

# Breast Cancer Phenotype

Columbia University

8/10/18 (Version 1)

4/9/19 (Version 1.1)

7/20/2019 (Version 2)

## Contacts:

- Ning “Sunny” Shang ([ns3026@cumc.columbia.edu](mailto:ns3026@cumc.columbia.edu))
- George Hripcsak ([gh13@cumc.columbia.edu](mailto:gh13@cumc.columbia.edu))
- Katherine Crew ([kd59@cumc.columbia.edu](mailto:kd59@cumc.columbia.edu))

## Update (Version 2 compared to V1.1)

- More detailed documentation.

## Update (Version 1.1 compared to V1)

- The manual review from Mayo found out that if a patient is only coded with one V10.3 (Personal history of malignant neoplasm of breast), the patient is more likely to be false positive. This leads to the revision of the algorithm which requires multiple occurrences if the patient has only breast cancer history codes.

## I. Background and Significance:

- Breast cancer is the most common cancer and the second leading cause of cancer-related death among women in the U.S. Known breast cancer risk factors include age, race/ethnicity, reproductive factors, and benign breast disease. Family history of breast cancer and hereditary cancer syndromes, such as *BRCA1/BRCA2* mutations, confer the strongest risk for this disease. Although there have been a number of genome-wide association studies (GWAS) to identify genetic predictors of breast cancer, most have focused on high-risk cohorts of women with a strong family history rather than population-based cohorts and few have looked at genetic predictors based upon breast cancer subtypes. For example, *BRCA1* mutation carriers tend to develop estrogen receptor (ER)-negative breast cancers, whereas the majority of breast tumors in *BRCA2* mutation carriers are ER-positive. The purpose of this algorithm is to identify breast cancer subtypes based upon tumor hormone receptor (HR) status. There are currently FDA-approved drugs, such as tamoxifen, which have been shown to reduce the incidence of ER-positive breast cancer by up to 50-65% among high-risk women. Identifying cohorts of women who are more likely to benefit from anti-estrogen therapy may lead to a more precision medicine approach to breast cancer prevention strategies.
- Our algorithm is based upon American Society of Clinical Oncology (ASCO)/College of American Pathologists (CAP) guidelines for defining hormone receptor status for breast cancer (<http://www.cap.org/ShowProperty?nodePath=/UCMCon/Contribution%20Folders/WebContent/pdf/er-pgr-summary.pdf>). Specifically, hormone receptor (HR)-positive status is defined as ER-positive and/or progesterone receptor (PR)-positive (defined as  $\geq 1\%$  of tumor cells are immunoreactive) on breast pathology from a core biopsy or surgical resection (or provided from tumor registry). For cases where tumor registry data or breast

pathology reports are unavailable, then the standard treatment for HR-positive breast cancers include the hormonal therapies such as tamoxifen (Nolvadex), anastrozole (Arimidex), letrozole (Femara), exemestane (Aromasin), and fulvestrant (Faslodex). Therefore, the presence of a HR-positive breast cancer may also be identified by the use of these medications after a breast cancer diagnosis. If this data is not available, then breast cancers may be classified as hormone receptor status “unknown”.

## **II. Development:**

- Breast cancer cases and controls will be separated by sex and initially determined by ICD-9/ICD-10 diagnostic codes, breast pathology reports, and medication lists.
- Validation is performed against expert opinion

## **III. Algorithm Definitions (Figure 1)**

### **1. Breast cancer case/control**

#### ***1.1 Female Breast Cancer Control***

- Female
- Age  $\geq 18$  years
- None of the female breast cancer diagnostic codes (Table 1) or the breast cancer history codes (Table 2)

#### ***1.2 Female Breast Cancer Case***

- Female
- Any age
- 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

#### ***1.3 Male Breast Cancer Control***

- Male
- Age  $\geq 18$  years
- None of the male breast cancer diagnostic codes (Table 3) or the breast cancer history codes (Table 2)

#### ***1.4 Male Breast Cancer Case***

- Male
- Any age
- At least one occurrence of the male breast cancer diagnostic codes (Table 3) 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

### ***2. Breast cancer case subtypes: further classification of male and female breast cancer cases by tumor hormone receptor status***

*The tumor hormone receptor status, estrogen receptor (ER) or progesterone receptor (PR), is provided by tumor registry or extracted from breast pathology report. A breast pathology report follows a biopsy or surgical resection.*

**A. Hormone receptor-positive**

- Any ER-positive ( $\geq 1\%$ ) OR PR-positive ( $\geq 1\%$ )
- If no tumor registry or breast pathology available, then use of at least one of the hormone therapy drugs (Table 4) after the date of diagnosis of breast cancer

**B. Hormone receptor-negative**

- All ER-negative ( $< 1\%$ ) AND PR-negative ( $< 1\%$ )

**C. Hormone receptor-unknown**

- No tumor registry/breast pathology available OR ER/PR status unknown (both unknown or one of them negative and the other one is unknown)
- No use of any of the hormone therapy drugs (Table 5) medications after the date of diagnosis of breast cancer

**IV. Breast Cancer Implementation (Figure 1)**

**1. Breast cancer algorithm related variables (coding see Table 1-5)**

- Patient demographics: current age, age at diagnosis (breast cancer cases only), sex
- Patient diagnosis: breast cancer ICD-9/10 diagnostic codes or breast cancer history ICD-9/10 codes
- ER/PR staining from patient pathology reports on breast biopsy or surgery (breast cancer cases only) OR ER/PR assay description from tumor registry
- Medication list: hormonal therapies after breast cancer diagnosis (cases only)

**2. Separate breast cancer cases and controls by sex, limit controls by age  $\geq 18$  years**

**3. For breast cancer cases, determine tumor hormone receptor status**

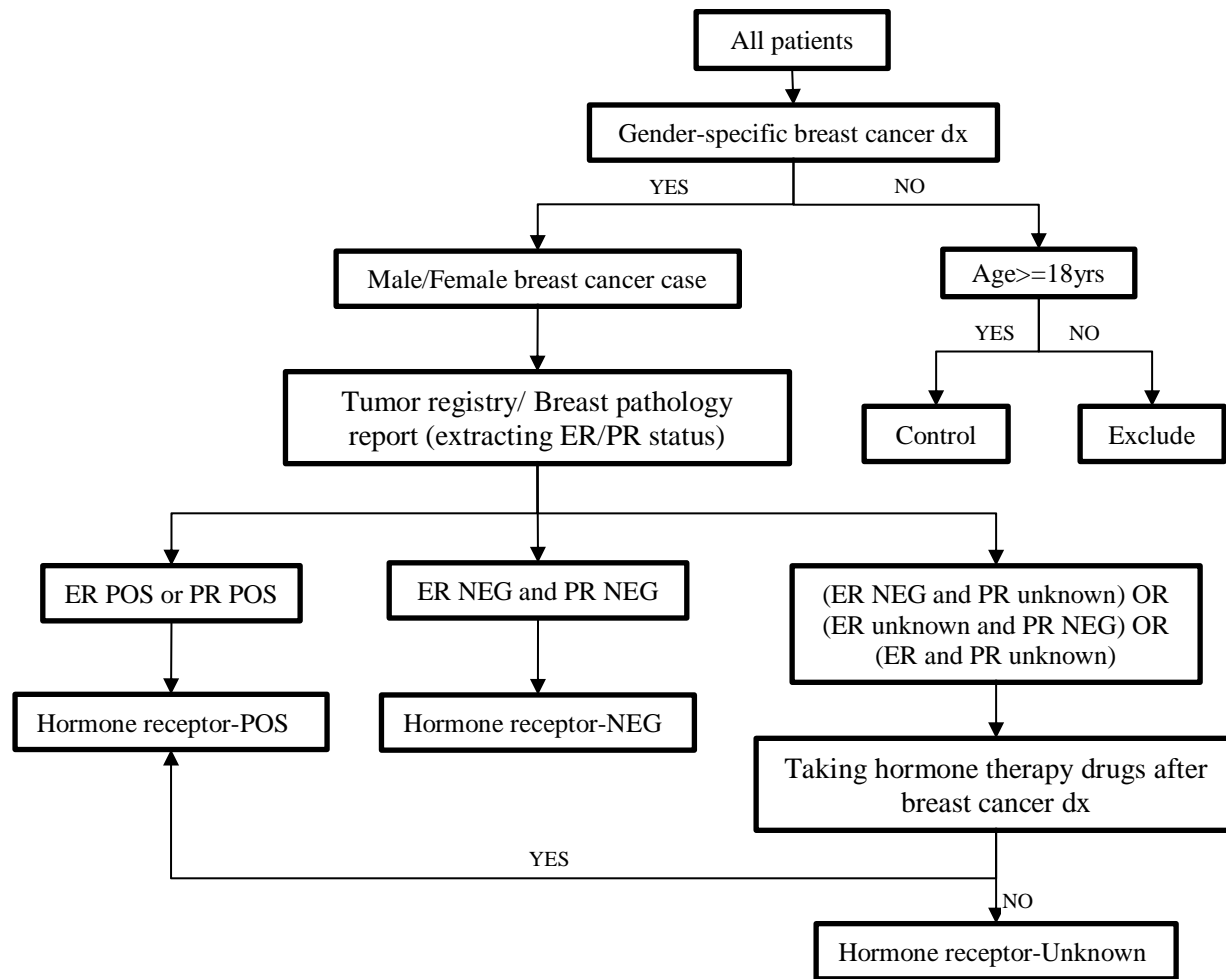
- Determine date of breast cancer diagnosis and age at diagnosis
- Obtain breast pathology reports or tumor registry data
- If breast pathology/tumor registry available, query for ER and/or PR status (POSITIVE = ER+ and/or PR+; NEGATIVE = ER- and PR-)
- If no breast pathology/tumor registry available or ER/PR status missing/unknown, query for hormonal therapies on medication list after breast cancer diagnosis (hormone receptor POSITIVE, if at least one medication listed)
- If no breast pathology/tumor registry or ER/PR status missing/unknown and no hormonal therapies identified, then classify as hormone receptor status UNKNOWN

**V. Collection of covariates**

**1. Demographic and phenotyping results**

- Current age
- Sex
- Race/ethnicity
- Breast cancer case/control status

- Hormone receptor status for breast cancer cases
  - Family history of breast cancer
2. BMI, and BMI's corresponding height and weight
  3. Breast cancer diagnosis (breast cancer cases only, ICD codes see Table 1, 2 and 3)
  4. Medications: hormonal therapies (cases only, RxNorm codes see Table 5)
  5. ER/PR status from tumor registry data or breast pathology report
  6. Breast density on mammogram report (radiology report): Based on Breast Imaging Reporting and Data System (BI-RADS), breast density is classified into 4 categories (<https://www.cancer.gov/types/breast/breast-changes/dense-breasts>):
    - Class A (or 1): Almost entirely fatty
    - Class B (or 2): Scattered fibroglandular density
    - Class C (or 3): Heterogeneously dense
    - Class D (or 4): Extremely dense



**Figure 1** Breast cancer phenotyping workflow

**Table 1** Female breast cancer diagnosis codes\*

| omop_domain_id | vocabulary_id | ICD_code | ICD_name  | ICD_omop_id | standard_omop_id | standard_omop_name  |
|----------------|---------------|----------|---|-------------|------------------|---|
| Condition      | ICD10CM       | C50.01   | Malignant neoplasm of nipple and areola, female                         | 1567539     | 4091464          | Malignant neoplasm of nipple and areola of female breast    |
| Condition      | ICD10CM       | C50.011  | Malignant neoplasm of nipple and areola, right female breast            | 45566599    | 4091464          | Malignant neoplasm of nipple and areola of female breast    |
| Condition      | ICD10CM       | C50.012  | Malignant neoplasm of nipple and areola, left female breast             | 45571519    | 4091464          | Malignant neoplasm of nipple and areola of female breast    |
| Condition      | ICD10CM       | C50.019  | Malignant neoplasm of nipple and areola, unspecified female breast      | 45595666    | 4091464          | Malignant neoplasm of nipple and areola of female breast    |
| Condition      | ICD10CM       | C50.1    | Malignant neoplasm of central portion of breast                         | 1567541     | 4092511          | Malignant neoplasm of central part of female breast         |
| Condition      | ICD10CM       | C50.11   | Malignant neoplasm of central portion of breast, female                 | 1567542     | 4092511          | Malignant neoplasm of central part of female breast         |
| Condition      | ICD10CM       | C50.111  | Malignant neoplasm of central portion of right female breast            | 45532911    | 4092511          | Malignant neoplasm of central part of female breast         |
| Condition      | ICD10CM       | C50.112  | Malignant neoplasm of central portion of left female breast             | 45537820    | 4092511          | Malignant neoplasm of central part of female breast         |
| Condition      | ICD10CM       | C50.119  | Malignant neoplasm of central portion of unspecified female breast      | 45556994    | 4092511          | Malignant neoplasm of central part of female breast         |
| Condition      | ICD10CM       | C50.2    | Malignant neoplasm of upper-inner quadrant of breast                    | 1567544     | 4187849          | Malignant neoplasm of breast upper inner quadrant           |
| Condition      | ICD10CM       | C50.21   | Malignant neoplasm of upper-inner quadrant of breast, female            | 1567545     | 4092512          | Malignant neoplasm of upper-inner quadrant of female breast |
| Condition      | ICD10CM       | C50.211  | Malignant neoplasm of upper-inner quadrant of right female breast       | 45561812    | 4092512          | Malignant neoplasm of upper-inner quadrant of female breast |
| Condition      | ICD10CM       | C50.212  | Malignant neoplasm of upper-inner quadrant of left female breast        | 45561813    | 4092512          | Malignant neoplasm of upper-inner quadrant of female breast |
| Condition      | ICD10CM       | C50.219  | Malignant neoplasm of upper-inner quadrant of unspecified female breast | 45586011    | 4092512          | Malignant neoplasm of upper-inner quadrant of female breast |
| Condition      | ICD10CM       | C50.3    | Malignant neoplasm of lower-inner quadrant of breast                    | 1567547     | 4188544          | Malignant neoplasm of breast lower inner quadrant           |
| Condition      | ICD10CM       | C50.31   | Malignant neoplasm of lower-inner quadrant of breast, female            | 1567548     | 4095740          | Malignant neoplasm of lower-inner quadrant of female breast |

|           |         |         |   |          |         |   |
|-----------|---------|---------|---|----------|---------|---|
|           |         |         |   |          |         | breast  |
| Condition | ICD10CM | C50.311 | Malignant neoplasm of lower-inner quadrant of right female breast       | 45561815 | 4095740 | Malignant neoplasm of lower-inner quadrant of female breast |
| Condition | ICD10CM | C50.312 | Malignant neoplasm of lower-inner quadrant of left female breast        | 45576325 | 4095740 | Malignant neoplasm of lower-inner quadrant of female breast |
| Condition | ICD10CM | C50.319 | Malignant neoplasm of lower-inner quadrant of unspecified female breast | 45566602 | 4095740 | Malignant neoplasm of lower-inner quadrant of female breast |
| Condition | ICD10CM | C50.4   | Malignant neoplasm of upper-outer quadrant of breast                    | 1567550  | 4160780 | Malignant neoplasm of breast upper outer quadrant           |
| Condition | ICD10CM | C50.41  | Malignant neoplasm of upper-outer quadrant of breast, female            | 1567551  | 4091465 | Malignant neoplasm of upper-outer quadrant of female breast |
| Condition | ICD10CM | C50.411 | Malignant neoplasm of upper-outer quadrant of right female breast       | 45537821 | 4091465 | Malignant neoplasm of upper-outer quadrant of female breast |
| Condition | ICD10CM | C50.412 | Malignant neoplasm of upper-outer quadrant of left female breast        | 45561817 | 4091465 | Malignant neoplasm of upper-outer quadrant of female breast |
| Condition | ICD10CM | C50.419 | Malignant neoplasm of upper-outer quadrant of unspecified female breast | 45600507 | 4091465 | Malignant neoplasm of upper-outer quadrant of female breast |
| Condition | ICD10CM | C50.5   | Malignant neoplasm of lower-outer quadrant of breast                    | 1567553  | 4187848 | Malignant neoplasm of breast lower outer quadrant           |
| Condition | ICD10CM | C50.51  | Malignant neoplasm of lower-outer quadrant of breast, female            | 1567554  | 4091466 | Malignant neoplasm of lower-outer quadrant of female breast |
| Condition | ICD10CM | C50.511 | Malignant neoplasm of lower-outer quadrant of right female breast       | 45590909 | 4091466 | Malignant neoplasm of lower-outer quadrant of female breast |
| Condition | ICD10CM | C50.512 | Malignant neoplasm of lower-outer quadrant of left female breast        | 45605276 | 4091466 | Malignant neoplasm of lower-outer quadrant of female breast |
| Condition | ICD10CM | C50.519 | Malignant neoplasm of lower-outer quadrant of unspecified female breast | 45547504 | 4091466 | Malignant neoplasm of lower-outer quadrant of female breast |
| Condition | ICD10CM | C50.6   | Malignant neoplasm of axillary tail of breast                           | 1567556  | 4155292 | Malignant neoplasm of axillary tail of breast               |

|           |         |         |  |          |         |  |
|-----------|---------|---------|--|----------|---------|--|
| Condition | ICD10CM | C50.61  | Malignant neoplasm of axillary tail of breast, female                | 1567557  | 4091467 | Malignant neoplasm of axillary tail of female breast |
| Condition | ICD10CM | C50.611 | Malignant neoplasm of axillary tail of right female breast           | 45586013 | 4091467 | Malignant neoplasm of axillary tail of female breast |
| Condition | ICD10CM | C50.612 | Malignant neoplasm of axillary tail of left female breast            | 45581222 | 4091467 | Malignant neoplasm of axillary tail of female breast |
| Condition | ICD10CM | C50.619 | Malignant neoplasm of axillary tail of unspecified female breast     | 45556995 | 4091467 | Malignant neoplasm of axillary tail of female breast |
| Condition | ICD10CM | C50.81  | Malignant neoplasm of overlapping sites of breast, female            | 1567560  | 133711  | Overlapping malignant neoplasm of female breast      |
| Condition | ICD10CM | C50.811 | Malignant neoplasm of overlapping sites of right female breast       | 45576327 | 133711  | Overlapping malignant neoplasm of female breast      |
| Condition | ICD10CM | C50.812 | Malignant neoplasm of overlapping sites of left female breast        | 45561819 | 133711  | Overlapping malignant neoplasm of female breast      |
| Condition | ICD10CM | C50.819 | Malignant neoplasm of overlapping sites of unspecified female breast | 45552262 | 133711  | Overlapping malignant neoplasm of female breast      |
| Condition | ICD10CM | C50.91  | Malignant neoplasm of breast of unspecified site, female             | 1567563  | 4157332 | Malignant neoplasm of female breast                  |
| Condition | ICD10CM | C50.911 | Malignant neoplasm of unspecified site of right female breast        | 45586014 | 4157332 | Malignant neoplasm of female breast                  |
| Condition | ICD10CM | C50.912 | Malignant neoplasm of unspecified site of left female breast         | 45556996 | 4157332 | Malignant neoplasm of female breast                  |
| Condition | ICD10CM | C50.919 | Malignant neoplasm of unspecified site of unspecified female breast  | 45600511 | 4157332 | Malignant neoplasm of female breast                  |
| Condition | ICD10CM | D05     | Carcinoma in situ of breast  | 1567735  | 81250   | Carcinoma in situ of breast                          |
| Condition | ICD10CM | D05.1   | Intraductal carcinoma in situ of breast                              | 1567737  | 4001670 | Intraductal carcinoma in situ of breast              |
| Condition | ICD10CM | D05.10  | Intraductal carcinoma in situ of unspecified breast                  | 45542678 | 4001670 | Intraductal carcinoma in situ of breast              |
| Condition | ICD10CM | D05.11  | Intraductal carcinoma in situ of right breast                        | 45600577 | 4001670 | Intraductal carcinoma in situ of breast              |
| Condition | ICD10CM | D05.12  | Intraductal carcinoma in situ of left breast                         | 45605341 | 4001670 | Intraductal carcinoma in situ of breast              |



|           |         |        |   |          |         |  |
|-----------|---------|--------|---|----------|---------|--|
| Condition | ICD10CM | D05.8  | Other specified type of carcinoma in situ of breast             | 1567738  | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.80 | Other specified type of carcinoma in situ of unspecified breast | 45547573 | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.81 | Other specified type of carcinoma in situ of right breast       | 45571587 | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.82 | Other specified type of carcinoma in situ of left breast        | 45561890 | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.9  | Unspecified type of carcinoma in situ of breast                 | 1567739  | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.90 | Unspecified type of carcinoma in situ of unspecified breast     | 45537886 | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.91 | Unspecified type of carcinoma in situ of right breast           | 45571588 | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD10CM | D05.92 | Unspecified type of carcinoma in situ of left breast            | 45552321 | 81250   | Carcinoma in situ of breast                                    |
| Condition | ICD9CM  | 174    | Malignant neoplasm of female breast                             | 44819434 | 4157332 | Malignant neoplasm of female breast                            |
| Condition | ICD9CM  | 174.0  | Malignant neoplasm of nipple and areola of female breast        | 44832128 | 4091464 | Malignant neoplasm of nipple and areola of female breast       |
| Condition | ICD9CM  | 174.1  | Malignant neoplasm of central portion of female breast          | 44829810 | 432845  | Primary malignant neoplasm of central portion of female breast |
| Condition | ICD9CM  | 174.2  | Malignant neoplasm of upper-inner quadrant of female breast     | 44819435 | 4092512 | Malignant neoplasm of upper-inner quadrant of female breast    |
| Condition | ICD9CM  | 174.3  | Malignant neoplasm of lower-inner quadrant of female breast     | 44834488 | 4095740 | Malignant neoplasm of lower-inner quadrant of female breast    |
| Condition | ICD9CM  | 174.4  | Malignant neoplasm of upper-outer quadrant of female breast     | 44827561 | 4091465 | Malignant neoplasm of upper-outer quadrant of female breast    |
| Condition | ICD9CM  | 174.5  | Malignant neoplasm of lower-outer quadrant of female breast     | 44820614 | 4091466 | Malignant neoplasm of lower-outer quadrant of female breast    |
| Condition | ICD9CM  | 174.6  | Malignant neoplasm of axillary tail of female breast            | 44834489 | 4091467 | Malignant neoplasm of axillary tail of female breast           |
| Condition | ICD9CM  | 174.8  | Malignant neoplasm of other specified sites of female breast    | 44834490 | 133711  | Overlapping malignant neoplasm of female breast                |

|           |        |       |  |          |        |   |
|-----------|--------|-------|--|----------|--------|---|
| Condition | ICD9CM | 174.9 | Malignant neoplasm of breast (female), unspecified | 44824029 | 137809 | Primary malignant neoplasm of female breast |
| Condition | ICD9CM | 233.0 | Carcinoma in situ of breast                        | 44819488 | 81250  | Carcinoma in situ of breast                 |

\*LCIS is considered benign and excluded

**Table 2 Breast cancer history codes**

| omop_domain_id           | vocabulary_id | ICD_code | ICD_name   | ICD_omop_id | standard_omop_id | standard_omop_name                     |
|--------------------------|---------------|----------|--|-------------|------------------|--|
| Observation <sup>#</sup> | ICD10CM       | Z85.3    | Personal history of malignant neoplasm of breast | 35225339    | 4214956          | History of clinical finding in subject |
| Observation <sup>#</sup> | ICD10CM       | Z86.000  | Personal history of in-situ neoplasm of breast   | 45537687    | 4214956          | History of clinical finding in subject |
| Observation <sup>#</sup> | ICD9CM        | V10.3    | Personal history of malignant neoplasm of breast | 44829584    | 4214956          | History of clinical finding in subject |

<sup>#</sup>if OMOP data model is used, please check observation table for finding relevant information following OMOP concept mapping rules and the ETL process (details of OMOP concept mapping see Table 4)

**Table 3 Male breast cancer diagnosis codes<sup>\*</sup>**

| omop_domain_id | vocabulary_id | ICD_code | ICD_name   | ICD_omop_id | standard_omop_id | standard_omop_name   |
|----------------|---------------|----------|--|-------------|------------------|--|
| Condition      | ICD10CM       | C50.02   | Malignant neoplasm of nipple and areola, male                    | 1567540     | 4091469          | Malignant neoplasm of nipple and areola of male breast           |
| Condition      | ICD10CM       | C50.021  | Malignant neoplasm of nipple and areola, right male breast       | 45566600    | 4091469          | Malignant neoplasm of nipple and areola of male breast           |
| Condition      | ICD10CM       | C50.022  | Malignant neoplasm of nipple and areola, left male breast        | 45576323    | 4091469          | Malignant neoplasm of nipple and areola of male breast           |
| Condition      | ICD10CM       | C50.029  | Malignant neoplasm of nipple and areola, unspecified male breast | 45542613    | 4091469          | Malignant neoplasm of nipple and areola of male breast           |
| Condition      | ICD10CM       | C50.1    | Malignant neoplasm of central portion of breast                  | 1567541     | 4092511          | Malignant neoplasm of central part of female breast <sup>§</sup> |
| Condition      | ICD10CM       | C50.12   | Malignant neoplasm of central portion of breast, male            | 1567543     | 4157447          | Malignant neoplasm of male breast                                |
| Condition      | ICD10CM       | C50.121  | Malignant neoplasm of central portion of right male breast       | 45532912    | 4157447          | Malignant neoplasm of male breast                                |
| Condition      | ICD10CM       | C50.122  | Malignant neoplasm of central portion of left male breast        | 45586010    | 4157447          | Malignant neoplasm of male breast                                |
| Condition      | ICD10CM       | C50.129  | Malignant neoplasm of central                                    | 45571520    | 4157447          | Malignant neoplasm of male                                       |

|           |         |         |   |          |         |   |
|-----------|---------|---------|---|----------|---------|---|
|           |         |         | portion of unspecified male breast                                    |          |         | breast  |
| Condition | ICD10CM | C50.2   | Malignant neoplasm of upper-inner quadrant of breast                  | 1567544  | 4187849 | Malignant neoplasm of breast upper inner quadrant |
| Condition | ICD10CM | C50.22  | Malignant neoplasm of upper-inner quadrant of breast, male            | 1567546  | 4187849 | Malignant neoplasm of breast upper inner quadrant |
| Condition | ICD10CM | C50.221 | Malignant neoplasm of upper-inner quadrant of right male breast       | 45547503 | 4187849 | Malignant neoplasm of breast upper inner quadrant |
| Condition | ICD10CM | C50.222 | Malignant neoplasm of upper-inner quadrant of left male breast        | 45595667 | 4187849 | Malignant neoplasm of breast upper inner quadrant |
| Condition | ICD10CM | C50.229 | Malignant neoplasm of upper-inner quadrant of unspecified male breast | 45561814 | 4187849 | Malignant neoplasm of breast upper inner quadrant |
| Condition | ICD10CM | C50.3   | Malignant neoplasm of lower-inner quadrant of breast                  | 1567547  | 4188544 | Malignant neoplasm of breast lower inner quadrant |
| Condition | ICD10CM | C50.32  | Malignant neoplasm of lower-inner quadrant of breast, male            | 1567549  | 4188544 | Malignant neoplasm of breast lower inner quadrant |
| Condition | ICD10CM | C50.321 | Malignant neoplasm of lower-inner quadrant of right male breast       | 45571521 | 4188544 | Malignant neoplasm of breast lower inner quadrant |
| Condition | ICD10CM | C50.322 | Malignant neoplasm of lower-inner quadrant of left male breast        | 45586012 | 4188544 | Malignant neoplasm of breast lower inner quadrant |
| Condition | ICD10CM | C50.329 | Malignant neoplasm of lower-inner quadrant of unspecified male breast | 45590907 | 4188544 | Malignant neoplasm of breast lower inner quadrant |
| Condition | ICD10CM | C50.4   | Malignant neoplasm of upper-outer quadrant of breast                  | 1567550  | 4160780 | Malignant neoplasm of breast upper outer quadrant |
| Condition | ICD10CM | C50.42  | Malignant neoplasm of upper-outer quadrant of breast, male            | 1567552  | 4160780 | Malignant neoplasm of breast upper outer quadrant |
| Condition | ICD10CM | C50.421 | Malignant neoplasm of upper-outer quadrant of right male breast       | 45561818 | 4160780 | Malignant neoplasm of breast upper outer quadrant |
| Condition | ICD10CM | C50.422 | Malignant neoplasm of upper-outer quadrant of left male breast        | 45600508 | 4160780 | Malignant neoplasm of breast upper outer quadrant |
| Condition | ICD10CM | C50.429 | Malignant neoplasm of upper-outer quadrant of unspecified male breast | 45590908 | 4160780 | Malignant neoplasm of breast upper outer quadrant |
| Condition | ICD10CM | C50.5   | Malignant neoplasm of lower-  | 1567553  | 4187848 | Malignant neoplasm of breast                      |

|           |         |         |   |          |         |   |
|-----------|---------|---------|---|----------|---------|---|
|           |         |         | outer quadrant of breast  |          |         | lower outer quadrant                              |
| Condition | ICD10CM | C50.52  | Malignant neoplasm of lower-outer quadrant of breast, male            | 1567555  | 4187848 | Malignant neoplasm of breast lower outer quadrant |
| Condition | ICD10CM | C50.521 | Malignant neoplasm of lower-outer quadrant of right male breast       | 45605277 | 4187848 | Malignant neoplasm of breast lower outer quadrant |
| Condition | ICD10CM | C50.522 | Malignant neoplasm of lower-outer quadrant of left male breast        | 45571522 | 4187848 | Malignant neoplasm of breast lower outer quadrant |
| Condition | ICD10CM | C50.529 | Malignant neoplasm of lower-outer quadrant of unspecified male breast | 45537822 | 4187848 | Malignant neoplasm of breast lower outer quadrant |
| Condition | ICD10CM | C50.6   | Malignant neoplasm of axillary tail of breast                         | 1567556  | 4155292 | Malignant neoplasm of axillary tail of breast     |
| Condition | ICD10CM | C50.62  | Malignant neoplasm of axillary tail of breast, male                   | 1567558  | 4155292 | Malignant neoplasm of axillary tail of breast     |
| Condition | ICD10CM | C50.621 | Malignant neoplasm of axillary tail of right male breast              | 45600510 | 4155292 | Malignant neoplasm of axillary tail of breast     |
| Condition | ICD10CM | C50.622 | Malignant neoplasm of axillary tail of left male breast               | 45576326 | 4155292 | Malignant neoplasm of axillary tail of breast     |
| Condition | ICD10CM | C50.629 | Malignant neoplasm of axillary tail of unspecified male breast        | 45537823 | 4155292 | Malignant neoplasm of axillary tail of breast     |
| Condition | ICD10CM | C50.82  | Malignant neoplasm of overlapping sites of breast, male               | 1567561  | 4003684 | Overlapping malignant neoplasm of male breast     |
| Condition | ICD10CM | C50.821 | Malignant neoplasm of overlapping sites of right male breast          | 45595668 | 4003684 | Overlapping malignant neoplasm of male breast     |
| Condition | ICD10CM | C50.822 | Malignant neoplasm of overlapping sites of left male breast           | 45566603 | 4003684 | Overlapping malignant neoplasm of male breast     |
| Condition | ICD10CM | C50.829 | Malignant neoplasm of overlapping sites of unspecified male breast    | 45537824 | 4003684 | Overlapping malignant neoplasm of male breast     |
| Condition | ICD10CM | C50.92  | Malignant neoplasm of breast of unspecified site, male                | 1567564  | 4157447 | Malignant neoplasm of male breast                 |
| Condition | ICD10CM | C50.921 | Malignant neoplasm of unspecified site of right male breast           | 45571523 | 4157447 | Malignant neoplasm of male breast                 |
| Condition | ICD10CM | C50.922 | Malignant neoplasm of unspecified site of left male                   | 45561820 | 4157447 | Malignant neoplasm of male breast                 |

|           |         |         |   |          |         |  |
|-----------|---------|---------|---|----------|---------|--|
|           |         |         | breast  |          |         |  |
| Condition | ICD10CM | C50.929 | Malignant neoplasm of unspecified site of unspecified male breast | 45590910 | 4157447 | Malignant neoplasm of male breast                      |
| Condition | ICD10CM | D05     | Carcinoma in situ of breast                                       | 1567735  | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.1   | Intraductal carcinoma in situ of breast                           | 1567737  | 4001670 | Intraductal carcinoma in situ of breast                |
| Condition | ICD10CM | D05.10  | Intraductal carcinoma in situ of unspecified breast               | 45542678 | 4001670 | Intraductal carcinoma in situ of breast                |
| Condition | ICD10CM | D05.11  | Intraductal carcinoma in situ of right breast                     | 45600577 | 4001670 | Intraductal carcinoma in situ of breast                |
| Condition | ICD10CM | D05.12  | Intraductal carcinoma in situ of left breast                      | 45605341 | 4001670 | Intraductal carcinoma in situ of breast                |
| Condition | ICD10CM | D05.8   | Other specified type of carcinoma in situ of breast               | 1567738  | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.80  | Other specified type of carcinoma in situ of unspecified breast   | 45547573 | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.81  | Other specified type of carcinoma in situ of right breast         | 45571587 | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.82  | Other specified type of carcinoma in situ of left breast          | 45561890 | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.9   | Unspecified type of carcinoma in situ of breast                   | 1567739  | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.90  | Unspecified type of carcinoma in situ of unspecified breast       | 45537886 | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.91  | Unspecified type of carcinoma in situ of right breast             | 45571588 | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD10CM | D05.92  | Unspecified type of carcinoma in situ of left breast              | 45552321 | 81250   | Carcinoma in situ of breast                            |
| Condition | ICD9CM  | 175     | Malignant neoplasm of male breast                                 | 44829811 | 4157447 | Malignant neoplasm of male breast                      |
| Condition | ICD9CM  | 175.0   | Malignant neoplasm of nipple and areola of male breast            | 44833299 | 4091469 | Malignant neoplasm of nipple and areola of male breast |
| Condition | ICD9CM  | 175.9   | Malignant neoplasm of other and unspecified sites of male breast  | 44826406 | 135489  | Primary malignant neoplasm of male breast              |
| Condition | ICD9CM  | 233.0   | Carcinoma in situ of breast                                       | 44819488 | 81250   | Carcinoma in situ of breast                            |

\*LCIS is considered benign and excluded; #if OMOP data model is used, please check observation table for finding relevant information following OMOP concept mapping rules and the ETL process (details of OMOP concept mapping see Table 3); §ICD code is not gender-specific

**Table 4** Breast cancer history codes in non-condition\_occurrence table (OMOP vocabulary)

| concept_code | vocabulary_id | concept_id | concept_name                                     | relationship_id | standard_concept_id | standard_concept_name                  |
|--------------|---------------|------------|--|-----------------|---------------------|--|
| V10.3        | ICD9CM        | 44829584   | Personal history of malignant neoplasm of breast | Maps to         | 4214956             | History of clinical finding in subject |
| V10.3        | ICD9CM        | 44829584   | Personal history of malignant neoplasm of breast | Maps to value   | 4112853             | Malignant tumor of breast              |
| Z85.3        | ICD10CM       | 35225339   | Personal history of malignant neoplasm of breast | Maps to         | 4214956             | History of clinical finding in subject |
| Z85.3        | ICD10CM       | 35225339   | Personal history of malignant neoplasm of breast | Maps to value   | 4112853             | Malignant tumor of breast              |
| Z86.000      | ICD10CM       | 45537687   | Personal history of in-situ neoplasm of breast   | Maps to         | 4214956             | History of clinical finding in subject |
| Z86.000      | ICD10CM       | 45537687   | Personal history of in-situ neoplasm of breast   | Maps to value   | 81250               | Carcinoma in situ of breast            |

**Table 5** Hormone therapy drugs for breast cancer\*

| Drug (Brand Name)      | Drug Name   | RxCUI  | OMOP Concept Id |
|------------------------|-------------|--------|-----------------|
| Tamoxifen (Nolvadex)   | Tamoxifen   | 10324  | 1436678         |
|                        | Nolvadex    | 281964 | 19004028        |
| Anastrozole (Arimidex) | Anastrozole | 84857  | 1348265         |
|                        | Arimidex    | 262485 | 19100842        |
| Letrozole (Femara)     | Letrozole   | 72965  | 1315946         |
|                        | Femara      | 203769 | 19048060        |
| Exemestane (Aromasin)  | Exemestane  | 258494 | 1398399         |
|                        | Aromasin    | 262105 | 19100386        |
| Goserelin (Zoladex)    | Goserelin   | 50610  | 1366310         |
|                        | Zoladex     | 58328  | 19079531        |
| Leuprolide (Lupron)    | Leuprolide  | 42375  | 1351541         |
|                        | Lupron      | 203852 | 19048354        |
| Fulvestrant (Faslodex) | Fulvestrant | 282357 | 1304044         |
|                        | Faslodex    | 203870 | 19048400        |
| Toremifene (Fareston)  | Toremifene  | 38409  | 1342346         |
|                        | Fareston    | 155082 | 19024254        |

\*Detail codes please check table5detail from BreastCancerPhenotype\_V2\_codes.xlsx

