

PROSPR II METRICS DATA DICTIONARY [VERSION 2.2]

OVERALL INSTRUCTIONS

The **Multilevel Optimization of the Cervical Cancer Screening 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** – an integrated county tax-supported public safety-net healthcare system; a not-for-profit integrated delivery system primarily represented by two sites; and a mixed-model healthcare system providing coverage to enrolled members in facilities that are owned and operated by the healthcare system and through providers, medical groups, and hospitals that are contracted to provide health care **services. Sites collected 10 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 2010-2019) 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 data harmonization nuances and context.

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; 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 Timeline and Screening Episode files.

SECTION 1: OVERVIEW – DESCRIPTION OF THE METRICS DATA CONTRIBUTING SITES, COHORT DEFINITIONS, DATA FILE LIST AND FLOW SCHEMATIC

METRICS DATA CONTRIBUTING SITES AND COHORT DEFINITIONS

Definition	Site A	Site B	Site C
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	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 on or after Jan. 1, 2010 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; resides in cancer registry catchment area.	Patients entered the study cohort on the date of their first visit to a system primary care clinic on or after January 1, 2010 if patient was of the female sex, age 18-89, and were county residents. Cohort members may have multiple cohort period if the cohort member moved out of the count for at least 6 months and then subsequently moved back into the county.	Patients entered the study cohort on the date of their first visit to a qualifying primary care clinic if patient was of the female sex and age 18-89.
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, 2019).	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, 2019). 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 may have multiple cohort periods in two select circumstances: first, if the cohort member moved out of the county for at least 6 months and then moved back into the county; and second, if the cohort member had a lapse in primary care utilization of 37 months and then had a subsequent primary care utilization. See Appendix: Cohort Exit due to Lack of Primary Care Utilization (p. 121) for further detail.	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 (i.e., December 31, 2019). Cohort members may have multiple cohort periods if there was a lapse in primary care utilization of 37 months and then a subsequent primary care utilization. See Appendix: Cohort Exit due to Lack of Primary Care Utilization (p. 121) for further detail.

DATA FILE LIST AND FLOW SCHEMATIC

File	File Name	Limited Consolidated Data Set (LCDS)	Limited Consolidated Data Subset (LCDSS) ^A
Participant	crv_ppt_YYYYMMDD	✓	✓
Calendar Year	crv_calyr_YYYYMMDD	✓	✓
Engagement	crv_enroll_YYYYMMDD	✓	✓
Pap Test	crv_paptest_YYYYMMDD	✓	
HPV Test	crv_hpvttest_YYYYMMDD	✓	
Procedures	crv_proc_YYYYMMDD	✓	
Cancer Registry	crv_cancer_YYYYMMDD	✓	✓
Prior	crv_prior_YYYYMMDD	✓	
HPV Vaccine	crv_hpvvac_YYYYMMDD	✓	
Encounter	crv_enc_YYYYMMDD	✓	✓
Pregnancy	crv_preg_YYYYMMDD	✓	
Provider	crv_prov_YYYYMMDD	✓	✓
Facility	crv_facil_YYYYMMDD	✓	✓
Social Determinants of Health	crv_sdoH_YYYYMMDD	✓	✓
Screening Episode (analytic file)	crv_screp_YYYYMMDD	✓	✓
Cervical Timeline File (analytic file)	crv_cervtimeline_YYYYMMDD	✓	✓

^AOnly a subset of variables in each of these files will be available due to the existing Data Use Agreements.

Limited Consolidated Data Subset (LCDSS): This pooled dataset includes the analytical Screening Episode and Cervical Timeline files and a subset of CDEs from eight CDE files (Participant, Calendar Year, Engagement, Cancer Registry, Encounter, Provider, Facility, and Social Determinants of Health). The NCI contractor received the LCDSS and will generate a subset of the LCDSS for external collaborators who have received approval from the PROSPR II Consortium's Data Request Review Committee.

SECTION 2: FILE DESCRIPTION – OVERVIEW OF ALL DATA FILES AND CDES

GLOBAL VALUES

Superscripts

¹ indicates derived variables, which were calculated by the PRC using the submitted CDEs.

² indicates variables provided to the NCI contractor.

Global Values

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

-99999 indicates missing or unknown quantitative variables 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.

Global Notes

Variable Category	Files/Variables Affected	Variable Description and Notes	DAU Harmonization Notes
PID	Participant, Calendar Year, Pap Test, HPV Test, Procedures, Prior, HPV Vaccine, Engagement, Encounter, Pregnancy, Cancer Registry	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. Cohort members included in multiple PROSPR organ cohorts (i.e. COLON, CERVICAL, BREAST, LUNG) have the same PROSPR ID.	PROSPR I PID is not necessarily the same as PROSPR II PID for all sites. Recommend using PROSPR II PIDs for analysis.
Dates	Participant CohortEntryDate Participant CutOffDate		Each date was parsed into its month, day, and year (e.g., CohortEntryDate was split into

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Category	Files/Variables Affected	Variable Description and Notes	DAU Harmonization Notes
	Participant CutOffDateSens Calendar Year WeightDate Calendar Year SmokeDate Pap Test PapDate HPV Test HPVDate HPV Vaccine HPVVacDate HPV Vaccine HPVVacRefDate Prior PCPVisitLatestDate Prior PCPVisitFirst3Date Prior CaCervDate Prior MedRec1stDate Encounter PCPVisitDate Engagement CohortEntryDate Engagement CutOffDate Pregnancy PregBeginDate Pregnancy PregEndDate Procedure ProcDate Cancer Registry DxDate Cancer Registry TxCalniDate		CohortEntryDateDay, CohortEntryDateMonth and CohortEntryDateYear). Recommend not using dates because of data restriction. Recommend using DSR in place of date to accurately date record. The LCDS contains both dates and DSR. Per the DUA, the LCDSS transmitted to the NCI contractor does not include dates (HIPAA precaution).
DSR (Days Since Reference)	Participant CohortEntryFirstDSR Participant CutOffLastDSR Participant CutOffSensDSR Calendar Year WeightDSR Calendar Year SmokeDSR PapTest PapDSR HPV Test HPVColIDSR HPV Vaccine HPVVacDSR HPV Vaccine HPVVacRefDSR Prior PCPVisitLatestDSR Prior PCPVisitFirst3DSR Prior CaCervDSR Prior MedRec1stDSR Encounter PCPVisitDSR Engagement CohortEntryDSR Engagement CutOffDSR Pregnancy PregBeginDSR Pregnancy PregEndDSR Procedure ProcDSR Cancer Registry DxDSR Cancer Registry TxCalniDSR	All DSR variables use birth date as the reference date.	Recommend using DSR version of date variable.
extractDate	Participant, Calendar Year, Pap Test, HPV Test, Procedures, Prior, HPV Vaccine,	Date each record was generated	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Category	Files/Variables Affected	Variable Description and Notes	DAU Harmonization Notes
	Engagement, Encounter, Pregnancy, Cancer Registry, Provider, Facility		
ProvidingSite	Participant, Calendar Year, Pap Test, HPV Test, Procedures, Prior, HPV Vaccine, Engagement, Encounter, Pregnancy, Cancer Registry, Provider, Facility, Cervical Timeline	The Site ID associated with the site providing data. There may be more than one Site ID associated with each PRC.	ProvidingSite was set as the text prefix to the PID number to generated PIDSite. The providingSite in the Participant file or Cervical Timeline file should be used in analyses to identify the site at which a cohort member entered the cohort.
PIDSite¹	Participant, Calendar Year, Pap Test, HPV Test, Procedures, Prior, HPV Vaccine, Engagement, Encounter, Pregnancy, Cancer Registry, Provider, Facility, Cervical Timeline	PROSPR Site ID Unique derived PROSPR ID METRICS PID combined with ProvidingSite at which participant entered the cohort	
ddVersion	Participant, Calendar Year, Pap Test, HPV Test, Procedures, Prior, HPV Vaccine, Engagement, Encounter, Pregnancy, Cancer Registry, Provider, Facility, Cervical Timeline	Data dictionary version number closest to submission date	

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 (see [METRICS Data Contributing Sites and Cohort Definitions: Cohort Entry](#) (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 then re-entered the study upon subsequent primary care utilization. See [Appendix: Cohort Member Study Periods for Primary and Secondary Analyses](#) (p. 121) for further detail. New Participant file variables (CutOffDateSensYear, CutOffDateSensDay, CutOffDateSensMonth, and CutOffSensDSR) provide the last date upon which a utilization cohort member exited the cohort due to either aging out, death, moving out of the geographic service area, or administrative cut-off due to either the end of the cohort period end (see [Engagement file: General Harmonization Notes](#) (p. 26) for further detail). These new 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., CutOffDSR 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 have to be identified in the Engagement file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
BirthMth	Numeric	Birth date, month An integer between 1-12.	
BirthDy	Numeric	Birth date, day An integer between 1-31	
BirthYr ²	Numeric	Birth date, year A 4 digit integer	
Hispanic ²	Numeric 0=No 1=Yes	Hispanic or Latino origin	
RaceWhite ²	Numeric 0=No 1=Yes	Race White	
RaceBlack ²	Numeric 0=No 1=Yes	Race Black or African-American	
RaceAsian ²	Numeric 0=No 1=Yes	Race Asian	
RaceAIAN ²	Numeric 0=No 1=Yes	Race American Indian or Alaska Native	
RacePI ²	Numeric 0=No 1=Yes	Race Native Hawaiian or Other Pacific Islander	
RaceMultipleNOS ²	Numeric 0=No 1=Yes	Race, Multiple (individual races not specified)	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
RaceOther ²	Numeric 0=No 1=Yes	Race Other	
HeightInch	Numeric	Height, inches Height closest to cohort entry date and prior to cut off date recorded, where the cohort member is >= 18 years of age. If the height was less than 48 inches or greater than 90 inches, recorded the next closest height to cohort entry date that fell between 48-90 inches. If no height was available in this range, recorded the original height closest to cohort entry date, and rounded to the nearest whole number.	Derived with Weight (WeightPound) variable from Calendar Year file as <u>BMI_Drv</u> (p. 24) variable in the Calendar Year file.
CohortEntryMonth	Numeric	Month of earliest cohort entry date See <u>Section 1: Overview: Cohort Entry</u> (p. 2) for cohort entry definition for each site.	See <u>Global Values: Dates</u> (p. 4) for further detail.
CohortEntryDay	Numeric	Day of earliest cohort entry date See <u>Section 1: Overview: Cohort Entry</u> (p. 2) for cohort entry definition for each site.	See <u>Global Values: Dates</u> (p. 4) for further detail.
CohortEntryYear	Numeric	Year of earliest cohort entry date See <u>Section 1: Overview: Cohort Entry</u> (p. 2) for cohort entry definition for each site.	See <u>Global Values: Dates</u> (p. 4) for further detail.
CohortEntryFirstDSR ^{1,2}	Numeric	Earliest cohort entry date, days since reference date	See <u>Global Values: DSR</u> (p. 5) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		See Section 1: Overview: Cohort Entry (p. 2) for cohort entry definition for each site.	
CutOffDateMonth	Numeric	Month of last cohort cut off date See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffDateDay	Numeric	Day of last cohort cut off date See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffDateYear	Numeric	Year of last cohort cut off date See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffLastDSR ^{1,2}	Numeric	Last cohort cut-off date, days since reference date See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: DSR (p. 5) for further detail.
CutOffDateSensMonth	Numeric	Month of last cohort cut off date for sensitivity analyses (does not include lack of primary care utilization cut-off) See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffDateSensDay	Numeric	Day of last cohort cut off date for sensitivity analyses (does not include lack of primary care utilization cut-off) See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
CutOffDateSensYear	Numeric	Year of last cohort cut off date for sensitivity analyses (does not include lack of primary care utilization cut-off) See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffSensDSR ^{1,2}	Numeric	Last cohort cut-off date, days since reference date for sensitivity analyses (does not include lack of primary care utilization cut-off) See Section 1: Overview: Cohort Exit (p. 2) for cohort exit definition for each site.	See Global Values: DSR (p. 5) for further detail.
RaceEth_Drv ^{1,2}	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.	
Age_17_Ind_Drv ¹	Numeric 0= No 1= Yes	Indicates age at cohort entry was 17 Select cohort members entered the cohort during the month but prior to their 18 th birthday.	Cohort entry was shifted forward in time to the cohort member's 18 th birthday in the Cervical Timeline File (p. 91).
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

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 last encounter within the calendar year, except where otherwise noted.

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. A new 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.

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 [METRICS Data Contributing Sites and Cohort Definitions: Cohort Entry](#) (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 then re-entered the study upon subsequent primary care utilization. See [Appendix: Cohort Member Study Periods for Primary and Secondary Analyses](#) (p. 121) for further detail. A new 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 [Engagement file: General Harmonization Notes](#) (p. 26) 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.

Ascertainment of Comorbidity Status via ICD-9 and ICD-10 Codes: Each site provided prevalent comorbidity data using inpatient and/or outpatient billing codes and/or problems lists at any point from the respective calendar year, regardless of cohort duration during the calendar year. The ICD-9 and ICD-10 codes used to flag individual comorbidities were identical to those presented in Quan *et al.*, 2005, except where deviations were noted in the variable description. Of note, HIV status, which is of particular relevance to cervical cancer screening guidelines for individuals at elevated risk, was identified across the sites using different data sources. It is recommended that the Cervical Timeline File variables EventHIV be used to record the earliest documented record of an HIV diagnosis and that StatusHIV be used to report whether a patient was HIV-positive at the time of event.

TABLE 1. ICD-9-CM and ICD-10 Coding Algorithms for Charlson Comorbidities			
Comorbidity	Charlson Weight Value	Comorbidity	Charlson Weight Value
Myocardial infarction	1	Diabetes without chronic complication	1
Congestive heart failure	1	Diabetes with chronic complication	2
Peripheral vascular disease	1	Hemiplegia or paraplegia	2
Cerebrovascular disease	1	Renal disease	2
Dementia	1	Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin	2
Chronic pulmonary disease	1	Moderate or severe liver disease	3

PROSPR METRICS DATA DICTIONARY [Version 2.2]

TABLE 1. ICD-9-CM and ICD-10 Coding Algorithms for Charlson Comorbidities			
Comorbidity	Charlson Weight Value	Comorbidity	Charlson Weight Value
Rheumatic disease	1	Metastatic solid tumor	6
Peptic ulcer disease	1	AIDS/HIV	6
Mild liver disease	1	Diabetes without chronic complication	

*Quan H, Sundararajan V, et al. Coding algorithms for defining Comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care.* 2005 Nov; 43(11): 1130-9

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
CalendarYr ²	Numeric	Calendar year	
WeightPound	Numeric	Weight, pounds Recorded 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 and note date value	Derived with Height (HeightInch) variable from Participant file as BMI_Drv (p. 24) variable in the Calendar Year file.
WeightDateMonth ²	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 and note date value	See Global Values: Dates (p. 4) for further detail.
WeightDateDay	Numeric	Day 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 and note date value	See Global Values: Dates (p. 4) for further detail.
WeightDateYear ²	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 and note date value	See Global Values: Dates (p. 4) for further detail.
WeightDSR ^{1,2}	Numeric	Date weight recorded, days since reference date	See Global Values: DSR (p. 5) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		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 and note date value	
PPTStFIPRes	Character	<p>Cohort member's state of residence, FIPS code</p> <p>See note in <u>Calendar Year File Record Structure</u> (p. 13) regarding multiple values in a given year.</p> <p>Entered valid FIPS code for the state, preserving leading 0s.</p>	
PPTCountyFIPRes	Character	<p>Cohort member's county of residence, FIPS code</p> <p>See note in <u>Calendar Year File Record Structure</u> (p. 13) regarding multiple values in a given year.</p> <p>Entered a valid FIPS code for the county, preserving leading 0s.</p>	
PPTZipRes	Character	<p>Cohort member's ZIP code of residence</p> <p>Enter a valid 5-digit numeric zip code and include leading 0s.</p> <p>For cohort entry year, used first value that occurred on or after cohort entry date that was within the same calendar year as cohort entry; if no value was available during the calendar year, used the closest prior recorded value. For subsequent calendar years used value closest to the beginning of year within the calendar year; if no value was available during the calendar year, used the closest prior recorded value.</p>	
ScreeningIntervention	Numeric		

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	0=No 1=Yes	Was the cohort member enrolled in an intervention that might affect screening behavior?	
Medicare ²	Numeric 0=No 1=Yes	Was the cohort member covered by Medicare?	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
Medicaid ²	Numeric 0=No 1=Yes	Was the cohort member covered by Medicaid?	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
InsOtherGov ²	Numeric 0=No 1=Yes	Was the cohort member covered by any other federal or state health insurance program?	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
InsCommerc ²	Numeric 0=No 1=Yes	Was the cohort member covered by commercial and/or private health insurance?	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
MedicalAssist ²	Numeric 0=No 1=Yes	Was the cohort member enrolled in a medical assistance charity program for the uninsured?	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
Uninsured ²	Numeric 0=No 1=Yes	Was the cohort member uninsured?	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
InsOther ²	Character	Health insurance coverage, other	Recommend using <u>CalYr_Ins_Drv</u> (p. 24) 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 <u>CalYr_Ins_Drv</u> (p. 24) variable to identify insurance during the calendar year.
ProviderIDPCP	Character	Primary care provider ID	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		<p>See note in Calendar Year File Record Structure (p.13) regarding multiple values in a given year.</p> <p>Indicates the primary care provider associated with the cohort member at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.</p>	
FacilityIDPCP	Character	<p>Primary care provider clinic ID</p> <p>Indicates the facility at which the primary care provider associated with the cohort member practiced at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.</p>	
ComorbMyocardial	Numeric 0=No 1=Yes	<p>Did the cohort member have a diagnosis code for myocardial infarction during the specified year or a prior calendar year within cohort window?</p> <p>ICD-9-CM codes: 410.x, 412.x ICD-10 codes: I21.x-I22.x, I25.2</p>	<p>Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.</p>
ComorbCongHeart	Numeric 0=No 1=Yes	<p>Did the cohort member have a diagnosis code for congestive heart failure during the specified year or a prior calendar year within cohort window?</p> <p>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</p>	<p>Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.</p>
ComorbVasPeriph	Numeric	<p>Did the cohort member have a diagnosis code for peripheral vascular disease during the</p>	<p>Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	0=No 1=Yes	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	CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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,	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
ComorbRheumatic	Numeric	Did the cohort member have a diagnosis code for rheumatic disease during the specified	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	0=No 1=Yes	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	CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
ComorbDiabetesComp	Numeric	Did the cohort member have a diagnosis code for diabetes with chronic complication during	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	0=No 1=Yes	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	CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
ComorbLiver	Numeric	Did the cohort member have a diagnosis code for moderate or severe liver disease during	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	0=No 1=Yes	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	CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
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	Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
ComorbHIV ²	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 <i>et al.</i> 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.	Derived as EventHIV (p. 94) variable in the Cervical Timeline file. Recommend using the Cervical Timeline File (p. 91) variables EventHIV (earliest documented record of patient having HIV) and StatusHIV (if patient was diagnosed with HIV at time of event). Recommend using the CharlsonIndex (p. 23), CharlsonIndexPrev_Drv (p. 24), and/or CharlsonIndexInc_Drv (p. 24) variables to identify composite comorbidity scores, respectively, within each calendar year based on analysis needs. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
EnrolledMonths ²	Numeric	Number of months enrolled in the health insurance plan during the calendar year Entered an integer between 1-12.	
Smoke ²	Numeric	Collected data from last value available in calendar year. If no value was recorded during	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	0=Never 1=Current 2=Former	the calendar year, recorded the last available value prior to cohort entry and note date value	
SmokeDateMonth ²	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 and note date value	See Global Values: Dates (p. 4) for further detail.
SmokeDateDay	Numeric	Day 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 and note date value	See Global Values: Dates (p. 4) for further detail.
SmokeDateYear ²	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 and note date value	See Global Values: Dates (p. 4) for further detail.
SmokeDSR ^{1,2}	Numeric	Date smoking status recorded, days since reference 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 and note date value.	See Global Values: DSR (p. 5) for further detail.
CalYrAgeBeg ¹	Numeric	Age as of January 1 st of the respective calendar year	
CharlsonIndex ²	Numeric	Charlson comorbidity index, based on HCSRN macro	Recommend using this variable for transPROSPR analyses for site A only.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
			See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
CharlsonIndexPrev_Drv ^{1,2}	Numeric	Charlson comorbidity index, based on prevalent comorbidities Comorbidity flags provided by sites were used to calculate prevalent Charlson Index score consistently across all sites.	See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
CharlsonIndexInc_Drv ^{1,2}	Numeric	Charlson comorbidity index, based on incident comorbidities The year in which a comorbidity flag shifts to indicating a comorbidity diagnosis based on comorbidities provided by sites were used to calculate incident Charlson Index score consistently across all sites.	Recommend using this variable for transPROSPR analyses in conjunction with the EncCalYr_Drv (p. 25) variable for sites B and C only. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
CalYr_Ins_Drv ^{1,2}	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. Insurance category was prioritized in descending order as follows: Multiple Insurance (if multiple of Medicaid, Medicare, Commercial, InsOtherGov are selected); single Insurance designation for Medicare, Medicaid, InsOtherGov, InsCommerc (considered single designation if selected in conjunction with MedicalAssist, Uninsured, InsHighDeductible); InsOther (combined with Medical Assistance and InsHighDeductible); Uninsured (even if InsHighDeductible is selected); and Unknown. Insurance designation was recorded from any time during the calendar year.	
BMI_Drv ^{1,2}	Numeric	Body Mass Index (NIH NHLBI scale)	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		<p>Rounded to the nearest tenth.</p> <p>Height and weight values used to calculate BMI were submitted within a limited range; 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).</p>	
EncCalYr_Drv ^{1,2}	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 ^{1,2}	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?	<p>This variable was added to permit analysis flexibility in inclusion of calendar year data for cohort members from sites B and C based on whether the cohort member had had a primary care encounter within 36 months of the last primary care encounter.</p> <p>See Appendix: Cohort Exit due to Lack of Primary Care Utilization (p. 121) for further detail.</p>
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

ENGAGEMENT FILE

Overview

This file contains one record for each cohort entry and exit and includes reason for exit. See [Section 1: Overview](#) (p. 2) for cohort entry and cohort exit definitions for each site.

Record Structure

One record per cohort entry event. All cohort members present in the Participant file have at least one record in this file.

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 (see [METRICS Data Contributing Sites and Cohort Definitions: Cohort Entry](#) (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 then re-entered the study upon subsequent primary care utilization. See [Appendix: Cohort Member Study Periods for Primary and Secondary Analyses](#) (p. 121) for further detail. A modification to an existing Engagement file variable (CutOffReason) indicates whether a cohort member left the cohort due to lack of utilization. Further, a new Engagement file variable (EnrollPeriods) flags the cohort period for each Engagement file record, permitting analysis restriction to select cohort study periods. The Engagement file supports primary analyses only and must be used in conjunction with data in the Participant and Calendar Year files to support sensitivity analyses.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
CohortEntryDateMonth ²	Numeric	Month of cohort entry date See Section 1: Overview - Cohort Entry (p. 2) for cohort entry definition for each site.	See Global Values: Dates (p. 4) for further detail.
CohortEntryDateDay	Numeric	Day of cohort entry date See Section 1: Overview - Cohort Entry (p. 2) for cohort entry definition for each site.	See Global Values: Dates (p. 4) for further detail.
CohortEntryDateYear ²	Numeric	Year of cohort entry date See Section 1: Overview - Cohort Entry (p. 2) for cohort entry definition for each site.	See Global Values: Dates (p. 4) for further detail.
CohortEntryDSR ^{1,2}	Numeric	Cohort entry date, days since reference date See Section 1: Overview - Cohort Entry (p. 2) for cohort entry definition for each site.	See Global Values: DSR (p. 5) for further detail.
CutOffReason ²	Numeric 1=Administrative cut-off 2=Age out 4=Disenrollment 5=Move out of coverage area 6=Attributed provider leaves healthcare system 7=Death 8=Death (Estimated)	Cut-off reason If the cohort member was still in the cohort on last day for data submission, entered 1. Cohort members aged out the day prior to their 90th birthday. See Section 1: Overview - Cohort Exit for cohort exit definition for each site.	
CutOffDateMonth ²	Numeric	Month of cut-off date See Section 1: Overview - Cohort Exit for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffDateDay	Numeric	Day of cut-off date See Section 1: Overview - Cohort Exit for cohort exit definition for each site.	See Global Values: Dates (p. 4) for further detail.
CutOffDateYear ²	Numeric	Year of cut-off date	See Global Values: Dates (p. 4) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		See Section 1: Overview - Cohort Exit for cohort exit definition for each site.	
CutOffDSR ^{1,2}	Numeric	Cut-off date, days since reference date See Section 1: Overview - Cohort Exit for cohort exit definition for each site.	See Global Values: DSR (p. 5) for further detail.
EnrollPeriods ¹	Numeric	Number of cohort member study periods See Section 1: Overview - Cohort Exit for cohort exit definition for each site.	
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

PAP TEST FILE

Overview

This file contains one record for each Pap test received by a cohort member, dating from at least 3 years prior to cohort entry to cohort exit. The data from this file must be considered in combination with data from the HPV Test file to determine how Pap and HPV tests and results are used to define cervical screening behavior and recommendations. The Cervical Timeline File permits the analysis of overall cervical screening behavior by including all Pap and HPV tests within the same file.

Record Structure

One record per cohort member per cervical or vaginal Pap test. Provides at least a 3-year look-back from cohort entry for each cohort member, though sites may report tests that occurred more than 3 years prior to cohort entry.

General Harmonization Notes

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See <u>Global Values: PID</u> (p. 4) for further detail.
PapInd	Numeric 1=Screening 2=Diagnostic 3=Repeat 4=Follow-up 5=Surveillance	Indication for Pap test Optional.	Pap indication derived according to algorithm as <u>PapInd</u> (p. 100) variable in the Cervical Timeline File when EventPap=1. See Pap indication definition and notes in <u>Cervical Timeline File</u> (p. 84). Derived in the Cervical Timeline file
PapDateMonth	Numeric	Month of Pap test collection date	See <u>Global Values: Dates</u> (p. 4) for further detail. Derived as <u>EventMth</u> (p. 92) variable in the Cervical Timeline file when EventPap=1.
PapDateDay	Numeric	Day of Pap test collection date	See <u>Global Values: Dates</u> (p. 4) for further detail.
PapDateYear	Numeric	Year of Pap test collection date	See <u>Global Values: Dates</u> (p. 4) for further detail. Derived as <u>EventYr</u> (p. 92) variable in the Cervical Timeline file when EventPap=1.
PapDSR ¹	Numeric	Date of Pap test collection, days since reference date	Derived as <u>EventDSR</u> (p. 92) variable in the Cervical Timeline file when EventPap=1.
PapSpecType	Numeric 1=Conventional smear 2=Liquid based 3=Other	Specimen collection type for Pap test	
PapSpecAdequacy	Numeric 1=Satisfactory 2=Unsatisfactory	Specimen adequacy of Pap test	Derived as <u>PapHPVResult</u> (p. 101) variable in the Cervical Timeline file when EventPap=1.
PapObscurFactors	Character 5=Drying artifact or poor fixation 7=Inflammation 10=Scant Cellularity 95=Other	Obscuring factors from Pap test Optional. Multiple options were allowed and were separated with a ' ' (pipe) character. Ordered numerically and not temporally.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PapEndocerv	Numeric 5=Endocervical or squamous metaplastic cell component present 6=Endocervical or squamous metaplastic cell component not present 98=Not applicable	Endocervical component of Pap test	
PapResult	Numeric 1=Negative for intraepithelial lesion or malignancy (NILM) 2=Atypical squamous cells of undetermined significance (ASC-US) 3=Low grade squamous intraepithelial lesion (LSIL) 4=Atypical squamous cells cannot exclude HSIL (ASC-H) 5=High grade squamous intraepithelial lesion (HSIL) 16=Squamous Cell Carcinoma 17=Abnormal Glandular Cells (AGC; including Atypical and Endocervical) 18=Adenocarcinoma in situ 19=Suspicious for malignancy, no further information 20=Adenocarcinoma 21=Carcinoma, other, NOS 22=Benign Glandular Cells 90=No result	Result of Pap Test Record most severe result. Report listed result regardless of test specimen adequacy. High-grade SIL includes squamous cell carcinoma in situ. NILM includes benign squamous cells. Benign glandular cells are more severe than NILM and less severe than ASC-US.	Derived as <u>PapHPVResult</u> (p. 101) variable in the Cervical Timeline file when EventPap=1.
PapType	Numeric	Specimen derived from a cervical or vaginal Pap test	Derived as <u>PapType</u> (p. 100) variable in the Cervical Timeline file when EventPap=1.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	1=Cervical 2=Vaginal	If record indicates both vaginal and cervical Pap test, record as cervical Pap test. If record does not indicate either vaginal or cervical test, and there is no history of hysterectomy, then record as cervical.	
ProvIDPerform	Character	Performing provider ID The provider who performed the procedure. Entered a de-identified code that links to the Provider file.	Derived as <u>PapProvIDPerform</u> (p. 102) in the Cervical Timeline file when EventPap=1.
FacilityIDPerform	Character	Performing facility ID The location where the procedure was performed. Entered a de-identified code that links to the Facility file.	Derived as <u>PapFacilIDPerform</u> (p. 102) variable in the Cervical Timeline file when EventPap=1.
ProviderIDPCP	Character	Primary care provider ID Indicates the primary care provider associated with the cohort member at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
FacilityIDPCP	Character	Primary care provider clinic ID Indicates the facility at which the primary care provider associated with the cohort member practiced at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
CalYr_Ins_Drv ¹	Numeric 1=Medicare	Insurance category was prioritized in descending order as follows: Multiple Insurance (if multiple of Medicaid, Medicare, Commercial, InsOtherGov are	This variable was derived from the <u>CalYr_Ins_Drv</u> (p. 24) variable in the Calendar Year file for all sites.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	2=Medicaid 3=InsOtherGov 4=InsCommerc 5=Uninsured 6=Medical Assistance 7=Multiple Insurance 8=InsOther	selected); single Insurance designation for Medicare, Medicaid, InsOtherGov, InsCommerc (considered single designation if selected in conjunction with MedicalAssist, Uninsured, InsHighDeductible); InsOther (combined with Medical Assistance and InsHighDeductible); Uninsured (even if InsHighDeductible is selected); and Unknown. Insurance designation could have been made at any time during the calendar year.	
extractDate	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ¹	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

HPV TEST FILE

Overview

This file contains one record for each HPV test received by a cohort member, dating from at least 3 years prior to cohort entry to cohort exit. The data from this file must be considered in combination with data from the Pap test file to determine how Pap and HPV tests and results are used to define cervical screening behavior and recommendations. The Cervical Timeline File permits the analysis of overall cervical screening behavior by including all Pap and HPV tests within the same file.

Record Structure

One record per cohort member per HPV test. Provides at least a 3-year look-back from cohort entry for each cohort member, though sites may report tests that occurred more than 3 years prior to cohort entry.

General Harmonization Notes

HPV Test Indication: All sites agreed that the best way to identify HPV test indication varied by site; thus, sites varied in the type of date that was reported, as this was intertwined with test indication assignment. At site C, natural language processing (NLP) of the Pap test report was found to be the most accurate method of identifying test indication. By contrast, a high degree of misclassification in the provider orders were identified at site B, and so test results were used in addition to NLP of the test report to identify HPV test indication, because test result was used to override the original test order in select circumstances. Site A used the distinction between test order dates, which yielded identical dates for Pap and HPV co-tests and distinct dates for reflex Pap and HPV tests, to identify test indication.

HPV Test Manufacturer: Different HPV genotyping assays were used across sites and over time as new assays became available (table below). The genotyping assays* used in this study across sites are as follows:

1. Digene Hybrid Capture 2 by Qiagen; approved by the FDA in 2000 for adjunct screening and ASC-US triage; detects HPV genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68 (pooled)
2. Cervista HPV HR with optional reflex genotyping by Hologic; approved by the FDA in 2009 for adjunct screening and ASC-US triage; detects HPV genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 (pooled) and HPV genotypes 16, 18, 45 (reflex genotyping);
3. Cobas HPV with optional reflex genotyping by Roche; approved by the FDA in 2011 for primary HPV screening, adjunct screening, and ASC-US triage; detects HPV genotypes 16, 18, 31, 33, 35, 39, 45, 52, 56, 59, 66, and 68 (pooled) and HPV genotypes 16, 18, and 45 (reflex genotyping)
4. Aptima HPV with optional reflex individual genotyping by Hologic; approved by the FDA 2011 for adjunct screening and ASC-US triage; detects HPV genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 (pooled) and HPV genotypes 16, 18/45 (reflex genotyping);
5. Onclarity with optional reflex individual genotyping by Becton-Dickinson; approved by the FDA in 2018; detects HPV genotypes 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68 (pooled) and HPV genotypes 16, 18/45 (reflex genotyping)

*from the FDA Executive Summary *New Approaches in the Evaluation for High-Risk Human Papillomavirus Nucleic Acid Detection Devices* prepared for the March 8, 2019 meeting of the Microbiology Devices Panel of the Medical Devices Advisory Committee

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
HPVInd	Numeric 1=Reflex (ASC-US) 2=Cotest 3=Primary Reflex	Indication for HPV test, check all that apply	Derived as HPVInd (p. 102) variable in the Cervical Timeline file when EventHPV=1.
HPVDateMonth	Numeric	Month of HPV test collection date	See Global Values: Dates (p. 4) for further detail. Derived as EventMth (p. 92) variable in the Cervical Timeline file when EventHPV=1.
HPVDateDay	Numeric	Day of HPV test collection date	See Global Values: Dates (p. 4) for further detail.
HPVDateYear	Numeric	Year of HPV test collection date	See Global Values: Dates (p. 4) for further detail. Derived as EventYr (p. 92) variable in the Cervical Timeline file when EventHPV=1.
HPVColIDSR ¹	Numeric	HPV test specimen collection date, days since reference date	Derived as EventDSR (p. 92) variable in the Cervical Timeline file when EventHPV=1.
HPVManufName	Numeric 1=Digene 2=Cervista 3=Aptima 4=Cobas 5=Becton Dickinson	HPV test manufacturer	
HPVGenotypePool	Numeric	HPV pooled genotypes	Derived as PapHPVResult (p. 101) variable in Cervical Timeline file when EventHPV=1.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	1=2 HR Types (16/18) 2=3 HR Types (16/18/45) 4=12 HR Types (31/33/35/39/45/51/52/56/58/59/66/68) 5=13 HR Types (16/18/31/33/35/39/45/51/52/56/58/59/68) 6=14 HR Types (16/18/31/33/35/39/45/51/52/56/58/59/66/68) 7=HR (Unspecified) 10=16 HR types (26/31/33/35/39/45/51/52/53/56/58/59/66/68/73/82) 11=3 HR types (51/56/66) 12=5 HR types (18/39/45/59/68) 13=6 HR types (16/31/33/35/52/58) 14=2 HR types (18/45) 15=4 HR types (31/33/35/39) 16=11 HR Types (31/33/35/39/45/51/ 52/56/58/59/68) 17= 12 HR Types (16/18/31/33/35/39/45/52/56/59/66/68) 18= 13 HR Types (16/18/31/33/35/39/45/52/56/58/59/66/68) 19= 12 HR Types (16/18/31/33/35/39/45/52/56/58/59/68) 98=Individual genotype	<p>If the record was for an individual genotype, entered 98 here and entered the genotype in “HPV test individual genotype.”</p> <p>If the HPV text indicates high-risk or at least one of the genotypes listed in '6', then select '7'. If a high-risk or low-risk type could not be determined, then select '99'. If the record was for an individual genotype, enter '98' and enter the genotype in HPVGenotypeIndiv field.</p>	
HPVGenotypeIndiv	Numeric	<p>HPV test individual genotype</p> <p>Each HPV result is listed as a separate record. (e.g., If there were 19 individual HPV genotype results, then there were 19 records, one for each result.)</p>	Derived as HPVGenotype variable in <u>PapHPVResult</u> (p. 101) variable in Cervical Timeline file when EventHPV=1.
HPVResult	Numeric 1=Positive 2=Negative 3=Insufficient	<p>HPV result</p> <p>A positive result for a pool indicates that one or more of the types was detected, but the specific type(s) is not known. A negative result for a pool indicates that none of the types in the pool were detected.</p>	Derived as PapHPVResult (p. 101) variable in Cervical Timeline file when EventHPV=1.
ProvIDPerform	Character	Performing provider ID	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		The provider who performed the procedure. This refers to the provider and procedure where the samples were collected. Entered a de-identified code that links to the Provider file.	Derived as <u>HPVProvIDPerform</u> (p. 102) variable in Cervical Timeline file when EventHPV=1.
FacilityIDPerform	Character	Performing facility ID The location where the procedure was performed. This refers to the facility and procedure where the samples were collected. Enter a de-identified code that links to the Facility file.	Derived as <u>HPVFacilIDPerform</u> (p. 102) variable in Cervical Timeline file when EventHPV=1.
ProviderIDPCP	Character	Primary care provider ID Indicates the primary care provider associated with the cohort member at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
FacilityIDPCP	Character	Primary care provider clinic ID Indicates the facility at which the primary care provider associated with the cohort member practiced at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
CalYr_Ins_Drv ¹	Numeric 1=Medicare 2=Medicaid 3=InsOtherGov 4=InsCommerc	Insurance category was prioritized in descending order as follows: Multiple Insurance (if multiple of Medicaid, Medicare, Commercial, InsOtherGov are selected); single Insurance designation for Medicare,	This variable was derived from the <u>CalYr_Ins_Drv</u> (p. 24) variable in the Calendar Year file for all sites.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	5=Uninsured 6=Medical Assistance 7=Multiple Insurance 8=InsOther	Medicaid, InsOtherGov, InsCommerc (considered single designation if selected in conjunction with MedicalAssist, Uninsured, InsHighDeductible); InsOther (combined with Medical Assistance and InsHighDeductible); Uninsured (even if InsHighDeductible is selected); and Unknown. Insurance designation could have been made at any time during the calendar year.	
extractDate	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ¹	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

PROCEDURES FILE

Overview

This file contains one record for each gynecological procedure received by a cohort member from up to 3 years prior to cohort entry to cohort exit.

Record Structure

One record per cohort member per procedure per day, with the most severe result associated with that procedure recorded. Provides at least a 3-year look-back from cohort entry for each cohort member, though sites may report procedures that occurred more than 3 years prior to cohort entry. Only the following data elements were ascertained for procedures that occurred prior to cohort entry: ProcDate, ProcTypeCerv, SpecUndist, ProcDx and NonCervCa.

General Harmonization Notes

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
ProcDateMonth	Numeric	Month of date of diagnostic or treatment procedure	See Global Values: Dates (p. 4) for further detail. Derived as EventMth (p. 92) variable in the Cervical Timeline file when EventProc=1.
ProcDateDay	Numeric	Day of date of diagnostic or treatment procedure	See Global Values: Dates (p. 4) for further detail.
ProcDateYear	Numeric	Year of date of diagnostic or treatment procedure	See Global Values: Dates (p. 4) for further detail. Derived as EventMth (p. 92) variable in the Cervical Timeline file when EventProc=1.
ProcDSR ¹	Numeric	Date of diagnostic or treatment procedure, days since reference date	Derived as EventDSR (p. 92) variable in the Cervical Timeline file when EventProc=1. Derived as EventMth (p. 92) variable in the Cervical Timeline file when EventProc=1.
ProcTypeCerv	Character 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 16=History of Cervix Removal	Procedure Type Source of procedure type prioritized in descending order as follows: pathology reports; CPT/ICD-9 codes; ICD-10 codes. If more than one of the following procedures were completed on the same day, enter only the most significant procedure (prioritized in descending order) and the worst result obtained from these procedures: LEEP, Cone, Excisional Procedure NOS, Colpo Biopsy, and Biopsy No Colpo. For all other procedures, enter every instance and worst result in a separate record. Does not include endometrial biopsies. If an endometrial biopsy occurred along with a cervical procedure, then only the cervical procedure from the billing codes/pathology report was included.	Derived as ProcType (p. 102) in the Cervical Timeline file when EventProc=1.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
SpecUndist	Numeric 0=No 1=Yes	Cannot distinguish between 2 Procedure types? For example, pathology report did not distinguish if result was from the ECC or biopsy. If this was the case, submitted one record for each procedure type, recorded the same result, and marked "Yes" for this variable.	
ProcDx	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	Cervical specimen result Worst result from the cervical tissue collected during the procedure. Descending severity hierarchy is as follows: invasive SCC, invasive adenosquamous, and/or invasive adenocarcinoma; other cervical cancer; CIN-III; CIN-II-III (keep separate from HSIL if reported in Bethesda schema); HSIL; CIN-II; CIN-I-II; LSIL; CIN-I; HPV/condylomata; atypical/atypia; normal/negative; and no biopsy taken. In the presence of a cervical specimen result that was a non-cervical cancer, results related specifically to the cervix were reported; in the absence of any information directly related the cervix, records set to '38'.	Derived as <u>ProcResult</u> (p. 103) variable in the Cervical Timeline file when EventProc=1.
NonCervCa	Numeric 0=No 1=Yes	Used to identify records associated with non-cervical cancer when clearly specified in the pathology report.	
ProvIDPerform	Character		

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		Performing provider ID The provider who performed the procedure. Enter a de-identified code that links to the Provider file.	Derived as <u>ProcProvIDPerform</u> (p. 104) variable in the Cervical Timeline file when EventProc=1.
FacilityIDPerform	Character	Performing facility ID The location where the procedure was performed. Entered a de-identified code that links to the Facility file.	Derived as <u>ProcFacIDPerform</u> (p. 104) variable in the Cervical Timeline file when EventProc=1.
ProviderIDPCP	Character	Primary care provider ID Indicates the primary care provider associated with the cohort member at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
FacilityIDPCP	Character	Primary care provider clinic ID Indicates the facility at which the primary care provider associated with the cohort member practiced at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
ProvIDPath	Character	Pathology provider ID The pathologist who evaluated specimens from the procedure. Entered a de-identified code that links to the Provider file.	
FacilityIDPath	Character	Pathology facility ID The location where the specimens from the procedure were evaluated. Entered a de-identified code that links to the Facility file.	
HysterIndPath ¹	Numeric	Hysterectomy Indication based on Pathology	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	1=HSIL, CIN-II, or higher 2=LSIL or abnormal cytology 3=No Abnormality	Derived by the PRC using submitted dates and histopathology results using the following descending hierarchy of severity of histopathology results occurring within 180 days of hysterectomy: high-grade HSIL, CINII, or higher; low-grade LSIL or abnormal cytology; and no cervical pathology indicating abnormality.	
HysterIndDiag	Numeric 0=Cervical Abnormality Not Present 1=Cervical Abnormality Present	Hysterectomy Indication based on Diagnostic Codes Enter based on presence of codes dropped at time of surgery that indicate cervical abnormality.	
CalYr_Ins_Drv ¹	Numeric 1=Medicare 2=Medicaid 3=InsOtherGov 4=InsCommerc 5=Uninsured 6=Medical Assistance 7=Multiple Insurance 8=InsOther	Insurance category was prioritized in descending order as follows: Multiple Insurance (if multiple of Medicaid, Medicare, Commercial, InsOtherGov are selected); single Insurance designation for Medicare, Medicaid, InsOtherGov, InsCommerc (considered single designation if selected in conjunction with MedicalAssist, Uninsured, InsHighDeductible); InsOther (combined with Medical Assistance and InsHighDeductible); Uninsured (even if InsHighDeductible is selected); and Unknown. Insurance designation could have been made at any time during the calendar year.	This variable was derived from the <u>CalYr_Ins_Drv</u> (p. 24) variable in the Calendar Year file for all sites.
extractDate	Numeric	Date each record was generated	See <u>Global Values: extractDate</u> (p. 5) for further detail.
providingSite	Numeric	PROSPR Site ID Entered the Site ID associated with the record. There may be more than one Site ID associated with each PRC.	See <u>Global Values: ProvidingSite</u> (p. 6) for further detail.
ddVersion	Numeric	Data dictionary version number closest to submission date	See <u>Global Values: ddVersion</u> (p. 6) for further detail.
PIDSite ¹	Character	Unique derived PROSPR ID	See <u>Global Values: PIDSite</u> (p. 6) for further detail.

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](#) concept indicating the order of cancer diagnoses) of primary tumor for cervical cancers diagnosed throughout cohort duration.

General Harmonization Notes

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See <u>Global Values: PID</u> (p. 4) for further detail.
DxSrc	Numeric 1=State 2=Local 3=SEER	Cancer Registry data source State cancer registries supplemented with local/hospital based registries when patient is not available in the state registry.	
SequenceNumber ²	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/380 or Hospital/560).	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PrimarySiteICD ²	Character	ICD Code, primary site NAACCR Item #400. Text string.	
DxDateMonth ²	Numeric	Month of cancer diagnosis date NAACCR item #390.	See Global Values: Dates (p. 4) for further detail. Derived as EventCervCa (p. 94) variable in the Cervical Timeline file.
DxDateDay	Numeric	Day of cancer diagnosis date NAACCR item #390.	See Global Values: Dates (p. 4) for further detail.
DxDateYear ²	Numeric	Year of cancer diagnosis date NAACCR item #390.	See Global Values: Dates (p. 4) for further detail. Derived as EventCervCa (p. 94) variable in the Cervical Timeline file.
DxDSR ^{1,2}	Numeric	Date of cancer diagnosis, days since reference date NAACCR item #390.	See Global Values: DSR (p. 5) for further detail. Derived as EventCervCa (p. 94) variable in the Cervical Timeline file.
TumorBehavior ²	Numeric 0=Benign 1=Uncertain/borderline 2=In situ 3=Malignant/invasive 99=Unknown	Tumor malignancy or behavior NAACCR Item #523.	
SEERStage ²	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 in 2018-2019.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
AJCCTDerived	Character 000=T0 010=Ta 050=Tis 060=Tispu 070=Tispd 100=T1 110=T1mic 120=T1a 121=T1a(s) 122=T1a(m) 130=T1a1 140=T1a2 150=T1b 151=T1b(s) 152=T1b(m) 160=T1b1 170=T1b2 180=T1c 181=T1d 191=T1 NOS(s) 192=T1 NOS(m) 199=T1NOS 200=T2 201=T2(s) 202=T2(m) 210=T2a 211=T2a1 212=T2a2 213=T2aNOS 220=T2b 230=T2c 240=T2d 299=T2NOS 300=T3 301=T3(s) 302=T3(m) 310=T3a 320=T3b 330=T3c 340=T3d 399=T3NOS 400=T4 410=T4a 411=T4a(s) 412=T4a(m) 420=T4b 421=T4b(s) 422=T4b(m) 430=T4c 440=T4d 450=T4e 491=T4 NOS(s) 492=T4 NOS(m) 499=T4NOS 800=T1aNOS 810=T1bNOS 888=NA 999=TX	Derived AJCC-7 T NAACCR Item #=3400: Text string, 3-digit integer, preserve leading 0s.	
AJCCTDescrip	Character c=Clinical stage p=Pathologic stage a=Autopsy stage y=Surg resection performed, tumor size/ext path-based evidence N=Not Applicable	Derived AJCC-7 T Descriptor Optional. NAACCR Item #=3402 Text string	
AJCCNDerived	Character	Derived AJCC-7 N	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	000=N0 010=N0(i-) 020=N0(i+) 030=N0(mo1-) 040=N0(mo1+) 099=N0NOS 100=N1 110=N1a 120=N1b 130=N1c 180=N1mi 199=N1NOS 200=N2 210=N2a 220=N2b 230=N2c 299=N2NOS 300=N3 310=N3a 320=N3b 330=N3c 399=N3NOS 400=N4 888=NA 999=NX	Optional NAACCR Item #=3410 Text string, 3-digit integer, preserve leading 0s.	
AJCCNDescrip	Character c=Clinical stage p=Pathologic stage a=Autopsy stage y=Lymph nodes removed, evaluation based on pathologic evidence N=Not Applicable	Derived AJCC-7 N Descriptor Optional. NAACCR Item #=3412 Text string	
AJCCMDerived	Character 000=M0 010=M0(i+) 100=M1 110=M1a 120=M1b 130=M1c 140=M1d 150=M1e 199=M1NOS 888=NA 999=MX	Derived AJCC-7 M Optional. NAACCR Item #=3420 Text string, 3-digit integer, preserve leading 0s.	
AJCCMDescrip	Character c=Clinical stage p=Pathologic stage a=Autopsy stage y=Pathologic examination, extension based on path evidence N=Not Applicable	Derived AJCC-7 M Descriptor Optional. NAACCR Item #=3422 Text string	
AJCCStageDerived	Character 000=0 010=0a 020=0is 100=I	Derived AJCC-7 stage group Optional.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	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	NAACCR Item #3430 Text string, 3-digit integer, preserve leading 0s	
Grade ²	Numeric 1=Well differentiated; Grade I 2=Moderately differentiated; Grade II 3=Poorly differentiated; Grade III 4=Undifferentiated; anaplastic; Grade IV 5=T-cell 6=B-cell; pre-B; B-precursor 7=Null cell; non T-non B 8=NK cell; natural killer cell (1995+) 9=Unknown	SEER Grade, differentiation or cell indicator NAACCR item #440 for cancers diagnosed 2010-2017. NAACCR item #3844 (pathological grade) if available and otherwise NAACCR Item #3843 (clinical grade) for cancers diagnosed 2018-2019.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
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. Enter a 3-digit code, preserving leading 0s.	
HistologyICD02or03 ²	Numeric	Tumor histology or cell type, as first 4 digits of the ICD-O-2 and ICD-O-3 morphology code	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		NAACCR Item #=522. Enter the first 4 digits of the ICD-0-2 or ICD-0-3 code.	
RegNodesExamined	Character 00=No nodes examined 01-89=1-89 nodes examined (code exact number of regional lymph nodes examined) 99=Unknown if nodes were examined; not applicable or negative Not documented in patient record 98=Regional lymph nodes surgically removed but number of lymph nodes unknown/not stated and not documented as sampling or dissection; nodes examined, but number unknown 97=Regional lymph node removal documented as dissection and number of lymph nodes unknown/not stated 96=Regional lymph node removal documented as a sampling and number of lymph nodes unknown/not stated 95=No regional nodes removed, but aspiration or core biopsy of regional nodes performed 90=Ninety or more regional lymph nodes examined	Number of regional nodes examined NAACCR Item #=830. Text string 2-digit integer, preserve leading 0s	
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 Text string 2-digit integer, preserve leading 0s	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
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 Summary of the status for all treatment modalities	
TxCaIniDateMonth	Numeric	Month of date therapy initiated	See Global Values: Dates (p. 4) for further detail.
TxCaIniDateDay	Numeric	Day of date therapy initiated	See Global Values: Dates (p. 4) for further detail.
TxCaIniDateYear	Numeric	Year of date therapy initiated	See Global Values: Dates (p. 4) for further detail.
TxCaIniDSR ¹	Numeric	Date therapy initiated, days since reference NAACCR Item #1260.	See Global Values: DSR (p. 5) for further detail.
SurgeryPrimarySite	Character 00=None; no surgical procedure of primary site; diagnosed at autopsy only 10-19=Site-specific codes. Tumor destruction; no pathologic specimen or unknown whether there is a pathologic specimen 20-80=Site-specific codes. Resection; pathologic specimen 90=Surgery, NOS. A surgical procedure to the primary site was done, but no information on the type of surgical procedure is provided. 98=special codes for hematopoietic, reticuloendothelial, immunoproliferative, myeloproliferative diseases; ill-defined sites; and unknown primaries (See site-specific codes for the sites and histologies), except death certificate only 99=Unknown if surgery performed	Surgery of Primary Site NAACCR Item #1290 Text string, 2-digit integer, preserve leading 0s	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
ScopeRegLNSurg	Numeric 0=No regional lymph nodes removed or aspirated; diagnosed at autopsy 1=Biopsy or aspiration of regional lymph node, NOS 2=Sentinel lymph node biopsy [only] 3=Number of regional lymph nodes removed unknown, not stated; regional lymph nodes removed, NOS 4=1 to 3 regional lymph nodes removed 5=4 or more regional lymph nodes removed 6=Sentinel node biopsy and code 3, 4, or 5 at same time or timing not noted 7=Sentinel node biopsy and code 3, 4, or 5 at different times 9=Unknown or not applicable	Scope of regional lymph node surgery NAACCR Item# =1292	
ReasonNoSurgery	Numeric 0=Surgery of the primary site was performed 1=Surgery of the primary site was not performed because it was not part of the planned first-course treatment 2=Surgery of the primary site was not recommended/performed because it was contraindicated due to patient risk factors (comorbid conditions, advanced age, etc.) 5=Surgery of the primary site was not performed because the patient died prior to planned or recommended surgery 6=Surgery of the primary site was not performed; it was recommended by the patient's physician, but was not performed as part of the first course of therapy. No reason was noted in the patient's record. 7=Surgery of the primary site was not performed; it was recommended by the	Reason for no surgery of primary site NAACCR Item# =1340	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	<p>patient's physician, but was refused by the patient, the patient's family member, or the patient's guardian. The refusal was noted in the patient record.</p> <p>8=Surgery of the primary site was recommended, but it is unknown if it was performed. Further follow up is recommended</p> <p>9=It is unknown if surgery of the primary site was recommended or performed; autopsy-only cases</p>		
TxIniRadiation	<p>Numeric</p> <p>0=None; diagnosed at autopsy</p> <p>1=Beam radiation</p> <p>2=Radioactive implants</p> <p>3=Radioisotopes</p> <p>4=Combination of 1 with 2 or 3</p> <p>5=Radiation, NOS – method or source not specified</p> <p>7=Patient or patient's guardian refused radiation therapy</p> <p>8=Radiation recommended, unknown if administered</p> <p>9=Unknown if radiation administered</p>	<p>Radiation, initial treatment</p> <p>NAACCR Item# =1360</p>	
TxIniRadiationSeq	<p>Numeric</p> <p>0=No radiation and/or no surgery; unknown if surgery and/or radiation given</p> <p>2=Radiation before surgery</p> <p>3=Radiation after surgery</p> <p>4=Radiation both before and after surgery</p> <p>5=Intraoperative radiation</p> <p>6=Intraoperative radiation with other radiation given before and/or after surgery</p> <p>7=Surgery both before and after radiation</p> <p>9=Sequence unknown, but both surgery and radiation were given</p>	<p>Radiation sequence with surgery</p> <p>NAACCR Item# =1380</p>	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
TxIniChemo	<p>Character</p> <p>00=None, chemotherapy was not part of the planned first course of therapy; diagnosed at autopsy</p> <p>01=Chemotherapy administered as first course therapy, but the type and number of agents is not documented in the patient record.</p> <p>02=Single agent chemotherapy administered as first course therapy.</p> <p>03=Multiagent chemotherapy administered as first course therapy.</p> <p>82=Chemotherapy was not recommended/administered because it was contraindicated due to patient risk factors (comorbid conditions, advanced age, etc.).</p> <p>85=Chemotherapy was not administered because the patient died prior to planned or recommended therapy.</p> <p>86=Chemotherapy was not administered. It was recommended by the patient's physician but was not administered as part of the first course of therapy. No reason was stated in patient record.</p> <p>87=Chemotherapy was not administered. It was recommended by the patient's physician, but the treatment was refused by the patient, a patient's family member, or the patient's guardian. The refusal was noted in the patient record.</p> <p>88=Chemotherapy was recommended, but it is unknown if it was administered.</p> <p>99=It is unknown whether a chemotherapeutic agent(s) was recommended or administered because it is not stated in the patient record.)</p>	<p>Chemotherapy, initial treatment</p> <p>NAACCR Item# =1390 Text string, 2-digit integer, preserve leading 0s Records the chemotherapy given as a part of the first course of treatment or the reason that chemotherapy was not given.</p>	
TxIniHormone	<p>Character</p>	Hormone therapy	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	<p>00=None, hormone therapy was not part of the planned first course of therapy; not usually administered for this type and/or stage of cancer; diagnosed at autopsy only</p> <p>01=Hormone therapy administered as first course therapy</p> <p>82=Hormone therapy was not recommended/administered because it was contraindicated due to patient risk factors (comorbid conditions, advanced age, etc.)</p> <p>85=Hormone therapy was not administered because the patient died prior to planned or recommended therapy</p> <p>86=Hormone therapy was not administered. It was recommended by the patient's physician but was not administered as part of the first course of therapy. No reason was stated in the patient record.</p> <p>87=Hormone therapy was not administered. It was recommended by the patient's physician, but this treatment was refused by the patient, a patient's family member, or the patient's guardian. The refusal was noted in the patient record.</p> <p>88=Hormone therapy was recommended, but it is unknown if it was administered.</p> <p>99=It is unknown whether a hormonal agent(s) was recommended or administered</p>	<p>NAACCR Item# =1400 Text string, 2-digit integer, preserve leading 0s Records therapy administered as first course treatment that affects cancer tissue by adding, blocking, or removing the action or production of hormones.</p>	
TxInImmuno	<p>Character</p> <p>00=None, immunotherapy was not part of the planned first course of therapy; not customary therapy for this cancer; diagnosed at autopsy only</p>	<p>Immunotherapy</p> <p>NAACCR Item# =1410 Text string, 2-digit integer, preserve leading 0s</p>	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	<p>01=Immunotherapy was administered as first course therapy</p> <p>82=Immunotherapy was not recommended/administered because it was contraindicated due to patient risk factors (comorbid conditions, advanced age, etc.)</p> <p>85=Immunotherapy was not administered because the patient died prior to planned or recommended therapy</p> <p>86=Immunotherapy was not administered; it was recommended by the patient's physician but was not administered as part of the first-course of therapy. No reason was noted in the patient's record.</p> <p>87=Immunotherapy was not administered. It was recommended by the patient's physician, but this treatment was refused by the patient, a patient's family member, or the patient's guardian. The refusal was noted in the patient record.</p> <p>88=Immunotherapy was recommended, but it is unknown if it was administered</p> <p>99=It is unknown if immunotherapy was recommended or administered because it is not stated in patient record.</p>		
TxIniTransplant	<p>Character</p> <p>00=None, transplant procedure or endocrine therapy was not a part of the first course of therapy; not customary therapy for this cancer; diagnosed at autopsy only</p> <p>10=Bone marrow transplant, NOS. A bone marrow transplant procedure was administered as first course therapy, but the type was not specified</p> <p>11=Bone marrow transplant autologous</p> <p>12=Bone marrow transplant allogeneic</p>	<p>Hematologic transplant and endocrine procedures</p> <p>NAACCR Item# =3250 Text string, 2-digit integer, preserve leading 0s Records systemic therapeutic procedures administered as part of the first course of treatment. These procedures include bone marrow transplants (BMT) and stem cell harvests with rescue (stem cell transplant), endocrine surgery and/or radiation performed for hormonal effect (when cancer originates at another site), and a</p>	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	<p>20=Stem cell harvest (stem cell transplant) and infusion</p> <p>30=Endocrine surgery and/or endocrine radiation therapy as first course therapy</p> <p>40=Combination of transplant procedure with endocrine surgery and/or endocrine radiation (Code 30 in combination with 10, 11, 12, or 20) as first course therapy</p> <p>82=Transplant procedure and/or endocrine therapy was not recommended/administered because it was contraindicated due to patient risk factors (comorbid conditions, advanced age, etc.)</p> <p>85=Transplant procedures and/or endocrine therapy was not administered because the patient died prior to planned or recommended therapy</p> <p>86=Transplant procedures and/or endocrine therapy was not administered; it was recommended by the patient's physician but was not administered as part of first course therapy. No reason was noted in the patient record.</p> <p>87=Transplant procedures and/or endocrine therapy were not administered; this treatment was recommended by the patient's physician but was refused by the patient, a patient's family member, or the patient's guardian. The refusal was noted in the patient record.</p> <p>88=Transplant procedures and/or endocrine therapy was recommended, but it is unknown if it was administered.</p> <p>99=It is unknown if a transplant procedure or endocrine therapy was recommended or administered because it is not stated in patient record</p>	combination of transplants and endocrine therapy.	
TxIniSequence	Numeric	Systemic treatment/surgery sequence	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	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	NAACCR Item# =1639 Records the sequence of any systemic therapy and surgery given as first course of therapy for those patients who had both systemic therapy and surgery	
TxIniOther	Numeric 0=None 1=Other 2=Other Experimental 3=Other-Double Blind 6=Other-Unproven 7=Refusal 8=Recommended 9=Unknown; unknown if administered	Other therapy NAACCR Item# =1420 Identifies treatment given that cannot be classified as surgery, radiation, systemic therapy, or ancillary treatment. This data item includes all complementary and alternative medicine used by the patient in conjunction with conventional therapy or in place of conventional therapy.	
DLFDateMonth	Numeric	Month of date of last follow-up or death NAACCR Item# =1750 Records the days since referent date of the date of last follow up or the date of death. SEER requires the registries to update the follow up information on all cases on an annual basis. The exception is carcinoma in situ of the cervix diagnosed on or after 1/1/1996.	See Global Values: Dates (p. 4) for further detail.
DLFDateDay	Numeric	Day of date of last follow-up or death NAACCR Item# =1750 Records the days since referent date of the date of last follow up or the date of death. SEER requires the registries to update the follow up information on all cases on an annual basis.	See Global Values: Dates (p. 4) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		The exception is carcinoma in situ of the cervix diagnosed on or after 1/1/1996.	
DLFDateYear	Numeric	<p>Year of date of last follow-up or death</p> <p>NAACCR Item# =1750 Records the days since referent date of the date of last follow up or the date of death. SEER requires the registries to update the follow up information on all cases on an annual basis. The exception is carcinoma in situ of the cervix diagnosed on or after 1/1/1996.</p>	See Global Values: Dates (p. 4) for further detail.
DLFDSR ¹	Numeric	<p>Date of last follow-up or death, days since reference</p> <p>NAACCR Item# =1750 Records the days since referent date of the date of last follow up or the date of death. SEER requires the registries to update the follow up information on all cases on an annual basis. The exception is carcinoma in situ of the cervix diagnosed on or after 1/1/1996.</p>	See Global Values: DSR (p. 5) for further detail.
VitalStatusSEER	Numeric 0=Dead (CoC) 1=Alive 4=Dead (SEER)	<p>Vital status (SEER)</p> <p>NAACCR Item# =1760</p>	
CauseOfDeathICD	Character	<p>Underlying Cause of Death (SEER)</p> <p>NAACCR Item# =1910 Text string, preserve leading 0s This is the official underlying cause of death coded from the death certificate using ICD-7, ICDA-8, ICD-9, or ICD-10 codes.</p> <p>Special codes: 0000=Patient alive at last contact 7777=State death certificate or listing not available 7797=State death certificate or listing available, but underlying cause of death not coded</p>	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
DeathCodeVersion	Numeric 0=Patient alive at last follow-up 1=ICD-10 7=ICD-7 8=ICDA-8 9=ICD-9	ICD Code revision used for cause of death NAACCR Item# =1920 SEER requires the registries to update the follow up information on all cases on an annual basis. The exception is carcinoma in situ of the cervix diagnosed on or after 1/1/1996	
MaritalStatusSEER	Numeric 1=Single (never married) 2=Married (including common law) 3=Separated 4=Divorced 5=Widowed 6=Unmarried or Domestic Partner (same sex or opposite sex, registered or unregistered, other than common law marriage) 9=Unknown 88=Refused	Marital status (SEER) NAACCR Item #=150 The patient's marital status at the time of diagnosis for the reportable tumor.	
PptSEERCatchment	Numeric 0=No 1=Yes	Did the cohort member reside in an area covered by a SEER registry (SEER catchment area)?	
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ^{1, 2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

PRIOR FILE

Overview

This file contains one record of known healthcare events prior to cohort entry for each cohort member.

Record Structure

One record per cohort member. Provides a 3-year look-back from cohort entry for each cohort member, except for the following variables with no look-back restrictions: CaCervEver and HIVEver.

General Harmonization Notes

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail
MonthsPriorToCohort	Numeric	Number of continuous months enrolled prior to cohort entry date This applies only to site A. Entered the number of months as an integer.	
PCPVisitLatestDateMonth	Numeric	Month of date of last primary care visit prior to cohort entry	See Global Values: Dates (p. 4) for further detail.
PCPVisitLatestDateDay	Numeric	Day of date of last primary care visit prior to cohort entry	See Global Values: Dates (p. 4) for further detail.
PCPVisitLatestDateYear	Numeric	Year of date of last primary care visit prior to cohort entry	See Global Values: Dates (p. 4) for further detail.
PCPVisitLatestDSR ¹	Numeric	Date of last primary care visit prior to cohort entry, days since reference date	See Global Values: DSR (p. 5) for further detail.
CaCervEver	Numeric 0=No 1=Yes	Ever diagnosed with invasive cervical cancer?	Derived as EventCervCa (p. 94) variable in the Cervical Timeline file when CaCervEver=1.
CaCervDateMonth	Numeric	Month of date of diagnosis for invasive cervical cancer	See Global Values: Dates (p. 4) for further detail. Derived as EventCervCa (p. 94) variable in the Cervical Timeline file when CaCervEver=1.
CaCervDateDay	Numeric	Day of date of diagnosis for invasive cervical cancer	See Global Values: Dates (p. 4) for further detail.
CaCervDateYear	Numeric	Year of date of diagnosis for invasive cervical cancer	See Global Values: Dates (p. 4) for further detail. Derived as EventCervCa (p. 94) variable in the Cervical Timeline file when CaCervEver=1.
CaCervDSR ¹	Numeric	Date of diagnosis for invasive cervical cancer, days since reference date	See Global Values: DSR (p. 5) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
			Derived as <u>EventCervCa</u> (p. 94) variable in the Cervical Timeline file when CaCervEver=1.
HIVEver	Numeric 0=No 1=Yes	Ever diagnosed with HIV? Enhanced ICD-9-CM codes 042.x–044.x and V08. ICD-10 codes: B20.x-B22.x, B24.x, B23.x and Z21.	Derived as <u>EventHIV</u> (p. 94) variable in the Cervical Timeline file when HIVEver=1. Recommend using the <u>Cervical Timeline File</u> (p. 91) variables EventHIV (earliest documented record of patient having HIV) and StatusHIV (if patient was diagnosed with HIV at time of event).
MEdRec1stDateMonth	Numeric	Month of date a medical record was first created This applies only to site B.	See <u>Global Values: Dates</u> (p. 4) for further detail.
MEdRec1stDateDay	Numeric	Day of date a medical record was first created This applies only to site B.	See <u>Global Values: Dates</u> (p. 4) for further detail.
MEdRec1stDateYear	Numeric	Year of date a medical record was first created This applies only to site B.	See <u>Global Values: Dates</u> (p. 4) for further detail.
MedRec1stDSR ¹	Numeric	Date a medical record was first created, days since reference date Reference date is date of birth.	See <u>Global Values: DSR</u> (p. 5) for further detail.
PCPVisitFirst3DateMonth	Numeric	Month of date of first primary care visit in three year period prior to cohort entry Add first PCP encounter within three years of first cohort entry.	
PCPVisitFirst3DateDay	Numeric	Day of date of first primary care visit in three year period prior to cohort entry Add first PCP encounter within three years of first cohort entry.	
PCPVisitFirst3DateYear	Numeric	Year of date of first primary care visit in three year period prior to cohort entry Add first PCP encounter within three years of first cohort entry.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PCPVisitFirst3DSR ¹	Numeric	Date of first primary care visit in three year period prior to cohort entry, days since reference date Add first PCP encounter within three years of first cohort entry.	
PriorCervRem	Numeric 0=No 1=Yes	Was the cohort member's cervix removed prior to cohort entry? Cervix was considered removed if a cohort member had any of the following procedures prior to cohort entry: hysterectomy, NOS; trachelectomy; total hysterectomy; or radical/modified radical hysterectomy.	Derived as <u>EventNoCervix</u> (p. 95) variable in the Cervical Timeline file when PriorCervRem=1.
extractDate	Numeric	Date each record was generated	See <u>Global Values: extractDate</u> (p. 5) for further detail.
providingSite	Numeric	PROSPR Site ID	See <u>Global Values: ProvidingSite</u> (p. 6) for further detail.
ddVersion	Numeric	Data dictionary version number closest to submission date	See <u>Global Values: ddVersion</u> (p. 6) for further detail.
PIDSite ¹	Character	Unique derived PROSPR ID	See <u>Global Values: PIDSite</u> (p. 6) for further detail.

HPV VACCINE FILE

Overview

This file contains one record for each HPV vaccine administration or HPV vaccine refusal for each cohort member prior to and during the cohort period.

Record Structure

One record per cohort member per HPV vaccine dose; records available as far back as possible for vaccines, regardless of cohort entry date.

General Harmonization Notes

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
HPVVacType	Numeric 1=Cervarix Bivalent 2=Gardasil Quadrivalent 3=Gardasil Nonavalent	Type of HPV vaccine	See Appendix: HPV Vaccination Schedule (p. 121) for approved vaccine schedule.
HPVVacDateMonth	Numeric	Month of date of HPV vaccine dose	See Global Values: Dates (p. 4) for further detail.
HPVVacDateDay	Numeric	Day of date of HPV vaccine dose	See Global Values: Dates (p. 4) for further detail.
HPVVacDateYear	Numeric	Year of date of HPV vaccine dose	See Global Values: Dates (p. 4) for further detail.
HPVVacDSR ¹	Numeric	Date of HPV vaccine dose, days since reference date	See Global Values: DSR (p. 5) for further detail.
HPVVacRefDateMonth	Numeric	Month of date of HPV vaccine dose refusal	See Global Values: Dates (p. 4) for further detail.
HPVVacRefDateDay	Numeric	Day of date of HPV vaccine dose refusal	See Global Values: Dates (p. 4) for further detail.
HPVVacRefDateYear	Numeric	Year of date of HPV vaccine dose refusal	See Global Values: Dates (p. 4) for further detail.
HPVVacRefDSR ¹	Numeric	Date of HPV vaccine dose refusal, days since reference date	See Global Values: DSR (p. 5) for further detail.
HPVVacPayment	Numeric	Source of payment for HPV vaccine	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	1= Commercial and/or private insurance 2=Medicaid 3=State Children's Health Insurance Program 4=Vaccines for Children program 5=Immunization Grant Program (317 funds) 6=State		
HPVVacInfoSource	Numeric 1=Patient self-report 4=Electronic medical record 5=State immunization registry 6=Claims-based	Source of information for HPV vaccine Provided definition of "Patient self-report" in the documentation for this variable.	
HPVVacInfoSourceOther	Character	Source of information for HPV vaccine, other, specify Enter text as needed, up to 200 characters.	
ProvIDPerform	Character	Performing provider ID The provider who performed the procedure. Entered a de-identified code that links to the Provider file.	
FacilityIDPerform	Character	Performing facility ID The location where the procedure was performed. Entered a de-identified code that links to the Facility file	
ProviderIDPCP	Character	Primary care provider ID Indicates the primary care provider associated with the cohort member at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
FacilityIDPCP	Character	Primary care provider clinic ID	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
		Indicates the facility at which the primary care provider associated with the cohort member practiced at the time closest to the test using a de-identified code that links to the Provider file. '-99999' can indicate unavailable data or cohort member known to not be assigned.	
extractDate	Numeric	Date each record was generated.	See Global Values: extractDate (p. 5) for further detail.
providingSite	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.
PIDSite ¹	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

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. Multiple encounters may occur on the same date.

General Harmonization Notes

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
PCPVisitDateMonth ²	Numeric	Month of Primary Care Visit Date	See Global Values: Dates (p. 4) for further detail.
PCPVisitDateDay	Numeric	Day of Primary Care Visit Date	See Global Values: Dates (p. 4) for further detail.
PCPVisitDateYear ²	Numeric	Year of Primary Care Visit Date	See Global Values: Dates (p. 4) for further detail.
PCPVisitDSR ^{1,2}	Numeric	Date of primary care visit, days since reference date	See Global Values: DSR (p. 5) for further detail.
ProviderIDEnc ²	Character	Provider ID for Encounter Entered a de-identified code that links to the Provider file.	
FacilityIDEnc ²	Character	Clinic ID where Encounter occurred The clinic/facility associated with the encounter. Entered a de-identified code that links to the Facility file.	
CalYr_Ins_Drv ¹	Numeric 1=Medicare 2=Medicaid 3=InsOtherGov 4=InsCommerc 5=Uninsured 6=Medical Assistance 7=Multiple Insurance 8=InsOther	Insurance category was prioritized in descending order as follows: Multiple Insurance (if multiple of Medicaid, Medicare, Commercial, InsOtherGov are selected); single Insurance designation for Medicare, Medicaid, InsOtherGov, InsCommerc (considered single designation if selected in conjunction with MedicalAssist, Uninsured, InsHighDeductible); InsOther (combined with Medical Assistance and InsHighDeductible); Uninsured (even if InsHighDeductible is selected); and Unknown. Insurance designation could have been made at any time during the calendar year.	This variable was derived from the CalYr_Ins_Drv (p. 24) variable in the Calendar Year file for all sites.
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
			See <u>Global Values: ddVersion</u> (p. 6) for further detail.
PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See <u>Global Values: PIDSite</u> (p. 6) for further detail.

PREGNANCY FILE

Overview

This file contains one record for each pregnancy for each cohort member from cohort entry through cohort exit.

Record Structure

One record per cohort member per pregnancy; only include pregnancies that were ongoing at some point during cohort period.

General Harmonization Notes

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID	Character	Unique participant ID	See Global Values: PID (p. 4) for further detail.
PregBeginDateDay	Numeric	Day of date pregnancy began	See Global Values: Dates (p. 4) for further detail.
PregBeginDateMonth	Numeric	Month of date pregnancy began	See Global Values: Dates (p. 4) for further detail.
PregBeginDateYear	Numeric	Year of date pregnancy began	See Global Values: Dates (p. 4) for further detail.
PregBeginDSR ¹	Numeric	Date pregnancy began, days since reference	See Global Values: DSR (p. 5) for further detail.
PregEndDateDay	Numeric	Day of date pregnancy ended	See Global Values: Dates (p. 4) for further detail.
PregEndDateMonth	Numeric	Month of date pregnancy ended	See Global Values: Dates (p. 4) for further detail.
PregEndDateYear	Numeric	Year of date pregnancy ended	See Global Values: Dates (p. 4) for further detail.
PregEndDSR ¹	Numeric	Date pregnancy ended, days since reference	See Global Values: DSR (p. 5) for further detail.
LiveBirth	Numeric 0=No 1=Yes	Live birth? Provide where possible, even if pregnancy start and/or stop dates are unknown.	
extractDate	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PIDSite ¹	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.

PROVIDER FILE

Overview

This file contains one record for each provider found in the following files: Calendar Year, Pap Test, HPV Test, HPV Vaccine, Encounter, and Procedures files.

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

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
ProviderID ²	Character	<p>Provider ID</p> <p>De-identified and unique to the provider within the PRC.</p> <p>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.</p>	
ProviderSpecialty ²	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	<p>Provider medical specialty</p> <p>Categories are new for 3rd data submission. Use a pipe " " symbol to separate multiple entries.</p>	
ProviderSpecialtyOther ²	Character	<p>Provider medical specialty (other, specify)</p> <p>Entered text as needed, up to 200 characters.</p>	
ProviderType ²	Numeric	Provider type	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
	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)		
ProviderTypeOther ²	Character	Provider type (other, specify) Entered text as needed, up to 200 characters.	
ProviderYears	Numeric	Total number of years as a provider in the health system	
ProviderIsPerson ²	Numeric 0=No	Is provider known to be a person?	
ProviderSex	Numeric 0=Male 1=Female 2=Non-binary	Sex that provider identifies as	
ProviderBirthYr	Numeric Enter a 4 digit value.	Year provider was born	
ProviderStartYr	Numeric Enter a 4 digit value.	Year provider started in present healthcare system	
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.

FACILITY FILE

Overview

This file contains one record for each facility ID found in the following files: Calendar Year, Pap Test, HPV Test, HPV Vaccine, Encounter, and Procedures files.

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](#) (p. 124) for facility level definition for each site.

General Harmonization Notes

The METRICS sites expended considerable effort harmonizing the level at which facilities were defined across the three disparate healthcare system to permit analysis execution at a consistent level. [Appendix: Facilities Harmonization](#) (p. 124) summarizes the method by which an analyst is able to identify facilities at a consistent level across all three METRICS sites.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
FacilityID ²	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 facility ID was used in both data sets.	See Section 3: Appendix - Facilities Harmonization (p. 124) for summary of how this variable can be harmonized by level for analyses across sites.
FacilityStFIP ²	Character	Facility state, FIPS code Indicates the state 2-digit FIPS code with leading zeroes included.	
FacilityCountyFIP ²	Character	Facility county, FIPS code Indicates the county 3 digit FIPS code with leading zeroes included.	
FacilityZip ²	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 ²	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	See Facility File: General Harmonization (p. 80) for definitions.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
FacilityTypeOther ²	Character	Facility type (other, specify) Indicates additional description of facility type not included in FacilityType variable up to 200 characters.	
FacilityNetworkStatus ²	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.
FacilityNCICC	Numeric 0=No 1=Yes	Reflects whether the facility was associated with a NCI-designated cancer center.	
FacilityIDRelatedPhys ²	Character	Associated physical larger facility ID Reflects the larger aggregating facility at which multiple co-located facility IDs may be rolled up as.	See Appendix: Facilities Harmonization (p. 124) for summary of how this variable was harmonized across sites.
FacilityIDRelatedAggr ²	Character	Facility aggregating value, PRC-defined Optional.	
FacilityPCWH	Numeric 1 = Primary Care 2 = Females's Health 3 = Both Primary Care and Females's Health 4 = HIV 5 = Other	Reflects whether a facility was a primary care clinic or a females's health clinic.	This variable aligns with the PCC 'FacilityPC' and 'FacilityWH' variables.
FacilityOwner	Numeric 0=Owned by provider organization 1=Owned by a hospital organization	Reflects whether the facility was owned by a hospital organization.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
FacilityLoc	Numeric 0=No 1=Yes	Reflects whether the facility was located on-site at hospital or in the community (not adjacent to hospital).	
FacilityRes	Numeric 0=No 1=Yes	Reflects whether residents rotated through the facility.	
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.

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

All race, education, Location Quotient, Index of Concentration, Local Exposure and Isolation, Local Isolation Score, and Yost Quintile were populated based on the FIPS code associated with cohort member's address using the "prospr.sdohealth.census.tract.level.data.beta.csv" file.

Cohort members that entered both the cervical and colorectal cohorts have the same geocoding information so long as the female 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.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PID		Unique participant ID	See Global Values: PID (p. 4) for further detail.
GeoAddrPOBox ²	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 ²	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).	
GeoAddrBeginDay ²	Numeric	Day 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 ²	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 ^{1,2}		Geocoding address date, days since reference	
pptZip ²	Character	Cohort member zip code Entered a valid 5-digit numeric zip code and included leading 0s.	
GeoCertainty ²	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	Geocoding certainty level Reflects standard output from geocoding software and varies by software.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	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		
GeoCertaintyOther ²	Character	Geocoding certainty level, Other (Specify)	
GeoMatchScore ²	Numeric	Geocoding match score	
GeoSoftwareName ²	Character	Geocoding software name	
GeoSoftwareVersion ²	Character	Geocoding software version	
Race_NHisp_White Race ²	Numeric	% White Alone, Non-Hispanic	
Race_NHisp_Black Race ²	Numeric	% Black or African American Alone, Non-Hispanic	
Race_NHisp_AIAN Race ²	Numeric	% American Indian and Alaska Native Alone, Non-Hispanic	
Race_NHisp_Asian ²	Numeric	% Asian Alone, Non-Hispanic	
Race_NHisp_NatHaw Race ²	Numeric	% Native Hawaiian and Other Pacific Islander Alone, Non-Hispanic	
Race_NHisp_Other Race ²	Numeric	% Some Other Race Alone, Non-Hispanic	
Race_NHisp_Multi Race ²	Numeric	% Two or More Races, Non-Hispanic	
Race_Hisp_White Race ²	Numeric	% White Alone, Hispanic	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Race_Hisp_Black Race ²	Numeric	% Black or African American Alone, Hispanic	
Race_Hisp_AIAN Race ²	Numeric	% American Indian and Alaska Native Alone, Hispanic	
Race_Hisp_Asian Race ²	Numeric	% Asian Alone, Hispanic	
Race_Hisp_NatHaw Race ²	Numeric	% Native Hawaiian and Other Pacific Islander Alone, Hispanic	
Race_Hisp_Other Race ²	Numeric	% Some Other Race Alone, Hispanic	
Race_Hisp_Multi Race ²	Numeric	% Two or More Races, Hispanic	
Educ_Less9th Education ²	Numeric	% less than 9th grade	
Educ_9th_12th Education ²	Numeric	% 9th-12th grade (no diploma)	
Educ_HSGrad Education ²	Numeric	% high school graduate	
Educ_SomeColl Education ²	Numeric	% some college, no degree	
Educ_AssocDeg Education ²	Numeric	% associate's degree	
Educ_BachDeg Education ²	Numeric	% bachelor's degree	
Educ_MastProfDeg Education ²	Numeric	% master's or professional school degree	
Educ_DoctoralDeg Education ²	Numeric	% doctorate degree	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Person_Below_Pov Poverty ²	Numeric	% of person below poverty level	
RUCA4A ²	Numeric	RUCA 4-level categorization 'A'	
Flag_Tract_Pop_Zero ²	Numeric 0=No 1=Yes	Flag: Census Tract Pop Zero	
LQ_White_Alone ²	Numeric	Location Quotient (LQ): White Alone	
LQ_Black_Alone ²	Numeric	Location Quotient (LQ): Black or African American Alone	
LQ_API_Alone ²	Numeric	Location Quotient (LQ): Asian or Pacific Islander Alone	
LQ_Hispanic ²	Numeric	Location Quotient (LQ): Hispanic	
LQ_NH_White_Alone ²	Numeric	Location Quotient (LQ): White Alone, Non-Hispanic	
ICE_Black_Alone_White_Alone ²	Numeric	Index of Concentration at Extremes (ICE): Black or African American Alone/White Alone	
ICE_API_Alone_White_Alone ²	Numeric	Index of Concentration at Extremes (ICE): Asian or Pacific Islander Alone/White Alone	
ICE_Hispanic_NH_White_Alone ²	Numeric	Index of Concentration at Extremes (ICE): Hispanic/White Alone, Non-Hispanic	
Lex_Is_Black_Alone_White_Alone ²	Numeric	Local Exposure and Isolation (Lex/IS): Black or African American Alone/White Alone	
Lex_Is_API_Alone_White_Alone ²	Numeric	Local Exposure and Isolation (Lex/IS): Asian or Pacific Islander Alone/White Alone	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Lex_Is_Hispanic_NH_White_Alone ²	Numeric	Local Exposure and Isolation (Lex/IS): Hispanic/White Alone, Non-Hispanic	
LIS_White_Alone ²	Numeric	Local Isolation Score: White Alone	
LIS_Black_Alone ²	Numeric	Local Isolation Score: Black or African American Alone	
LIS_API_Alone ²	Numeric	Local Isolation Score: Asian or Pacific Islander Alone	
LIS_Hispanic ²	Numeric	Local Isolation Score: Hispanic	
LIS_NH_White_Alone ²	Numeric	Local Isolation Score: White Alone, Non-Hispanic	
Yost_Overall_Quintile ²	Numeric	Yost Quintile (Total US)	
Yost_State_Quintile ²	Numeric	Yost Quintile (by state)	
Enc_St ²	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 ²	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 ²	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 ²	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.	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.

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. The data in this file derive from the Engagement, Calendar Year, Pap Test, HPV Test, HPV Vaccine, Procedures, Pregnancy, and Prior CDE files. See data source listed in 'DAU Notes about Harmonization' section for individual variables.

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. Each record in this file represents a distinct date on which at least one event occurred; events include cohort entry or exit (see [Appendix: Cohort Member Study Periods for Primary and Secondary Analyses](#) (p. 121) for further detail), an eligibility status change, a Pap and/or HPV test, and/or a procedure. Each event record contains information on the date of the event, cohort member eligibility status, and any available screen or procedure information (where applicable). Eligibility status variables were defined in each record both beginning and conclusion of the events that occurred at each record. Status variables include risk, pregnancy, HIV, cervical cancer, absent cervix, and abnormal test or procedure results.

The Cervical Timeline file supports primary analyses only and must be used in conjunction with data in the Participant and Calendar Year files to support sensitivity analyses.

In addition to presenting data submitted by the sites in the original CDE files, the following three derived variables were created to convey relevant screening process measures: risk status, which conveys screening eligibility and anticipated follow-up (see [Appendix: Cervical Timeline File Risk Status Assignment](#) (p. 125) for further detail); screening modality, which conveys the type of screening used (see [Appendix: Cervical Timeline File Screening Modality](#) (p. 126) for further detail); and screening results, which conveys the combined screen results relevant to follow-up recommendations (see [Appendix: Cervical Timeline File Screening Result](#) (p. 128) for further detail).

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.
EventAge ²	Numeric	Event Age Age of the cohort member at the time of the event.	This variable was calculated as follows: EventDSR/365.25, then rounded to the lowest integer. Cohort members age 17 at PROSPR cohort entry rounded up to age 18. No other data element was modified including EventDSR.
EventDSR ^{1,2}	Numeric	Event Days Since Reference Date	Days since date of birth. See Global Values: DSR (p. 5) for further detail.
EventMth ²	Numeric Integer between 1-12	Event Month Month event occurred.	This variable was determined as follows: SAS Formula: Month(EventDate) See Global Values: Dates (p. 4) for further detail.
EventYr ²	Numeric 4-digit integer	Event Year Year that event occurred in.	This variable was determined as follows: SAS Formula: Year(EventDate) See Global Values: Dates (p. 4) for further detail.
EventCE ²	Numeric 0=No 1=Yes	Cohort Entry Flag Event flagging record at which cohort member entered the PROSPR cohort. Cohort member will have multiple cohort entries if multiple records exist in the Engagement file.	
EventPap ²	Numeric 0=No 1=Yes	Pap Test Flag Event flagging record when patient had a Pap test.	This variable was derived from the PapDate and PapDSR variables in the PapTest file. If this variable has a value of '1', then the following variables will have a corresponding value: PapInd, PapResult, PapProvidPerform, PapFacilityIDPerform, PapInsur, PapHPVModalityMod. All sites now submit specimen collection dates; thus, the majority of Pap tests will occur on the same day as HPV tests when either a co-test or a reflex test. However, there are rare instances in

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			which an HPV test linked with a Pap test was reported on a later date. To identify the features of an HPV test with a separate event date from the linked Pap test, recorded the PapHPV14daysDSR variable in the Pap test event to identify associated HPV tests occurring on a later date.
EventHPV ²	Numeric 0=No 1=Yes	HPV Test Flag Event flagging record at which patient had a HPV test.	<p>This variable was derived from the HPVDate and HPVColIDSR variables in the HPVTest file.</p> <p>If this variable has a value of '1' then the following variables will have a corresponding value: HPVInd, HPVResult, HPVgenoType, HPVProvIDPerform, HPVFacilityIDPerform, HPVInsur.</p> <p>All sites now submit specimen collection dates; thus, the majority of Pap tests will occur on the same day as HPV tests when either a co-test or a reflex test. However, there are rare instances in which an HPV test linked with a Pap test was reported on a later date. To identify the features of an HPV test with a separate event date from the linked Pap test, recorded the PapHPV14daysDSR variable in the Pap test event to identify associated HPV tests occurring on a later date.</p>
EventProc ²	Numeric 0=No 1=Yes	Procedure Flag Event flagging record at which patient had a procedure. Procedure includes any colposcopy/biopsy and treatment.	<p>This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file.</p> <p>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.</p>
EventColpoBx ²	Numeric 0=No 1=Yes	Colpo/Biopsy Flag Event flagging record at which patient had a colposcopy/biopsy procedure.	<p>This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 1=Colpo, No Biopsy, 2=Colpo, Biopsy, 3=Endocervical Curretage (Brush), 4=Endometrial Biopsy, 15=Biopsy, No Colpo</p> <p>If this variable has a value of '1', then EventProc will have a value of '1'.</p>
EventTx ²	Numeric	Treatment Flag	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	0=No 1=Yes	Event flagging record at which patient had a treatment procedure.	file where the variable value was any of the following: 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. If this variable has a value of '1', then EventProc will have a value of '1'.
EventHIV ²	Numeric 0=No 1=Yes	First HIV positive Flag Event flagging record at which patient is first documented as being diagnosed with HIV.	This variable was derived from the HIVEver variable in the Prior file or the ComorbHIV variable in the Calendar Year file. 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. If the cohort member was diagnosed with HIV during the study period, the EventDSR associated with this variable will correspond to January 1st of the calendar year in which the ComorbHIV variable was first flagged in the Calendar Year file.
EventPregBegin ²	Numeric 0=No 1=Yes	Pregnancy Begin Flag Event flagging record at which patient documented as being pregnant.	This variable was derived from the PregBeginDate and PregBeginDSR variables in the Pregnancy file. Pregnancy begin date was not truncated if preceded cohort entry. 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 ²	Numeric 0=No 1=Yes	Pregnancy End Flag Event flagging record at which patient is documented as no longer being pregnant.	This variable was derived from the PregEndDate and PregEndDSR variables in the Pregnancy file. 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 ²	Numeric	First Cancer Diagnosis Flag	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	0=No 1=Yes	Event flagging record at which patient is first diagnosed with cervical cancer.	This file was derived from the CaCervDate and CaCervDSR variables in the Prior file or the DxDate and DxDSR variables in the Cancer file.
EventNoCervix ²	Numeric 0=No 1=Yes	Cervix Not Present Flag Event flagging record at which patient is first categorized as not having a cervix.	This variable was derived from the PriorCervRem variable in the Prior file or the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 10=Hysterectomy, NOS, 12=Trachelectomy, 13=Total hysterectomy, 14=Radical hysterectomy or modified radical hysterectomy. 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 procedure was used.
EventCEnd ²	Numeric 0=No 1=Yes	Cohort Exit Flag Event flagging record at which cohort member exited the PROSPR cohort. Cohort member will have multiple cohort exits if multiple records exist in the Engagement file.	This variable was derived from the CutOffDate and CutOffDSR variables in the Engagement file.
Event18 ²	Numeric 0=No 1=Yes	Age 18 Flag Event flagging record at which cohort member turned age 18.	This variable was derived from the BirthMth, BirthDy, and BirthYr variables in the Participant file. This variable was determined as follows: SAS Formula: Age18DSR=INTNX('year',BirthDate,18,"same")-BirthDate
Event21 ²	Numeric 0=No 1=Yes	Age 21 Flag Event flagging record at which cohort member turned age 21.	This variable was derived from the BirthMth, BirthDy, and BirthYr variables in the Participant file. This variable was determined as follows: SAS Formula: Age21DSR=INTNX('year',BirthDate,21,"same")-BirthDate
Event25 ²	Numeric 0=No 1=Yes	Age 25 Flag Event flagging record at which cohort member turned age 25.	This variable was derived from the BirthMth, BirthDy, and BirthYr variables in the Participant file. This variable was determined as follows:

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			SAS Formula: Age25DSR=INTNX('year',BirthDate,25,"same")-BirthDate
Event30 ²	Numeric 0=No 1=Yes	Age 30 Flag Event flagging record at which cohort member turned age 30.	This variable was derived from the BirthMth, BirthDy, and BirthYr variables in the Participant file. This variable was determined as follows: SAS Formula: Age30DSR=INTNX('year',BirthDate,30,"same")-BirthDate
Event66 ²	Numeric 0=No 1=Yes	Age 66 Flag Event flagging record at which cohort member turned age 66.	This variable was derived from the BirthMth, BirthDy, and BirthYr variables in the Participant file. This variable was determined as follows: SAS Formula: Age66DSR=INTNX('year',BirthDate,66,"same")-BirthDate
EventHPVVac ²	Numeric 0=No 1=Yes	Vaccination Flag Event flagging record at which patient had vaccination administered.	This variable was derived from the HPVVacDate and HPVVacDSR variables in the HPV Vaccine file. 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 ²	Numeric	First Primary Care Encounter Flag	This variable was derived from the PCPVisitDSR variable in the Encounters file.
StatusRiskPrior ^{1,2}	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.	This variable was derived from the StatusPapAbnpriorEver, StatusHPVPosPriorEver, StatusProcPriorEver, StatusCervCancer, StatusScrnSurvPrior, StatusScrnSurv, EventAge, and StatusHIV variables in the Cervical Timeline file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			See Appendix: Cervical Timeline File Risk Status Assignment (p. 125) for further detail.
StatusRisk ^{1,2}	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.	This variable was derived from the StatusPapAbnpriorEver, StatusHPVPosPriorEver, StatusProcPriorEver, StatusCervCancer, StatusScrnSurvPrior, StatusScrnSurv, EventAge, and StatusHIV variables in the Cervical Timeline file. See Appendix: Cervical Timeline File Risk Status Assignment (p. 125) for further detail.
StatusHIV ²	Numeric 0=No 1=Yes	HIV+ Status at Event Status flagging HIV diagnosis at end of record.	This variable was derived from the HIVEver variable in the Prior file or the ComorbHIV variable in the Calendar Year file. Record was flagged as HIV-positive on and after EventHIV record.
StatusPreg ²	Numeric 0=No 1=Yes	Pregnant at Event Status flagging pregnancy at end of record.	This variable was derived from the PregBeginDate, PregBeginDSR, PregEndDate, and PregEndDSR variables in the Pregnancy file.
StatusNoCervix ²	Numeric 0=No 1=Yes	Cervix Not Present at Event Status flagging whether cohort member has a cervix at end of record.	This variable was derived from the PriorCervRem variable in the Prior file or the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 10=Hysterectomy, NOS, 12=Trachelectomy, 13=Total hysterectomy, 14=Radical hysterectomy or modified radical hysterectomy. 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 procedure was used.
StatusHPVVacDose ^{1,2}	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.	This variable was derived from the HPVVacDate and HPVVacDSR variables in the HPV Vaccine file. 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

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			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 ²	Numeric 0=No 1=Yes	Invasive Cervical Cancer at Event Status flagging invasive cervical cancer diagnosis at end of record	This filed was derived from the CaCervDate and CaCervDSR variables in the Prior file or the DxDate and DxDSR variables in the Cancer Registry file.
StatusHPVPosPrior ³	Numeric 0=No 1=Yes	HPV positive Status 3 Years Prior to Event Status flagging HPV diagnosis within prior 3 years of record.	This variable was derived from the HPVDate, HPVColIDSR, and HPVResult variables in the HPVTest file.
StatusHPVPosPriorEver ²	Numeric 0=No 1=Yes	HPV positive Status Ever Prior to Event Status flagging HPV diagnosis ever prior to record.	This variable was derived from the HPVDate, HPVColIDSR, and HPVResult variables in the HPVTest file.
StatusPapAbnprior ³	Numeric 0=No 1=Yes	Pap Abnormal Status 3 Years Prior to Event Status flagging abnormal Pap test result within prior 3 years of record.	This variable was derived from the PapDate, PapDSR, PapResult, and PapUnsat variables of the PapTest file. A Pap test result was defined as abnormal under the following circumstances: for Pap alone tests, a result at least as severe as an ASC-US; for co-tests, a result more severe than ASC-US if HPV-negative or a result more severe than NILM if HPV-positive; and for primary HPV tests, HPV-positive.
StatusPapAbnpriorEver ²	Numeric 0=No 1=Yes	Pap Abnormal Status Ever Prior to Event Status flagging abnormal Pap test result ever prior to record..	This variable was derived from the PapDate, PapDSR, PapResult, and PapUnsat variables of the PapTest file. A Pap test result was defined as abnormal under the following circumstances: for Pap alone tests, a result at least as severe as an ASC-US; for co-tests, a result more severe than ASC-US if HPV-negative or a result more severe than NILM if HPV-positive; and for primary HPV tests, HPV-positive.
StatusProcPrior ³	Numeric	Any Procedure Status 3 Years Prior to Event Status flagging procedure within prior 3 years of record.	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	0=No 1=Yes	Procedure includes any colposcopy/biopsy and treatment.	If this variable has a value of '1', then either the EventColpoBx or EventTx variable should have a value of '1', based on whether the procedure was a colposcopy/biopsy or a treatment.
StatusProcPriorEver ²	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.	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file. If this variable has a value of '1', then either the EventColpoBx or EventTx variable should have a value of '1', based on whether the procedure was a colposcopy/biopsy or a treatment.
StatusColpoBxPrior3 ²	Numeric 0=No 1=Yes	Any Colpo Biopsy Status 3 Years Prior to Event Status flagging colposcopy/biopsy within prior 3 years of record.	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 1=Colpo, No Biopsy, 2=Colpo, Biopsy, 3=Endocervical Curretage (Brush), 4=Endometrial Biopsy, 15=Biopsy, No Colpo If this variable has a value of '1', then EventProc should have a value of '1'.
StatusColpoBxPriorEver ²	Numeric 0=No 1=Yes	Any Colpo Biopsy Status Ever Prior to Event Status flagging colposcopy/biopsy ever prior to record.	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 1=Colpo, No Biopsy, 2=Colpo, Biopsy, 3=Endocervical Curretage (Brush), 4=Endometrial Biopsy, 15=Biopsy, No Colpo If this variable has a value of '1', then EventProc should have a value of '1'.
StatusTxPrior3 ²	Numeric 0=No 1=Yes	Any Treatment Status 3 Years Prior to Event Status flagging treatment within prior 3 years of record.	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 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.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			If this variable has a value of '1', then EventProc should have a value of '1'.
StatusTxPriorEver ²	Numeric 0=No 1=Yes	Any Treatment Status Ever Prior to Event Status flagging treatment ever prior to record.	This variable was derived from the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 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. If this variable has a value of '1', then EventProc should have a value of '1'.
PapInd ^{1,2}	Numeric 1 = Screening 2 = Repeat 3 = Diagnostic 4 = Surveillance	Pap Indication Variable flagging Pap test indication for test record.	This variable was derived from the PapDate, PapDSR, PapResult, and PapUnsat variables in the Pap Test file, the HPVDate and HPVDSR variables in the HPV Test file, and the ProcDate and ProcDSR variables in the Procedures file. Pap indication category was prioritized as follows: if the Pap test occurred within 4 months of a prior unsatisfactory Pap test, then assigned repeat; if the Pap test occurred on the same date as a procedure (EventProc=1), then assigned diagnostic; if a Pap test occurred following an HPV+ diagnosis, prior abnormal Pap test, or procedure, then assigned surveillance; otherwise, assigned screening.
PapType ^{1,2}	Numeric 1 = Cervical 2 = Vaginal 3 = Procedure Vaginal	Specimen derived from a cervical or vaginal Pap test If record indicates both vaginal and cervical Pap test, record set to '1 = Cervical'; if there is a history of hysterectomy and the submitted PapType was either 'Cervical' or 'Unknown', record set to '3 = Procedure Vaginal'; and if record does not indicate either vaginal or cervical test, and there is no history of hysterectomy, record set to '99 = Unknown'.	This variable was derived from the PapType variable in the Pap Test file and either the PriorCervRem variable in the Prior file, or the ProcTypeCerv, ProcDate, and ProcDSR variables in the Procedures file where the variable value was any of the following: 10=Hysterectomy, NOS, 12=Trachelectomy, 13=Total hysterectomy, 14=Radical hysterectomy or modified radical hysterectomy.
PapHPVModalityMod ^{1,2}	Numeric 1=Screening Pap alone/reflex	Screening Modality Screening modality calculated using algorithm that	This variable was derived from the PapDate, PapDSR, PapResult, and PapUnsat variables in the Pap Test file, the HPVDate, HPVDSR, HPVResult,

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	<p>2=Diagnostic/Surveillance Pap alone/reflex 3=Surveillance co-test 4=Diagnostic co-test 5=Screening co-test 6=Surveillance HPV alone 7=Primary HPV alone 8=Screening Other/Unknown 9=Surveillance/Diagnostic Other/Unknown</p>	<p>incorporates time between Pap and HPV tests and site-submitted HPV indication.</p>	<p>and HPVInd variables in the HPV Test file, and the PapInd derived variable.</p> <p>See Appendix: Cervical Timeline File Screening Modality (p. 126) for further detail.</p>
PapHPV14daysDSR ^{1,2}	<p>Numeric</p>	<p>Date of HPV test associated with Pap test, days since reference</p> <p>This variable will be populated with the date and HPV test occurred if an HPV test occurred 1-14 days of a Pap test. HPV tests on the same day as the Pap test are assigned the EventDSR value.</p>	<p>This variable was derived from the PapDSR variable in the Pap Test file and the HPVDSR in the HPV Test file.</p> <p>All sites now submit specimen collection dates; thus, the majority of Pap tests will occur on the same day as HPV tests when either a co-test or a reflex test. However, there are rare instances in which an HPV test linked with a Pap test was reported on a later date. To identify the features of an HPV test with a separate event date from the linked Pap test, recorded the PapHPV14daysDSR variable in the Pap test event to identify associated HPV tests occurring on a later date.</p>
PapHPVResult ^{1,2}	<p>Numeric</p> <p>1=NILM only 2=NILM/HPV- (Any) 3=NILM/HPV+ (Pooled/Other) 4=NILM/HPV+ (16/18) 5=ASC-US only 6=ASC-US/HPV- 7=ASC-US/HPV+ (Any) 8=LSIL only OR LSIL/HPV- OR LSIL/HPV insufficient 9=LSIL/HPV+ (Any) 10=HSIL only OR HSIL/HPV- 11=HSIL/HPV+ (Any) 12=AGC only OR AGC/HPV- 13=AGC/HPV+ (Any) 14=ASC-H only OR ASC-H/HPV- 15=ASC-H/HPV+ (Any) 16=Suspicious for cancer regardless of HPV</p>	<p>Screen Result, with associated Pap test and HPV test results combined</p> <p>The most severe Pap and HPV test results that occurred on the same day were reported.</p> <p>Pap test result and HPV test result combined when HPV test completed within 14 days of Pap test.</p>	<p>This variable was included to provide a single screen result for all Pap tests and HPV tests.</p> <p>This variable will be non-missing when the PapHPVModalityMod variable is non-missing.</p> <p>See Appendix: Cervical Timeline File Screening Result (p. 128) for further detail.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	17=HPV- Only 18=HPV+ (16/18) Only 19=HPV+ (Pooled) Only 20=HPV + (Other) Only 21=Insufficient (Pap or HPV combo that must be repeated) 22=Unknown Pap/HPV+(16/18) 23=ASC-US / HPV unknown 99=Pap Only Unknown and/or HPV Unknown (Except + 16/18)		
PapProvIDPerform ²	Character	Pap Performing Provider The provider that performed the Pap test for test record.	This variable was derived from the ProvIDPerform variable in the Pap Test file.
PapFacilityIDPerform ²	Character	Pap Performing Facility ID The clinic/facility associated with the provider that performed the Pap test for test record.	This variable was derived from the FacilityIDPerform variable in the Pap Test file.
HPVInd	Numeric 1=Reflex (ASC-US) 2=Cotest 3=Primary Reflex	Indication for HPV test, check all that apply	This variable was derived from the HPVInd variable in the HPV Test file.
HPVProvIDPerform ²	Character	Event HPV Performing Provider The provider that performed the HPV test for test record.	This variable was derived from the ProvIDPerform variable in the HPV Test file.
HPVFacilityIDPerform ²	Character	Event HPV Performing Facility ID The clinic/facility associated with the provider that performed the HPV test for test record.	This variable was derived from the FacilityIDPerform variable in the HPV Test file.
ProcType ²	Numeric 1=Colpo, No Biopsy 2=Colpo, Biopsy 3=Endocervical Curretage (Brush) 4=Endometrial Biopsy 5=LEEP 6=Cone 7=Cryotherapy	Procedure Type Variable flagging procedures for test record. Procedures include all procedure types	This variable was derived from the ProcTypeCerv in the Procedures file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	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		
ProcResult ²	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	Procedure Result Variable flagging procedure pathology results for test record.	This variable was derived from the ProcDx in the Procedures file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	38=Non-Cervical Cancer of the Cervix, No Other Information Available 40=No biopsy		
ProcProvIDPerform ²	Character	Event Procedure Performing Provider The provider that performed the procedure for test record.	This variable was derived from the ProvIDPerform variable in the Procedures file.
ProcFacilityIDPerform ²	Character	Event Procedure Performing Facility ID The clinic/facility associated with the provider that performed the procedure for test record.	This variable was derived from the FacilityIDPerform variable in the Procedures file.
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID All records set to site at which cohort member entered the cohort. Thus, the providingSite recorded for each cohort member in the Cervical Timeline file should match the providingSite recorded for each cohort member in the Participant file.	See Global Values: ProvidingSite (p. 6) for further detail.
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See Global Values: ddVersion (p. 6) for further detail.

SCREENING EPISODE FILE

Overview

This analytic file contains a record of every screening episode for each eligible cohort member. The goal of this file is to connect all events related to a single screening event in one record.

Record Structure

One record per cohort member per screening episode. A screening episode always begins at a Pap/HPV test at which a female enters under average risk or unknown risk status and includes any subsequent procedures that occur prior to the next Pap/HPV test and within 13 months of the incident Pap/HPV test. This file only includes information from first cohort entry through first cohort exit based on primary analysis definitions (see [Appendix: Cohort Member Study Periods for Primary and Secondary Analyses](#) (p. 121) for further detail). If a cohort member subsequently re-entered the study, then the subsequent cohort period is not included.

General Harmonization Notes

The Screening Episode file supports primary analyses only and must be used in conjunction with data in the Participant and Calendar Year files to support sensitivity analyses.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

Variable Name	Variable Type, Formatting, Permissible Values	Description	DAU Harmonization Notes
PIDSite ^{1,2}	Character	Unique derived PROSPR ID	See Global Values: PIDSite (p. 6) for further detail.
BirthYr ²	Numeric	Birth date, year	This variable was derived from the BirthYr variable in the Participant file (p. 8).
Sex ²	Numeric 1=Male 2=Female 3=Other	Cohort member Sex All cohort member records set to '2'.	See METRICS Data Contributing Sites and Cohort Definitions: Cohort Entry (p. 2) for further detail.
RaceEthAggr ^{1,2}	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.	This variable was derived from the RaceEth_Drv (p. 11) variable in the Participant file.
Hispanic ²	Numeric 0=No 1=Yes	Hispanic or Latino origin	This variable was derived from the Hispanic (p. 8) variable in the Participant file.
RaceWhite ²	Numeric 0=No 1=Yes	Race White	This variable was derived from the RaceWhite (p. 8) variable in the Participant file.
RaceBlack ²	Numeric 0=No 1=Yes	Race Black or African-American	This variable was derived from the RaceBlack (p. 8) variable in the Participant file.
RaceAsian ²	Numeric 0=No 1=Yes	Race Asian	This variable was derived from the RaceAsian (p. 8) variable in the Participant file.
RaceAIAN ²	Numeric	Race American Indian or Alaska Native	

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	0=No 1=Yes		This variable was derived from the <u>RaceAIAN</u> (p. 8) variable in the Participant file.
RacePI ²	Numeric 0=No 1=Yes	Race Native Hawaiian or Other Pacific Islander	This variable was derived from the <u>RacePI</u> (p. 8) variable in the Participant file.
RaceMultipleNOS ²	Numeric 0=No 1=Yes	Race, Multiple (individual races not specified)	This variable was derived from the <u>RaceMultipleNOS</u> (p. 8) variable in the Participant file.
RaceOther ²	Numeric 0=No 1=Yes	Race Other	This variable was derived from the <u>RaceOther</u> (p. 9) variable in the Participant file.
CohortEntryMth ²	Numeric	Earliest cohort entry month	This variable was derived from the <u>CohortEntryMonth</u> (p. 9) variable in the Participant file.
CohortEntryYr ²	Numeric	Earliest cohort entry year	This variable was derived from the <u>CohortEntryYear</u> (p. 9) variable in the Participant file.
CohortEntryFirstDSR ²	Numeric	Earliest cohort entry date, days since reference date	Days since actual date of birth. See <u>Global Values: DSR</u> (p. 5) for further detail. This variable was derived from the <u>CohortEntryFirstDSR</u> (p. 9) variable in the Participant file.
CutOffMth ²	Numeric	Last cohort cut-off month	This variable was derived from the <u>CutOffDateMonth</u> (p. 10) variable in the Participant file.
CutOffYr ²	Numeric	Last cohort cut-off year	This variable was derived from the <u>CutOffDateYear</u> (p. 10) variable in the Participant file.
CutOffLastDSR ²	Numeric	Last cohort cut-off date, days since reference date	Days since actual date of birth. See <u>Global Values: DSR</u> (p. 5) for further detail. This variable was derived from the <u>CutOffSensDSR</u> (p. 11) variable in the Participant file.
CalendarYr ²	Numeric	Calendar year	This variable was derived from the <u>CalendarYear</u> (p. 15) variable in the Calendar Year file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

InsAggr ^{1,2}	Numeric 1=Medicare 2=Medicaid 3=InsOtherGov 4=InsCommerc 5=Uninsured/Medical Assistance 8=InsOther	Insurance (aggregate)	This variable was derived from the <u>CalYr_Ins_Drv</u> (p. 24) variable in the Calendar Year file.
Medicare ²	Numeric 0=No 1=Yes	Was the cohort member covered by Medicare?	This variable was derived from the <u>Medicare</u> (p. 17) variable in the Calendar Year file.
Medicaid ²	Numeric 0=No 1=Yes	Was the cohort member covered by Medicaid?	This variable was derived from the <u>Medicaid</u> (p. 17) variable in the Calendar Year file.
InsOtherGov ²	Numeric 0=No 1=Yes	Was the cohort member covered by any other federal or state health insurance program?	This variable was derived from the <u>InsOtherGov</u> (p. 17) variable in the Calendar Year file.
InsCommerc ²	Numeric 0=No 1=Yes	Was the cohort member covered by commercial and/or private health insurance?	This variable was derived from the <u>InsCommerc</u> (p. 17) variable in the Calendar Year file.
MedicalAssist ²	Numeric 0=No 1=Yes	Was the cohort member enrolled in a medical assistance charity program for the uninsured?	This variable was derived from the <u>MedicalAssist</u> (p. 17) variable in the Calendar Year file.
Uninsured ²	Numeric 0=No 1=Yes	Was the cohort member covered by high deductible insurance (as defined by the U.S. IRS (Pub 969)?	This variable was derived from the <u>MedicalAssist</u> (p. 17) variable in the Calendar Year file.
InsOther ²	Numeric 0=No 1=Yes	Was the cohort member covered by insurance other than those specified above?	This variable was derived from the <u>InsOther</u> (p. 17) variable in the Calendar Year file.
ProviderIDPCP ²	Numeric 0=No 1=Yes	Primary care provider ID	This variable was derived from the <u>ProviderIDPCP</u> (p. 17) variable in the Calendar Year file.
FacilityIDPCP ²	Numeric	Primary care provider clinic ID	This variable was derived from the <u>FacilityIDPCP</u> (p. 18) variable in the Calendar Year file.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	0=No 1=Yes		
CharlsonIndex ^{1,2}	Numeric	Charlson comorbidity index	This variable was derived from the CharlsonIndex (p. 23) and the CharlsonIndexInc_Drv (p. 24) variables. See Calendar Year File General Harmonization Notes (p. 13) for method by which comorbidity scores were calculated and analysis recommendations.
Smoke ²	Numeric 0=Never 1=Current 2=Former	Smoking Status - Social History	This variable was derived from the Smoke (p. 22) variable in the Calendar Year file.
AgeScreen ^{1,2}	Numeric	Age at screen in years	This variable was calculated as follows: ScrnDSR/365.25, then rounded to the lowest integer. Cohort members age 17 at PROSPR cohort entry rounded up to age 18.
DxDSR ²	Numeric	Date of cancer diagnosis, days since reference date	Days since actual date of birth. See Global Values: DSR (p. 5) for further detail. This variable was derived from the EventDSR variable in the Cervical Timeline file (p. 91) when EventCervCancer=1.
CohortEntryStatusRiskPrior ^{1,2}	Numeric 1= Surveillance 2=Not Screen Eligible (No Cervix) 3=Alternate Risk Schedule 4=Unknown Risk 5=Average Screen Risk	Screening prior risk status at PROSPR Cohort Entry	This variable was derived from the StatusRiskPrior variable in the Cervical Timeline file (p. 91) when EventCE=1. Status is assigned in the following descending order: Surveillance schedule, which includes cohort members with a history of abnormal Pap and/or HPV test results (NILM/HPV+ or worse) or a history of procedure (including colposcopy, regardless of pathology results. This includes those high-risk cohort members that were age eligible at time of surveillance, outside of the average risk screening guidelines (<21 or >65 years old), and/or with particular medical conditions (previous cancer or HIV diagnosis, pregnancy, or prior hysterectomy).

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			<p>Cohort members with a history of cancer or abnormal test results are included regardless of the presence of the cervix.</p> <p>Not Screen-Eligible schedule, which includes cohort members without a cervix and with no history of abnormal test results.</p> <p>Alternate Risk schedule, which includes cohort members that were not under surveillance and were either 18-20 years old, >65 years old, or diagnosed with HIV.</p> <p>Unknown Risk schedule, which includes cohort members with no history of a Pap or HPV test result or a colposcopy or diagnostic procedure as well as those cohort members with an NILM Pap test result with an unknown HPV test result.</p> <p>Average Screen Risk schedule, which includes cohort members that were screen-eligible or low-risk pregnant.</p>
ScrnStatusRiskPrior ^{1,2}	<p>Numeric</p> <p>1= Surveillance 2=Not Screen Eligible (No Cervix) 3=Alternate Risk Schedule 4=Unknown Risk 5=Average Screen Risk</p>	Risk (Pap/HPV) status prior Pap/HPV result	<p>This variable was derived from the StatusRiskPrior variable in the Cervical Timeline file (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5.</p> <p>Status is assigned in the following descending order:</p> <p>Surveillance schedule, which includes cohort members with a history of abnormal Pap and/or HPV test results (NILM/HPV+ or worse) or a history of procedure (including colposcopy, regardless of pathology results. This includes those high-risk cohort members that were age eligible at time of surveillance, outside of the average risk screening guidelines (<21 or >65 years old), and/or with particular medical conditions (previous cancer or HIV diagnosis, pregnancy, or prior hysterectomy). Cohort members with a history of cancer or abnormal test results are included regardless of the presence of the cervix.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			<p>Not Screen-Eligible schedule, which includes cohort members without a cervix and with no history of abnormal test results.</p> <p>Alternate Risk schedule, which includes cohort members that were not under surveillance and were either 18-20 years old, >65 years old, or diagnosed with HIV.</p> <p>Unknown Risk schedule, which includes cohort members with no history of a Pap or HPV test result or a colposcopy or diagnostic procedure as well as those cohort members with an NILM Pap test result with an unknown HPV test result.</p> <p>Average Screen Risk schedule, which includes cohort members that were screen-eligible or low-risk pregnant.</p>
ScrnDSR ²	Numeric	Screening (Pap/HPV), days since reference	<p>Days since actual date of birth.</p> <p>See Global Values: DSR (p. 5) for further detail.</p> <p>This variable was derived from EventDSR variable in the Cervical Timeline file (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5.</p>
ScrnPapHPVModalityMod ^{1,2}	Numeric 1=Screening Pap alone/reflex 2=Diagnostic/Surveillance Pap alone/reflex 3=Surveillance co-test 4=Diagnostic co-test 5=Screening co-test 6=Surveillance HPV alone 7=Primary HPV alone 8=Screening Other/Unknown 9=Surveillance/Diagnostic Other/Unknown	Screening Modality Screening modality calculated using algorithm that incorporates time between Pap and HPV tests and site-submitted HPV indication.	<p>This variable was derived from the PapHPVResult variable in the Cervical Timeline file (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5.</p> <p>See Appendix: Cervical Timeline File Screening Modality (p. 126) for further detail.</p>
ScrnPapHPVResult ^{1,2}	Numeric 1=NILM only 2=NILM/HPV- (Any)	Screen Result, with associated Pap test and HPV test results combined	<p>This variable was derived from the PapHPVResult variable in the Cervical Timeline file (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	<p>3=NILM/HPV+ (Pooled/Other) 4=NILM/HPV+ (16/18) 5=ASC-US only 6=ASC-US/HPV- 7=ASC-US/HPV+ (Any) 8=LSIL only OR LSIL/HPV- OR LSIL/HPV insufficient 9=LSIL/HPV+ (Any) 10=HSIL only OR HSIL/HPV- 11=HSIL/HPV+ (Any) 12=AGC only OR AGC/HPV- 13=AGC/HPV+ (Any) 14=ASC-H only OR ASC-H/HPV- 15=ASC-H/HPV+ (Any) 16=Suspicious for cancer regardless of HPV 17=HPV- Only 18=HPV+ (16/18) Only 19=HPV+ (Pooled) Only 20=HPV + (Other) Only 21=Insufficient (Pap or HPV combo that must be repeated) 22=Unknown Pap/HPV+(16/18) 23=ASC-US / HPV unknown 99=Pap Only Unknown and/or HPV Unknown (Except + 16/18)</p>	<p>The most severe Pap and HPV test results that occurred on the same day were reported.</p> <p>Pap test result and HPV test result combined when HPV test completed within 14 days of Pap test.</p>	<p>This variable was included to provide a single screen result for all Pap tests and HPV tests.</p> <p>See Appendix: Cervical Timeline File Screening Result (p. 128) for further detail.</p>
ScrnPapHPVProvIDPerform ²	Character	Screening performing (ordering) provider ID	<p>This variable was derived from the PapProvIDPerform and/or the HPVProvIDPerform variables in the Cervical Timeline file (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5. In the event of distinct Pap test and HPV test providers, the Pap test provider was recorded unless the test was a primary HPV screen, in which case the HPV test provider was recorded.</p>
ScrnPapHPVFacilityIDPerform ²	Character	Screening performing (ordering) facility ID	<p>This variable was derived from the PapProvIDPerform and/or the HPVProvIDPerform variables in the Cervical Timeline file (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5. In the event of distinct Pap test and HPV test facilities, the Pap test facility was recorded unless the test was a primary HPV screen, in which case the HPV test facility was recorded.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

ScrnStatusRisk ^{1,2}	<p>Numeric</p> <p>1= Surveillance 2=Not Screen Eligible (No Cervix) 3=Alternate Risk Schedule 4=Unknown Risk 5=Average Screen Risk</p>	Risk (Pap/HPV) status after Pap/HPV result	<p>This variable was derived from the StatusRisk variable in the <u>Cervical Timeline file</u> (p. 91) when EventPap=1 or EventHPV=1 and StatusRiskPrior=4 or StatusRiskPrior=5.</p> <p>Status is assigned in the following descending order:</p> <p>Surveillance schedule, which includes cohort members with a history of abnormal Pap and/or HPV test results (NILM/HPV+ or worse) or a history of procedure (including colposcopy, regardless of pathology results. This includes those high-risk cohort members that were age eligible at time of surveillance, outside of the average risk screening guidelines (<21 or >65 years old), and/or with particular medical conditions (previous cancer or HIV diagnosis, pregnancy, or prior hysterectomy). Cohort members with a history of cancer or abnormal test results are included regardless of the presence of the cervix.</p> <p>Not Screen-Eligible schedule, which includes cohort members without a cervix and with no history of abnormal test results.</p> <p>Alternate Risk schedule, which includes cohort members that were not under surveillance and were either 18-20 years old, >65 years old, or diagnosed with HIV.</p> <p>Unknown Risk schedule, which includes cohort members with no history of a Pap or HPV test result or a colposcopy or diagnostic procedure as well as those cohort members with an NILM Pap test result with an unknown HPV test result.</p> <p>Average Screen Risk schedule, which includes cohort members that were screen-eligible or low-risk pregnant.</p>
ScrnStatusFUp ^{1,2}	<p>Numeric</p> <p>1=Return for screening 3 or 5 years depending on modality</p>	Screening follow-up recommendation (for screening status only)	This variable was derived from the PapHPVResult variable in the <u>Cervical Timeline file</u> (p. 91) when EventPap=1 or EventHPV=1.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	<p>2=Repeat pap (unsatisfactory) 3=Surveillance co-test in 1 year 4=Immediate colposcopy (may also need treatment) 5=Provider/Site choice</p>		<p>If the screen result was NILM, NILM HPV-, or HPV-, then return to routine screening recommended. If the screen result was NILM HPV+ (pooled/other strain), ASC-US only or ASC-US HPV-, or HPV+ (pooled/other strain), then surveillance co-test in one year recommended. If the screen result was as or more severe than ASC-US HPV+ (any strain) or HPV+ (16/18) (irrespective of cytology), then immediate colposcopy recommended. If the screen result was insufficient, then repeat Pap test recommended. If the screen result was ASC-US HPV Unknown, then follow-up recommendation was left to provider discretion (provider/site choice).</p>
DiagType ²	<p>Numeric</p> <p>1=Colpo, no biopsy 2=Colpo, biopsy/ECC 3=Cone</p>	Diagnostic/surveillance procedure type	<p>This variable was derived from the ProcType variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p>
DiagDSR ²	<p>Numeric</p>	Diagnostic/surveillance procedure, days since reference	<p>Days since actual date of birth.</p> <p>See Global Values: DSR (p. 5) for further detail.</p> <p>This variable was derived from EventDSR variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p>
DiagStatusRiskPrior ^{1,2}	<p>Numeric</p> <p>1= Surveillance 2=Not Screen Eligible (No Cervix) 3=Alternate Risk Schedule 4=Unknown Risk 5=Average Screen Risk</p>	Diagnostic/surveillance risk prior status	<p>This variable was derived from the StatusRiskPrior variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p> <p>Status is assigned in the following descending order:</p> <p>Surveillance schedule, which includes cohort members with a history of abnormal Pap and/or HPV test results (NILM/HPV+ or worse) or a history of procedure (including colposcopy, regardless of pathology results. This includes those high-risk cohort members that were age eligible at time of surveillance, outside of the average risk screening guidelines (<21 or >65 years old), and/or with particular medical conditions (previous cancer or HIV diagnosis, pregnancy, or prior hysterectomy). Cohort members with a history of cancer or</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			<p>abnormal test results are included regardless of the presence of the cervix.</p> <p>Not Screen-Eligible schedule, which includes cohort members without a cervix and with no history of abnormal test results.</p> <p>Alternate Risk schedule, which includes cohort members that were not under surveillance and were either 18-20 years old, >65 years old, or diagnosed with HIV.</p> <p>Unknown Risk schedule, which includes cohort members with no history of a Pap or HPV test result or a colposcopy or diagnostic procedure as well as those cohort members with an NILM Pap test result with an unknown HPV test result.</p> <p>Average Screen Risk schedule, which includes cohort members that were screen-eligible or low-risk pregnant.</p>
DiagResult ²	<p>Numeric</p> <p>1=Normal, no tissue sample 2=Normal, tissue examined 3=HPV reactive changes 4=LSIL (CIN1) 5=HSIL (CIN2 or worse) 6=AIS, CIS 7=Invasive cancer 8=Other (non-cervical) 9=Insufficient for diagnosis/unsatisfactory tissue</p>	Diagnostic/surveillance result	<p>This variable was derived from the ProcResult variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p>
DiagProvIDPerform ²		Diagnostic/surveillance procedure Performing provider ID	<p>This variable was derived from the ProcProvIDPerform variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p>
DiagFacilityIDPerform ²		Diagnostic/surveillance procedure Performing facility ID	<p>This variable was derived from the ProcFacilityIDPerform variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p>
DiagStatusRisk ^{1,2}	<p>Numeric</p> <p>1= Surveillance 2=Not Screen Eligible (No</p>	<p>Diagnostic/surveillance risk status after procedure</p> <p>Cohort member risk status will always be set to '1' after a procedure.</p>	<p>This variable was derived from the StatusRisk variable in the Cervical Timeline file (p. 91) when EventColpoBx=1.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

	<p>Cervix)</p> <p>3=Alternate Risk Schedule</p> <p>4=Unknown Risk</p> <p>5=Average Screen Risk</p>		<p>Status is assigned in the following descending order:</p> <p>Surveillance schedule, which includes cohort members with a history of abnormal Pap and/or HPV test results (NILM/HPV+ or worse) or a history of procedure (including colposcopy, regardless of pathology results. This includes those high-risk cohort members that were age eligible at time of surveillance, outside of the average risk screening guidelines (<21 or >65 years old), and/or with particular medical conditions (previous cancer or HIV diagnosis, pregnancy, or prior hysterectomy). Cohort members with a history of cancer or abnormal test results are included regardless of the presence of the cervix.</p> <p>Not Screen-Eligible schedule, which includes cohort members without a cervix and with no history of abnormal test results.</p> <p>Alternate Risk schedule, which includes cohort members that were not under surveillance and were either 18-20 years old, >65 years old, or diagnosed with HIV.</p> <p>Unknown Risk schedule, which includes cohort members with no history of a Pap or HPV test result or a colposcopy or diagnostic procedure as well as those cohort members with an NILM Pap test result with an unknown HPV test result.</p> <p>Average Screen Risk schedule, which includes cohort members that were screen-eligible or low-risk pregnant.</p>
TxType ²	<p>Numeric</p> <p>1=LEEP</p> <p>2=Cryotherapy</p> <p>3=Laser</p> <p>4=Hysterectomy, partial</p> <p>5=Hysterectomy, total</p>	Treatment procedure type	<p>This variable was derived from the ProcType variable in the <u>Cervical Timeline file</u> (p. 91) when EventTx=1.</p>
TxDSR ²	<p>Numeric</p>	Treatment procedure, days since reference	<p>Days since actual date of birth.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			<p>See Global Values: DSR (p. 5) for further detail.</p> <p>This variable was derived from the EventDSR variable in the Cervical Timeline file (p. 91) when EventTx=1.</p>
TxResult ²	<p>Numeric</p> <p>1=Normal, no tissue sample 2=Normal, tissue examined 3=HPV reactive changes 4=LSIL (CIN1) 5=HSIL (CIN2 or worse) 6=AIS, CIS 7=invasive cancer 8=Other (non-cervical) 9=Insufficient for diagnosis/unsatisfactory tissue</p>	Treatment procedure result	<p>This variable was derived from the ProcResult variable in the Cervical Timeline file (p. 91) when EventTx=1.</p>
TxProvIDPerform ²	Character	Treatment procedure Performing provider ID	<p>This variable was derived from the ProcProvIDPerform variable in the Cervical Timeline file (p. 91) when EventTx=1.</p>
TxFacilityIDPerform ²	Character	Treatment procedure Performing facility ID	<p>This variable was derived from the ProcFacilityIDPerform variable in the Cervical Timeline file (p. 91) when EventTx=1.</p>
TxStatusRisk ^{1,2}	<p>Numeric</p> <p>1= Surveillance 2=Not Screen Eligible (No Cervix) 3=Alternate Risk Schedule 4=Unknown Risk 5=Average Screen Risk</p>	<p>Treatment risk status after procedure</p> <p>Cohort member risk status will always be set to '1' after a treatment procedure.</p>	<p>This variable was derived from the StatusRisk variable in the Cervical Timeline file (p. 91) when EventTx=1.</p> <p>Status is assigned in the following descending order:</p> <p>Surveillance schedule, which includes cohort members with a history of abnormal Pap and/or HPV test results (NILM/HPV+ or worse) or a history of procedure (including colposcopy, regardless of pathology results. This includes those high-risk cohort members that were age eligible at time of surveillance, outside of the average risk screening guidelines (<21 or >65 years old), and/or with particular medical conditions (previous cancer or HIV diagnosis, pregnancy, or prior hysterectomy). Cohort members with a history of cancer or abnormal test results are included regardless of the presence of the cervix.</p>

PROSPR METRICS DATA DICTIONARY [Version 2.2]

			<p>Not Screen-Eligible schedule, which includes cohort members without a cervix and with no history of abnormal test results.</p> <p>Alternate Risk schedule, which includes cohort members that were not under surveillance and were either 18-20 years old, >65 years old, or diagnosed with HIV.</p> <p>Unknown Risk schedule, which includes cohort members with no history of a Pap or HPV test result or a colposcopy or diagnostic procedure as well as those cohort members with an NILM Pap test result with an unknown HPV test result.</p> <p>Average Screen Risk schedule, which includes cohort members that were screen-eligible or low-risk pregnant.</p>
ScrnStatusFUp13Mth ^{1,2}	<p>Numeric</p> <p>1=Complete, normal screen 2=Complete, now under surveillance 3=Incomplete, missing diagnostic or treatment</p>	Episode status at the end of 13 month interval post-screen	<p>This variable was derived from the ScrnStatusFUp, DiagResult, and TxResult variables in the Screening Episode file.</p> <p>If the screen status follow-up recommendation was to return to routine screening, then episode status was complete normal screen. If the screen status was surveillance co-test, immediate colposcopy, provider discretion, or repeat screen and the diagnostic surveillance or treatment result were known (i.e., not insufficient or unknown), then episode status was complete and now under surveillance. If the screen status follow-up recommendation was surveillance co-test, immediate colposcopy, provider discretion, or repeat screen and the diagnostic surveillance and treatment result was insufficient or unknown, then episode status was incomplete.</p>
extractDate ²	Numeric	Date each record was generated	See Global Values: extractDate (p. 5) for further detail.
providingSite ²	Numeric	PROSPR Site ID	See Global Values: ProvidingSite (p. 6) for further detail.

PROSPR METRICS DATA DICTIONARY [Version 2.2]

		All records set to site at which cohort member entered the cohort. Thus, the providingSite recorded for each cohort member in the Cervical Timeline file should match the providingSite recorded for each cohort member in the Participant file.	
ddVersion ²	Numeric	Data dictionary version number closest to submission date	See <u>Global Values: ddVersion</u> (p. 6) for further detail.

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

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 and utilization cohorts 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, has to 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 restrictions, 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 [METRICS Data Contributing Sites and Cohort Definitions: Cohort Entry](#) (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. Utilization cohort members were permitted to re-enter the study upon subsequent primary care utilization. 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](#) (p. 7), which conveys the cohort exit incorporating the lack of primary care utilization cut-off (CutOffDateYear, CutOffDateMonth, CutOffDateDay, CutOffDSR) and the cohort exit date that does not incorporate the lack of primary care utilization cut-off (CutOffDateSensYear, CutOffDateSensDay, CutOffDateSenseMonth, 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. Importantly, for the enrollment-based cohort, all variables related to cohort exit will be equivalent (e.g., CutOffDSR will be the same as CutOffSensDSR).

The [Calendar Year file](#) (p. 13), 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.

The [Engagement file](#) (p. 26), which conveys the updated cohort exit date (CutOffDateMonth, CutOffDateDay, CutOffDateYear, and CutOffDSR, where applicable), and reason for cohort exit (CutOffReason, where applicable) as well as additional cohort periods (where applicable).

The [Cervical Timeline file](#) (p. 91), which conveys every cohort entry and exit noted in the Engagement file (EventCE and EventCEnd).

The [Screening Episode file](#) (p. 91), which conveys the first cohort entry and last exit noted in the Participant file (CohortEntryFirstDSR and CutOffLastDSR).

For primary analyses, utilization member cohort entry and exit dates can be identified using the Participant (CohortEntryFirstDSR, CutOffDSR), Engagement (CohortEntryDSR, CutOffDSR), Cervical Timeline (EventCE and EventCEnd), and Screening Episode file (CohortEntryFirstDSR and CutOffLastDSR) files. Records for calendar years in which the cohort member was in the study based on utilization can be identified (UtilCalYr_Drv=1). Of note, only the first cohort entry and last cohort exit is noted in the Participant file, while the Engagement and Cervical Timeline files report all cohort entries and exits.

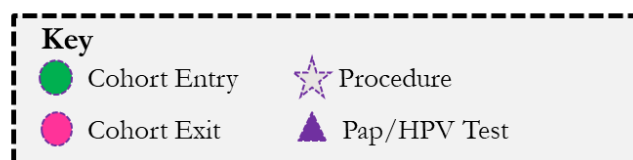
For secondary analyses, utilization member first cohort entry and last cohort exit dates can be identified using the Participant file (CohortEntryFirstDSR, CutOffSensDSR). Intermittent cohort entries and exits can be identified in the Engagement file by looking for all cohort exits and subsequent cohort entries that have a cut off reason that corresponds to death, aging out, or moving out of geographic service area.

See graphic below for an example of how a cohort member's study period shifts with the introduction of a lack of primary care utilization cut-off.

Site-Submitted Data with Single Cohort Entry and Exit



Incorporation of 37 Month 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 2019 due to administrative cut-off at the end of the study period (December 31st, 2019). For secondary 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 2019 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, CutOffDSR), Engagement (CohortEntryDSR, CutOffDSR), Cervical Timeline (EventCE and EventCEnd), and Screening Episode file (CohortEntryFirstDSR, CutOffLastDSR) 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). Thus, the two fundamental changes for this cohort member were (1) the division of the originally submitted cohort period into two cohort periods, and (2) the reduction in the final cohort exit from the end of the study period.

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

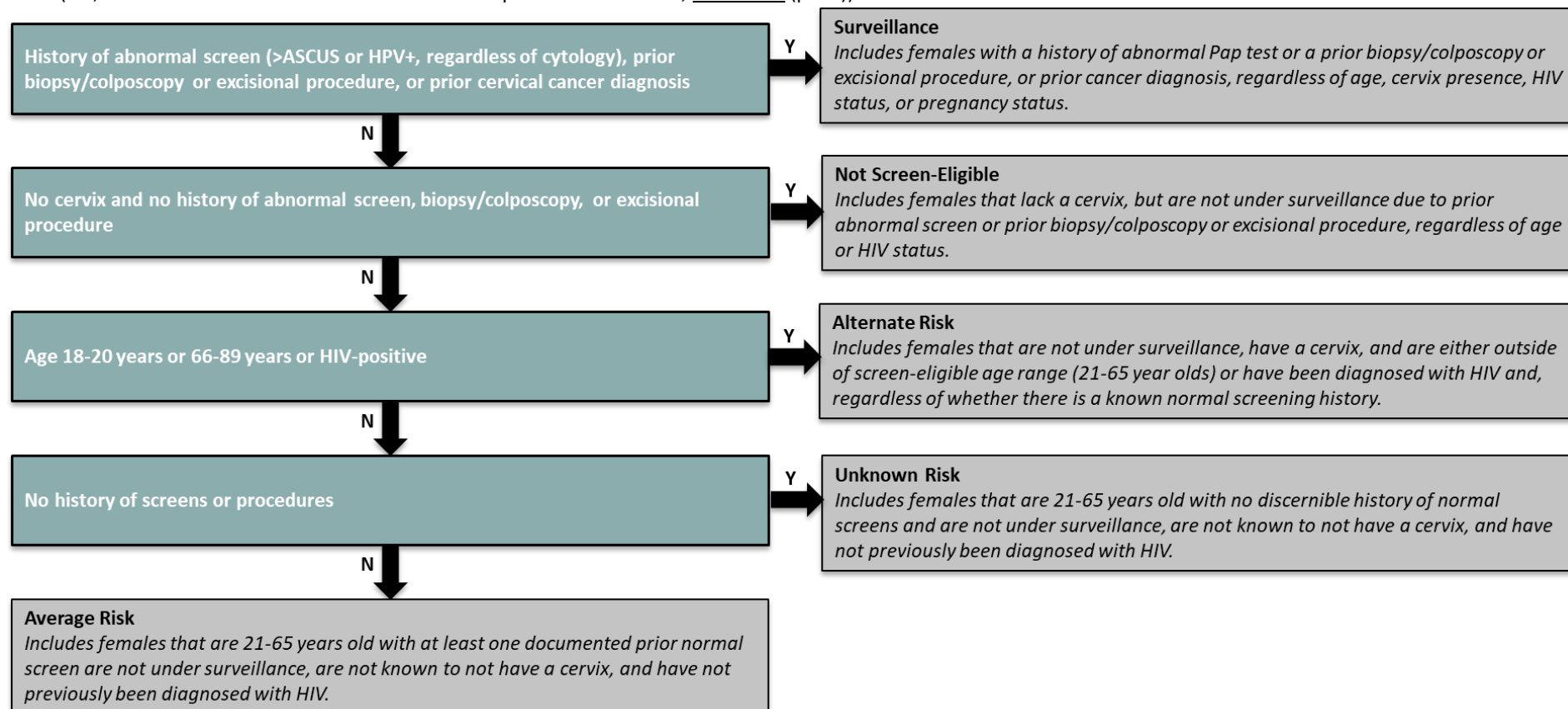
The METRICS sites expended considerable effort harmonizing the level at which facilities were defined across the three disparate healthcare system. The goals for harmonizing the Facility file were (1) to understand at which level organizational policies are established for delivery of screening (clinical co-investigator input and qualitative data documented that this varies across sites); and (2) to execute analyses, including nested multi-level analyses, at a consistent level.

FacilityID represents the following: for site A, a distinct campus/medical center/building location (level 2); for site B, a distinct clinic/specialty location (level 3); and for site C, either a distinct clinic/specialty location (level 3) in the Calendar Year, HPV Vaccine, Encounter, and Procedures files and a distinct campus/medical center location (level 2) in the Pap Test and HPV Test files. FacilityIDRelatedPhys represents a distinct medical center/campus/building (level 2) for sites B and C, allowing an analyst to “roll-up” distinct FacilityIDs representing clinic/specialty location (level 3) into medical center/campus/building locations (level 2) to permit harmonization of the facility level concept across sites. Future data submissions will include the clinic/specialty location (level 3) at each encounter to permit harmonization of the clinic/specialty location (level 3) concept across all three sites.

CERVICAL TIMELINE FILE

Risk Status Assignment

Risk status is a reflection of a female's screening eligibility and conveys anticipated screening type and frequency as well as the type of follow-up recommended based on screening outcome. As described in the diagram below, risk status was algorithmically assigned based on a female'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 (p. 96)) and at the event of an event (i.e., risk status established as a result of the completion of the event, StatusRisk (p. 97)).



Once a female enters the 'Surveillance' risk status, the female remains under this status for the duration of the cohort period. Likewise, a female that enters the 'Not Screen-Eligible' risk status may only transition to the 'Surveillance' risk status. A female that enters the 'Alternate Risk' status due to being under age 21 years old may transition to any of the other statuses; however, if a female enters the 'Alternate Risk' status due to being over age 65 or HIV diagnosis may only transition to the 'Not Screen-Eligible' or 'Surveillance' risk statuses. A female that enters the 'Unknown Risk' status may transition to any of the other statuses. A female that enters the 'Average Risk' status may transition to the 'Surveillance', 'Not Screen-Eligible', or 'Alternate Risk' statuses, as the documented normal screen required to enter the 'Average Risk' status precludes 'Unknown Risk' status assignment.

Screening Modality

Test modality describes the type of test used to screen a female subset on the rationale for screening. Test modality is determined based on the indication for the Pap test as well as the HPV test. There are currently three broad classes of guideline-approved modalities available for females ages 30-65; only the first modality (Pap alone/reflex) is recommended for females ages 21-29:

1. **Pap alone/reflex** – Clinicians review only Pap test results to determine next clinical step, unless the findings are ASC-US, in which the lab performs a reflex HPV test to help the clinician manage findings (**Note:** in some systems, LSIL results in females over age 50 are also reflexed for management);
2. **Co-test** – Clinicians review both Pap and HPV test results to determine the next clinical step; and
3. **Primary HPV screening** – Clinicians review HPV test results; if positive for HPV 16/18, immediately route to colposcopy, and if positive for another high-risk HPV type, then lab performs a reflex Pap test to help clinician manage findings.

The broad classes of test modality are subset based on the rationale for screening as follows:

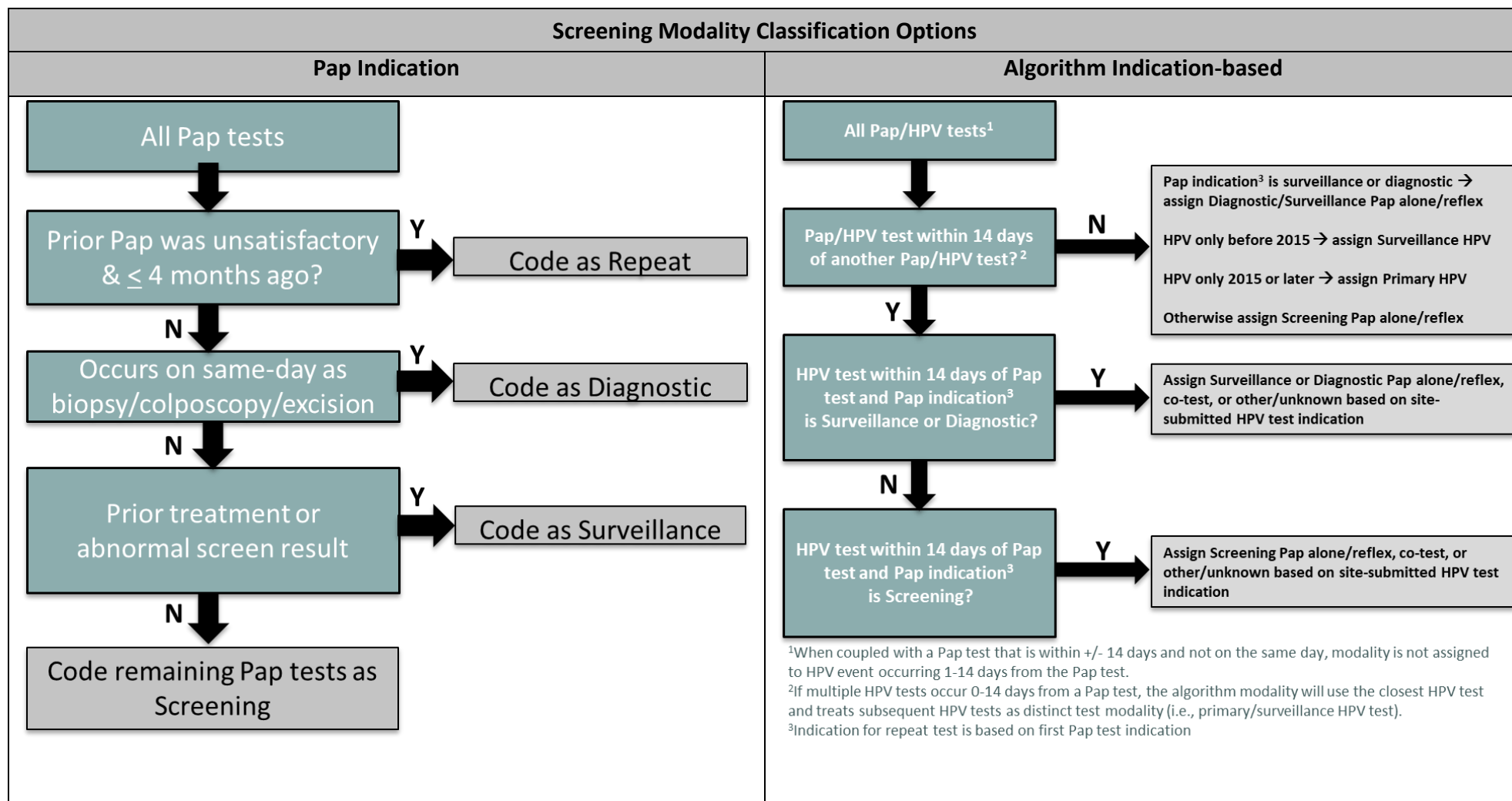
1. **Repeat** – Test occurs within four months of an unsatisfactory screen;
2. **Diagnostic** – Test occurs concomitant with a procedure to diagnose observed symptoms or confirm prior abnormal result before procedure;
3. **Surveillance** – Test is a part of elevated risk screening due to prior history of abnormality; and
4. **Screening** – Test is a part of routine screening and does not indicate prior or current history of abnormality.

Pap test indication was first assigned based on a female's prior history (see diagram below, left column). If a prior unsatisfactory Pap test occurred within four months of the present Pap test, then the Pap test indication was set to 'Repeat'. Otherwise, if a Pap test was completed on the same day as a biopsy, colposcopy, or excisional procedure, then the Pap test indication was set to 'Diagnostic'. Otherwise, if a female had a prior history of procedures (biopsy, colposcopy, or excisional treatment) or abnormal screen results (HPV+ or at least as severe as ASC-US), then the Pap test indication was set to 'Surveillance'. Otherwise, the Pap test indication was set to 'Screening'.

Test modality was then determined by combining Pap test indication and site-submitted HPV test indication to distinguish reflex, co-test, and other/unknown modalities (see diagram below, right column). If an HPV test did not occur within 14 days of a Pap test, the test modality was assigned as either 'Surveillance HPV' or 'Primary HPV' based on the prior Pap test indication. Otherwise, if an HPV test occurred within 14 days of a Pap test and the Pap test indication was 'Surveillance' or 'Diagnostic', then the modality was assigned as 'Surveillance Co-Test', 'Surveillance Pap alone/Reflex', 'Surveillance Other/Unknown', 'Diagnostic Co-Test', 'Diagnostic Pap alone/Reflex', or 'Diagnostic Other/Unknown' based on site-submitted HPV test indication. Otherwise, if an HPV test occurred within 14 days of a Pap test, the Pap test indication was 'Screening', then the modality was assigned as 'Screening Co-Test', 'Screening Pap alone/Reflex', or 'Screening Other/Unknown'.

Importantly, once a female enters the 'Surveillance' risk status due to a history of a cervical abnormality (abnormal prior screen or biopsy, colposcopy, excisional procedure), then the female is always under surveillance in the present dataset, and the only recommended testing option is 'Surveillance Co-test'.

When an HPV test is coupled with a Pap test and not on the same day, screening modality is not assigned to HPV event occurring 1-14 days from the Pap test. If multiple HPV tests occur 0-14 days from a Pap test, the algorithm modality will recorded the closest HPV test and treats subsequent HPV tests as distinct test modality (i.e., primary/surveillance HPV test). Screening modality for a repeat test is based on the most proximal prior Pap test indication that is not repeat.



Screening Result

Screen results were reported based on whether a Pap test and/or HPV test was completed. Result combinations are not exhaustive and were collapsed to reflect follow-up guidelines.

For co-tests and reflex tests, the strain of HPV was noted based on the Pap test cytology results. HPV strain was only noted for normal cytology results (including negative for intraepithelial lesion or malignancy [NILM] and benign glandular cells [BGC]), 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 ASC-US. Of note, ASC-US cytology results were separately distinguished as HPV positive, HPV negative, HPV not tested, and HPV unknown to reflect clinical follow-up guidance.

For primary HPV tests and normal cytology results, HPV genotype was assigned based on decreasing order of severity. An HPV test was considered to be 16/18 positive if the test result was positive either for individual genotyping for strains 16 and/or 18 or 16/18, 16/18/45, or 18/45 pooled strains. Otherwise, an HPV test was considered to be pooled 16/18 positive if the test result was positive for pooled genotypes that included 16 and/or 18 strains. Otherwise, an HPV test was considered to be positive for other strains if the test result was positive for pooled genotypes that did not include 16 or 18 strains. Otherwise, an HPV test was noted to be negative, with the strain tested unspecified.