Real-World Molecular Subtypes and Characteristics of Breast Cancer in Belgium: Insights from Routinely Collected Data [BE-REAST]

First published: 03/02/2026

Last updated: 03/02/2026

Study Finalised

Administrative details

EU PAS number

EUPAS1000000398

Study ID

1000000398

DARWIN EU® study

No

Study countries

Belgium

Study description

This multicenter study describes molecular subtypes and characteristics of Belgian patients with breast cancer by leveraging natural language processing and the OMOP common data model.

Study status

Finalised

Research institutions and networks

Institutions

AZ Maria Middelares General Hospital

Belgium

First published: 01/02/2024

Last updated: 01/02/2024

Institution | Hospital/Clinic/Other health care facility

AstraZeneca

First published: 01/02/2024

Last updated: 01/02/2024

Institution

Universitair Ziekenhuis Antwerpen (UZA)

Belgium

First published: 11/06/2025

Last updated: 16/06/2025

Institution | Hospital/Clinic/Other health care facility | Laboratory/Research/Testing facility | ENCePP partner

LynxCare Clinical Informartics

Belgium

First published: 20/10/2025

Last updated: 20/10/2025

Institution | Other

Algemeen Ziekenhuis Groeninge

Contact details

Study institution contact

Katoo Muylle katoo.muylle@astrazeneca.com

Study contact

katoo.muylle@astrazeneca.com

Primary lead investigator

Katoo Muylle 0000-0003-1117-6709

Primary lead investigator

ORCID number:

0000-0003-1117-6709

Study timelines

Date when funding contract was signed
Actual: 16/11/2022

Study start date
Actual: 27/10/2023

Date of final study report
Planned: 31/12/2023

Actual: 10/06/2024

Sources of funding

- Pharmaceutical company and other private sector

More details on funding

AstraZeneca BeLux

Regulatory

Was the study required by a regulatory body?

No

Is the study required by a Risk Management Plan (RMP)?

Not applicable

Methodological aspects

Study type

Study type list

Study topic:

Disease /health condition

Study type:

Non-interventional study

Scope of the study:

Disease epidemiology

Data collection methods:

Secondary use of data

Study design:

Multicenter study (n = 3) analyzing routinely collected data from pseudonymized electronic health records (EHRs) using natural language processing (NLP) for unstructured sources. The algorithm mapped variables to a standard terminology, generating OMOP CDM databases, validated per hospital.

Main study objective:

- Assess the feasibility of systematically capturing stage and receptor status across hospitals
- Assess contemporary real-world molecular subtype distributions in early and advanced disease (eBC and aBC)
- Preliminarily assess subtype stability at progression/relapse

Study Design

Non-interventional study design

Cohort

Study drug and medical condition

Medical condition to be studied

Breast cancer

Population studied

Short description of the study population

Breast cancer patients (all stages) diagnosed between 1 January 2018 and 31 December 2022.

Estimated number of subjects

2.000

Study design details

Setting

Belgium, diagnosis between 2018 and 2022

Data management

ENCePP Seal

> The use of the ENCePP Seal has been discontinued since February 2025. The ENCePP Seal fields are retained in the display mode for transparency but are no longer maintained.

Data sources

Data source(s)

LynxCare

Use of a Common Data Model (CDM)

CDM mapping

Yes

CDM Mappings

CDM name

OMOP

CDM website

https://www.ohdsi.org/Data-standardization/

CDM version

5.4

Data quality specifications

Check conformance

Yes

Check completeness

Yes

Check stability

Yes

Check logical consistency

Yes

Data characterisation

Data characterisation conducted

Yes

Data characterisation moment

after data extraction

after extract-transform-load to a common data model

after creation of study variables