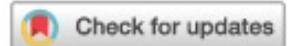




scientific reports



OPEN

Trends in incidence, prevalence, and survival of breast cancer in the United Kingdom from 2000 to 2021

Nicola L. Barclay¹, Edward Burn¹, Antonella Delmestri¹, Talita Duarte-Salles^{2,3}, Asieh Golozar^{4,5}, Wai Yi Man¹, Eng Hooi Tan¹, Ilona Tietzova⁶, OPTIMA Consortium*, Daniel Prieto-Alhambra^{1,3}✉ & Danielle Newby¹

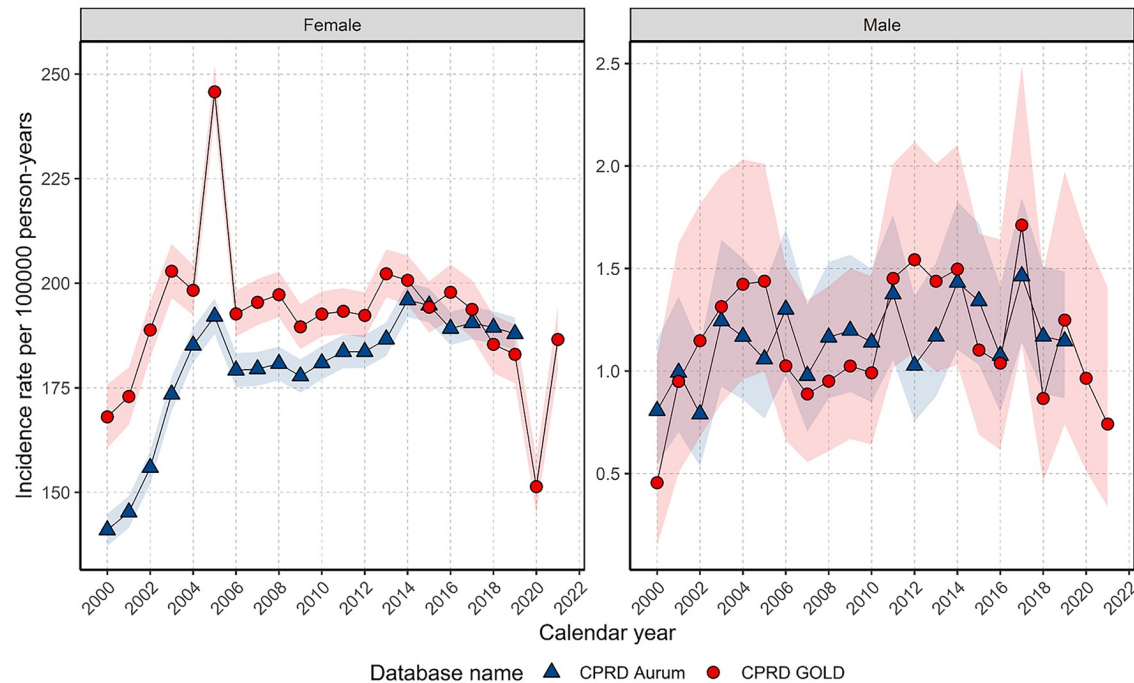


What did we do and what did we find?

- Population-based cohort study using UK primary care CPRD GOLD and Aurum to describe crude incidence, prevalence and survival from breast cancer from 2000 to 2021
- Overall findings:
 - 5,848,436 females and 5,539,681 males aged 18+ years, with \geq one year of prior data
 - Crude incidence rates of breast cancer was in line with national statistics
 - Prevalence of breast cancer in females and for most age groups in males increased across the study period
 - Survival probability improved from 2000 to 2021 for female.



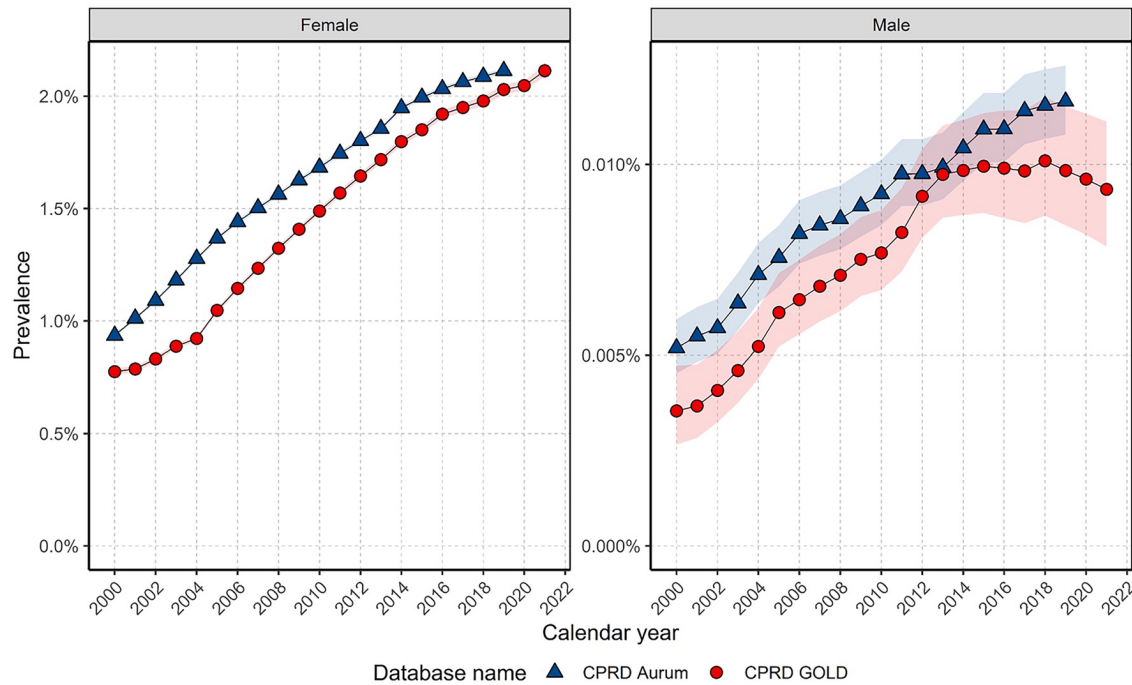
Breast Cancer Incidence



- Crude IR: 194.4 per 100,000 person-years for females and 1.16 for males.
- Incidence increased for females across the study period before dropping dramatically during 2020 – coinciding with the COVID-19 pandemic – and returning to expected levels in 2021.
- Increase in breast cancer incidence reflects the success of national breast cancer screening programs



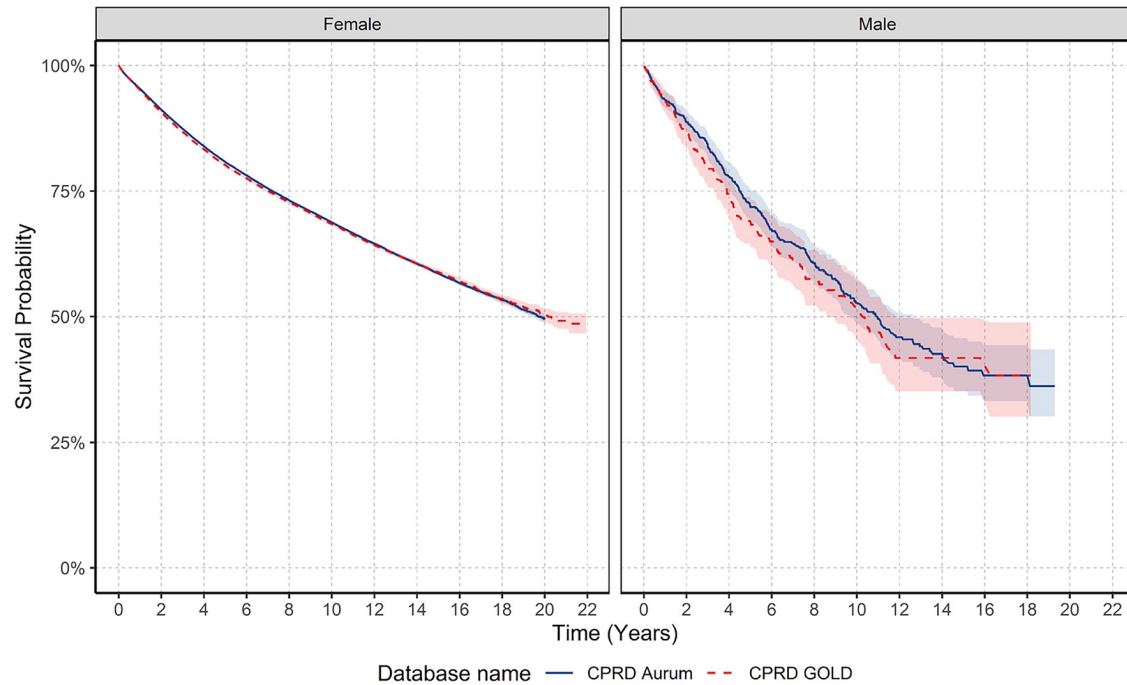
Breast Cancer Prevalence



- Crude prevalence in 2021 was 2.1% for females peaking in those aged 70–79 years and 0.009% for males.
- ~2.5-fold increase in prevalence across time.



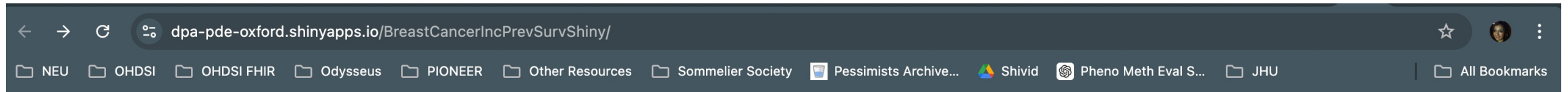
Breast Cancer Survival



- 1-, 5-, and 10-years survival probability after diagnosis:
 - 95.1%, 80.2%, and 68.4% for women
 - 92.9%, 69.0%, and 51.3% for men.
- Survival improvement observed over the past 20 years in line with National Cancer Registration and Analysis Service data



Where can you find the results?



Incidence, Prevalence and Survival of Breast Cancer in the United Kingdom between 2000 and 2021

Background

[Population Prevalence](#)

[Population Incidence](#)

[Whole Population Survival](#)

[Calendar Year Population Survival](#)

[Population Characteristics](#)

Background

This app is a companion to the study focussing on determining the incidence, prevalence and survival for breast cancer between 2000 to 2021 using Primary care GP records from the United Kingdom ([Clinical Practice Research Datalink](#) (CPRD) GOLD). We additionally compared all results using CPRD Aurum between 2000 to 2019

In the following pages you can find information on annualised period prevalence, annualised and overall incidence, survival for the whole population, calendar year survival and a description of the characteristics of the study population of breast cancer patients. All results have been stratified by age group and sex.

The results can be found published in the following journal:

1. **Barclay et al (2023) PREPRINT** Incidence, prevalence, and survival of breast cancer in the United Kingdom from 2000-2021: a population-based cohort study ([Paper Link](#))

The analysis code used to generate these results can be found [here](#) .The cohort diagnostics including the clinical codelists for breast cancer can be found [here](#)

Any questions regarding these studies or problems with the app please contact [Danielle Newby](#)

<https://dpa-pde-oxford.shinyapps.io/BreastCancerIncPrevSurvShiny/>



What is next?



Ongoing Oncology Database Test Query

- **Problem:** What information about cancer do databases contain?
 - General cancer (histology, topology, mets, nodes, grades, stages)
 - Genomic (small variants, large variants)
 - Episode (treatment, disease)
- **Solution:** Query across the Oncology Network in the Oncology WG
 - Source to standard concept mapping correct?
 - Source and standard concepts correctly placed? (source_concept_id, standard_concept_id)
 - Which vocabularies used?
 - Conventions followed (e.g. genomic data in OMOP Genomic)?
 - Domain applied correctly (e.g. topology and histology in Condition and metastases in Measurement)?
 - Concept prevalence, to give a general idea to the size and depth of the database?
 - **NO patient information of any kind, no aggregation of results by patients**
- **Queries:**
 - general.sql: for general cancer concepts: diagnoses, treatments, other mgt, 280k concepts
 - genomic.sql: for genomic concepts: small (usually SNPs), large (e.g. fusion proteins), DNA, RNA, protein level, 590k concepts
 - episode.sql: for disease (progression, remission) and treatment (regimen) episodes

- **Result format:**

domain	source_concept_id	concept_id	count
m	35919362	35957667	6469
m	3017600	3017600	5



How can you participate?

- Queries are available in Teams (Oncology WG Channel [Oncology Data Readiness Assessment](#))
- Run the query and send your results and all your questions to golozar@ohdsi.org
- Join us on October 24 to review findings

	Timeline
Develop cancer specific concept prevalence query	Sunday, Sep 21
Execute the query and share the results	Friday, Oct 4
Aggregate results	Monday, October 14
Develop high level summary	Friday, October 18
Review the results together at the workshop	Thursday, October 24
Develop and publish the list of impediments and solutions	Friday, November 8

