

## Prostate Cancer

We are using the prostate cancer definition used by the eMERGE program, with a slight modification to the control definition (**prostate\_cancer\_status\_emerge\_mod\_1**). You can review the eMERGE documentation for reference, but please follow this document for implementation.

- Modifications:
  - Combine “Control A” and “Control B” groups from the provided schematic figure into a single control group for analysis
  - Controls should be filtered based on age  $\geq$  the minimum age of identified cases to the age distributions of cases and controls are comparable within each study
  - We recommend using age at the sample draw used for genotyping for cases and controls to adjust for age in analyses. If this is not available, age at enrollment or consent would suffice. Traditionally, age at case diagnosis is often used in conjunction with control current age or enrollment age; however, this does not properly account for true age differences between cases and controls and can lead to age-associated somatic variation (e.g., CHIP) differing between cases and controls).

## Codes:

- See linked Excel document for Prostate Cancer Phenotyping Algorithm codes for the Tables mentioned in the definitions.

## Case definition:

- All of the following:
  - Male
  - Age  $\geq$  18 years
  - At least two occurrences of malignant neoplasm of prostate codes on different dates (Table 1)

## Control definition:

- All of the following:
  - Male
  - No occurrences of malignant neoplasm of prostate codes (Table 1)
  - No occurrences of non-malignant prostate neoplasm codes (Table 2)

- No prostatectomy codes (Table 3)
- At least one encounter
- Note that the flowchart distinguishes “Control A” and “Control B” groups based on presence of PSA testing (Table 4). Please report the count of each, but we anticipate that we will simply combine these into a single control group for analysis.
- Controls should be filtered based on age  $\geq$  the minimum age of identified cases to the age distributions of cases and controls are comparable within each study

**Covariate data to collect:**

- Age – please collect data (if available) on all ages listed below:
  - For cases:
    - Age at sample draw for genotyping (preferred)
    - Age at enrollment
    - Age at diagnosis
  - For controls:
    - Age at sample draw for genotyping (preferred)
    - Age at enrollment
    - Age at last contact/EHR record

### PrCa Phenotyping Algorithm

