Breast Cancer

We are using the breast cancer definition used by the eMERGE program, with some slight modifications (**breast_cancer_status_emerge_mod_1).** You can review the eMERGE documentation for reference, but please follow this document for implementation.

- Modifications:

    - We will only include cis females in the analyses

    - The standard definition requires controls be age 18+. If the age distribution of cases within a study is much older (e.g. 30+), then studies may filter controls based on age >= the minimum age of identified cases

    - 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).

    - We are not considering cancer subtypes

    - The only covariate to be collected is age (see below)

Codes:

- See linked excel file for codes referenced (Tables 1-5)

Case definition:

- All of the following:

    - Female

    - Age >= 18 years

    - At least one occurrence of the female breast cancer diagnostic codes (Table 1) or at least two occurrences (from distinct calendar days) of breast cancer history codes (Table 2) if the patient has no breast cancer diagnostic codes

Control definition:

- All of the following:

    - Female

- Age ≥18 years
    - However, if the age distribution of cases within a study is much older (e.g. 30+), then studies may filter controls based on age >= the minimum age of identified cases
- None of the female breast cancer diagnostic codes (Table 1) or the breast cancer history codes (Table 2)

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