

Summary of the number of cases for Confluence Project

Updated on 13th March 2025

The size of the full cohort is **513228**, which includes male and female.

The total number of participants with genotype data is **100706**.

The endpoints data contains information of breast cancer (ICD10 C50; diy_p1232015483) and Carcinoma in situ of breast (D05; diy_p1781594777). The total number of rows in “endpoints” file for the full cohort is **513228**.

(i) the full cohort (513228)

we can count the number of breast cancer prevalence with ((df['cancer_diag'] == 1) & (df['cancer_Breast'] == 1)):

- For women, the count is **567**.
- For men, the count is **1**.

(ii) the full cohort (513228)

breast cancer (ICD10 C50; diy_p1232015483; ep_diy_p1232015483_combined_ep == 1):

- For women, the count is **3113**.
- For men, the count is **57**.

carcinoma in situ of breast (D05; diy_p1781594777; ep_diy_p1781594777_combined_ep == 1):

- For women, the count is **81**.
- For men, the count is **4**.

For women, the overlap between C50 and D05 is **67**.

For men, the overlap between C50 and D05 is **1**.

(iii) the participants with genetic data (100706)

breast cancer (ICD10 C50; diy_p1232015483; ep_diy_p1232015483_combined_ep == 1):

- For women, the count is **574**.
- For men, the count is **12**.

carcinoma in situ of breast (D05; diy_p1781594777; ep_diy_p1781594777_combined_ep == 1):

- For women, the count is **14**.
- For men, the count is **1**.

For women, the overlap between C50 and D05 is **11**.

For men, the overlap between C50 and D05 is **0**.

(iv) the population-based subset of the genetics (77176)

breast cancer (ICD10 C50; ep_diy_p1232015483_combined_ep == 1 & is_in_gwas_population_subset == 1):

- For women, the count is **510**.
- For men, the count is **7**.

carcinoma in situ of breast (D05; diy_p1781594777;

ep_diy_p1781594777_combined_ep == 1 & is_in_gwas_population_subset == 1):

- For women, the count is **12**.
- For men, the count is **1**.

For women, the overlap between C50 and D05 is **9**.

For men, the overlap between C50 and D05 is **0**.

Note that due to the small number of cases, in our previous summary statistics by SAIGE and REGENIE, we also considered the situation where we included the cases even if these participants are in the non-population subset. The number of breast cancers cases (breast_cancer_prevalence == 1 & ep_diy_p1232015483_combined_ep == 1 & is_in_gwas_population_subset == 0) in the non-population subset is **71** in our previous analysis.