PROSPR II METRICS PUDS2 Data Dictionary

# Overall Instructions

The **M**ultil**E**vel *Op****T****imization of the* Ce**R**v**I**cal **C**ancer **S**creening Process *in* Diverse Settings & Populations (METRICS) *PROSPR II* *Research* *Center* (PRC) was designed to elucidate multilevel factors that hamper or facilitate the cervical cancer screening process and reduce disparities. METRICS has three data-contributing sites from diverse healthcare settings – 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. 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 PROSPR II METRICS public use dataset (PUDS2) contains a subset of the METRICS LCDSS and excludes site identifiers.

This data dictionary provides information and guidance for using the METRICS PUDS2 in analyses. The METRICS PUDS2 should be considered for pilot use only and is not appropriate for publication due to the inability to adjust for site in analyses, as there are known differences by site within the dataset. Hence, the METRICS sites strongly caution against publishing any analyses using these data. Data users wishing to analyze data for publication should apply to use the LCDSS through the PROSPR DataShare site (<https://healthcaredelivery.cancer.gov/prospr/datashare.html>). Appropriate use-case scenarios for the METRICS PUDS2 include the following:

1. Descriptive analyses for grant submissions;
2. Descriptive analyses to determine value in pursuing collaboration with PROSPR; and
3. Descriptive analyses for an educational project

While the following datasets are large, there is the potential for a data user to cross-tab variables in such a way that leads to small cell values (fewer than 5 people). The METRICS sites strongly encourage data users to collapse categories to prevent reporting small cell values.

The METRICS PUDS2 data dictionary is divided into the following three sections:

1. Section 1: Overview – A summary of the METRICS PUDS2 cohort entry and exit criteria as well as data sources;
2. Section 2: File Description – A description of the Cohort and Episode data files and common data elements (CDEs); and
3. Section 3: Appendices – A list of abbreviations used throughout the document, exclusions consort, and decision-making guidance for select data elements.

# Section 1: Overview – Description of the METRICS Data Sources and Cohort Criteria

## METRICS PUDS2 Cohort Entry and Exit Criteria

### **METRICS Cohort Definition**

Females entered the METRICS cohort from January 1st, 2010 through December 31st, 2019 if 18-89 years old and met additional site-specific criteria as follows: visited a system primary care clinic or elected/attributed a system primary care providers and/or lived within the system’s geographic service area.

Females exited the METRICS cohort at the earliest occurrence of any of the following:

1. Aged out (90 years old);
2. Death; or
3. Moved out of the system’s geographic service area, disenrolled from the system, stopped using the primary care system for three years, or did not elect/attribute a system primary care provider for more than three months, depending on the site-specific requirements.

### **METRICS PUDS2 Cohort Definition**

METRICS cohort members entered the METRICS PUDS2 cohort on their birthdate in 2017 if the following conditions were met:

1. 21-65 years old as of December 31st, 2017;
2. Enrolled/engaged in the METRICS study cohort; and
3. Either under surveillance due to prior abnormality, were of unknown risk due to absence of prior screening history, or were of average risk based on prior normal screening history.

METRICS cohort members exited the METRICS PUDS2 cohort at the earliest occurrence of any of the following:

1. Reached the end of the cohort period (i.e., two years passed since cohort entry);
2. Became no longer cohort-eligible due to turning 66 years old, becoming diagnosed with HIV, or having a cervix removal without surveillance indication prior to the end of the cohort period;
3. Left the cohort following disenrollment, moving out of the coverage area, or three years lapsed without primary care utilization prior to the end of the cohort period; or
4. Death

# Section 2: File Description – Overview of All Data Files and CDEs

## Cohort File

### **Overview**

This file contains one record of static covariates for every METRICS PUDS2 cohort member.

### **Record Structure**

One record per METRICS PUDS2 cohort member.

Data derive from site electronic health record systems that were either built and maintained by the respectice healthcare system or by Epic.

Blank indicates missing or not applicable data. See CDE Notes for the meaning of blank data for specific variables.

### General Harmonization Notes

BMI, comorbidity score, and insurance information are reflective of data collected over the entirety of the calendar year in which the female entered the cohort.

C**omorbidity data were collected from 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\*. 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.** The PRC calculated the derived comorbidity scores as of December 31st of the calendar year using all available data from the entire year (e.g., for a calendar year record for 2012, use data from 1/1/2012 through 12/31/2012 to calculate the score for the year 2012). If a cohort member had 'Diabetes with Chronic Complication' and 'Diabetes without Chronic Complication', only the weight value for 'Diabetes with Chronic Complication' (+2, not +3) was counted. If a cohort member had 'Metastatic Solid Tumor' and 'Any Malignancy', only the weight value for 'Metastatic Solid Tumor' (+6, not +8) was counted. If a cohort member had 'Moderate or Severe Liver Disease' and 'Mild Liver Disease', only the weight value for 'Moderate or Severe Liver Disease' (+3, not +5) was counted.

\**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*

| **CDE Name** | **CDE Concept** | **CDE Type, Formatting, Permissible Values** | **CDE Notes** |
| --- | --- | --- | --- |
| UID | Unique METRICS PUDS2 Cohort Member ID | Character, 10 characters in length |  |
|
|
|
|
| BirthYear | Birth date, year | Numeric, continuous 4-digit integer |  |
|
|
|
|
|
|
| EntryYear | Cohort Entry Date, Year | Numeric, continuous 4-digit integer |  |
|
|
| EntryDSR | Cohort Entry Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate |  |
|
|
| EntryAge | Age at Cohort Entry | Numeric, categorical 2-digit integer |  |
| EntryRisk | Risk Status at Cohort Entry | Numeric, categorical 1-digit integer  1 = Surveillance 2 = Unknown Risk 3 = Average Screen Risk | Reflects risk status at cohort entry, prior to any events that may have also occurred on the cohort entry date.  See [Appendix: Episode File Risk Status Assignment](#AppRiskStat) (p. 16) for further detail. |
|
|
| EntryIns | Insurance at Cohort Entry | Numeric, categorical 1-digit integer  1 = Medicare 2 = Medicaid 3 = Other Government or Uninsured 4 = Commercial 5 = Other/Unknown | Reflects insurance noted during calendar year of cohort entry.  If multiple insurance designations were made within a single calendar year, assigned Medicaid, then Medicare, then Commercial, then Other Government.  See [Cohort File: General Harmonization Notes](#_General_Harmonization_Notes) (p. 3) for further detail. |
|
|
| EntryBMI | BMI at Cohort Entry | Numeric, categorical 1-digit integer  1 = <18.5 2 = 18.5–24.9 3 = 25.0–29.9 4 = >30.0 9 = Unknown | Reflects body mass index (BMI, NIH NHLBI scale) noted during calendar year of cohort entry.  See [Cohort File: General Harmonization Notes](#_General_Harmonization_Notes) (p. 3) for further detail. |
|
|
| EntryComorb | Comorbidity at Cohort Entry | Numeric, categorical 1-digit integer  0 = 0 1 = 1 2 = 2+ 9 = Unknown | Reflects all comorbidities cohort member was diagnosed with during calendar year of cohort entry.  See [Cohort File: General Harmonization Notes](#_General_Harmonization_Notes) (p. 3) for further detail. |
|
|
| ExitYear | Cohort Exit Date, Year | Numeric, continuous 4-digit integer |  |
| ExitDSR | Cohort Exit Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate |  |
| ExitAge | Age at Cohort Exit | Numeric, categorical 2-digit integer |  |
| ExitRisk | Risk Status at Cohort Exit | Numeric, categorical 1-digit integer  1 = Surveillance 2 = Unknown Risk 3 = Average Screen Risk | Reflects risk status at cohort exit, after all events that occurred on or prior to cohort exit date.  See [Appendix: Episode File Risk Status Assignment](#AppRiskStat) (p. 16) for further detail. |
| ExitReason | Cohort Exit Reason | Numeric, categorical 1-digit integer  1 = End of cohort period 2 = No longer cohort-eligible 3 = Left cohort 4 = Death | End of cohort period indicates two years from cohort entry passed.  See [Overview: METRICS PUDS2 Cohort Definition](#OverviewCohortEx) (p. 2) for further detail. |
| RaceEth | Race/Ethnicity | Numeric, categorical 1-digit integer  1 = White  2 = Black  3 = Asian  4 = Hispanic  9 = Other | Race/Ethnicity category was prioritized in descending order as follows: Hispanic; any single race designation; and if no designation made, other.  Asian includes those people who identified as Asian, Native Hawaiian, or Pacific Islander.  Other includes those people who identified as Native American, Alaskan Native, multiple single race designations, other, and unknown. |
|
|
|
|
|
|
| CohortPreg | Cohort Pregnancy | Numeric, categorical 1-digit integer  0 = No  1 = Yes | Indicates whether a cohort member was pregnant at any point from cohort entry through cohort exit |
| PriorTestDSR | Last Prior Test Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate | If no prior test was documented, then this variable will be blank. |
|
|
| PriorTestRes | Last Prior Test Result | Numeric, categorical 1-digit integer  0 = No Prior Test  1 = Normal  2 = HPV+  3 = Abnormal, Low-Grade  4 = Abnormal, High-Grade  9 = Other | Reflects most significant test result identified in cytology report  No prior test indicates that the cohort member did not have a known test result prior to PUDS2 cohort entry.  Normal indicates Pap alone result of NILM, co-test result of NILM HPV-, and primary HPV result of HPV-.  HPV+ indicates co-test result of NILM HPV+ (pooled high-risk) or primary HPV result of HPV+ (pooled high-risk).  Abnormal, Low-Grade indicates Pap alone result of ASC-US or reflex/co-test result of ASC-US HPV- or ASC-US HPV unknown.  Abnormal, High-Grade indicates co-test result of ASC-US HPV+, Pap alone/co-test result of LSIL or worse, regardless of HPV test, or co-test or primary HPV result of HPV 16/18+, regardless of cytology.  Other indicates that a cytology and/or HPV test result was not available or that the specimen was not deemed satisfactory to evaluate. |
|
|
| PC2017Ct | Primary Care Visits in 2017 | Numeric, categorical 1-digit integer  0 = 0  1 = 1  2 = 2  3 = 3+ | Reflects the number of primary care encounters that occurred in calendar year 2017 on or after cohort entry and before cohort exit |
|
|
| PC2018Ct | Primary Care Visits in 2018 | Numeric, categorical 1-digit integer  0 = 0  1 = 1  2 = 2  3 = 3+ | Reflects the number of primary care encounters that occurred in calendar year 2018 on or after cohort entry and before cohort exit  Note that 2018 will be the most reflective of primary care engagement for most PUDS2 cohort members, because this is the only full calendar year in which PUDS2 cohort members can be in the cohort. |
|
|
| PC2019Ct | Primary Care Visits in 2019 | Numeric, categorical 1-digit integer  0 = 0  1 = 1  2 = 2  3 = 3+ | Reflects the number of primary care encounters that occurred in calendar year 2019 on or after cohort entry and before cohort exit |
|
|

## Episode File

### Overview

This file contains one record for every test episode for every cohort member present in the Cohort file that received a test during the cohort period.

### Record Structure

One record per cohort member per test episode. A test episode always begins at a Pap/HPV test at which a female enters under surveillance, 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.

Data derive from site electronic health record systems that were either built and maintained by the respectice healthcare system or by Epic. Cancer diagnosis data derived from SEER, state, and/or hospital cancer registries and pathology data.

Blank indicates missing or not applicable data. See CDE Notes for the meaning of blank data for specific variables.

### General Harmonization Notes

**The Episode file supports primary analyses only and must be used in conjunction with data in the Cohort file to support sensitivity analyses.**

Information on cancer diagnoses derives from state registries and hospital pathology reports and are known to be complete through the end of 2018. The PUDS2 will be updated with 2019 Cancer Registry information upon availability from the states in which the METRICS data contributing sites reside.

| **CDE Name** | **CDE Concept** | **CDE Type, Formatting, Permissible Values** | **CDE Notes** |
| --- | --- | --- | --- |
| UID | Unique METRICS PUDS2 Cohort Member ID | Character, 10 characters in length |  |
|
|
| CancerDSR | Cancer Diagnosis Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate | Reflects date at which cervical cancer (CIN III/AIS+) was diagnosed on or after cohort entry  If no cancer diagnosis was documented, then this variable will be blank. |
| CancerYear | Cancer Diagnosis Date, Year | Numeric, continuous 4-digit integer | Reflects date at which cervical cancer (CIN III/AIS+) was diagnosed on or after cohort entry  If no cancer diagnosis was documented, then this variable will be blank. |
| TestDSR | Test Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate | Reflects date of test occurring on or after cohort entry |
|
|
| TestYear | Test Date, Year | Numeric, continuous 4-digit integer | Reflects year of test occurring on or after cohort entry |
|
|
| TestAge | Age at Test | Numeric, categorical 2-digit integer | Reflects age at test occurring on or after cohort entry |
|
|
|
|
| TestStatusRiskPrior | Risk Status Prior to Testing Episode | Numeric, categorical 1-digit integer  1 = Surveillance 2 = Unknown Risk 3 = Average Screen Risk | Reflects risk status at the beginning of the test, prior to any events that occurred within the testing episode   See [Appendix: Episode File Risk Status Assignment](#AppRiskStat) (p. 16) for further detail. |
|
|
| TestStatusRisk | Risk Status at the End of Testing Episode | Numeric, categorical 1-digit integer  1 = Surveillance 2 = Unknown Risk 3 = Average Screen Risk | Reflects risk status at the end of the test  See [Appendix: Episode File Risk Status Assignment](#AppRiskStat) (p. 16) for further detail. |
|
|
| TestPreg | Pregnant at Test | Numeric, categorical 1-digit integer  0 = No 1 = Yes | Reflects whether a cohort member was pregnant at the test |
| TestMod | Test Modality | Numeric, categorical 1-digit integer  1 = Screening Pap alone/reflex  2 = Diagnostic/Surveillance Pap alone/reflex  3 = Surveillance co-test  4 = Diagnostic co-test 5 = Screening co-test  9 = Other | Reflects test modality of the test  Test modality was determined by combining algorithmically-determined Pap test indication and site-submitted HPV test indication.  Other indicates primary HPV alone, surveillance HPV alone, or unknown.  See [Appendix: Episode File Test Modality](#_Test_Modality) (p. 17) for further detail. |
|
|
| TestResult | Test Result | Numeric, categorical 1-digit integer  1 = Normal  2 = HPV+  3 = Abnormal, Low-Grade  4 = Abnormal, High-Grade  9 = Other | Reflects most significant test result identified in cytology report  Normal indicates Pap alone result of NILM, co-test result of NILM HPV-, and primary HPV result of HPV-.  HPV+ indicates co-test result of NILM HPV+ (pooled high-risk) or primary HPV result of HPV+ (pooled high-risk).  Abnormal, Low-Grade indicates Pap alone result of ASC-US or reflex/co-test result of ASC-US HPV- or ASC-US HPV unknown.  Abnormal, High-Grade indicates co-test result of ASC-US HPV+, Pap alone/co-test result of LSIL or worse, regardless of HPV test, or co-test or primary HPV result of HPV 16/18+, regardless of cytology.  Other indicates that a cytology and/or HPV test result was not available or that the specimen was not deemed satisfactory to evaluate. |
|
|
| TestFURec | Test Follow-Up Recommendation | Numeric, categorical 1-digit integer  1 = Return for test in 3 or 5 years depending on modality 2 = Repeat test (unsatisfactory) 3 = Surveillance co-test in 1 year 4 = Immediate colposcopy | Reflects the test follow-up recommendation based on the test only (i.e., if a procedure followed the test, the results of the procedure were not incorporated into the test follow-up recommendation)  See [Appendix: Test Follow-Up Recommendation and Status](#AppTestFURecStat) (p. 18) for further detail. |
|
|
| DiagDSR | Diagnostic Procedure Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate | Reflects date at which a diagnostic procedure was performed on or after test date  If no diagnostic procedure was documented during the test episode, then this variable will be blank. |
|
|
| DiagYear | Diagnostic Procedure Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer | Reflects date at which a diagnostic procedure was performed on or after test date  If no diagnostic procedure was documented during the test episode, then this variable will be blank. |
|
|
| DiageAge | Age at Diagnostic Procedure | Numeric, categorical 2-digit integer | Reflects age at diagnostic procedure occurring on or after test date  If no diagnostic procedure was documented during the test episode, then this variable will be blank. |
|
|
| DiagPreg | Pregnant at Diagnostic Procedure | Numeric, categorical 1-digit integer  0 = No 1 = Yes | Reflects whether a cohort member was pregnant at the diagnostic procedure  If no diagnostic procedure was documented during the test episode, then this variable will be blank. |
| DiagType | Diagnostic Procedure Type | Numeric, categorical 1-digit integer  1 = No biopsy 2 = Biopsy/ECC | Reflects type of diagnostic procedure, which may have occurred on the same date as the test or a treatment procedure  No biopsy indicates that a colposcopy was completed, but no sample was collected. Biopsy/ECC indicates that a biopsy was collected with or without a colposcope and/or an endocervical curretage was completed.  If no diagnostic procedure was documented during the test episode, then this variable will be blank. |
|
|
| DiagResult | Diagnostic Procedure Result | Numeric, categorical 1-digit integer  1 = Normal  2 = Low-grade  3 = High-grade  4 = Cancer  9 = Other | Reflects results of diagnostic procedure  No tissue sample indicates that no tissue sample was collected during the diagnostic procedure.  Normal indicates that the tissue biopsy collected was normal, had mild inflammation, and/or atypia.  Low-grade indicates reactive changes consistent with HPV/condylomata, low-grade SIL, or CIN-I/mild dysplasia.  HSIL indicates high-grade squamous intraepithelial lesion, CIN I-II, CIN-I/mild dysaplsia in which high-grade dysplasia cannot be ruled out, CIN-II/moderate dysplasia, and CIN-II-III  Cancer indicates adenocarcinoma in situ of the cervix, CIN-III/severe dysplasia/carcinoma in situ, invasive cervical squamous cell carcinoma, invasive cervical adenocarcinoma, invasive cervical adenosquamous carcinoma, or other cervical cancer.  Other indicates that either no tissue sample was collected, a pathology result was not available, the specimen was not deemed satisfactory to evaluate, or another result not captured in the above options. Other indicates that a pathologyresult was not available, that the specimen was not deemed satisfactory to evaluate, or another result not captured in the above options.  If no diagnostic procedure was documented during the test episode, then this variable will be blank. |
|
|
| TxDSR | Treatment Procedure Date, Days Since Reference | Numeric, continuous 4-to-5-digit integer  Reference is birthdate | Reflects date at which a treatment procedure was performed on or after testing date  If no treatment procedure was documented during the test episode, then this variable will be blank. |
|
|
| TxYear | Treatment Procedure Date, Year | Numeric, continuous 4-digit integer | Reflects year of treatment procedure occurring on or after cohort entry  If no treatment procedure was documented during the test episode, then this variable will be blank. |
|
|
|
|
| TxAge | Age at Treatment Procedure | Numeric, categorical 2-digit integer | Reflects age at treatment procedure occurring on or after test date  If no treatment procedure was documented during the test episode, then this variable will be blank. |
|
|
| TxType | Treatment Procedure Type | Numeric, categorical 1-digit integer  1 = Excision/Ablation 2 = Hysterectomy | Reflects type of treatment procedure, which may have occurred on the same date as the test or a diagnostic procedure  Excision/ablation indicates LEEP, cone, excisional assay not otherwise specified, cryotherapy, or laser treatment.  Hysterectomy indicates partial/subtotal/supracervical hysterectomy, total, radical, or modified radical hysterectomy, trachelectomy, hysterectomy not otherwise specified; date of hysterectomy indicates either date hysterectomy was completed or date hysterectomy was first noted in patient surgical history.  If no treatment procedure was documented during the test episode, then this variable will be blank. |
|
|
| TxResult | Treatment Procedure Result | Numeric, categorical 1-digit integer  1 = Normal  2 = Low-grade  3 = High-grade  4 = Cancer  9 = Other | Reflects results of diagnostic procedure  No tissue sample indicates that no tissue sample was collected during the diagnostic procedure.  Normal indicates that the tissue biopsy collected was normal, had mild inflammation, and/or atypia.  Low-grade indicates reactive changes consistent with HPV/condylomata, low-grade SIL, or CIN-I/mild dysplasia.  HSIL indicates high-grade squamous intraepithelial lesion, CIN I-II, CIN-I/mild dysaplsia in which high-grade dysplasia cannot be ruled out, CIN-II/moderate dysplasia, and CIN-II-III  Cancer indicates adenocarcinoma in situ of the cervix, CIN-III/severe dysplasia/carcinoma in situ, invasive cervical squamous cell carcinoma, invasive cervical adenocarcinoma, invasive cervical adenosquamous carcinoma, or other cervical cancer.  Other indicates that either no tissue sample was collected, a pathology result was not available, the specimen was not deemed satisfactory to evaluate, or another result not captured in the above options.  If no treatment procedure was documented during the test episode, then this variable will be blank. |
|
|
| TestFUStatus | Episode Status at the End of 13-Month Interval Post-Test | Numeric, categorical 1-digit integer  1 = Complete, Normal 2 = Complete, Start Surveillance 3 = Complete, Continue Surveillance  4 = Incomplete | Reflects cohort member status at the end of the 13-month period starting with the test.  See [Appendix: Test Follow-Up Recommendation and Status](#AppTestFURecStat) (p. 18) for further detail. |
| TestEpComp | Episode Record Complete | Numeric, categorical 1-digit integer  0 = No 1 = Yes | Reflects whether the test episode was completely reflected in the episode record  No indicates that the next diagnostic or treatment procedure occurred after cohort exit but within 13 months of the test and so the record does not indicate what is known about patient follow-up. |
|
|
|
|
|
|

# Section 3: Appendices

## Abbreviations

**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

**PCC**: PROSPR Coordinating Center

**PCP**: primary care provider

**PRC**: PROSPR Research Center

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

**PUDS2:** PROSPR II Public Use Dataset

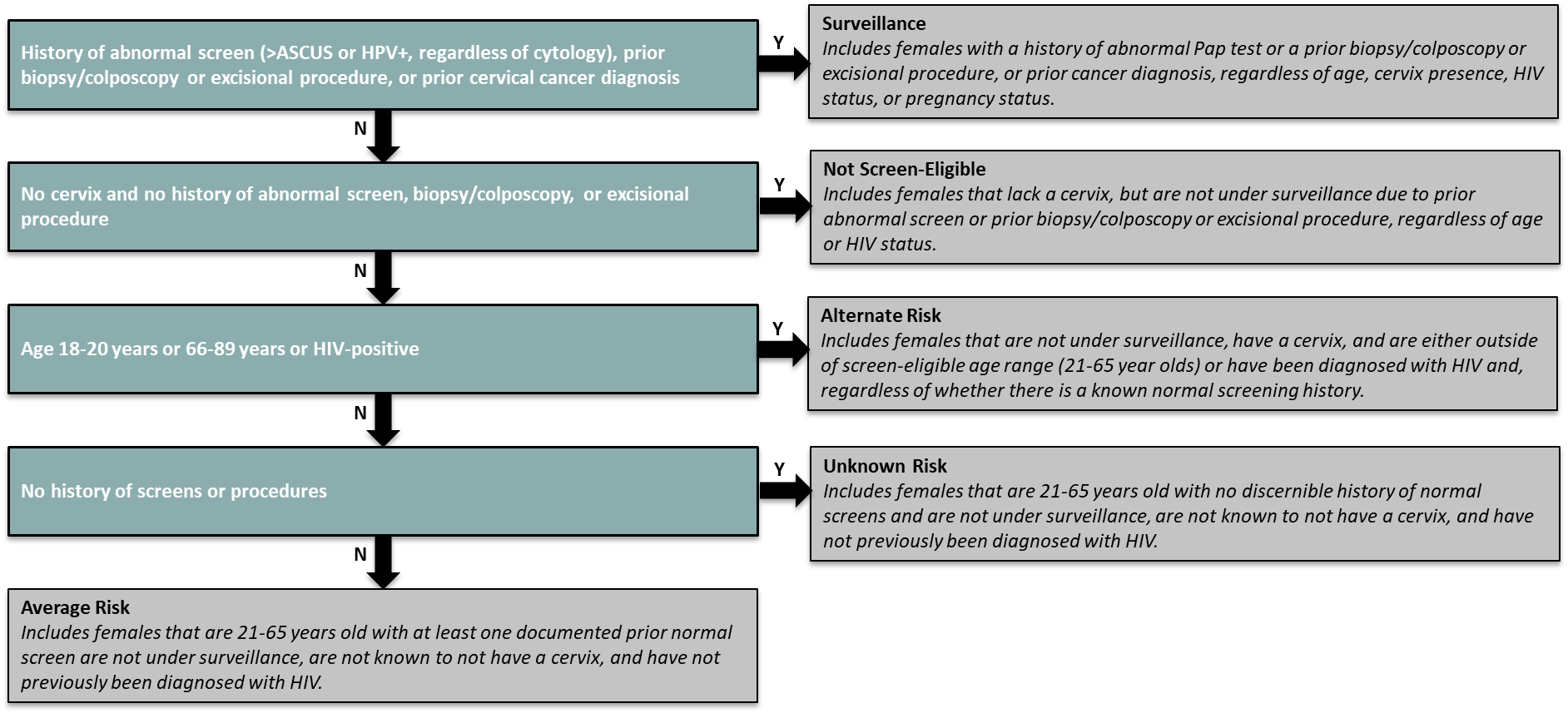
## Cohort File

### Exclusions Consort Diagram



## Episode File

### Risk Status Assignment

Risk status is a reflection of a person’s screening eligibility and conveys anticipated test type and frequency as well as the type of follow-up recommended based on test outcome. As described in the diagram below, risk status was algorithmically assigned based on a female’s prior testing and procedure history, age, cervix presence, and prior HIV diagnosis. Of note, at least a three-year look-back period was available for each cohort member from which prior tests, procedures, and diagnoses could be used to identify risk status; however, it is possible that a cohort member experienced these events at systems in which data was not available in the medical record. Risk status was determined at the start of an event (i.e., risk status established prior to an event, including [EntryRisk](#CohortEntryRisk) (p. 4) and [TestStatusRiskPrior](#TestStatusRiskPrior) (p. 9)) and at the event of an event (i.e., risk status established as a result of the completion of the event, including [ExitRisk](#CohortExitRisk) (p. 5) and [TestStatusRisk](#TestStatusRIsk) (p. 9)). ****

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. The present dataset only includes females that were under surveillance or were of unknown or average risk; if a female transitioned from unknown or average risk to not screen-eligible or alternate risk statuses, the female was removed from the cohort due to no longer being cohort-eligible (see [ExitReason](#CohortExitReason) (p. 5)).

### Test Modality

Test modality describes the type of test used to screen a female subset on the rationale for testing. 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 testing 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. 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 test 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. 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 (prior abnormal test result 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’.

### Test Follow-Up Recommendation and Status

Test follow-up recommendation was determined based on the cohort member’s risk status immediately prior to the test and the cytology/genotyping results of the test as follows:

1. If the test result was normal and the risk status at the beginning of the episode was average or unknown risk, then return to routine testing recommended;
2. If the test result was Normal HPV+/HPV+ or Abnormal, Low-Grade, then surveillance co-test in one year recommended;
3. If the test result was Abnormal, High-Grade, then immediate colposcopy recommended; and
4. If the test result was unsatisfactory or unknown, then repeat Pap test recommended.

Test follow-up status at the end of the 13-month period following the test was determined based on the test follow-up recommendation and procedures completed after the test as follows:

1. Complete, Normal indicates that the test follow-up recommendation (TestFURec) was to return to routine testing;
2. Complete, Start Surveillance indicates the following:
   1. Risk status at the beginning of the episode (TestStatusRIskPrior) was either Unknown or Average Risk; and
   2. Risk status at the end of the episode (TestStatusRisk) was surveillance; and either
   3. Test follow-up recommendation (TestFURec) was surveillance co-test or provider discretion, and if a procedure was completed, there was a known procedure result (i.e., not insufficient or unknown); or
   4. Test follow-up recommendation (TestFURec) was immediate colposcopy and there was a known procedure result (i.e., not insufficient or unknown).
3. Complete, Continue Surveillance indicates all of the following:
   1. Risk status at the beginning of the episode (TestStatusRIskPrior) and end of the episode (TestStatusRisk) were both surveillance; and either
   2. Test follow-up recommendation (TestFURec) was surveillance co-test or provider discretion, and if a procedure was completed, there was a known procedure result (i.e., not insufficient or unknown); or
   3. Test follow-up recommendation (TestFURec) was immediate colposcopy and there was a known procedure result (i.e., not insufficient or unknown).
4. Incomplete indicates any of the following:
   1. Test status follow-up recommendation (TestFURec) was repeat test; or
   2. Test status follow-up recommendation (TestFURec) was immediate colposcopy or provider discretion and the diagnostic and/or treatment results (DiagResult, TxResult) were insufficient or unknown (regardless of risk status at the beginning of the episode).