited the cohort at the earliest date as of any of the following: date as of age 90; death; or administrative cut-off due to the end of the cohort period end (December 31, 2020). |
| Primary Care Visit | Completed in-person (2010-2020) or telehealth (2019-2020) visit with a physician (MD/DO), nurse practitioner, physician’s assistant, or unknown provider in a primary care setting, including family practice, internal medicine, pediatrics, obstetrics and gynecology, gerontology, and adolescent medicine. | Completed in-person (2010-2020) or telehealth (2019-2020) office visit/appointment with a physician, nurse practitioner, or physician’s assistant at a community health, family practice, internal medicine, women’s health, geriatrics, gynecology, or HIV clinic. | Completed in-person (2010-2020) or telehealth (2019-2020) primary care or women’s health office visit |

# Section 2: File Description – Overview of All Data Files and CDEs with Site-Specific Notes and Exceptions

## Global Values

### Global Values

95 indicates other value available in the record, but value does not conform to permissible value structure.

-99999 indicates missing or unknown continuous numeric or small cell data (i.e., if the data for a variable is so small that it could potentially be used to identify the PROSPR cohort member).

99 indicates missing or unknown categorical data.

Blank indicates missing or unknown text-based, date, or days since reference data.

## Participant File

### **Overview**

This file contains one record of static covariates for every cohort member.

### **Record Structure**

One record per cohort member.

### General Harmonization Notes

**Identification of cohort member study periods for primary and sensitivity analyses. To make cohort periods more conceptually analogous between the enrollment and utilization cohorts, cohort members from the utilization cohorts (Site B and Site C; see** [Section 1: Overview - Description of METRICS Data Contributing Sites & Cohort Definitions](#CohortEntry) (p. 2) **for further detail) that had a lapse in primary care utilization were administratively cut-off from the cohort 37 months after the last primary care encounter. Cohort members from utilization cohorts,then re-entered the study upon subsequent primary care utilization. See** [Section 3: Appendices - Cohort Member Study Periods for Primary and Secondary Analyses](#HPVTestDate) (p. 66) **for further detail. The provided date values permit sensitivity analyses beyond the primary analyses, which incorporate the lack of primary care utilization cut-off, and present cohort periods structured similarly to other utilization cohorts throughout the PROSPR II consortium. Of note, enrollment cohort members have the same dates listed for these variables (e.g., CutOffLastDSR is equivalent to CutOffSensDSR).** **Importantly, while the Participant file supports both primary and sensitivity analysis goals, the file only reports the last cohort exit (either using the primary or sensitivity analysis cohort exit criteria**)**, such that no events occur after this exit; thus, if a cohort member left and re-entered the cohort multiple times, cohort exits that occurred prior to the last cohort exit must be identified in the Engagement file.**

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSIte | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
|
|
|
|
| BirthYr | **Numeric** | Birth date, year  A 4 digit integer |  |
|
|
| Hispanic | **Numeric**  0 = No  1 = Yes | Hispanic or Latino origin |  |
|
|
| RaceWhite | **Numeric**  0 = No  1 = Yes | Race White |  |
|
|
| RaceBlack | **Numeric**  0 = No  1 = Yes | Race Black or African-American |  |
|
|
| RaceAsian | **Numeric**  0 = No  1 = Yes | Race Asian |  |
|
|
| RaceAIAN | **Numeric**  0 = No  1 = Yes | Race American Indian or Alaska Native |  |
|
|
| RacePI | **Numeric**  0 = No  1 = Yes | Race Native Hawaiian or Other Pacific Islander |  |
|
|
| RaceMultipleNOS | **Numeric**  0 = No  1 = Yes | Race, Multiple (individual races not specified) |  |
|
|
| RaceOther | **Numeric**  0 = No  1 = Yes | Race Other |  |
|
|
|
|
|
|
|
|
|
|
| CohortEntryFirstDSR | **Numeric** | Earliest cohort entry date, days since reference date (birth date) |  |
|
|
|
|
|
|
| CutOffLastDSR | **Numeric** | Last cohort cut-off date, days since reference date (birth date) |  |
|
|
| CutOffSensDSR | **Numeric** | Last cohort cut-off date, days since reference date (birth date) for sensitivity analyses (does not include lack of primary care utilization cut-off) |  |
|
|
| RaceEth\_Drv | **Numeric**  1 = White  2 = Black  3 = Asian  4 = Native American/Alaskan Native  5 = Native Hawaiian/Pacific Islander  6 = Other  7 = Multiple Races  8 = Hispanic | Race/Ethnicity  Race/Ethnicity category was prioritized in descending order as follows: Hispanic; multiple race designations (not including Hispanic) and/or designation of multiple race NOS; any single race designation; and if no designation made, unknown. |  |
|
|
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|
|
|

## Calendar Year File

### Overview

This file contains a record of time-varying covariates for every year a cohort member remained in the cohort.

### Record Structure

One record per cohort member per calendar year in the cohort. When multiple visits occurred during the calendar year, data derive from either the last encounter within the calendar year (weight/BMI, smoke) or tallied across all visits (comorbidities, insurance types).

### General Harmonization Notes

**Identification of cohort members based on healthcare utilization. To permit maximum analysis flexibility, all calendar year records were included in the Calendar Year file, even if a cohort member did not utilize primary care during the calendar year. The Calendar Year file variable (**EncCalYr\_Drv**) flags whether a cohort member had a primary care encounter during the respective calendar year, permitting exclusion of calendar year data for cohort members that did not have a qualifying primary care encounter. See** [Section 3: Appendices - Cohort Member Study Periods for Primary and Secondary Analyses](#HPVTestDate) (p. 66) for more information.

**Identification of cohort member study periods for primary and sensitivity analyses. To make cohort periods more conceptually analogous between the enrollment and utilization cohorts, cohort members from the utilization cohorts (see** [Section 1: Overview - Description of METRICS Data Contributing Sites & Cohort Definitions](#CohortEntry) (p. 2) **for further detail) that had a lapse in primary care utilization were administratively cut-off from the cohort 37 months after the last primary care encounter. Cohort members at ulitization-based sites then re-entered the study upon subsequent primary care utilization. See** [Section 3: Appendices - Cohort Member Study Periods for Primary and Secondary Analyses](#HPVTestDate) (p. 66) **for further detail. The Calendar Year file variable (**UtilCalYr\_Drv**) flags whether a cohort member was considered to be in the cohort for any part of the calendar year based on cohort exit due to lack of utilization (see** [Section 2: Engagement file - General Harmonization Notes](#_General_Harmonization_Notes_1) (p. 21) **for further detail), permitting exclusion of any calendar year records that occurred following cohort exit due to lack of utilization. Thus, the Calendar Year file supports both primary and sensitivity analyses.**

**Calculation of Composite Comorbidity Score: There are several different published methods used to summarize comorbidity data. To facilitate analysis flexibility, three different comorbidity scores based on the Charlson Index were included in the Calendar Year file as follows:**

1. **Comorbidity Index (ComorbIndex)** **using the HCSRN macro, which reports comorbidities recorded at in-person (2010-2020) and telehealth (2019-2020) encounters in the respective calendar year (only available for Site A);**
2. **In-Person Comorbidity Index (ComorbIndexIP)** **using the HCSRN macro, which reports comorbidities recorded at in-person encounters in the respective calendar year (only available for Site A);**
3. **De**rived prevalent Comorbidity Index (ComorbIndexPrev\_Drv), which reports comorbidities recorded at in-person (2010-2020) and telehealth (2019-2020) encounters in the respective calendar year or preceding calendar years (**only available for** Site B and Site C); and

**All scores use the same weights for each comorbidity and differ by the time interval from which comorbidities were collected and the data source used to identify each comorbidity (e.g., billing codes, problem lists, etc).**

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
| CalendarYr | **Numeric** | Calendar year |  |
|
|
|
|
| WeightDateMonth | **Numeric** | Month of weight recorded date  Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
|
|
| WeightDateYear | **Numeric** | Year of weight recorded date  Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
| WeightDSR | **Numeric** | Date weight recorded, days since reference date (birth date)  Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
|
|
|
|
| Medicare | **Numeric**  0 = No  1 = Yes | Was the cohort member covered by Medicare? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| Medicaid | **Numeric**  0 = No  1 = Yes | Was the cohort member covered by Medicaid? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| InsOtherGov | **Numeric**  0 = No  1 = Yes | Was the cohort member covered by any other federal or state health insurance program? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| InsCommerc | **Numeric**  0 = No  1 = Yes | Was the cohort member covered by commercial and/or private health insurance? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| MedicalAssist | **Numeric**  0 = No  1 = Yes | Was the cohort member enrolled in a medical assistance charity program for the uninsured? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| Uninsured | **Numeric**  0 = No  1 = Yes | Was the cohort member uninsured? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| InsOther | **Character** | Health insurance coverage, other | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
| InsHighDeductible | **Numeric**  0 = No  1 = Yes | Was the cohort member covered by high deductible insurance (as defined by the U.S. IRS (Pub 969)? | Recommend using [CalYr\_Ins\_Drv](#CalYrIns) (p. 19) variable to identify insurance during the calendar year. |
|
|
|
|
| ComorbMyocardial | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for myocardial infarction during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 410.x, 412.x  ICD-10 codes: I21.x-I22,x, I25.2  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbMyocardialIP | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for myocardial infarction during the specified year at an in-person encounter?  This variable is only available for Site A.  Identified based on codes used in HCSRN macro.  Weighted as +1 in overall comorbidity score |  |
| ComorbCongHeart | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for congestive heart failure during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4–425.9, 428.x  ICD-10 codes: I09.9, I11.0, I13.0, I13.2, I25.5, I42.0 ,I42.5-I42.9, I43.x, I50.x, P29.0  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
|
| ComorbCongHeartIP | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for congestive heart failure during the specified year at an in-person encounter?  This variable is only available for Site A.  Identified based on codes used in HCSRN macro.  Weighted as +1 in overall comorbidity score |  |
| ComorbVasPeriph | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for peripheral vascular disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 093.0, 437.3, 440.x, 441.x, 443.1–443.9, 447.1, 557.1, 557.9, V43.4  ICD-10 codes: I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbVasCerebro | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for cerebrovascular disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 362.34, 430.x–438.x  ICD-10 codes: G45.x, G46.x, H34.0, I60.x-I69.x  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbDementia | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for dementia during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 290.x, 294.1, 331.2  ICD-10 codes: F00.x-F03.x, G30.x, F05.1, G31.1  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbPulmChronic | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for chronic pulmonary disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 416.8, 416.9, 490.x–505.x, 506.4, 508.1, 508.8  ICD-10 codes: I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbRheumatic | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for rheumatic disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 446.5, 710.0–710.4, 714.0– 714.2, 714.8, 725.x  ICD-10 codes: M05.x, M06.x, M31.5, M32.x-M34.x, M35.1, M35.3, M36.0  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbPepticUlcer | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis for peptic ulcer disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 531.x–534.x ICD-10 codes: K25.x - K28.x  Weighted as +1 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbLiverMild | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis for mild liver disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 570.x, 571.x, 573.3, 573.4, 573.8, 573.9, V42.7  ICD-10 codes: B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4  Weighted as +1 in overall comorbidity score if ComorbLiver = 0; otherwise weighted as 0 | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbDiabetes | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis for diabetes without chronic complication during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 250.0–250.3, 250.8, 250.9 ICD-10 codes: E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9  Weighted as +1 in overall comorbidity score if ComorbDiabetesComp = 0; otherwise weighted as 0 | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbDiabetesIP | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis for diabetes without chronic complication during the specified year at an in-person encounter?  This variable is only available for Site A.  Identified based on codes used in HCSRN macro.  Weighted as +1 in overall comorbidity score if ComorbDiabetesComp = 0; otherwise weighted as 0 |  |
| ComorbDiabetesComp | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for diabetes with chronic complication during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 250.4–250.7 ICD-10 codes: E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7  Weighted as +2 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbDiabetesCompIP | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for diabetes with chronic complication during the specified year at an in-person encounter?  This variable is only available for Site A.  Identified based on codes used in HCSRN macro.  Weighted as +2 in overall comorbidity score |  |
| ComorbHemiplegia | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for hemiplegia or paraplegia during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 334.1, 342.x, 343.x, 344.0– 344.6, 344.9  ICD-10 codes: G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9  Weighted as +2 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbRenal | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for renal disease during the specified year or a prior calendar year within cohort window?  Enhanced ICD-9-CM codes: 403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 582.x, 583.0–583.7, 585.x, 586.x, 588.0, V42.0, V45.1, V56.x   ICD-10 codes: I12.0, I13.1, N03.2-N03.7, N05.2-N05.7, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2  Weighted as +2 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbMalignancy | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for malignancy, including lymphoma and leukemia, except malignant neoplasm of skin during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 140.x–172.x, 174.x–195.x, 200.x–208.x, 238.6  ICD-10 codes: C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C81.x-C85.x, C88.x, C90.x-C97.x  Weighted as +2 in overall comorbidity score if ComorbTumor = 0; otherwise weighted as 0 | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbMalignancyIP | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for malignancy, including lymphoma and leukemia, except malignant neoplasm of skin during the specified year at an in-person encounter?  This variable is only available for Site A.  Identified based on codes used in HCSRN macro.  Weighted as +2 in overall comorbidity score if ComorbTumor = 0; otherwise weighted as 0 |  |
| ComorbLiver | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for moderate or severe liver disease during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 456.0–456.2, 572.2–572.8 ICD-10 codes: I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7  Weighted as +3 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbTumor | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for metastatic solid tumor during the specified year or a prior calendar year within cohort window?  ICD-9-CM codes: 196.x–199.x  ICD-10 codes: C77.x-C80.x  Weighted as +6 in overall comorbidity score | Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| ComorbTumorIP | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for metastatic solid tumor during the specified year at an in-person encounter?  This variable is only available for Site A.  Identified based on codes used in HCSRN macro.  Weighted as +6 in overall comorbidity score |  |
| ComorbHIV | **Numeric**  0 = No  1 = Yes | Did the cohort member have a diagnosis code for AIDS/HIV during the specified year or a prior calendar year within cohort window?  Only record new diagnoses that are not recorded in the Prior file, which contains all diagnoses recorded prior to cohort entry.  ICD-9-CM codes 042.x–044.x and V08. ICD-10 codes: B20.x-B22.x, B24.x, B23.x and Z21.  Codes for asymptomatic HIV were not originally included in Quan *et al.* 2005 study were included (V08 and Z21). Sites used different data sources, including billing codes, laboratory data reports, and problem lists, to identify HIV diagnosis based on site-specific nuances.  Weighted as +6 in overall comorbidity score | Derived as [EventHIV](#CTFHIV) (p. 54) variable in the Cervical Timeline file.  Recommend using the [Cervical Timeline File](#CTF) (p. 52) variables EventHIV (earliest documented record of patient having HIV) and StatusHIV (if patient was diagnosed with HIV at time of event).  Recommend using the [ComorbIndex](#CalYrCharlKP) (p. 19) and [ComorbIndexPrev\_Drv](#CalYrCharlPrev) (p. 19) to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. |
|
|
| EnrolledMonths | **Numeric** | Number of months enrolled in the health insurance plan during the calendar year  Entered an integer between 1-12. |  |
|
|
| Smoke | **Numeric**  0=Never  1 = Current  2 = Former | Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
| SmokeDateMonth | **Numeric** | Month of smoking status recorded date  Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
|
|
| SmokeDateYear | **Numeric** | Year of smoking status recorded date  Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
| SmokeDSR | **Numeric** | Date smoking status recorded, days since reference date (birth date)  Collected data from last value available in calendar year. If no value was recorded during the calendar year, recorded the last available value prior to cohort entry. |  |
|
|
|
|
| ComorbIndex | **Numeric** | Comorbidity index, based on HCSRN macro  This variable is only available for Site A. |  |
|
|
| ComorbIndexIP2 | **Numeric** | Comorbidity index from in-person encounters only, based on HCSRN macro  This variable is only available for Site A. |  |
| ComorbIndexPrev\_Drv | **Numeric** | Comorbidity index, based on prevalent comorbidities  This variable is only available for Site B and Site C.  Comorbidity flags provided by Site B and Site C were used to calculate prevalent scores. | Recommend using this variable for analyses in conjunction with the [EncCalYr\_Drv](#CalYrEncDrv) (p. 20) variable to select for cohort members with at least one primary care encounter within the respective calendar year. |
|
|
|
|
| CalYr\_Ins\_Drv | **Numeric**  1 = Medicare  2 = Medicaid 3 = InsOtherGov  4 = InsCommerc 5 = Uninsured/Medical Assistance  8 = InsOther | Cohort member insurance for the calendar year  Insurance designation made at any time during the calendar year.  If multiple insurance designations of the following insurance types were made in a calendar year, then record set to ‘7’: Medicaid, Medicare, Commercial, Other Government Insurance, and Other Insurance. |  |
|
|
| BMI\_Drv | **Numeric** | Body Mass Index (NIH NHLBI scale)  Rounded to the nearest tenth.  Height and weight values used to calculate BMI were submitted within a limited range (48-90 inches and 50-600 lb). However, height and/or weight values at the respective range limit(s) may lead to extreme BMI values (e.g., above 54 or under 19). |  |
|
|
| EncCalYr\_Drv | **Numeric**  0 = No  1 = Yes | Did the cohort member have a primary care visit within the calendar year? | This variable was added to permit analysis flexibility in inclusion of calendar year data based on whether the patient had a primary care encounter. |
|
|
| UtilCalYr\_Drv | **Numeric**  0 = No  1 = Yes | Was the cohort member in the study for at least one day during the calendar year based on a 36-month utilization period from last primary care encounter? | This variable was added to permit analysis flexibility in inclusion of calendar year data for cohort members from Site B and Site C based on whether the cohort member had had a primary care encounter within 36 months of the last primary care encounter.  **See** [Section 3: Appendices - Cohort Exit due to Lack of Primary Care Utilization](#_Cohort_Exit_due) (p. 66) **for further detail.** |
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|
|
|

## Engagement File

### Overview

This file contains one record for each cohort entry and exit and includes reason for exit. See [Section 1: Overview - Description of METRICS Data Contributing Sites & Cohort Definitions](#CohortEntry) (p. 2) for cohort entry and cohort exit definitions for each site.

### Record Structure

One record per cohort entry event.

### General Harmonization Notes

**Identification of cohort member study periods for primary and sensitivity analyses. To make cohort periods more conceptually analogous between the enrollment and utilization cohorts, cohort members from the utilization cohorts (Site B and Site C; see** [Section 1: Overview - Description of METRICS Data Contributing Sites & Cohort Definitions](#CohortEntry) (p. 2) **for further detail) that had a lapse in primary care utilization were administratively cut-off from the cohort 37 months after the last primary care encounter. Cohort members from utilization cohort then re-entered the study upon subsequent primary care utilization. See** [Section 3: Appendices - Cohort Member Study Periods for Primary and Secondary Analyses](#HPVTestDate) (p. 66) **for further detail.**

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
| CohortEntryDateMonth | **Numeric** | Month of cohort entry date |  |
|
|
|
|
| CohortEntryDateYear | **Numeric** | Year of cohort entry date |  |
|
|
| CohortEntryDSR | **Numeric** | Cohort entry date, days since reference date (birth date) |  |
|
|
| CutOffReason | **Numeric**  1 = Administrative cut-off  2 = Age out  4 = Disenrollment  5 = Move out of coverage area  6 = Attributed provider leaves healthcare system  7 = Death  9 = Lack of Primary Care Utilization | Cut-off reason  If the cohort member was still in the cohort on last day for data submission, record set to ‘1’.  Cohort members aged out the day prior to their 90th birthday. |  |
|
|
| CutOffDateMonth | **Numeric** | Month of cut-off date |  |
|
|
|
|
| CutOffDateYear | **Numeric** | Year of cut-off date |  |
|
|
| CutOffLastDSR | **Numeric** | Cut-off date, days since reference date (birth date) |  |
|
|
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|
|
|

## Cancer Registry File

### Overview

This file contains one record for each cervical cancer diagnosis for each cohort member from cohort entry through cohort exit based on available cancer registry data.

### Record Structure

One record per cohort member per sequence number (a [NAACCR](https://www.naaccr.org/data-standards-data-dictionary/) concept indicating the order of cancer diagnoses) of primary tumor for cervical cancers diagnosed throughout cohort duration. Cancer diagnoses made prior to cohort entry are not present in this file but are present in the Cerivcal Timeline file.

### General Harmonization Notes

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
|
|
| SequenceNumber | **Character**  2-digit Text string  00=One primary in the patient’s lifetime  01 = First of two or more primaries  02 = Second of two or more primaries  03-34 = (Actual number of this primary)  35 = Thirty-fifth of thirty-five or more primaries  99 = Unspecified or unknown sequence number of Federally required in situ or malignant tumors. Sequence number 99 can be used if there is a malignant tumor and its sequence number is unknown. (If there is known to be more than one malignant tumor, then the tumors must be sequenced.)  88 = Unspecified or unknown sequence number of non-malignant tumor or central-registry defined neoplasms. (Sequence number 88 can be used if there is a non-malignant tumor and its sequence number  98 = Cervix carcinoma in situ (CIS/CIN III, Diagnosis Years 1996-2002)  60=Only one non-malignant tumor or central registry-defined neoplasm 61 = First of two or more non-malignant tumors or central registry-defined neoplasms  62 = Second of two or more non-malignant tumors or central registry-defined neoplasms | Sequence number of primary tumor at this reporting facility  NAACCR Item #380 or #560 (if both NAACCR items available, populated with the item that corresponds to the source registry [Central or Hospital]) |  |
|
|
| PrimarySiteICD | **Character** | ICD Code, primary site  NAACCR Item #400 |  |
|
|
| DxDateMonth | **Numeric** | Month of cancer diagnosis date  NAACCR item #390 |  |
|
|
|
|
| DxDateYear | **Numeric** | Year of cancer diagnosis date  NAACCR item #390 |  |
|
|
| DxDSR | **Numeric** | Date of cancer diagnosis, days since reference date (birth date)  NAACCR item #390 |  |
|
|
| TumorBehavior | **Numeric**  0=Benign  1 = Uncertain/borderline  2 = In situ  3 = Malignant/invasive  99 = Unknown | Tumor malignancy or behavior  NAACCR Item #523 |  |
|
|
| SEERStage | **Numeric**  0=In situ  1 = Localized only  2 = Regional by direct extension only  3 = Regional lymph nodes involved only  4 = Regional by BOTH direct extension AND lymph node involvement  5 = Regional, NOS (Not Otherwise Specified)  7 = Distant site(s)/node(s) involved  9 = Unknown/unstaged/unspecified/DCO  98 = N/A | Cancer stage (SEER)  NAACCR Items #759, #762, #764, or #3020, (whichever was available from registry)  Used SEER stage 2000 for cancers diagnosed 2010-2018 and SEER stage 2018 for cancers diagnosed 2018-2019. |  |
|
|
| AJCC7StageDerived | **Character**  000 = 0 010 = 0a  020 = 0is 100 = I  110 = INOS 120 = IA  121 = IANOS 130 = IA1  140 = IA2 150 = IB  151 = IBNOS 160 = IB1  170 = IB2 180 = IC  190 = IS 200 = IEA  210 = IEB 220 = IE  230 = ISA 240 = ISB  300 = II 310 = IINOS  320 = IIA 321 = IIANOS  322 = IIA1 323 = IIA2  330 = IIB 340 = IIC  350 = IIEA 360 = IIEB  370 = IIE 380 = IISA  390 = IISB 400 = IIS  410 = IIESA 420 = IIESB  430 = IIES 500 = III  510 = IIINOS 520 = IIIA  530 = IIIB 540 = IIIC  541 = IIIC1 542 = IIIC2  550 = IIIEA 560 = IIIEB  570 = IIIE 580 = IIISA  590 = IIISB 600 = IIIS  610 = IIIESA 620 = IIIESB  630 = IIIES 700 = IV  710 = IVNOS 720 = IVA  721 = IVA1 722 = IVA2  730 = IVB 740 = IVC  888 = NA 900 = OCCULT  999 = UNK Stage | AJCC-7 Stage Group, Derived  NAACCR item #910, #970, #3430, or #3000  Pathological (preferred) or clinical AJCC stage group at diagnosis for cancers diagnosed 2010-2017. |  |
|
|
| AJCC7PathStage | **Character** | AJCC-7 Stage Group, pathologic  NAACCR item #910 |  |
|
|
| AJCC7ClinStage | **Character** | AJCC-7 Stage Group, clinical  NAACCR item #970 |  |
|
|
| AJCC7Stage | **Character** | AJCC-7 Stage  NAACCR item #3430 |  |
|
|
|
|
|
|
|
|
|
|
| TumorSizeMm | **Character**  000 = No mass/tumor found  001-988 = 001-988 millimeters (mm)(Code exact size in mm)  989 = 989 mm or larger  990=Microscopic focus or foci only and no size of focus given For Breast: Microinvasion Microscopic focus or foci only and no size given, Described as "less than 1 mm", Stated as T1mi with no other information on tumor size For Breast: Stated as T1b with no other information on tumor size  991 = Described as "less than 1 centimeter (cm)"  992 = Described as "less than 2 cm," or "greater than 1 cm," or "between 1 cm and 2 cm" For Breast: Stated as T1 [NOS] or T1c [NOS] with no other information on tumor size  993 = Described as "less than 3 cm," or "greater than 2 cm," or "between 2 cm and 3 cm"  994 = Described as "less than 4 cm," or "greater than 3 cm," or "between 3 cm and 4 cm"  995 = Described as "less than 5 cm," or "greater than 4 cm," or "between 4 cm and 5 cm" 996 = For Breast: Mammographic/xerographic diagnosis only, no size given; clinically not palpable 997 = For Breast: Paget disease of nipple with no demonstrable tumor  998 = For Breast: Diffuse For colon: Familial/multiple polyposis  999 = Unknown; size not stated Size of tumor cannot be assessed, Not documented in patient record | Tumor size (mm)  NAACCR Item #2800 |  |
|
|
| HistologyICD02or03 | **Numeric** | Tumor histology or cell type, as first 4 digits of the ICD-O-2 and ICD-O-3 morphology code  NAACCR Item #=522 |  |
|
|
|
|
| RegNodesPositive | **Character**  00=All nodes examined negative  01-89 = 1-89 nodes positive (code exact number of nodes positive)  90=90 or more nodes positive  95 = Positive aspiration or core biopsy of lymph node(s)  97 = Positive nodes but number of positive nodes not specified  98 = No nodes examined  99 = UNKNOWN if nodes are positive or negative; not applicable | Number of regional nodes positive  NAACCR Item #820 |  |
|
|
| TxIniStatus | **Numeric**  0 = No treatment given  1 = Treatment given  2 = Active surveillance (watchful waiting)  9 = Unknown if treatment was given | Treatment status  NAACCR Item #1285 |  |
|
|
| TxCaIniDateMonth | **Numeric** | Month of date therapy initiated |  |
|
|
|
|
| TxCaIniDateYear | **Numeric** | Year of date therapy initiated |  |
|
|
| TxCaIniDSR | **Numeric** | Date therapy initiated, days since reference  NAACCR Item #1260 | All DSR variables use birth date as the reference date. |
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
| TxIniSequence | **Numeric**  0 = No systemic therapy and/or surgical treatment; Unknown if surgery and/or systemic therapy given  2 = Systemic therapy before surgery  3 = Systemic therapy after surgery  4 = Systemic therapy both before and after surgery  5 = Intraoperative systemic therapy  6 = Intraoperative systemic therapy with other systemic therapy administered before or after surgery  7 = Surgery both before and after systemic therapy (effective for cases diagnosed 1/1/2012 and later)  9 = Sequence unknown | Systemic treatment/surgery sequence  NAACCR Item# =1639 |  |
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
| Laterality | **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 | Side of the body on which the tumor originated  NAACR item #410 |  |
| AJCCTNMClinStage | **Character** | AJCC TNM Clinical Stage Group  NAACCR item #1004 |  |
| AJCCTNMPathStage | **Character** | AJCC TNM Pathologic Stage Group  NAACCR item #1014 |  |
| AJCCTNMPostTxStage | **Character** | AJCC TNM Post-Therapy Stage Group  NAACCR item #1024 |  |
| EOD2018Stage | **Character** | EOD 2018 Stage Group, Derived  NAACCR item #818  Available for cancers diagnosed from 2018-2020 |  |
| AJCCTNMPostTxClinStage | **Character** | AJCC TNM Post-Therapy Clinical (yc) Stage Group  NAACCR item #1067 |  |
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|
|
|

## Encounter File

### Overview

This file contains one record for every primary care visit for each cohort member from cohort entry to cohort exit.

### Record Structure

One record per primary care encounter from cohort entry through cohort exit for each cohort member (see [Section 1: Overview - Description of METRICS Data Contributing Sites & Cohort Definitions](#PCV) (p. 3) for primary care visit definitions for each site.). Multiple encounters may occur on the same date.

### General Harmonization Notes

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
| PCPVisitDateMonth | **Numeric** | Month of Primary Care Visit Date |  |
|
|
|
|
| PCPVisitDateYear | **Numeric** | Year of Primary Care Visit Date |  |
|
|
| PCPVisitDSR | **Numeric** | Date of primary care visit, days since reference date (birth date) |  |
|
|
| PCPVisitType | **Numeric**  1 = In-Person  2 = Telehealth – Scheduled Phone  3 = Telehealth – Scheduled Virtual  4 = Telehealth – Chat  5 = Telehealth – Text  6 = Telehealth – Other  7 = Telehealth – Unknown | Type of primary care visit  All primary care visits collected prior to 2019 were in-person. |  |
| ProviderIDEnc | **Character** | Provider ID for Encounter  Entered a de-identified code that links to the Provider file. |  |
|
|
| FacilityIDEnc | **Character** | Clinic ID where Encounter occurred  The clinic/facility associated with the encounter. Entered a de-identified code that links to the Facility file. |  |
|
|
| FacilityPerformType | **Numeric**  1 = Primary Care  2 = Women’s Health/OB Gyn  3 = Both Primary Care and Women’s Health/OB Gyn  4 = HIV | Performing facility type  The type of facility at which the encounter occurred |  |
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|
|
|

## Death File

### Overview

This file contains one record for each cohort member known to have died during the cohort period.

### Record Structure

One record per known cohort member death during the cohort period through the end of 2020. Note that death due to cervical cancer is noted for cohort members who were diagnosed prior to cohort entry and died from cervical cancer after cohort entry; these cohort members do not appear in the Cancer Registry file and do appear in the Cervical Timeline file. Also note that some cohort members who were removed from the cohort due to lack of utilization for primary analyses and were known to have died later and prior to removal for other reasons used for secondary analyses are included in this file (**see** [Section 3: Appendices - Cohort Member Study Periods for Primary and Secondary Analyses](#HPVTestDate) (p. 66) **for further detail.**)

### General Harmonization Notes

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
| DeathDateMonth | **Numeric** | Month of death date |  |
|
|
| DeathDateYear | **Numeric** | Year of death date |  |
|
|
| DeathDSR | **Numeric** | Death date, days since reference |  |
|
|
| DeathDateType | **Numeric**  1 = Known 2 = Estimate | Death date type  If the death date was specifically reported in a registry, health plan, state, or EMR, set to '1'.  If the cohort member is known to be deceased but the exact death date is unknown, set to '2'. |  |
|
|
| DeathSource | **Numeric**  1 = Registry 2 = Health plan 3 = EMR | Death data source |  |
|
|
| CauseofDeath | **Numeric**  1 = Cervical Cancer | Cause of death  Death due to cervical cancer noted for cohort members diagnosed prior to cohort entry and during the cohort period.  All known causes of death not attributed to cervical cancer were set to Other. |  |
|
|
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|
|
|

## Provider File

### Overview

This file contains one record for each provider found in the following files: Calendar Year, Encounter, and Cervical Timeline.

### Record Structure

One record per unique provider ID. All provider IDs recorded in any other files have a record in this file.

### General Harmonization Notes

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| ProviderID | **Character** | Provider ID | De-identified and unique to the provider within the PRC. If a provider is a member of both the colorectal and cervical cohorts, sites used the same unique provider ID for that provider in all files in both data sets. |
|
|
| ProviderSpecialty | **Character**  21 = Anesthesiology  22 = Emergency Medicine  23 = Family Medicine  24 = Internal Medicine, General internal medicine  25 = Internal Medicine, Geriatrics  26 = Internal Medicine, Gastroenterology  27 = Internal Medicine, Oncology  28 = Internal Medicine, Other  29 = Internal Medicine, Sub-specialty unknown or no sub-specialty  30 = Midwifery  31 = Nursing  32 = Obstetrics and Gynecology  33 = Pathology  34 = Pediatrics  35 = Radiology  36 = Surgery | Provider medical specialty  Multiple entries indicated by a pipe "|" symbol. |  |
|
|
| ProviderSpecialtyOther | **Character** | Provider medical specialty (other, specify) |  |
|
|
| ProviderType | **Numeric**  1 = Administrative staff  2 = Fellows (includes MDs and DOs)  3 = Licensed practical nurse  4 = Medical assistant  5 = Nurse practitioner  6 = Physician (includes MDs and DOs)  7 = Physician´s assistant  8 = Registered nurse  9 = Resident physician (includes MDs and DOs) | Provider type |  |
|
|
| ProviderTypeOther | **Character** | Provider type (other, specify) |  |
|
|
|
|
| ProviderIsPerson | **Numeric**  0 = No  1 = Yes | Is provider known to be a person? |  |
|
|
| ProviderSex | **Numeric**  0 = Male 1 = Female 2 = Non-binary | Sex that provider identifies as |  |
|
|
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|

## Facility File

### Overview

This file contains one record for each facility ID found in the following files: Calendar Year, Encounter, and Cervical Timeline.

### Record Structure

One record per unique facility ID. All facility IDs recorded in any other files have a record in this file. See [Section 3: Appendix - Facilities Harmonization](#FacilHarm) (p. 69) for facility level definition for each site.

### General Harmonization Notes

The METRICS sites sought to harmonize the level at which facilities were defined across the three disparate healthcare system to permit analysis execution at a consistent level. [Appendix: Facilities Harmonization](#_Facilities_Harmonization) (p. 69) summarizes the method by which an analyst can identify facilities at a consistent level across all three METRICS sites. Direct comparison at level 2 and level 3 should not be conducted without direct input from the data contributing sites

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| FacilityID | **Character** | Facility ID | De-identified and unique to the facility within the PRC. If a facility was present in both the colorectal and cervical cohorts, the same unique faciliity ID was used in both data sets. See [Section 3: Appendix - Facilities Harmonization](#FacilHarm) (p. 69) for summary of how this variable can be harmonized by level for analyses across sites. |
|
|
| FacilityStFIP | **Character** | Facility state, FIPS code  Indicates the state 2-digit FIPS code with leading zeroes included. |  |
|
|
| FacilityCountyFIP | **Character** | Facility county, FIPS code  Indicates the county 3 digit FIPS code with leading zeroes included. |  |
|
|
| FacilityZip | **Character** | Facility ZIP code  Indicates the 5-digit zip code with leading zeroes included. | Zip codes not recognized by SAS/USPS set to ‘99’ for this variable. |
|
|
| FacilityType | **Numeric**  21 = Medical Center  22 = Hospital  23 = Emergency Room – Hospital  24 = Urgent Care Facility  25 = Ambulatory Surgical Center  26 = Office or Clinic  27 = Public Health Clinic  28 = Rural Health Clinic  29 = Federally Qualified Health Center  30=Indian Health Service facility  31 = Tribal Facility  32 = Mobile unit  33 = Laboratory  97 = Other, specify | Facility type |  |
|
|
| FacilityTypeOther | **Character** | Facility type (other, specify)  Indicates additional description of facility type not included in FacilityType variable up to 200 characters. |  |
|
|
| FacilityNetworkStatus | **Numeric**  1 = Always owned and/or operated by your health care organization  2 = Always an external facility (includes contract facilities)  3 = Was an owned facility, most recently an external facility  4 = Was an external facility, most recently an owned facility | Facility status in health system  Historical relationship with the healthcare system was not available to any site and so known facilities will only be flagged as ‘1’ or ‘2’ across all sites. | This variable aligns with the PCC ‘Relationship’ and ‘Relationship\_History’ variables. |
|
|
|
|
| FacilityIDRelatedPhys | **Character** | Associated physical larger facility ID  Reflects the larger aggregating facility at which multiple co-located facility IDs may be rolled up as. | See [Section 3: Appendices - Facilities Harmonization](#FacilHarm) (p. 69) for summary of how this variable was harmonized across sites. |
|
|
| FacilityIDRelatedAggr | **Character** | Facility aggregating value, PRC-defined  Optional. |  |
|
|
|
|
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|

## Social Determinants of Health

### Overview

This file contains one record for every cohort member in the cohort at cohort entry.

### Record Structure

One record per cohort member that entered the cohort.

### General Harmonization Notes

Cohort members that entered both the cervical and colorectal cohorts have the same geocoding information so long as the cohort member entered both cohorts at the same address. Select cohort members present in both the cervical and colorectal datasets entered at different timepoints and so may have different geocoding information due to a change in address between cohort entries.

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
| GeoAddrPOBox | **Numeric**  0 = No 1 = Yes | Is the address a PO Box?  If ‘1 = Yes’ then all other data fields assigned missing, except PID, GeoAddrPOBox, address date begin and end, extractDate, providingSite, and ddVersion. |  |
|
|
| GeoAddrBeginMth | **Numeric** | Month of geocoding address date  Date may be reflective of single date associated with geocoding address or earliest date in interval associated with geocoding address (i.e., address is reflective of beginning to end of specific time interval). |  |
|
|
|
|
| GeoAddrBeginYr | **Numeric** | Year of geocoding address date  Date may be reflective of single date associated with geocoding address or earliest date in interval associated with geocoding address (i.e., address is reflective of beginning to end of specific time interval). |  |
|
|
| GeoAddrBeginDSR |  | Geocoding address date, days since reference |  |
|
|
|
|
| GeoCertainty | **Numeric**  1 = Census tract based on complete and valid address;  2 = Census tract based on ZIP+4;  3 = Census based on residence ZIP+2; 4 = Census tract based on ZIP only;  5 = Census tract based on ZIP of PO box;  6 = Census based on residence city/ZIP, city/ZIP has only one census tract;  7 = Other | Geocoding certainty level  Reflects standard output from geocoding software and varies by software. |  |
|
|
| GeoCertaintyOther | **Character** | Geocoding certainty level, Other (Specify) |  |
|
|
| GeoMatchScore | **Numeric** | Geocoding match score |  |
|
|
| GeoSoftwareName | **Character** | Geocoding software name |  |
|
|
| GeoSoftwareVersion | **Character** | Geocoding software version |  |
|
|
| Race\_NHisp\_White Race | **Numeric** | % White Alone, Non-Hispanic |  |
|
|
| Race\_NHisp\_Black Race | **Numeric** | % Black or African American Alone, Non-Hispanic |  |
|
|
| Race\_NHisp\_AIAN Race | **Numeric** | % American Indian and Alaska Native Alone, Non-Hispanic |  |
|
|
| Race\_NHisp\_Asian | **Numeric** | % Asian Alone, Non-Hispanic |  |
|
|
| Race\_NHisp\_NatHaw Race | **Numeric** | % Native Hawaiian and Other Pacific Islander Alone, Non-Hispanic |  |
|
|
| Race\_NHisp\_Other Race | **Numeric** | % Some Other Race Alone, Non-Hispanic |  |
|
|
| Race\_NHisp\_Multi Race | **Numeric** | % Two or More Races, Non-Hispanic |  |
|
|
| Race\_Hisp\_White Race | **Numeric** | % White Alone, Hispanic |  |
|
|
| Race\_Hisp\_Black Race | **Numeric** | % Black or African American Alone, Hispanic |  |
|
|
| Race\_Hisp\_AIAN Race | **Numeric** | % American Indian and Alaska Native Alone, Hispanic |  |
|
|
| Race\_Hisp\_Asian Race | **Numeric** | % Asian Alone, Hispanic |  |
|
|
| Race\_Hisp\_NatHaw Race | **Numeric** | % Native Hawaiian and Other Pacific Islander Alone, Hispanic |  |
|
|
| Race\_Hisp\_Other Race | **Numeric** | % Some Other Race Alone, Hispanic |  |
|
|
| Race\_Hisp\_Multi Race | **Numeric** | % Two or More Races, Hispanic |  |
|
|
| Educ\_Less9th Education | **Numeric** | % less that 9th grade |  |
|
|
| Educ\_9th\_12th Education | **Numeric** | % 9th-12th grade (no diploma) |  |
|
|
| Educ\_HSGrad Education | **Numeric** | % high school graduate |  |
|
|
| Educ\_SomeColl Education | **Numeric** | % some college, no degree |  |
|
|
| Educ\_AssocDeg Education | **Numeric** | % associate's degree |  |
|
|
| Educ\_BachDeg Education | **Numeric** | % bachelor's degree |  |
|
|
| Educ\_MastProfDeg Education | **Numeric** | % master's or professional school degree |  |
|
|
| Educ\_DoctDeg Education | **Numeric** | % doctorate degree |  |
|
|
| Person\_Below\_Pov Poverty | **Numeric** | % of person below poverty level |  |
|
|
| RUCA4A | **Numeric** | RUCA 4-level categorization 'A' |  |
|
|
| Flag\_Tract\_Pop\_Zero | **Numeric**  0 = No 1 = Yes | Flag: Census Tract Pop Zero |  |
|
|
| LQ\_White\_Alone | **Numeric** | Location Quotient (LQ): White Alone |  |
|
|
| LQ\_Black\_Alone | **Numeric** | Location Quotient (LQ): Black or African American Alone |  |
|
|
| LQ\_API\_Alone | **Numeric** | Location Quotient (LQ): Asian or Pacific Islander Alone |  |
|
|
| LQ\_Hispanic | **Numeric** | Location Quotient (LQ): Hispanic |  |
|
|
| LQ\_NH\_White\_Alone | **Numeric** | Location Quotient (LQ): White Alone, Non-Hispanic |  |
|
|
| ICE\_Black\_Alone\_White\_Alone | **Numeric** | Index of Concentration at Extremes (ICE): Black or African American Alone/White Alone |  |
|
|
| ICE\_API\_Alone\_White\_Alone | **Numeric** | Index of Concentration at Extremes (ICE): Asian or Pacific Islander Alone/White Alone |  |
|
|
| ICE\_Hispanic\_NH\_White\_Alone | **Numeric** | Index of Concentration at Extremes (ICE): Hispanic/White Alone, Non-Hispanic |  |
|
|
| Lex\_Is\_Black\_Alone\_White\_Alone | **Numeric** | Local Exposure and Isolation (Lex/IS): Black or African American Alone/White Alone |  |
|
|
| Lex\_Is\_API\_Alone\_White\_Alone | **Numeric** | Local Exposure and Isolation (Lex/IS): Asian or Pacific Islander Alone/White Alone |  |
|
|
| Lex\_Is\_Hispanic\_NH\_White\_Alone | **Numeric** | Local Exposure and Isolation (Lex/IS): Hispanic/White Alone, Non-Hispanic |  |
|
|
| LIS\_White\_Alone | **Numeric** | Local Isolation Score: White Alone |  |
|
|
| LIS\_Black\_Alone | **Numeric** | Local Isolation Score: Black or African American Alone |  |
|
|
| LIS\_API\_Alone | **Numeric** | Local Isolation Score: Asian or Pacific Islander Alone |  |
|
|
| LIS\_Hispanic | **Numeric** | Local Isolation Score: Hispanic |  |
|
|
| LIS\_NH\_White\_Alone | **Numeric** | Local Isolation Score: White Alone, Non-Hispanic |  |
|
|
| Yost\_Overall\_Quintile | **Numeric** | Yost Quintile (Total US) |  |
|
|
| Yost\_State\_Quintile | **Numeric** | Yost Quintile (by state) |  |
|
|
| Enc\_St | **Character** | State (Pseudocode)  Pseudocodes are preceded by letters that denote the site from which the record derives as follows: ‘A’ for site A; ‘B’ for site B; and ‘D’ for site C. |  |
|
|
| Enc\_County | **Character** | County (Pseudocode)  Pseudocodes are preceded by letters that denote the site from which the record derives as follows: ‘A’ for site A; ‘B’ for site B; and ‘D’ for site C. |  |
|
|
| Enc\_Tract | **Character** | Census Tract (Pseudocode)  Pseudocodes are preceded by letters that denote the site from which the record derives as follows: ‘A’ for site A; ‘B’ for site B; and ‘D’ for site C. |  |
|
|
| Enc\_FIPS | **Character** | FIPS code (Pseudocode)  Pseudocodes are preceded by letters that denote the site from which the record derives as follows: ‘A’ for site A; ‘B’ for site B; and ‘D’ for site C. |  |
|
|
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|

## Cervical Timeline File

### Overview

This analytic file contains a record of every event that determines the cervical cancer screening timeline for each cohort member.

### Record Structure

One record per cohort member per date on which an event occurs, where multiple events can occur on the date. Where multiple events can occur on the same date, mutually exclusive data utilizes information from the more severe event.

### General Harmonization Notes

The goal of the Cervical Timeline file is to consolidate screening measures across multiple files into one modular time-series that reflects the entirety of the cervical cancer screening process for each cohort member (see [Section 3: Appendices - Cervical Timeline File Data Structure](#_Data_Structure) (p. 70) for further detail)

In addition to presenting data submitted by the sites in the original CDE files, the following derived variables were created to convey relevant screening process measures: risk status, which conveys screening eligibility and anticipated follow-up (see [Section 3: Appendices - Cervical Timeline File Risk Status Assignment](#_Risk_Status_Assignment) (p. 71) for further detail); testing strategy, which conveys test modality (see [Section 3: Appendices - Cervical Timeline File Pap/HPV Test Strategy](#APPCTFTest) (p. 72) for further detail); and test results, which conveys the individual and combined screen results relevant to follow-up recommendations (see [Section 3: Appendices - Cervical Timeline File Test Results](#AppCTFRes) (p. 74) for further detail).

**The Cervical Timeline file supports primary analyses only (i.e., those incorporating the lack of utilization cut-off for cohort exit for Site B and Site C); if executing sensitivity analyses, use the Cervical Timeline file in conjunction with cohort entry/exit information in the Participant and Calendar Year files (see** [Section 3: Appendices - Cohort Member Study Periods for Primary and Secondary Analyses](#HPVTestDate) (p. 66)).

| **Variable Name** | **Variable Type, Formatting, Permissible Values** | **Description** | **DAU Harmonization Notes** |
| --- | --- | --- | --- |
| PIDSite | **Character** | Unique participant ID | De-identified and unique to the cohort member within the PRC. Cohort members in PROSPR I cohort either have the same ID in PROSPR II or a cross-walk is available to connect IDs. |
|
|
| EventAge | **Numeric** | Event Age | This variable was calculated as follows:  EventDSR/365.25, then rounded to the lowest integer. |
|
|
| EventDSR | **Numeric** | Event Days Since Reference Date (birth date) |  |
|
|
| EventMth | **Numeric**  Integer between 1-12 | Event Month  Month in which event occurred |  |
|
|
| EventYr | **Numeric**  4-digit integer | Event Year  Year in which event occurred |  |
|
|
|
|
| EventCE | **Numeric**  0 = No 1 = Yes | Cohort Entry Flag  Event flagging record at which cohort member entered the PROSPR cohort | Cohort members that entered the METRICS cohort multiple times will have multiple records for which EventCE = 1. |
|
|
| EventPap | **Numeric**  0 = No 1 = Yes | Pap Test Flag  Event flagging record when patient had a Pap test | All Pap tests and HPV tests that occurred within a 14-day window were grouped together into one event record. If multiple Pap tests occurred within a 14-day window, the most severe result was used, and the date of the associated Pap test was used (if the HPV test occurred on a later date). |
|
|
| EventHPV | **Numeric**  0 = No 1 = Yes | HPV Test Flag  Event flagging record at which patient had a HPV test. | All Pap tests and HPV tests that occurred within a 14-day window were grouped together into one event record. If multiple HPV tests occurred within a 14-day window, the most severe result was used, and the date of the associated Pap test was used (if the HPV test occurred on a later date). |
|
|
| EventProc | **Numeric**  0 = No 1 = Yes | Procedure Flag  Event flagging record at which patient had a procedure. Procedure includes any colposcopy/biopsy and treatment. | If this variable has a value of ‘1’, then either the EventColpoBx or EventTx variable will have a value of ‘1’, based on whether the procedure was a colposcopy/biopsy or a treatment. |
|
|
| EventColpoBx | **Numeric**  0 = No 1 = Yes | Colpo/Biopsy Flag  Event flagging record at which patient had a colposcopy/biopsy procedure, including endocervical curettage. | If this variable has a value of ‘1’, then EventProc will have a value of ‘1’. |
|
|
| EventTx | **Numeric**  0 = No 1 = Yes | Treatment Flag  Event flagging record at which patient had a treatment procedure, including LEEP, cone, cryotherapy, laser, excision NOS, partial/subtotal/supracervical hysterectomy, trachelectomy, or total/radical hysterectomy. | If this variable has a value of ‘1’, then EventProc will have a value of ‘1’. |
|
|
| EventHIV | **Numeric**  0 = No 1 = Yes | First HIV positive Flag  Event flagging record at which patient is first documented as being diagnosed with HIV. | If the cohort member entered the cohort with a prior HIV diagnosis, the EventDSR associated with this variable will correspond to 1 day prior to the EventDSR associated with the EventCE variable. |
|
|
| EventPregBegin | **Numeric**  0 = No 1 = Yes | Pregnancy Begin Flag  Event flagging record at which patient documented as being pregnant. | If pregnancy begin date was unknown, pregnancy begin date was calculated from the pregnancy end date as follows: for a live birth, pregnancy end date - 270 days; for a non-live birth, pregnancy end date - 112 days. |
|
|
| EventPregEnd | **Numeric**  0 = No 1 = Yes | Pregnancy End Flag  Event flagging record at which patient is documented as no longer being pregnant. | If pregnancy end date was unknown, pregnancy end date was calculated from the pregnancy begin date as follows: for a live birth, pregnancy begin date + 270 days; for a non-live birth, pregnancy begin date + 112 days. |
|
|
| EventCervCancer | **Numeric**  0 = No 1 = Yes | First Cancer Diagnosis Flag  Event flagging record at which patient is first diagnosed with cervical cancer. |  |
|
|
| EventNoCervix | **Numeric**  0 = No 1 = Yes | Cervix Not Present Flag  Event flagging record at which patient is first categorized as not having a cervix. | If the cohort member was noted as having had a cervix removal prior to cohort entry but no surgical date was available, the EventDSR associated with this variable will correspond to 1 day prior to the EventDSR associated with the EventCE variable. Hysterectomy NOS was presumed to remove the cervix.  If a cohort member underwent multiple surgical procedures, the first surgery in which a cervix was removed was used. If the first surgery was a partial/sub-cervical hysterectomy, the next hysterectomy procedure was used. |
|
|
| EventCEnd | **Numeric**  0 = No 1 = Yes | Cohort Exit Flag  Event flagging record at which cohort member first exited the METRICS cohort. |  |
|
|
| Event18 | **Numeric**  0 = No 1 = Yes | Age 18 Flag  Event flagging record at which cohort member turned age 18. |  |
|
|
| Event21 | **Numeric**  0 = No 1 = Yes | Age 21 Flag  Event flagging record at which cohort member turned age 21. |  |
|
|
| Event25 | **Numeric**  0 = No 1 = Yes | Age 25 Flag  Event flagging record at which cohort member turned age 25. |  |
|
|
| Event30 | **Numeric**  0 = No 1 = Yes | Age 30 Flag  Event flagging record at which cohort member turned age 30. |  |
|
|
| Event66 | **Numeric**  0 = No 1 = Yes | Age 66 Flag  Event flagging record at which cohort member turned age 66. |  |
|
|
| EventHPVVac | **Numeric**  0 = No 1 = Yes | Vaccination Flag  Event flagging record at which patient had vaccination administered. | Records were excluded under the following circumstances: unknown administration date; vaccination date prior to 2006; less than 8 years of age at vaccination; vaccines beyond the fourth dose for a given cohort member; if the second and third vaccine doses, if the age at administration was less than 15 and the difference in dates between first and second vaccination is less than 5 months (150 days), then the vaccine record was excluded; for the second vaccine, if the age at administration is greater than or equal to 15 and the difference in dates between first and second vaccination were less than 1 month (30), then the vaccine record was excluded. |
|
|
| EventEncFirst | **Numeric**  0 = No 1 = Yes | First Primary Care Encounter Flag  Event flagging record at which patient had first primary care encounter in the cohort period |  |
| StatusRiskPrior | **Numeric**  1 = Surveillance 2 = Not Screen-Eligible  3 = Alternate Risk  4 = Unknown Risk  5 = Average Screen Risk | Risk Status at Event  Status flagging screening risk at record creation. Tests, procedures, and diagnoses that occur at record may lead to different status at end of record. | See [Section 3: Appendices - Cervical Timeline File Risk Status Assignment](#APPCTFRisk) (p. 71) for further detail. |
|
|
| StatusRisk | **Numeric**  1 = Surveillance 2 = Not Screen-Eligible  3 = Alternate Risk  4 = Unknown Risk  5 = Average Screen Risk | Risk Status at Event  Status flagging screening risk at end of record based on tests, procedures, and diagnoses that occur during record. | See [Section 3: Appendices - Cervical Timeline File Risk Status Assignment](#APPCTFRisk) (p. 71) for further detail. |
|
|
| StatusHIV | **Numeric**  0 = No 1 = Yes | HIV+ Status at Event  Status flagging HIV diagnosis at end of record. | Record was flagged as HIV-positive on and after record at which EventHIV = 1. |
|
|
| StatusPreg | **Numeric**  0 = No 1 = Yes | Pregnant at Event  Status flagging pregnancy at end of record. | Record was flagged as pregnant on and after record at which EventPregBegin = 1 and through all records up to and on record at which EventPregEnd = 1. |
|
|
| StatusNoCervix | **Numeric**  0 = No 1 = Yes | Cervix Not Present at Event  Status flagging whether cohort member has a cervix at end of record. | Record was flagged as no cervix on and after record at which EventNoCervix = 1. |
|
|
| StatusHPVVacDose | **Numeric**  0 = None/Unknown 1 = 1 2 = 2 3 = 3 4 = 4 or more | HPV Vaccination Dose Administered Status at Record Event  Status noting number of HPV vaccination doses administered HPV vaccination at end of record. | Record was flagged as HPV vaccine dosed and after record at which EventHPVVac = 1.  Vaccinations were counted as follows: if the initial vaccination occurred prior to age 15, then the next dose was counted as the second dose if it was administered at least 6 months later; if the initial vaccination occurred at age 15 or later, the next dose was counted as the second if it occurred at least 1 month later than the initial vaccine, and the next dose was counted as the third dose if it was administered at least 6 months after the second dose.  Immunocompromised status was not considered. |
|
|
| StatusCervCancer | **Numeric**  0 = No 1 = Yes | Invasive Cervical Cancer at Event  Status flagging invasive cervical cancer diagnosis at end of record | Record was flagged as cervical cancer on and after record at which EventCervCancer = 1. |
|
|
| StatusHPVPosPrior3 | **Numeric**  0 = No 1 = Yes | HPV positive Status 3 Years Prior to Event  Status flagging HPV diagnosis within prior 3 years of record. |  |
|
|
| StatusHPVPosPriorEver | **Numeric**  0 = No 1 = Yes | HPV positive Status Ever Prior to Event  Status flagging HPV diagnosis ever prior to record. |  |
|
|
| PapResultAbn | **Numeric**  0 = No 1 = Yes | Abnormal Pap Result at Event  Variable flagging abnormal (ASC-US worse cytology and/or HPV-positive) Pap test result at end of record. |  |
|
|
| StatusPapAbnprior3 | **Numeric**  0 = No 1 = Yes | Pap Abnormal Status 3 Years Prior to Event  Status flagging abnormal (ASC-US worse cytology and/or HPV-positive) Pap test result within prior 3 years of record. |  |
|
|
| StatusPapAbnpriorEver | **Numeric**  0 = No 1 = Yes | Pap Abnormal Status Ever Prior to Event  Status flagging abnormal (ASC-US worse cytology and/or HPV-positive) Pap test result ever prior to record. |  |
|
|
| StatusProcPrior3 | **Numeric**  0 = No 1 = Yes | Any Procedure Status 3 Years Prior to Event  Status flagging procedure within prior 3 years of record. Procedure includes any colposcopy/biopsy and treatment. |  |
|
|
| StatusProcPriorEver | **Numeric**  0 = No 1 = Yes | Any Procedure Status Ever Prior to Event  Status flagging procedure ever prior to record. Procedure includes any colposcopy/biopsy and treatment. |  |
|
|
| StatusColpoBxPrior3 | **Numeric**  0 = No 1 = Yes | Any Colpo Biopsy Status 3 Years Prior to Event  Status flagging colposcopy/biopsy, including endocervical curettage, within prior 3 years of record. |  |
|
|
| StatusColpoBxPriorEver | **Numeric**  0 = No 1 = Yes | Any Colpo Biopsy Status Ever Prior to Event  Status flagging colposcopy/biopsy, including endocervical curettage, ever prior to record. |  |
|
|
| StatusTxPrior3 | **Numeric**  0 = No 1 = Yes | Any Treatment Status 3 Years Prior to Event  Status flagging treatment including LEEP, cone, cryotherapy, laser, excision NOS, partial/subtotal/supracervical hysterectomy, trachelectomy, or total/radical hysterectomy, within prior 3 years of record. |  |
|
|
| StatusTxPriorEver | **Numeric**  0 = No 1 = Yes | Any Treatment Status Ever Prior to Event  Status flagging treatment including LEEP, cone, cryotherapy, laser, excision NOS, partial/subtotal/supracervical hysterectomy, trachelectomy, or total/radical hysterectomy, ever prior to record. |  |
|
|
| PapInd | **Numeric**  1 = Screening 2 = Repeat 3 = Diagnostic 4 = Surveillance  5 = Reflex to Primary HPV | Test Indication  Reflects the test indication for test record, regardless of whether the test was a Pap test and/or an HPV test  Records for which the Pap test indication was unknown set to ’-99999’  Records for which there was no Pap test or HPV test set to ‘-77777’  Records for which the Pap test was vaginal set to ‘95’ | See [Section 3: Appendices - Cervical Timeline File Pap/HPV Test Strategy](#APPCTFTest) (p. 72) for further detail |
|
|
| PapType | **Numeric**  1 = Cervical 2 = Vaginal 3 = Procedure Vaginal | Cervical or vaginal Pap test  Reflects whether the sample specimen derived from a cervical or vaginal Pap test  Records for which the Pap test type was unknown set to ’-99999’  Records for which there was no Pap test set to ‘-77777’ | See [Section 3: Appendices - Cervical Timeline File Pap/HPV Test Strategy](#APPCTFTest) (p. 72) for further detail |
|
|
| TestModality | **Numeric**  1 = Pap Alone (ASC-US Reflex)  2 = Co-Test  3 = Primary HPV (Reflex Cytology) | Test Modality  Reflects the type of Pap and/or HPV test(s) completed  Records for which the HPV test indication was unknown set to ’-99999’  If no Pap or HPV test was noted on this date (i.e., EventPap = 0 and EventHPV = 0), records set to ‘-77777’  Records for which the Pap test was vaginal set to ‘95’. | See [Section 3: Appendices - Cervical Timeline File Pap/HPV Test Strategy](#APPCTFTest) (p. 72) for further detail. |
|
|
|
| HPVGenotype | **Numeric**  1 = Reflex genotyping  2 = Concurrent genotyping  3 = General pooled assay | HPV Strain Genotypes  Reflects the HPV strains that may have been tested based on the manufacturer assay that was used and the year in which the assay was executed.  Reflex genotyping indicates that pooled genotyping inclusive of 16/18/45 was run first, and that if pooled genotyping assay was positive, then HPV 16/18/45 genotyping was run; this includes those pooled genotyping assays inclusive of 16/18/45 that were negative but would have had a 16/18/45 genotyping completed if the pooled test was positive.  Concurrent genotyping indicates that both a pooled hrHPV (which may or may not include 16/18/45) and HPV 16/18/45 genotyping were run at the same time.  General pooled assay indicates that only a pooled genotyping assay inclusive of 12-14 hrHPV strains was run, with no further genotyping for 16/18/45 available in the event of a positive test result.  Records for which the HPV test manufacturer was unknown set to ’-99999’.  Records for which there was no HPV test set to ‘-77777’.  Records for which site-submitted test manufacturer was ‘Other’ set to ‘95’. Site-specific nuances are described in the notes section. | See [Section 3: Appendices - Cervical Timeline File Pap/HPV Test Strategy](#APPCTFTest) (p. 72) for further detail |
|
|
| PapResult | **Numeric**  1 = NILM  2 = ASC-US  3 = LSIL  4 = HSIL  5 = AGC  6 = ASC-H  7 = Suspicious for Cancer  9 = Cytology Insufficient | Pap Test Result  Reports the most severe cytology result  Records for which the Pap test occurred but the result was unknown set to ’-99999’  Records for which there was no Pap test set to ‘-77777’ | This information can be used in combination with the HPVResult variable to describe the full test results observed from the same visit.  See [Section 3: Appendices - Cervical Timeline File Screening Result](#AppScrnRes) (p. 70) for further detail. |
| HPVResult | **Numeric**  1 = Genotype HPV 16/18/45+  2 = Genotype HPV 16/18/45- Pooled hrHPV+  3 = Genotype HPV 16/18/45- Pooled hrHPV-  4 = Pooled HPV+  5 = Pooled HPV-  9 = HPV insufficient | HPV Test Result  Reports most severe strain(s) for which the person tested positive to support analyses focused on reporting HPV prevalence  Note that Genotype HPV 16/18/45- includes those Pooled hrHPV- test results where a reflex genotyping assay was run, and the pooled hrHPV test result was negative  Records for which an HPV test occurred but the result was unknown set to ‘-99999’  Records for which there was no HPV test set to’-77777’ | This information can be used in combination with the PapResult variable to describe the full test results observed from the same visit.  See [Section 3: Appendices - Cervical Timeline File Screening Result](#AppScrnRes) (p. 70) for further detail. |
| PapHPVResult | **Numeric**  1 = NILM  2 = NILM HPV Unknown  3 = NILM HPV-  4 = NILM HPV+ (Pooled, Inclusive of 16/18/45)  5 = NILM HPV+ (Pooled, Exclusive of 16/18/45)  6 = NILM HPV+ (16/18/45)  7 = ASC-US  8 = ASC-US HPV unknown  9 = ASC-US HPV-  10 = ASC-US HPV+ (Pooled, Inclusive of 16/18/45)  11 = ASC-US HPV+ (Pooled, Exclusive of 16/18/45)  12 = ASC-US HPV+ (16/18/45)  13 = LSIL  14 = LSIL HPV Unknown  15 = LSIL HPV-  16 = LSIL HPV+ (Pooled, Inclusive of 16/18/45)  17 = LSIL HPV+ (Pooled, Exclusive of 16/18/45)  18 = LSIL HPV+ (16/18/45)  19 = HSIL or HSIL HPV- or HSIL HPV Unknown  20 = HSIL HPV+ (any)  21 = AGC or AGC HPV- or AGC HPV Unknown  22 = AGC HPV+ (any)  23 = ASC-H or ASC-H HPV- or ASC-H HPV Unknown  24 = ASC-H HPV+ (any)  25 = Suspicious for Cancer  26 = HPV-  27 = HPV+ (16/18) or HPV+ (16/18) with Unknown Pap  28 = HPV+ (Pooled)  29 = HPV + (Other)  30 = Insufficient (Pap and HPV combo that must be repeated) | Screen Result, with associated Pap test and HPV test results combined based on 2012 and 2019 management guidelines  The most severe Pap and HPV test results that occurred on the same day were reported.  Note that ASC-US HPV Unknown indicates that an HPV test was run, but the result was unknown, which is distinct from an insufficient HPV test result.  Note that ASC-US HPV+ (Pooled, Exclusive of 16/18/45) includes those reflex genotyping results that were positive for the pooled test and negative for the individual genotyping as well as those co-test results that were pooled positive (exclusive of 16/18/45) and negative for individual strains.  Records for which a Pap test and/or HPV test occurred but the result was unknown set to ’-99999’  Records for which there was no Pap or HPV test set to’-77777’ | See [Section 3: Appendices - Cervical Timeline File Screening Result](#AppScrnRes) (p. 70) for further detail. |
|
|
| PapProvIDPerform | **Character** | Pap-HPV Performing Provider  The provider that performed the Pap and/or HPV test  Records for which performing provider was unknown set to’-99999’  Records for which there was no Pap or HPV test set to’-77777’  Note that each provider has an ID that is unique within each site but that these IDs may be duplicated across sites; data users should combine the provider ID with site to obtain a unique identifier across sites |  |
|
|
| PapFacilityIDPerform | **Character** | Pap-HPV Performing Facility ID  The clinic/facility associated with the provider that performed the Pap and/or HPV test for test record  Records for which performing facility was unknown set to’-99999’  Records for which there was no Pap or HPV test set to’-77777’ |  |
|
|
| PapHPVFacilityPerformType | **Numeric**  1 = Primary Care  2 = Women’s Health/OB Gyn  3 = Both Primary Care and Women’s Health/OB Gyn  4 = HIV | Pap-HPV Performing facility type  Records for which performing facility type was unknown set to’-99999’  Records for which there was no Pap or HPV test set to’-77777’  Records for which performing facility type was a hospital set to ‘95’ |  |
|
|
|
|
| ProcType | **Numeric**  1 = Colpo, No Biopsy  2 = Colpo, Biopsy  3 = Endocervical Curettage (Brush)  5 = LEEP  6 = Cone  7 = Cryotherapy  8 = Laser  9 = Excisional Procedure, NOS  10 = Hysterectomy, NOS  11 = Partial/Subtotal/Supracervical  12 = Trachelectomy  13 = Total hysterectomy  14 = Radical hysterectomy or modified radical hysterectomy  15 = Biopsy, No Colpo | Procedure Type  If a pathology report from a cervical specimen with a result was available, but the procedure type was not specified, record set to ‘15 = Biopsy, No Colpo’.  Records for which there was no procedure set to’-77777’ |  |
|
|
| ProcResult | **Numeric**  20 = Insufficient for diagnosis/unsatisfactory tissue  21 = Normal/benign reaction/inflammation  22 = Atypical/atypia  23 = HPV/condylomata  24 = Low grade SIL  25 = CIN I/mild dysplasia  26 = CIN I-II  27 = CIN I/mild dysplasia, cannot rule out high grade dysplasia, detached fragments, cannot assess grade  28 = High grade SIL  29 = CIN II/moderate dysplasia  30 = CIN II-III  31 = CIN III/severe dysplasia/Carcinoma in situ (Stage 0)  32 = Adenocarcinoma In Situ of the cervix (AIS)  33 = Invasive Cervical Squamous Cell Carcinoma  34 = Invasive Cervical Adenocarcinoma  35 = Invasive Cervical Adenosquamous  36 = Other cervical cancer, including NOS  37 = Cancer of unclear origin  38 = Non-Cervical Cancer of the Cervix, No Other Information Available  40 = No biopsy | Procedure Result  Most severe pathology result noted for the procedure  Records for which the pathology result was unknown set to’-99999’  Records for which there was no procedure set to’-77777’ |  |
|
|
| ProcProvIDPerform | **Character** | Event Procedure Performing Provider  Records for which performing provider was unknown set to’-99999’  Records for which there was no procedure set to ’-77777’  Note that each provider has an ID that is unique within each site but that these IDs may be duplicated across sites; data users should combine the provider ID with site to obtain a unique identifier across sites |  |
|
|
| ProcFacilityIDPerform | **Character** | Event Procedure Performing Facility ID  The clinic/facility associated with the provider that performed the procedure |  |
|
|
| ProcFacilityPerformType | **Numeric**  1 = Primary Care  2 = Women’s Health/OB Gyn  3 = Both Primary Care and Women’s Health/OB Gyn  4 = HIV | Procedure performing facility type  Records for which performing facility type was unknown set to’-99999’  Records for which there was no procedure set to’-77777’  Records for which performing facility type was a hospital set to ‘95’ |  |
|
|
| extractDate | **Numeric** | Date each record was generated |  |
|
|
| providingSite | **Numeric** | PROSPR Site ID |  |
|
|
| ddVersion | **Numeric** | Data dictionary version number closest to submission date |  |
|
|

# Section 3: Appendices

## Abbreviations

**BGC**: Benign Glandular Cells

**BMI**: Body Mass Index

**CDE**: common data element

**CPT**: Current Procedural Terminology

**DRP**: data request packet

**DSR**: days since reference

**DUA**: data use agreement

**EHR**: electronic health record

**EMR**: electronic medical record

**HIV**: Human Immunodeficiency Virus

**HPV**: Human Papilloma Virus

**ICD**: International Statistical Classification of Diseases and Related Health Problems

**IMS**: Information Management Services

**LCDSS**: limited consolidated data-subset

**METRICS**: MultilEvel OpTimization of the CeRvIcal Cancer Screening Process in Diverse Settings & Populations

**NILM**: Negative for Intraepithelial Lesion or Malignancy

**PCC**: PROSPR Coordinating Center

**PCP**: primary care provider

**PRC**: PROSPR Research Center

**PROSPR II:** Population-based Research to Optimize the Screening Process

## Cohort Member Study Periods for Primary and Secondary Analyses

**Enrollment cohorts (Site A) and utilization cohorts (Site B and Site C) differ most fundamentally in the method by which a person engages with the healthcare system. An enrollment cohort member does not have to have any primary care utilization to remain in the cohort, but any utilization will be visible in the dataset so long as the cohort member remains enrolled in the healthcare plan. A utilization cohort member, by contrast, must have at least one primary care utilization in the healthcare system to enter the cohort, but may choose to utilize other healthcare systems that are not visible in the dataset. Utilization cohort members were initially only removed from the cohort due to aging out of eligibility, death, administrative cut-off at the end of the study period, or, at select sites, moving outside of the geographic service area. Thus, with the aforementioned cohort removal definition, it remained possible that a utilization cohort member only engaged with the healthcare system once and yet remained in the cohort until the end of the study, presenting a false study period and obscuring survival analyses.**

**To make the pattern of healthcare system engagement more analogous between the enrollment and utilization cohorts, utilization cohort members (see** [Section 1: Overview - Description of the METRICS Data Contributing Sites and Cohort Definitions](#CohortEntry) (p. 2) **for further detail) were administratively cut-off from the cohort after a 37-month lapse since the last primary care encounter. Utilization cohort members were permitted to re-enter the study upon subsequent primary care utilization or Pap/HPV test. An analyst is thus able to construct primary analyses that incorporate lack of primary care utilization as well as sensitivity analyses that do not incorporate lack of primary care utilization. This concept is conveyed in the following files and variables:**

**The** [Participant file](#_Participant_File) (p. 5)**, which conveys the cohort exit incorporating the lack of primary care utilization cut-off (CutOffDateYear, CutOffDateMonth, CutOffDateDay, CutOffLastDSR) and the cohort exit date that does not incorporate the lack of primary care utilization cut-off (**CutOffDateSensYear, CutOffDateSensDay, CutOffDateSensMonth, and CutOffSensDSR**). These concepts were introduced to permit sensitivity analyses beyond the primary analyses, which incorporate the lack of primary care utilization cut-off, and present cohort periods structured similarly to other utilization cohorts throughout the PROSPR II consortium. For the enrollment-based cohort, all variables related to cohort exit will be equivalent (e.g., CutOffLastDSR will be the same as CutOffSensDSR).**

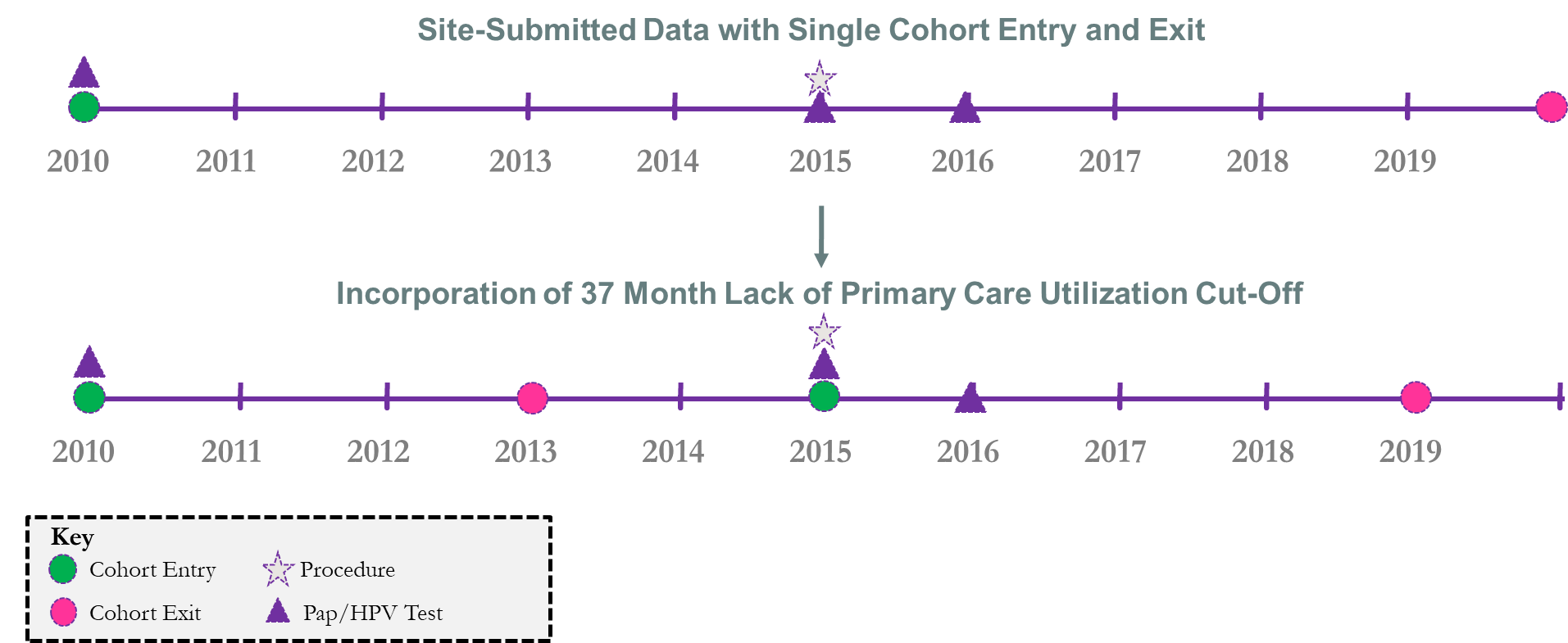
**The** [Calendar Year file](#_Calendar_Year) (p. 9)**, which conveys whether the calendar year record occurred during a year in which the cohort member was in the study based on utilization (**UtilCalYr\_Drv)**. This concept was introduced to retain site-submitted data and to permit an analyst to exclude Calendar Year data if executing primary analyses based on lack of primary care utilization and so only applies to utilization cohort members.**

**The** [Engagement file](#_Enrollment_File) (p. 21)**, which conveys the cohort exit date (CutOffDateMonth, CutOffDateDay, CutOffDateYear, and CutOffLastDSR, where applicable), and reason for cohort exit (CutOffReason, where applicable) as well as additional cohort periods (where applicable).**

**The** [Cervical Timeline file](#_Cervical_Timeline_File_1) (p. 52)**, which conveys every cohort entry and exit noted in the Engagement file (EventCE and EventCEnd).**

**For primary analyses, utilization cohort entry and exit dates can be identified using the Participant (CohortEntryFirstDSR, CutOffLastDSR), Engagement (CohortEntryDSR, CutOffLastDSR), and Cervical Timeline (EventCE and EventCEnd) files. Records for calendar years in which the cohort member was in the study based on utilization can be identified (UtilcalYr\_Drv = 1). Only the first cohort entry and last cohort are noted in the Participant file, while the Engagement and Cervical Timeline files report all cohort entries and exits.**

**For secondary analyses, utilization-based cohort members’ first cohort entry and last cohort exit dates can be identified using the Participant file (CohortEntryFirstDSR, CutOffSensDSR). Intermittent cohort entries and exits can be identified in the Engagement file by looking for all cohort exits and subsequent cohort entries that have a cut off reason that corresponds to death, aging out, or moving out of the geographic service area. See graphic below for an example of how a utilization-based cohort member’s study period shifts with the introduction of a lack of primary care utilization cut-off.**



In this example, a utilization cohort member’s timeline is presented as originally submitted by the site (top timeline) and as transformed by the introduction of the lack of primary care utilization cut-off (bottom timeline). The site-submitted timeline shows that the cohort member entered the study at a primary care utilization in 2010, had another primary care utilization in 2015 and 2016, and then left the cohort at the end of 2020 due to administrative cut-off at the end of the study period (December 31st, 2020). For sensitivity analyses, the cohort entry and exit dates can be identified using the Participant file (CohortEntryFirstDSR, CutOffSensDSR). With the introduction of the lack of primary care utilization cut-off, this same cohort member now has two cohort periods: the first cohort period begins in 2010 at the first primary care utilization, then exits the cohort in 2013 after 37 months of no primary care utilization; the second period begins in 2015 at a subsequent primary care utilization, then exits the cohort in 2020 after 37 months of no primary care utilization following the last utilization in 2016. For primary analyses, the cohort entry and exit dates can be identified using the **Participant (CohortEntryFirstDSR, CutOffLastDSR), Engagement (CohortEntryDSR, CutOffLastDSR), and Cervical Timeline (EventCE and EventCEnd) files; the Calendar Year file can then be used to identify calendar years in which the cohort member was in the study based on utilization (UtilcalYr\_Drv = 1).**

## HPV Vaccination Schedule

HPV vaccination is approved for children and adults ages 9 through 26 years; adults ages 27 through 45 years may also receive the vaccine at clinician discretion. The following three HPV formulations have been approved for use in the United States:

**Gardasil-4 (4vHPV, Merck & Co.)** is a quadrivalent virus-like particle vaccine directed against HPV strains 16, 18, 6, and 11. This vaccine was approved by the FDA in 2006 for use in both females and males.

**Cervarix (2vHPV, GlaxoSmithKline)** is a bivalent virus-like particle vaccine directed against HPV strains 16 and 18. This vaccine was approved by the FDA in 2009 for use in females.

**Gardasil-9 (9vHPV, Merck & Co.)** is a nonavalent virus-like particle vaccine directed against HPV strains 16, 18, 6, 11, 31, 33, 45, 52, and 58. This vaccine was approved by the FDA in 2014 for use in both females and males. As of the end of 2016, Gardasil-9 is the only vaccine used for routine vaccination in the United States.

As of late 2016, a person is to receive two doses of the HPV vaccine, with the second dose occurring within 6-12 months of the first dose. Prior to late 2016, vaccine dosage and schedule was determined by the age at first vaccination. If the first dose was received before the person’s 15th birthday, the second dose was to occur 6-12 months after the first dose. If the first dose was received on or after the person’s 15th birthday, the second dose was to occur 1-2 months after the first dose, and the third dose was to occur six months after the first dose. Repeated vaccinations are not required following vaccination schedule interruptions. Gardasil-9 can be used to complete vaccination schedules initiated with Gardasil-4 or Cervarix.

## Facilities Harmonization

The METRICS sites sought to harmonize the level at which facilities were defined across the three disparate healthcare systems. The goals for harmonizing the Facility file were (1) to understand at which level organizational policies were established for delivery of screening (clinical co-investigator input and qualitative data documented that this varies across sites); and (2) to enable conduct of analyses, including nested multi-level analyses, at a consistent level across sites. Facilities were defined at three different levels across the healthcare systems.

Level 1 represents the unique data-submitting site. There are three level-1 facilities (i.e., data-contributing sites) in the dataset. This is the most harmonized level in the dataset, but it provides the least granularity for most analyses and is most appropriate for stratifying data. This can be identified through the providingSite variable for all sites across all files.

Level 2 represents either the network, medical center, or building at which an event occurred and is based on distinct geographic entities or separate network relevant to each data-contributing site. This the most conceptually heterogeneous level in the dataset. This can be identified by site as follows:

* Site A – distinct medical center or building identifiers for which the facility is a primary care, specialty care, urgent care, or care clinic [all non-missing FacilityID where FacilityIDRelatedAggr (Facility file) is set to 1 or 3]; missing values indicate that the facility could not be identified
* Site B – distinct medical center or building identifiers [FacilityIDRelatedPhys (Facility file)]; missing values indicate that the facility could not be identified
* Site C – distinct medical center or building identifiers [FacilityIDRelatedPhys (Facility file)]; missing values indicate that the facility is independent and not affiliated with a level 1 network (FacilityIDRelatedPhys)

Level 3 represents the represents the clinic, department, or specialty of the facility at which an event occurred and is based on distinct entities at which operational policies are thought to be implemented for 2/3 sites. This can be identified by site as follows:

* Site A – unique combinations of non-missing medical center or building identifiers and non-missing facility types for which the facility is a primary care, specialty care, urgent care, or care clinic [all non-missing FacilityID where FacilityIDRelatedAggr (Facility file) is set to 1 or 3 and non-missing FacilityPerformType (Pap Test, HPV Test, Procedures, or Encounters files)]; missing values indicate that the facility type could not be identified
* Site B – distinct clinic/specialty identifiers [all non-missing FacilityID]; missing values indicate that the facility could not be identified
* Site C – distinct clinic/specialty identifiers for which there is a network/medical center identifier [all non-missing FacilityID with non-missing FacilityIDRelatedPhys (Facility file)]; missing values indicate that the facility could not be identified

## Cervical Timeline File

### Data Structure

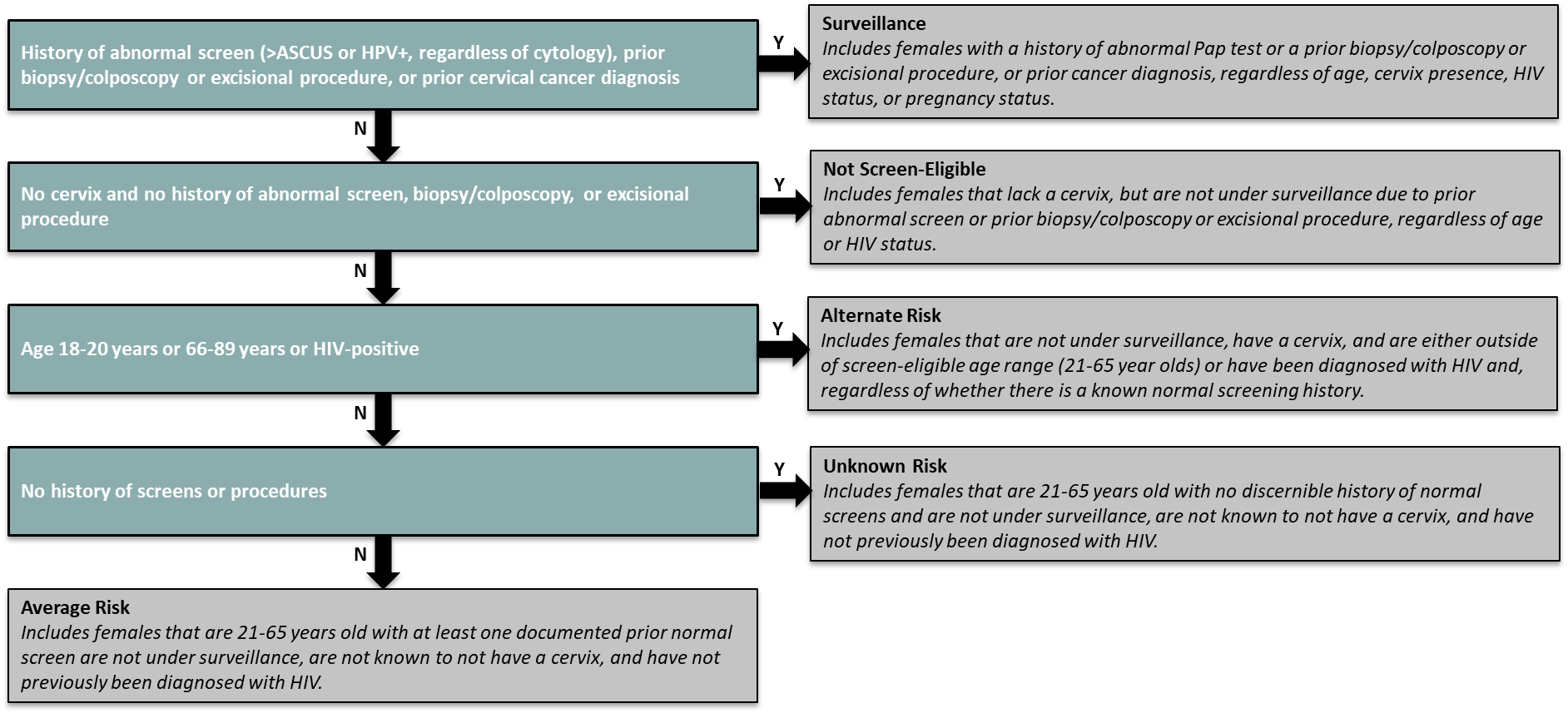
The Cervical Timeline file consolidates screening measures across multiple files into one modular time-series that reflects the entirety of the cervical cancer screening process for each cohort member.

Each record in this file represents at least one event occurring on a distinct date; all events that occur on the same date will be noted on the same record. Events include cohort entry or exit, an eligibility status change, a Pap and/or HPV test, and/or a procedure.

Each event record also contains information on the date of the event, cohort member eligibility status, and any available screen or procedure information (where applicable). Cohort member eligibility status is reported most directly by risk status, which is defined at the beginning and conclusion of the event, because the event itself can lead to eligibility changes. Other variables that report various aspects of cohort member eligibility status include pregnancy, HIV, cervical cancer, absent cervix, and abnormal test or procedure results.

Note that the event date for tests indicates the date that the test was ordered, which is synonymous with the visit at which the test was ordered across METRICS sites. Additionally, to avoid overcounting tests, all Pap tests and HPV tests that occurred within a 14-day window were grouped together into one event record. If multiple Pap tests occurred within a 14-day window, the date from the first Pap test and the last cytology result noted among the multiple Pap tests were used. If multiple HPV tests occurred within a 14-day window, the most severe result was used, and the date of the associated Pap test was used (if the HPV test occurred on a later date). Lastly, the most severe procedure that was completed on a given date was noted, along with the most severe pathology result, regardless of whether the most severe pathology yielded the worst result.

### Risk Status Assignment

Risk status reflects a person’s eligibility for screening and conveys the anticipated test type and frequency as well as the type of follow-up recommended based on the test outcome. As described in the diagram below, risk status was algorithmically assigned based on a cohort member’s prior screening and procedure history, age, cervix presence, prior HIV diagnosis. Risk status was determined at the start of an event (i.e., risk status established prior to an event, [StatusRiskPrior](#CTFStatRiskPri) (p. 56)) and at the event of an event (i.e., risk status established as a result of the completion of the event, [StatusRisk](#CTFStatRisk) (p. 56)). ****

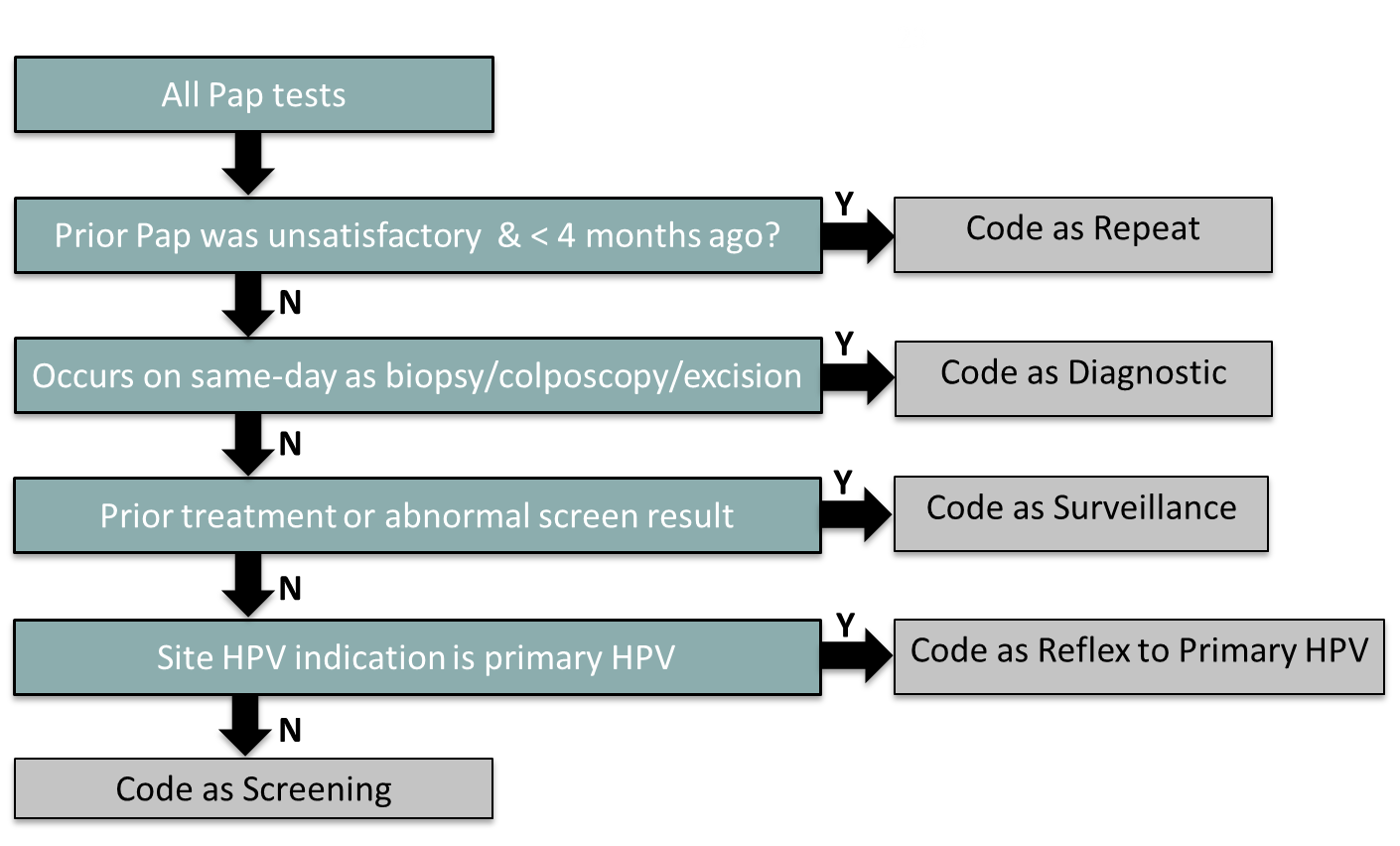
Once a cohort member entered the ‘Surveillance’ risk status, they remained under this status for the duration of the cohort period. Likewise, a cohort member that entered the ‘Not Screen-Eligible’ risk status could only transition to the ‘Surveillance’ risk status. A cohort member that entered the ‘Alternate Risk’ status due to being under age 21 years old may transition to any of the other statuses; however, if a cohort member entered the ‘Alternate Risk’ status due to being over age 65 or HIV diagnosis, they only transitioned to the ‘Not Screen-Eligible’ or ‘Surveillance’ risk statuses. A cohort member that entered the ‘Unknown Risk’ status could transition to any of the other statuses. A cohort member that entered the ‘Average Risk’ status could transition to the ‘Surveillance’, ‘Not Screen-Eligible’, or ‘Alternate Risk’ statuses, as the documented normal screen required to enter the ‘Average Risk’ status precluded ‘Unknown Risk’ status assignment.

### Pap/HPV Test Strategy

The indication for a Pap test and/or an HPV test as well as the HPV strains that were tested can be used to describe the overall testing strategy. The following four variables were used to inform the test indication (PapInd), testing modality (TestModality), and HPV strains that were genotyped (HPVGenotype).

Pap test indication ([PapInd](#CTFPapInd) (p. 58)) refers to the rationale for testing, regardless of whether the test was a primary Pap or HPV test or a co-test. These data were identified algorithmically as follows:

1. **Repeat** – Test occurred within four months of an unsatisfactory screen; if not, then
2. **Diagnostic** – Test occurred concomitant with a procedure to diagnose observed symptoms or confirm prior abnormal result before procedure; if not, then
3. **Surveillance** – Test occurred after a prior history of abnormality; if not, then
4. **Reflex to Primary HPV –** Test occurred after a primary HPV test with a positive test result; if not, then
5. **Screening** – Test occurred as a part of routine screening with no indication of prior or current history of abnormality.



HPV test indication refers to the timing of an HPV test relative to a Pap test. These data were submitted by the sites as follows:

1. **Primary HPV screening** – an HPV test occurred without a Pap test noted within 14 days;
2. **Pap alone** – a Pap test occurred without an HPV test observed within 14 days; or
3. **Reflex or Co**-**test** – an HPV test occurred within 14 days of a Pap test, then site-submitted HPV test indication was used to assign the ‘Pap alone/reflex’ or ‘Co-test’ modality.

Test modality ([TestModality](#CTFPapHPVModAlg) (p. 59)) was then determined based on HPV test indication as well as the time between a Pap test and an HPV test and whether the test was a cervical or vaginal Pap test. Note that Pap alone and reflex were ultimately grouped together, as the decision of whether to run an HPV test is based on the result of the cytology and thus does not represent a distinct clinical decision.

1. **Pap alone/reflex** – if a Pap test is reported and either no HPV test is noted within a 14-day window or if an HPV test occurred and the HPV test indication was reported as reflex to Pap test
2. **Co-Test** – if both a Pap and an HPV test were reporting within a 14-day window and the HPV test indication was reported as co-test
3. **Primary HPV** – if an HPV test is reported and no Pap test is noted within a 14-day window and primary HPV testing was known to be available at the site at the time of the HPV test
4. **Other** – includes vaginal Pap tests and tests that did not meet site-specific criteria for reporting as other modalities
5. **Unknown –** includes tests for which an HPV test occurred within a 14-day window of a Pap test but not HPV test indication was not documented

Lastly, HPV Genotype ([HPVGenotype](#CTFHPVGen) (p. 59)) indicates the strains of HPV that were tested and the order in which they were tested. These data were identified based on the HPV test manufacturer, calendar year, and site, as follows:

1. **Reflex genotyping** – a pooled genotyping reaction inclusive of 16/18/45 was run first, and if positive, then HPV 16/18/45 individual genotyping was run
2. **Concurrent genotyping** – both a pooled hrHPV (which may or may not include 16/18/45) and HPV 16/18/45 individual genotyping were run at the same time
3. **General pooled assay** – only a pooled genotyping reaction inclusive of 12-14 high-risk HPV strains was run, with no further HPV 16/18/45 individual genotyping completed

### Test Results

All results from tests ordered at a visit will be noted on the same record. Test results are reported in two formats as follows: the most severe cytology result ([PapResult](#CTFPapRes) (p. 60)) and/or most severe HPV test result ([HPVResult](#CTFHPVRes) (p. 60)); and the combined Pap and HPV results ([PapHPVResult](#CTFPapHPVRes) (p. 60)), which were collapsed to reflect both 2012 and 2019 management guidelines.

For the combined result variable, the strain of HPV was noted based on the Pap test cytology results. HPV strain was only noted for normal cytology results (negative for intraepithelial lesion or malignancy [NILM]), low-grade cytology results (atypical squamous cells, undetermined significant [ASC-US] or low-grade squamous intraepithelial lesions [LSIL]), primary HPV tests, and when the HPV test was positive for the 16/18 strain and the cytology result was insufficient or unknown. Otherwise, HPV status was only noted as either positive or negative/unknown/not tested for cytology results as or more severe than LSIL.

HPV result was assigned based on decreasing order of severity. An HPV test was 16/18/45-positive if the test result was positive for individual genotyping for strains 16, 18, and/or 45. Otherwise, an HPV test was 16/18/45- and pooled high-risk HPV+ if a concurrent or reflex genotyping assay had negative individual genotyping and a positive pooled HPV test, regardless of whether the pooled HPV test was inclusive or exclusive of 16/18/45. Otherwise, an HPV test was 16/18/45- and pooled high-risk HPV- if a concurrent assay had negative individual genotyping and a negative pooled high-risk HPV test, as well as those reflex genotyping tests in which the pooled high-risk HPV test was negative and so reflex genotyping was not run. Otherwise, an HPV test was pooled HPV-positive or HPV-negative if only a pooled HPV test was completed. An HPV test was reported as unknown or insufficient as follows: if all HPV tests run were unknown/insufficient; if a 16/18/45 genotyping assay result was unknown/insufficient either following a positive pooled high-risk HPV test or along with a negative concurrent pooled high-risk HPV test that was exclusive of 16/18/45; or if a 16/18/45 genotyping assay result was negative and a pooled high-risk HPV test result was unknown/insufficient.