



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

Breast cancer is the most frequently diagnosed cancer in females globally. However, we know relatively little about trends in males. This study describes United Kingdom (UK) secular trends in breast cancer from 2000 to 2021 for both sexes. We describe a population-based cohort study using UK primary care Clinical Practice Research Datalink (CPRD) GOLD and Aurum databases. There were 5,848,436 eligible females and 5,539,681 males aged 18+ years, with \geq one year of prior data availability in the study period. We estimated crude breast cancer incidence rates (IR), prevalence and survival probability at one-, five- and 10-years after diagnosis using the Kaplan–Meier method. Analyses were further stratified by age. Crude IR of breast cancer from 2000 to 2021 was 194.4 per 100,000 person-years for females and 1.16 for males. Crude prevalence in 2021 was 2.1% for females and 0.009% for males. Both sexes have seen around a 2.5-fold increase in prevalence across time. Incidence increased with age for both sexes, peaking in females aged 60–69 years and males 90+. There was a drop in incidence for females aged 70–79 years. From 2003–2019, incidence increased > twofold in younger females (aged 18–29: IR 2.12 in 2003 vs. 4.58 in 2018); decreased in females aged 50–69 years; and further declined from 2015 onwards in females aged 70–89 years. Survival probability for females after one-, five-, and ten-years after diagnosis was 95.1%, 80.2%, and 68.4%, and for males 92.9%, 69.0%, and 51.3%. Survival probability at one-year increased by 2.08% points, and survival at five years increased by 5.39% from 2000–2004 to 2015–2019 for females, particularly those aged 50–70 years. For males, there were no clear time-trends for short-term and long-term survival probability. Changes in incidence of breast cancer in females largely reflect the success of screening programmes, as rates rise and fall in synchronicity with ages of eligibility for such programmes. Overall survival from breast cancer for females has improved from 2000 to 2021, again reflecting the success of screening programmes, early diagnosis, and improvements in treatments. Male breast cancer patients have worse survival outcomes compared to females, highlighting the need to develop male-specific diagnosis and treatment strategies to improve long-term survival in line with females.

Female breast cancer has been the leading cause of global cancer incidence in recent years, with an estimated number of new cases of 2.3 million in 2020 alone¹. Male breast cancer accounts for around 1% of all diagnosed cases², though incidence and survival trends are infrequently investigated given its rarity. Whilst the incidence of female breast cancer is stabilising or decreasing in certain age groups due to earlier detection and improved

¹Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK. ²Fundació Institut Universitari Per a La Recerca a L'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain. ³Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands. ⁴Odysseus Data Service, Cambridge, MA, USA. ⁵OHDSI Center at the Roux Institute, Northeastern University, Boston, MA, USA. ⁶First Department of Tuberculosis and Respiratory Diseases, First Faculty of Medicine, Charles University, Katerinska 1660/32, 121 08 Prague, Czech Republic. *A list of authors and their affiliations appears at the end of the paper. ✉email: daniel.prietoalhambra@ndorms.ox.ac.uk

treatments³, male breast cancer has been increasing from the 1980s to 2000s at least in the United Kingdom (UK) and the United States (US)^{4,5}.

Breast cancer risk increases with age across both sexes, though males tend to be older at the age of diagnosis⁵. Other risk factors in both males and females include family history, the risk for which is doubled in males with a first-degree relative with the disease⁶; genetic mutations in *BRCA1/2* genes and others^{2,7}; elevated estrogen levels⁸, and lifestyle factors such as alcohol consumption^{9,10}, obesity¹¹ and radiation exposure¹².

Compared to females, males with breast cancer are more likely to be estrogen-receptor positive, androgen-receptor positive, Hormone receptor-positive with Human Epidermal Growth Factor 2-negative, and present with regional nodal metastases². Males are also more likely to present at more advanced stage of disease than females², which is likely to impact their survival.

Recent evidence from 500,000 females with early invasive breast cancer in England suggests breast cancer survival has improved over time, with five-year risk of death reducing from 14 to 5% from the 1990s to 2015¹³. Whilst there is relatively little evidence for males, one study shows mortality from breast cancer reduced by nearly 40% in North-Western Europe from the early 2000s to 2017¹⁴. At least for females, the improved prognosis is likely driven by the success of national screening programmes aiding early detection, whereas the lack of such routine screening in males precludes this explanation. Improvements in males are likely a reflection of improved local and systemic treatments which have substantially improved over recent decades¹⁵.

A comprehensive assessment of the disease burden and survival of breast cancer across both sexes in the UK will inform future decisions regarding screening, prevention, treatment, and disease management in both females and males. However, much of our understanding of the disease burden of breast cancer has been derived from cancer registries. Cancer registry analyses do not have a general population denominator to estimate incidence, prevalence and survival, but rather use national general population statistics as their denominator population—methods which can introduce biases^{16,17}.

Therefore, the aim of the present study was to describe the breast cancer trends from 2000 to 2021 in the UK for both females and males in terms of incidence, prevalence and survival probability (at one-, five- and ten-years after diagnosis) using nationally representative, routinely collected electronic health record data from primary care. Additionally, incidence and prevalence analyses were stratified by age, and survival probability estimates were stratified by calendar time to investigate age and time trends.

Ethical approval

The protocol for this research was approved by the Research Data Governance (RDG) Board of the Medicine and Healthcare products Regulatory Agency database research (protocol number 22_001843).

Results

Patient populations and characteristics

There were 5,848,436 and 5,539,681 eligible female and male patients 18 years and older, with at least one year of data availability prior to diagnosis from January 2000 to December 2021 for CPRD GOLD. Attrition tables for this study can be found in the supplementary information (Supplement S2). A summary of study patient characteristics of those with a diagnosis of breast cancer stratified by sex is shown in Table 1.

Overall, the majority of those with breast cancer were female, with a median age of 63 years across both databases. Males only made up 0.6% of cancer diagnoses with an older median age of 70 years. In females, the highest percentage of patients were those aged 60–69 years, contributing to 26% of diagnosed patients, whereas for males, those aged 70–79 years contributed to 32% of diagnosed patients. Overall, males were more likely to have comorbidities compared to females apart from depressive disorders which were higher in females. The patient characteristics in Aurum can be found in the supplementary information (Supplement S3).

Overall and annualised crude incidence rates for study population stratified by age and sex across databases

Overall Crude Incidence Rates

Table 2 shows the overall incidence rates (IR) of breast cancer stratified by age and sex. For females, the overall IR per 100,000 person-years (pys) of breast cancer from 2000 to 2021 was 194.4 (95% CI 193.1–195.7) in GOLD, with slightly lower results in Aurum (180.4; 95% CI 179.5–181.3). For males, the overall IR was 1.16 (95% CI 1.07–1.28) in GOLD, with the same results in Aurum. When stratifying by age, the overall IR for females increased with age, peaking in those 60–69 years (IR: 381.0) before dropping in those aged 70–79 years (IR: 343.9), increasing in those aged 80–89 years (IR: 366.9), and with a final decrease in those 90+ years (IR: 357.6). This trend was similar in both databases. For males, overall IR was higher with increasing age up to 90+ years (IR: 6.7) in GOLD and up to 80–89 years (IR: 5.2) in Aurum. The biggest increase in overall IR for females was between those aged 30–39 years (IR: 38.3) to 40–49 years (IR: 143.7) with an increase in IR of 3.75-fold; whereas the biggest difference in IR for males was between those aged 50–59 (IR: 0.9) and 60–69 (IR: 2.2) years with a 2.43-fold increase.

Annualised crude incidence rates

For females, annualised IRs rapidly increased to 2004 before a sharp peak and gradual increase up to 2014 before declining (Fig. 1). In GOLD, IRs dropped in 2020 but recovered in 2021. For males, IRs gradually increased to a small extent over the study period but had high variability due to small sample numbers (Fig. 1).

When stratifying by age group, annualised IRs over the study period showed different trends in females depending on age of diagnosis (Fig. 2). For those aged 18–29 years, despite an initial peak IR of 5.06 in 2000, IRs increased over the study period (from a low of 2.1 in 2003 to a high of 4.6 in 2018), whereas IRs in those

	Female	Male
Number of patients	85,400	507
Age in years (Median [IQR])	63 (52 to 73)	70 (61 to 78)
Age Groups in years N (%)		
18–29	278 (0.3%)	5 (1.0%)
30–39	2860 (3.3%)	9 (1.8%)
40–49	11,298 (13.2%)	29 (5.7%)
50–59	21,496 (25.2%)	67 (13.2%)
60–69	22,273 (26.1%)	129 (25.4%)
70–79	15,090 (17.7%)	163 (32.1%)
80–89	9665 (11.3%)	87 (17.2%)
90+	2440 (2.9%)	18 (3.6%)
Prior history in the database (data availability), days		
Median [IQR]	3189 (1637–4992)	3337 (1763–4970)
Smoking status (any time 5 years prior)		
Non-smoker	41,286 (48.3%)	210 (41.4%)
Former smoker	453 (0.5%)	< 5
Smoker	13,636 (16.0%)	73 (14.4%)
Missing	30,025 (35.2%)	222 (43.8%)
General conditions (any time prior)		
Atrial fibrillation	2797 (3.3%)	51 (10.1%)
Cerebrovascular disease	2676 (3.1%)	36 (7.1%)
Chronic liver disease	216 (0.3%)	< 5
Chronic obstructive lung disease	2751 (3.2%)	38 (7.5%)
Coronary arteriosclerosis	313 (0.4%)	< 5
Dementia	1082 (1.3%)	14 (2.8%)
Depressive disorder	12,862 (15.1%)	53 (10.5%)
Diabetes mellitus	5728 (6.7%)	63 (12.4%)
Gastrointestinal haemorrhage	4403 (5.2%)	29 (5.7%)
Heart failure	1380 (1.6%)	20 (3.9%)
Hyperlipidemia	5865 (6.9%)	37 (7.3%)
Hypertensive disorder	17,772 (20.8%)	142 (28.0%)
Ischemic heart disease	3878 (4.5%)	62 (12.2%)
Osteoarthritis	14,817 (17.4%)	92 (18.1%)
Peripheral vascular disease	668 (0.8%)	12 (2.4%)
Pulmonary embolism	582 (0.7%)	< 5
Renal impairment	6576 (7.7%)	57 (11.2%)
Venous thrombosis	3485 (4.1%)	25 (4.9%)

Table 1. Baseline characteristics of breast cancer patients at time-of-diagnosis stratified by sex from CPRD GOLD. IQR: interquartile range.

aged 30–39 years declined from 2000–2011 (from 42.4 in 2000 to 32.3 in 2011) before a gradual increase from 2012–2019.

For those aged 40–49 and 70–79/80–89 years, IRs increased rapidly between 2000 (and 2004 before gradually increasing and peaking around 2011–2014 (peaking at 156.7 in 2011 for 40–49 years; and peaking at 379.4 in 2014 for 70–79 years; and 402.8 in 2014 for 80–89 years) then declining (to a low of 121.5 in 2020 for 40–49; 255.2 in 2020 for 70–79; 296.8 in 2020 for 80–89 years).

For those aged 50–69 years, IRs increased from 2000–2005 (from 296.8 in 2000 to 467.0 in 2004 for 50–59; and from 294.3 to 556.8 in 2005 for 60–69 years) before a gradual decline from 2006–2020.

For those aged 90+ years, there were differences between the two databases with IRs in GOLD declining but with a peak in 2013 (474.0); whereas IRs in Aurum increased over the study period, peaking in 2018 (394.1). For all age groups, IRs decreased in 2020 before increasing in 2021, apart from those aged 30–39 years. For males, there were not enough cases per age group to assess trends in annualised IR across age groups apart from those aged 70–79 years, which shows the stability of IRs over the study period (Supplement S4).

All results for this study can be found and downloaded in a user-friendly interactive web application: <https://dpa-pde-oxford.shinyapps.io/BreastCancerIncPrevSurvShiny/>.

Sex	Database	Age Group (years)	n cases	n population	person years	Incidence (per 100,000 person years)
Female	CPRD GOLD	18–29	278	2,034,573	7,907,080	3.52 (3.11–3.95)
		30–39	2860	1,819,975	7,470,652	38.3 (36.9–39.7)
		40–49	11,298	1,562,588	7,860,585	143.7 (141.1–146.4)
		50–59	21,496	1,403,726	7,144,276	300.9 (296.9–304.9)
		60–69	22,273	1,131,038	5,845,509	381.0 (376.0–386.1)
		70–79	15,090	876,415	4,388,250	343.9 (338.4–349.4)
		80–89	9665	604,872	2,634,476	366.9 (359.6–374.3)
	90+	2440	223,748	682,359	357.6 (343.5–372.1)	
	CPRD Aurum	18–29	485	4,901,661	15,388,591	3.15 (2.88–3.45)
		30–39	5888	4,085,576	15,968,841	36.9 (35.9–37.8)
		40–49	22,064	3,084,523	15,854,144	139.2 (137.3–141.0)
		50–59	38,135	2,609,920	13,699,589	278.4 (275.6–281.2)
		60–69	38,685	2,027,941	10,923,859	354.1 (350.6–357.7)
		70–79	28,048	1,603,425	8,358,509	335.6 (331.7–339.5)
80–89		18,696	1,131,288	5,152,779	362.8 (357.7–368.1)	
90+	4806	452,649	1,557,892	308.5 (299.8–317.3)		
Male	CPRD GOLD	18–29	5	1,836,496	8,110,918	0.06 (0.02–0.14)
		30–39	9	1,861,598	7,627,706	0.12 (0.05–0.22)
		40–49	29	1,680,277	8,204,536	0.35 (0.24–0.51)
		50–59	67	1,469,617	7,443,890	0.90 (0.70–1.14)
		60–69	129	1,143,727	5,879,519	2.19 (1.83–2.61)
		70–79	163	787,598	3,893,012	4.19 (3.57–4.88)
		80–89	87	408,602	1,723,760	5.05 (4.04–6.23)
	90+	18	94,907	267,380	6.73 (3.99–10.6)	
	CPRD Aurum	18–29	7	4,336,623	16,551,872	0.04 (0.02–0.09)
		30–39	19	4,204,747	16,690,740	0.11 (0.07–0.18)
		40–49	70	3,420,966	16,956,055	0.41 (0.32–0.52)
		50–59	141	2,797,488	14,679,807	0.96 (0.81–1.13)
		60–69	268	2,098,707	11,213,511	2.39 (2.11–2.69)
		70–79	302	1,471,532	7,588,042	3.98 (3.54–4.46)
80–89		181	789,877	3,508,496	5.16 (4.43–5.97)	
90+	31	202,246	666,457	4.65 (3.16–6.60)		

Table 2. Overall crude incidence rates of breast cancer, stratified by age and sex in CPRD GOLD/ Aurum.

Overall and annualised crude prevalence for the study population stratified by age and sex across databases

Crude prevalence

In GOLD, the crude prevalence for breast cancer in 2021 was 2.11% (2.09%–2.14%) for females and 0.009% (0.008%–0.011%) for males, which is equivalent to 2,110 cases per million people for females, and 90 cases per million people for males. Similar prevalence was obtained in 2019 when comparing GOLD and Aurum across sexes. When stratifying by age, prevalence in GOLD in 2021 peaked in females aged 70–79 years (5.39%) and in males 90+ years (0.06%) with similar trends in 2019 when comparing both databases (Supplement S5).

Annualised crude prevalence

In GOLD, prevalence increased from 2000–2013 before stabilising to 2018 in females and declining to the end of the study period in males; whereas in Aurum, prevalence increased each year over the study period for females and males (Fig. 3). Both sexes have seen around 2.5-fold increase in PP across the study period in both databases (Fig. 3). The annual percent change was significantly different from zero across all timepoints, except for between 2013–2018 for males in CPRD GOLD (supplementary figures S6–S9).

When stratifying by age, prevalence trends over time for females showed some differences per age group (Fig. 4). Those aged 40–49 and 60–69 years showed increases in prevalence over time until 2014 where prevalence stabilised to the end of the study period, with a small decline in those aged 40–49 years. For all other age groups prevalence increased over the study period.

For males, there again were differences in prevalence trends over the study period depending on age (Fig. 5). For those aged 40–49 years, prevalence was stable for most of the study period with an increase in prevalence from 2012/14. For those aged 50–79 years, prevalence increased over the study period in both databases. In GOLD, prevalence increased between 2000 and 2018 and declined thereafter; whereas in aurum, prevalence increased over time for those aged 80–89 years. for those aged 30–39 years and 90+ years there was not enough

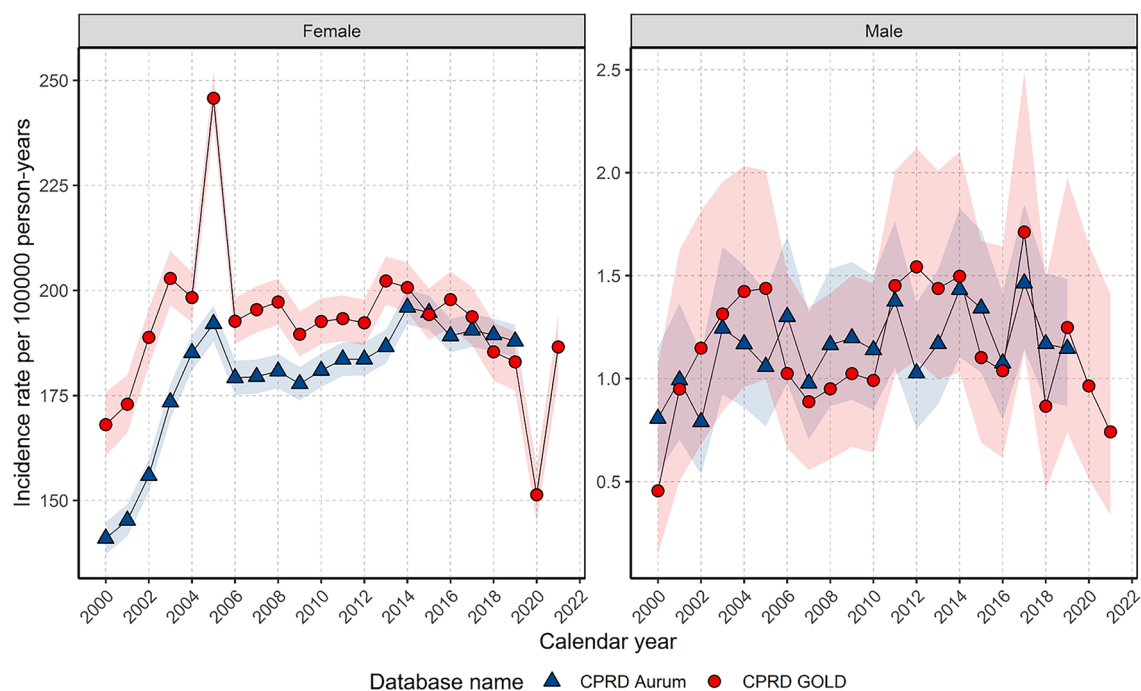


Figure 1. Annualised crude incidence rates for breast cancer stratified by database and sex.

data to assess trends in gold; however, in aurum prevalence trends indicated stability in those aged 30–39 years and an increase in prevalence over time in those aged over 90 years.

Overall survival probability for breast cancer population stratified by age, sex and calendar year

For females, there were 84,984 patients, 19,974 deaths and a median follow-up of 4.7 years; and for males, there were 505 patients, 173 deaths and a median follow-up of 3.8 years in GOLD (Fig. 6). Number of patients at risk, died and censored across the follow-up period is indicated in Supplement S10. Median survival was not reached for females within the specified follow-up period, whereas for males, median survival was between 10–11 years across both databases. What this means is that, on average, females lived beyond the duration of the follow-up period; whereas males survived for a period of time falling within the range of 10 to 11 years during the specified follow-up period. Indeed, survival probability for females after one-, five-, and ten-years after diagnosis was 95.1%, 80.2%, and 68.4%, and for males 92.9%, 69.0%, and 51.3% in GOLD, with similar results in Aurum. Long-term survival probability at five- and ten-years was higher in females than males across both databases (Supplement S11).

For females, when stratifying by age group, for those aged 18–69 years median survival was not reached. For those aged older than 70 years, median survival decreased with increasing age. Median survival decreased from 11–12 years to 2.5 for those older than 90 years. For males, median survival was not achieved in those aged 40–69 and 90+ years. However, median survival was lower in those aged 80–89 years compared to those aged 70–79 years across both databases.

For females, one-year survival probability for those aged 18–69 years was similar (97–98%), peaking in those aged 40–59 years, and declining from 70 years (Table 3). After five- and ten-years, survival rates increased from 18–29 years peaking in those aged 50–59 years before declining. For short- and long-term survival probability for males, results indicate that those aged 80 years and older had lower short- and long-term survival probability compared to younger age groups, however, sample numbers were small.

To investigate if survival probability has changed over time, we stratified by calendar time of cancer diagnosis in five-year windows. Figure 7 shows the KM survival curves stratified by sex and calendar year. For females, short- and long-term survival probability increased over calendar time. Survival probability at one-year increased by 2.1%, and survival probability at five-years increased by 5.4% from 2000–2004 to 2015–2019 in GOLD (note that survival data stratified by calendar year was not available for Aurum). For males, when comparing survival probability between those diagnosed in 2000–2004 with those diagnosed in 2015–2019, there was no clear pattern over calendar time for short-term and long-term survival probability due to small sample numbers.

Supplement S12 shows the short- and long-term survival probabilities and 95% confidence intervals stratified by calendar year of diagnosis, age and sex. Short-term survival probability in the different age groups showed that in females those aged between 50–69 years had increases in survival probability over time when comparing those diagnosed in 2000–2004 to those diagnosed between 2015–2019 (one-year survival probability of 97.4% vs. 99.0% in those aged 50–59 years; and 95.6% vs. 97.8% in those aged 60–69 years from 2000–2004 vs. 2015–2019). There was a similar pattern of increased survival probability over time for long-term (five-year) survival probability in

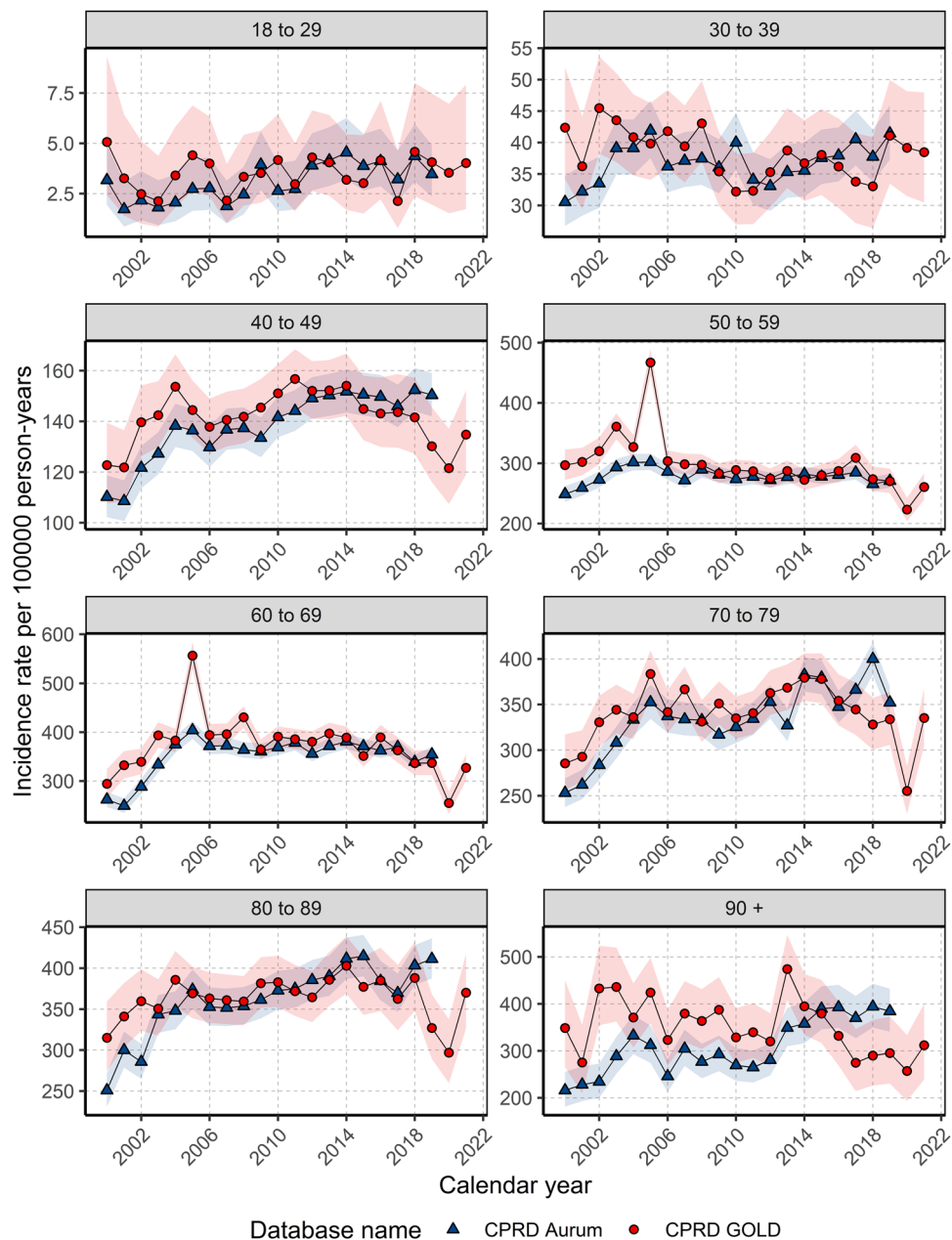


Figure 2. Annualised crude incidence rates for females stratified by database and age group.

those aged 50–69 years when comparing those diagnosed in 2000–2004 to those diagnosed between 2015–2019 (five-year survival probability of 88.2% vs. 92.4% in those aged 50–59 years; and 84.5% vs. 89.2% in those aged 60–69 years from 2000–2004 vs. 2015–2019). For those aged 80–89 years results showed an improvement in short- and long-term survival probability over calendar time (one-year survival was 84.3% vs. 88.1% for those diagnosed in 2000–2004 vs. 2015–2019). There was no significant change in short-term survival probability in any age groups in those diagnosed in 2020–2021 compared to previous years. For males, trends stratified by age did not show improvement over time in short- and long-term survival probability, however, sample numbers were small and for certain age groups not enough data to assess secular trends.

Discussion

This study demonstrates trends of breast cancer incidence, prevalence and survival probability in the UK in females and males. Below is a summary of the key findings in the context of previous research.

Overall incidence and prevalence for study population stratified by age and sex

Overall crude incidence rates of breast cancer in females (IR: 194 per 100,000 person-years) and males (IR: 1.2 per 100,000 person-years) were in line with national statistics (IR: 166 and 1.1 for females and males, respectively,

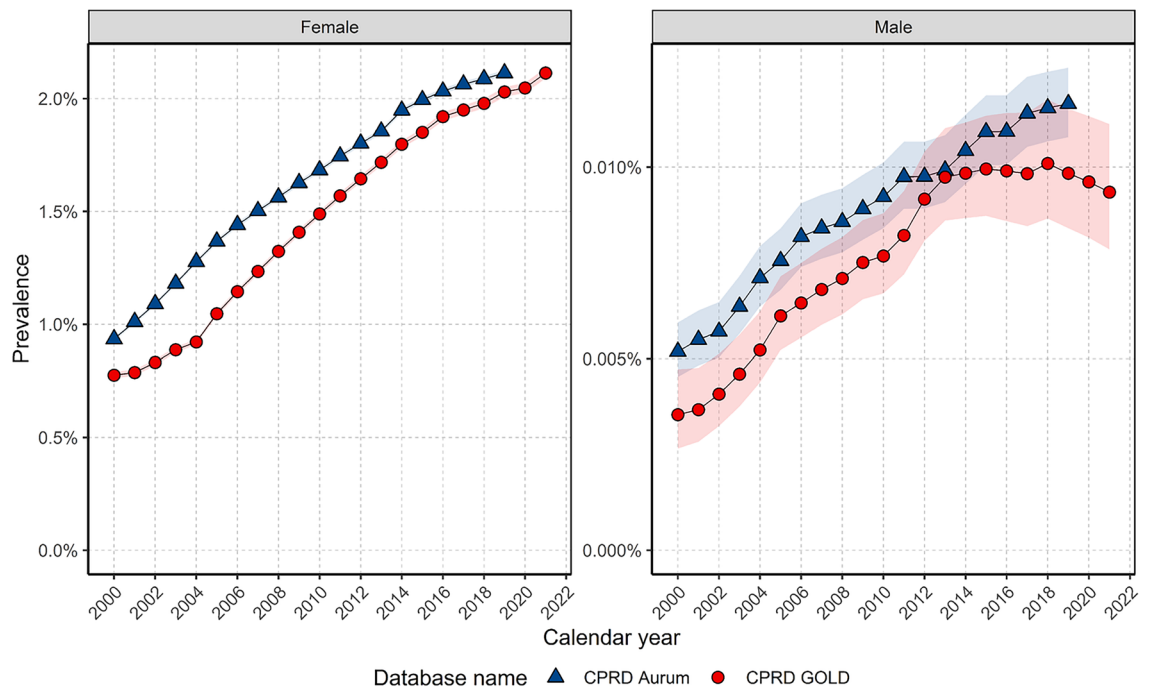


Figure 3. Annualised crude prevalence stratified by database and sex.

between 2016–2018 from Cancer Research UK)³. Incidence of breast cancer increased with age in both females and males peaking in those aged 60–69 years in females and those aged 90+ in males. These estimates deviate slightly with national statistics, which demonstrate a peak in females aged 90+ years and in males 85–89 years³. The drop in overall incidence for females aged 70–79 years observed here is likely to coincide with the ending of routine breast cancer screening in the UK (women are eligible for the breast cancer screening programme between the ages of 50 and 70 years¹⁸). Thus, this is likely a compensatory decrease in incidence as screening has advanced the detection of cases among women in this age bracket³. This also explains why the incidence subsequently reverts to somewhat higher rates among those aged 80–89 years. Overall incidence in those 90+ years is lower than younger ages, which could indicate reduced diagnostic activity, perhaps due to general ill health in this age group.

National data on prevalence of breast cancer is scarce. In this study prevalence of breast cancer at the end of the study period was 2.1% in females, peaking in those aged 70–79 years, and 0.009% in males, peaking in those aged 90+ years. That said, these could be overestimates of population prevalence as in this study anyone with a diagnosis of breast cancer was included until the end of their observation period. Patients with survival over 5–10 years who are discharged would still be contributing to the prevalence estimate. Furthermore, the increase of early-stage breast cancers in the context of screening programmes is likely to drive this overestimation further due to patients surviving longer. However, while many of these cases may be considered cured after five years and no longer being actively treated, people in this survivorship phase may have long-term medical needs and accordingly, it is important to provide accurate counts to allow for healthcare planning.

Trends in incidence and prevalence over time for females and males

In terms of trends over time, 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. Largely speaking, changes in incidence of breast cancer over time are likely explained by many factors, including increased screening and diagnostic activity, but also increases in risk factors in the population (such as obesity^{19,20}). Another time trend of note is the sharp spike in incidence in females in 2004–2005. One possible explanation for this is that the Quality and Outcomes Framework (QOF), which assesses performance of general practices on several key disease areas (including cancer), and provides financial incentives for achieving specified quality targets, was introduced in 2004 (of note, there were substantially more patients from Scotland with a date of diagnosis in April 2004, which is the start of QOF reporting period). Thus, screening, diagnostics and reporting of cancer diagnoses may have been greater during this time-period. Additionally, many screening units in the UK extended the screening of women aged 65–70 years between 2001 and 2004²¹, and it is possible that this extension accounted for the large spike observed.

When examining incidence trends over time by age group, three key findings are highlighted. First is the increase in incidence over the study period for younger women (aged 18–29; and 30–39). Several possible explanations include: increasing awareness of breast symptoms leading to more women being diagnosed; but also risk factor exposures in early life such as earlier age of thelarche (pubertal phase of breast development) and menarche than previous decades^{22,23}, leading to increased cumulative exposure to oestrogen; and increased use

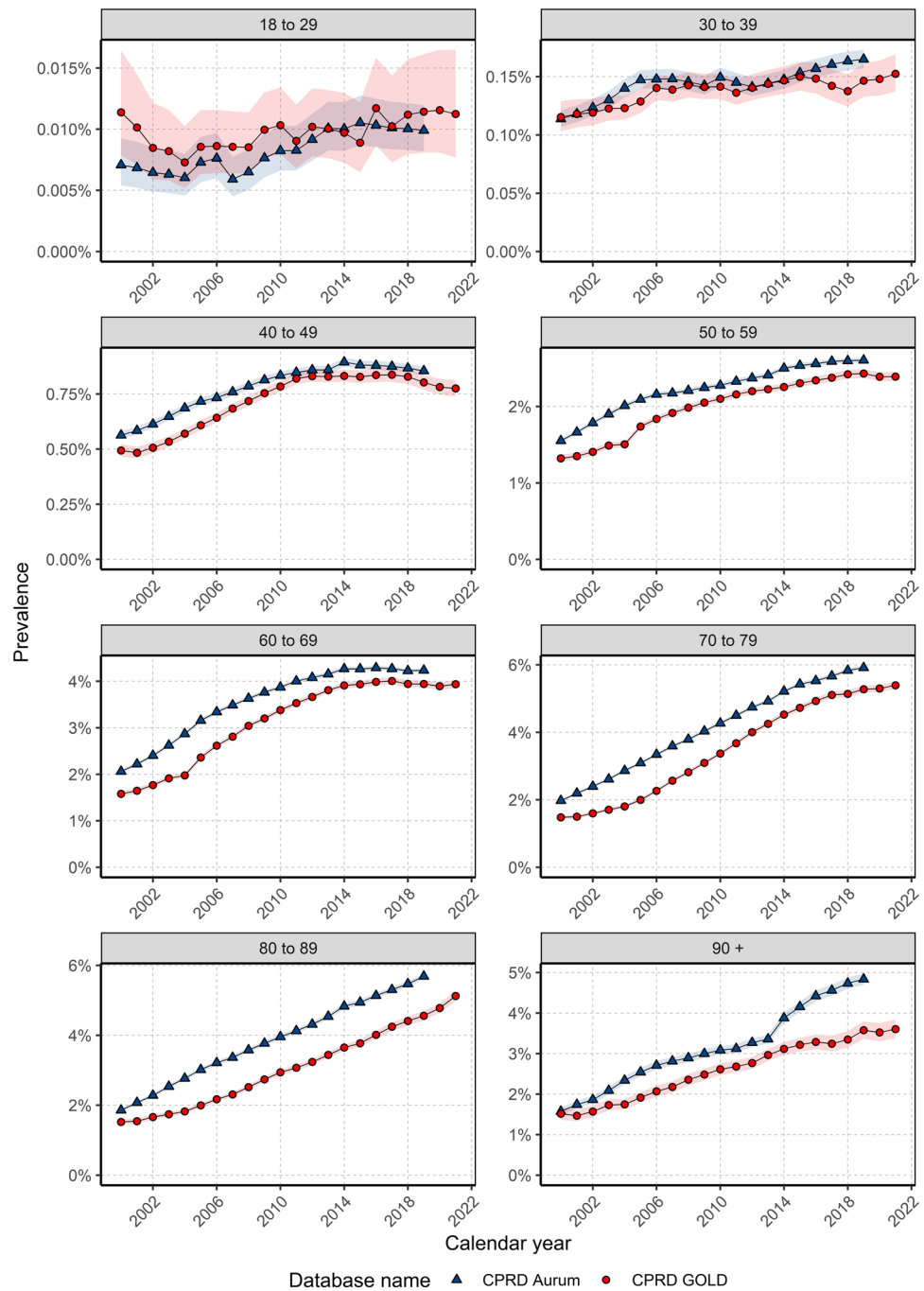


Figure 4. Annualised crude prevalence for females stratified by database and age.

of hormonal contraceptives which pose an elevated risk for breast cancer²⁴. Second, the decline in incidence for women aged 50–69 years from 2005 to 2019 may be a reflection of the success of screening programmes; and third, the decline from 2015 onwards for women aged 70–89 years coincides with the launch of the “Be Clear on Cancer” campaign aimed at women in this age group²⁵.

Prevalence of breast cancer in females and for most age groups in males increased across the study period, which is likely a reflection of increased survival due to the success of screening programmes (for females at least), early diagnosis and effective treatments^{26,27}.

Differences in short-term and long-term survival probability in different age groups in females and males

Short-term (one-year) survival probability in females was similar across the age groups from 18–69 years; whereas long-term survival probability at five-years was low in younger age groups, and highest in those aged 50–59 years

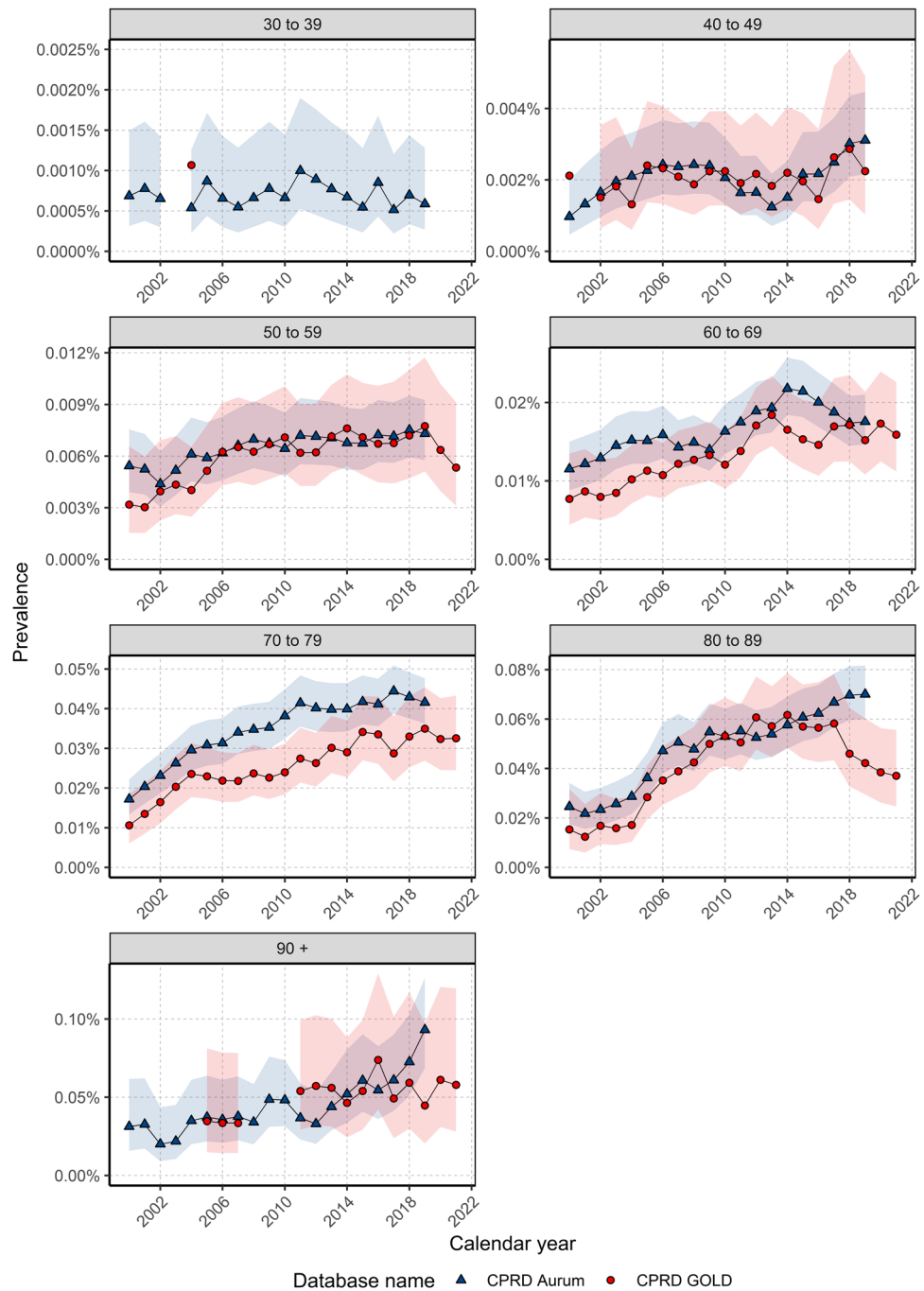


Figure 5. Annualised crude prevalence for males stratified by database and age.

likely due to the eligibility of females into national breast cancer screening programmes in the UK (which starts at 50 years of age)¹⁸. It should also be noted that women typically transition through menopause from age 50 years, and breast cancer that develops during menopause typically progresses more slowly and is less aggressive than earlier onset breast cancer, which may account for age-related differences in survival²⁸.

Generally speaking, males have lower long-term survival compared to females which is in line with previous studies²⁹. This could be due to several factors such as age and disease severity. Males tend to be older when diagnosed compared to females and it is likely that older males present with more comorbidities and medication use making treatment decisions more complex. Males also tend to present with more advanced disease likely due to the rarity of the disease and consequential delays in diagnosis^{5,30}. Furthermore, as males are underrepresented in trials, treatment recommendations follow those for postmenopausal women³¹. Therefore, the current management of male breast cancer might not be ideal and could explain the lower long-term survival compared to females. Additionally, males lack breast cancer screening and it is clear that at least in females this has contributed

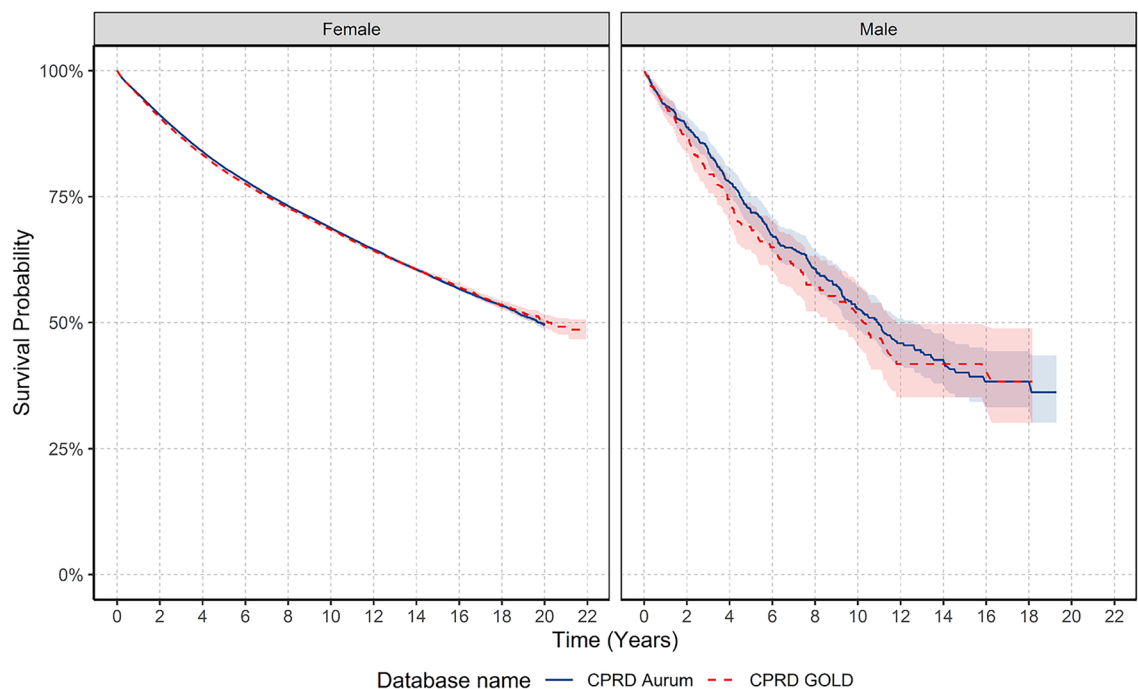


Figure 6. Kaplan–Meier survival curve of breast cancer stratified by database and sex.

Age Group (years)	One-year survival (%)		Five-year survival (%)		Ten-year survival (%)	
	GOLD	Aurum	GOLD	Aurum	GOLD	Aurum
Females						
18–29	98.0 (96.3–99.8)	97.0 (95.3–98.6)	84.2 (79.2–89.5)	80.9 (76.7–85.5)	75.3 (68.1–83.2)	71.3 (65.1–78.0)
30–39	97.4 (96.8–98.0)	98.1 (97.8–98.5)	83.7 (82.2–85.3)	83.9 (82.8–85.0)	74.2 (72.1–76.4)	75.4 (74.0–76.9)
40–49	98.4 (98.2–98.7)	98.6 (98.5–98.8)	87.9 (87.2–88.6)	89.6 (89.1–90.0)	80.5 (79.5–81.4)	82.0 (81.3–82.6)
50–59	98.2 (98.1–98.4)	98.3 (98.2–98.4)	90.6 (90.1–91.0)	90.4 (90.1–90.8)	83.3 (82.7–84.0)	83.4 (82.9–84.0)
60–69	97.2 (97.0–97.4)	97.5 (97.3–97.7)	87.2 (86.7–87.7)	87.5 (87.1–87.9)	77.0 (76.3–77.8)	77.2 (76.7–77.8)
70–79	93.2 (92.8–93.6)	94.1 (93.8–94.4)	73.3 (72.5–74.1)	75.8 (75.2–76.3)	55.1 (54.0–56.2)	57.1 (56.3–57.9)
80–89	86.7 (86.0–87.4)	87.4 (86.9–87.9)	50.9 (49.8–52.2)	53.8 (52.9–54.6)	24.6 (23.2–26.1)	27.1 (26.1–28.2)
90+	74.4 (72.6–76.3)	73.6 (72.3–74.9)	23.2 (21.0–25.6)	24.2 (22.6–25.8)	2.6 (1.4–4.7)	4.42 (3.33–5.86)
Males						
18–29	–	–	–	–	–	–
30–39	100	89.5 (76.7–100)	100	89.5 (76.7–100)	–	–
40–49	100	94.2 (88.8–99.9)	90.4 (78.6–100)	81.4 (71.4–92.8)	80.4 (61.4–100)	70.3 (56.9–86.9)
50–59	95.2 (89.0–100)	97.8 (95.4–100)	80.6 (70.4–92.3)	88.6 (82.6–95.0)	72.0 (59.6–86.9)	81.0 (72.9–90.1)
60–69	93.6 (89.4–98.0)	96.1 (93.7–98.5)	73.6 (65.6–82.5)	78.6 (73.2–84.3)	61.6 (52.0–72.9)	65.4 (58.4–73.1)
70–79	95.5 (92.2–98.8)	93.7 (91.0–96.6)	68.5 (60.4–77.6)	72.3 (66.7–78.3)	41.5 (30.9–55.6)	43.0 (35.7–51.6)
80–89	84.4 (76.9–92.6)	87.6 (82.9–92.6)	49.3 (38.3–63.6)	44.5 (36.4–54.4)	22.8 (12.1 – 43.0)	15.7 (8.35–29.6)
90+	81.9 (65.2–100)	68.1 (51.9–89.3)	23.6 (7.5–74.5)	45.3 (27.7–74.0)	–	–

Table 3. Overall survival rates after breast cancer diagnosis stratified by database age and sex.

a marked decrease in female breast cancer mortality since the late 1980s³². The National Comprehensive Cancer Network (NCCN) recommend that males aged 35 years and older with *BRCA* mutations receive self-examination training for breast cancer alongside annual clinical breast examination³³. Yet, the sporadic nature of genetic testing for such mutations may impede the practical realisation of this recommendation.

Survival probability over calendar time for whole population and age strata

For females, one-year survival probability increased by 2.08% and five-year survival probability increased by 5.39% from 2000–2004 to 2015–2019. Improvements in survival over the past 20 years are echoed in data from the National Cancer Registration and Analysis Service which showed annual mortality rates of ~4% for females diagnosed between 1993–1999 reduced to around 1% for those diagnosed between 2010–2015. Similarly, 5-year

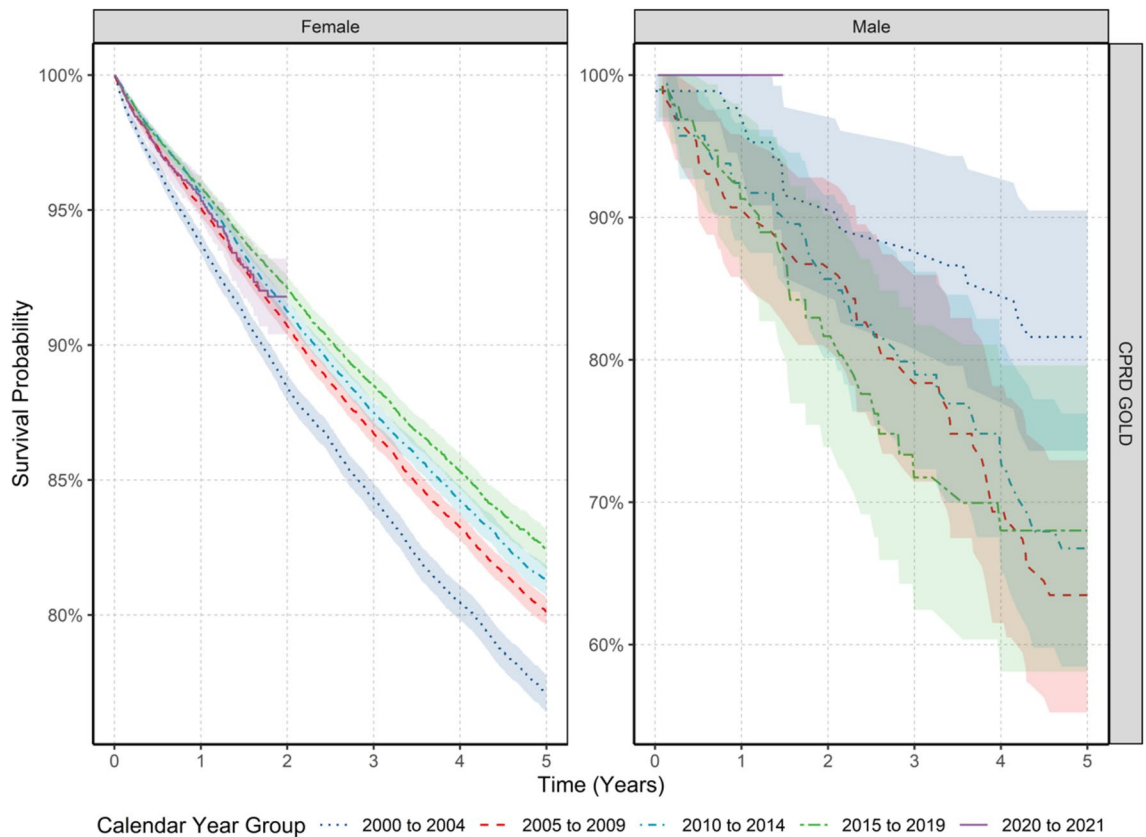


Figure 7. Kaplan–Meier survival curve of breast cancer stratified by sex and calendar year of diagnosis.

cumulative mortality risk reduced from 14.4% for females diagnosed between 1993–1999 to 4.9% for those diagnosed between 2010–2015¹³. In the current study, survival probability particularly improved for females aged 50–70 years, not surprisingly coinciding with eligibility into national screening programmes. Reassuringly, there did not appear to be an effect of the COVID-19 pandemic on short-term survival probability for those diagnosed between 2020–2021 compared to those diagnosed in the years prior. This is somewhat surprising, given data that suggests that screening, diagnosis and treatments were impacted by the pandemic^{34–36}.

For males, both short-term and long-term survival probability did not show improvements across calendar periods, but this is likely driven by small sample sizes. Other data shows death rates in North-Western Europe decreased between 10%–40% from 2000–2004 to 2015–2017¹⁴, and so further data is required before we have clear evidence on the survival trends from male breast cancer in the UK.

Strengths and limitations

The main strength of this study is the use of two large primary care databases covering the whole of the UK. CPRD GOLD covers primary care practices from England, Wales, Scotland and Northern Ireland (with greater representation from Scotland rather than England), whereas CPRD Aurum covers primary care practices in England. The similarity between the results in both databases provides increased generalizability across the UK. Nevertheless, there were a few discrepancies in results between the two databases which can partly be explained by differences in observation period for patients across the study period. In GOLD, the number of people in the database steadily increased from 2000 up to 2006, then remained stable until 2011 before a gradual decline. This gradual decline is likely due to GP practices in England moving EMIS clinical systems. Furthermore, over time the demographic representation of GOLD has changed which could explain differences in results. The advent of the CPRD Aurum database saw some practices transferred from GOLD to Aurum. Across our observation period practices from England and Northern Ireland reduced, whilst practices from Scotland and Wales increased.

Another strength of our study is the inclusion of a complete study population database for the assessment of incidence and prevalence. In contrast, cancer registry studies extrapolate the registry data to the whole population using national population statistics, potentially introducing biases^{16,17}. The high validity and completeness of mortality data with over 98% accuracy compared to national mortality records³⁷ allowed us to examine the impact of calendar time on overall survival probability—one of the key outcomes in cancer care.

Our study had limitations. First, we used primary care data without linkage to cancer registry potentially leading to misclassification and delayed recording of diagnoses. However, previous validation studies have shown high accuracy and completeness of cancer diagnoses in primary care records³⁸. Second, our use of primary care records precluded us from studying tumour histology, genetic mutations, staging or cancer therapies, which can all impact breast cancer survival. Therefore, our survival probability estimates may overestimate survival

in those with higher staging as well those with specific genetic mutations such as *BRCA 1/2*³⁹. Other factors such as socio-economic status and ethnicity could also result in different values for incidence, prevalence and survival^{40–42}. Third, in this study we calculated overall survival probability which does not differentiate between deaths caused by cancer vs. other causes. Therefore, it is a broad measure of overall survival probability rather than specifically cancer mortality.

Conclusion

Our study demonstrates that changes in incidence of breast cancer in females largely reflect the success of national breast cancer screening programmes, as rates rise and fall in synchronicity with ages of eligibility for such programmes. Overall survival probability for females from breast cancer has improved from 2000 to 2021, again reflecting the success of screening programmes, early diagnosis, and improvements in treatments. Male breast cancer patients, however, have worse survival outcomes compared with those of female patients. This highlights the need to develop male-specific treatment strategies and promote education and self-examination recommendations of breasts in males, given there are no screening programmes, to improve long-term survival in line with females.

Methods

Study design, setting, and data sources

We carried out a population-based cohort study using routinely collected primary care data from the United Kingdom (UK). People with a diagnosis of breast cancer and a background cohort were identified from Clinical Practice Research Datalink (CPRD) GOLD to estimate overall survival probability, incidence, and prevalence. We additionally carried out this study using CPRD Aurum and compared to the results in GOLD. Both these databases contain pseudo-anonymised patient-level information on demographics, lifestyle data, clinical diagnoses, prescriptions, and preventive care provided by GPs and collected by the NHS as part of their care and support. CPRD GOLD contains data from across the UK, representing 6.9% of the UK population⁴³, whereas Aurum only contains data from England, but represents 13% of the population of England⁴⁴. CPRD GOLD and Aurum have been mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to facilitate replication of analyses^{45,46}.

The protocol for this research was approved by the Research Data Governance (RDG) Board of the Medicine and Healthcare products Regulatory Agency database research (protocol number 22_001843). The data is provided by patients to their GPs and collected by the NHS as part of their care and support, and so consent is provided to GPs prior to inclusion in this study. All methods were carried out in accordance with ethical principles outlined in the Declaration of Helsinki; and the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance.

Study participants and time at risk

All participants were required to be aged 18 years or older and have at least one year of data availability prior to diagnosis, and information on age and sex. For the incidence and prevalence analysis, the study cohort consisted of individuals present in the database from 1st January 2000. For CPRD GOLD, these individuals were followed up to whichever came first: diagnosis of breast cancer, exit from the database, date of death, or the 31st of December 2021 (the end of the study period), whereas for Aurum, the end of the study period was 31st of December 2019 (due to data availability). For the survival analysis, only individuals with a newly diagnosed cancer were included. These individuals were followed up from the date of their diagnosis to either date of death, exit from the database, or end of the study period. Any patients whose death date and cancer diagnosis date occurred on the same date were removed from the survival analysis.

Outcome definitions

We used Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) diagnostic codes to identify breast cancer events (see Supplementary Table S1). Diagnostic codes indicative of either non-malignant cancer or metastasis were excluded (apart from prevalence analyses), as well as diagnosis codes indicative of melanoma, sarcoma, lymphoma, and other tumors not originating from breast tissue. Note that prior diagnoses of other cancers were not excluded. The study outcome cancer definition was reviewed using the CohortDiagnostics R package⁴⁷. This package was used to identify additional codes of interest and to remove those highlighted as irrelevant based on feedback from clinicians with oncology expertise through an iterative process during the initial stages of analysis. For survival analysis, mortality was defined as all-cause mortality based on records of date of death. Mortality data in CPRD GOLD has been previously validated and shown to be over 98% accurate³⁷.

Statistical methods

The population characteristics of patients with a diagnosis of breast cancer were summarised on a range of comorbid conditions using standardised SNOMED codes, with median and interquartile range (IQR) used for continuous variables and counts and percentages used for categorical variables.

We calculated the overall and annualised crude incidence rates (IR) and annualised crude prevalence for breast cancer from 2000 to 2021. For incidence, the number of events, the observed time at risk, and the incidence rate per 100,000 person-years (pys) were summarised along with 95% confidence intervals. Annualised incidence rates were calculated as the number of incident cancer cases as the numerator and the recorded number of person-years in the general population within that year as the denominator, whereas overall incidence was calculated from 2000 to 2021.

Prevalence was calculated on 1st January for the years 2000 to 2021 with the number of patients aged ≥ 18 years fulfilling the case definition for breast cancer as the numerator. The denominators were the participants ≥ 18 years on 1st January in the respective years for each database. The number of events and prevalence (%) were summarised along with 95% confidence intervals. Joinpoint regression using the National Cancer Institute's Joinpoint Trend analysis software⁴⁸ was utilised to determine the significance of changes in prevalence across time.

For survival probability analysis, we used the Kaplan–Meier (KM) method to estimate the overall survival probability from observed survival times with 95% confidence intervals⁴⁹. We estimated the median survival and survival probability one-, five-, and ten-year after diagnosis.

All results were stratified by database, age (ten-year age bands apart from the first and last age bands which were 18–29 years and 90+ years, respectively) and sex. For survival analysis, we additionally stratified by calendar time of cancer diagnosis (2000–2004, 2005–2009, 2010–2014, 2015–2019 and 2020–2021) allowing a maximum of five years follow-up from cancer diagnosis. To avoid re-identification, we do not report results with fewer than five cases.

To compare with results obtained from GOLD, the same statistical analyses were performed in Aurum using data from 1st January 2000 to 31st December 2019 (note that data from Aurum beyond 2019 was not available).

The statistical software R version 4.2.3 was used for analyses. For calculating incidence and prevalence, we used the IncidencePrevalence R package⁵⁰. For survival analysis we used the survival R package. The analytic code to perform the study is available at <https://github.com/oxford-pharmacoepi/EHDENcancerIncidencePrevalence>.

Data availability

This study is based in part on data from the Clinical Practice Research Datalink (CPRD) obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone. Patient level data used in this study was obtained through an approved application—the CPRD (application number 22_001843) and is only available following an approval process—safeguard the confidentiality of patient data. Details on how to apply for data access can be found at <https://cprd.com/data-access>.

Received: 17 January 2024; Accepted: 30 July 2024

Published online: 17 August 2024

References

- Sung, H. *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **71**, 209–249 (2021).
- Giordano, S. H. Breast cancer in men. *N. Engl. J. Med.* **378**, 2311–2320 (2018).
- Cancer Research UK. *Breast cancer incidence (invasive) statistics*, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-Zero> (2021).
- Speirs, V. & Shaaban, A. M. The rising incidence of male breast cancer. *Breast Cancer Res. Treat.* **115**, 429–430 (2009).
- Giordano, S. H., Cohen, D. S., Buzdar, A. U., Perkins, G. & Hortobagyi, G. N. Breast carcinoma in men: a population-based study. *Cancer* **101**, 51–57 (2004).
- Brinton, L. A. *et al.* Prospective evaluation of risk factors for male breast cancer. *J. Natl. Cancer Inst.* **100**, 1477–1481 (2008).
- Ding, Y. C., Steele, L., Kuan, C.-J., Greilac, S. & Neuhausen, S. L. Mutations in BRCA2 and PALB2 in male breast cancer cases from the United States. *Breast Cancer Res. Treat.* **126**, 771–778 (2011).
- Brinton, L. A. *et al.* Prediagnostic sex steroid hormones in relation to male breast cancer risk. *J. Clin. Oncol.* **33**, 2041 (2015).
- Macke, A. J. & Petrosyan, A. Alcohol and prostate cancer: Time to draw conclusions. *Biomolecules* **12**, 375 (2022).
- Shield, K. D., Soerjomataram, I. & Rehm, J. Alcohol use and breast cancer: a critical review. *Alcohol. Clin. Exp. Res.* **40**, 1166–1181 (2016).
- Pati, S., Irfan, W., Jameel, A., Ahmed, S. & Shahid, R. K. Obesity and cancer: A current overview of epidemiology, pathogenesis, outcomes, and management. *Cancers* **15**, 485 (2023).
- Ron, E., Ikeda, T., Preston, D. L. & Tokuoka, S. Male breast cancer incidence among atomic bomb survivors. *J. Natl. Cancer Inst.* **97**, 603–605 (2005).
- Taylor, C. *et al.* Breast cancer mortality in 500 000 women with early invasive breast cancer in England, 1993–2015: Population based observational cohort study. *BMJ* **381**, 1 (2023).
- Pizzato, M. *et al.* Trends in male breast cancer mortality: A global overview. *Eur. J. Cancer* **30**, 472–479 (2021).
- Eggemann, H. *et al.* Survival benefit of tamoxifen in male breast cancer: Prospective cohort analysis. *Br. J. Cancer* **123**, 33–37 (2020).
- Sarfati, D., Blakely, T. & Pearce, N. Measuring cancer survival in populations: Relative survival vs cancer-specific survival. *Int. J. Epidemiol.* **39**, 598–610 (2010).
- Swerdlow, A. Cancer registration in England and Wales: Some aspects relevant to interpretation of the data. *J. R. Stat. Soc.* **149**, 146–160 (1986).
- Cancer Research UK. *Breast Screening*, [https://www.cancerresearchuk.org/about-cancer/breast-cancer/getting-diagnosed/screening-breast#:~:text=Who%20has%20breast%20screening%3F,are%2052%20or%2053%20years](https://www.cancerresearchuk.org/about-cancer/breast-cancer/getting-diagnosed/screening-breast#:~:text=Who%20has%20breast%20screening%3F,are%2052%20or%2053%20years.). (2023).
- Woolcott, O. O. & Seuring, T. Temporal trends in obesity defined by the relative fat mass (RFM) index among adults in the United States from 1999 to 2020: A population-based study. *BMJ open* **13**, e071295 (2023).
- Malik, V. S., Willet, W. C. & Hu, F. B. Nearly a decade on—Trends, risk factors and policy implications in global obesity. *Nature Reviews Endocrinology* **16**, 615–616 (2020).
- Bennett, R., Blanks, R. & Moss, S. Evaluation of extension of breast screening to women aged 65–70 in England using screening performance measures. *Br. J. Cancer* **100**, 1043–1047 (2009).
- Eckert-Lind, C. *et al.* Worldwide secular trends in age at pubertal onset assessed by breast development among girls: A systematic review and meta-analysis. *JAMA Pediatr.* **174**, e195881–e195881 (2020).
- Morris, D. H., Jones, M. E., Schoemaker, M. J., Ashworth, A. & Swerdlow, A. J. Secular trends in age at menarche in women in the UK born 1908–93: Results from the Breakthrough Generations Study. *Paediatr. Epidemiol.* **25**, 394–400 (2011).
- Fitzpatrick, D., Pirie, K., Reeves, G., Green, J. & Beral, V. Combined and progestagen-only hormonal contraceptives and breast cancer risk: A UK nested case–control study and meta-analysis. *Plos Medicine* **20**, e1004188 (2023).

25. Public Health England. *Public Health England launches nationwide breast cancer campaign*, <https://www.gov.uk/government/news/public-health-england-launches-nationwide-breast-cancer-campaign#:~:text=The%20Be%20Clear%20on%20Cancer%20campaign%20is%20part%20of%20the,us%20to%20improve%20cancer%20survival>. (2015).
26. Arnold, M. *et al.* Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURV-MARK-2): A population-based study. *Lancet Oncol* **20**, 1493–1505 (2019).
27. Cancer Research UK. *Cancer mortality statistics*, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality#heading-Zero> (2023).
28. Benz, C. C. Impact of aging on the biology of breast cancer. *Crit. Rev. Oncol./Hematol.* **66**, 65–74 (2008).
29. Liu, N., Johnson, K. J. & Ma, C. X. Male breast cancer: An updated surveillance, epidemiology, and end results data analysis. *Clin. Breast Cancer* **18**, e997–e1002 (2018).
30. Miao, H. *et al.* Incidence and outcome of male breast cancer: an international population-based study. *J. Clin. Oncol.* **29**, 4381–4386 (2011).
31. Korde, L. A. *et al.* Multidisciplinary meeting on male breast cancer: summary and research recommendations. *J. Clin. Oncol.* **28**, 2114 (2010).
32. Cancer Research UK. *Breast cancer survival statistics*, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival#heading-Zero> (2014).
33. NCCN Clinical Practice Guidelines in Oncology. *Genetic/Familial High-Risk Assessment: Breast, Ovarian, and Pancreatic*, https://www.nccn.org/professionals/physician_gls/pdf/genetics_bop.pdf (2023).
34. Barclay, N. L. *et al.* The impact of the COVID-19 pandemic on incidence and short-term survival for common solid tumours in the United Kingdom: A cohort analysis. *Clin. Epidemiol.* **1**, 417–429 (2024).
35. Barclay, N. L. *et al.* Collateral effects of the COVID-19 pandemic on endocrine treatments for breast and prostate cancer in the UK: A cohort study. *Ther. Adv. Med. Oncol.* **16**, 17588359241253116 (2024).
36. Barclay, N. L. *et al.* The impact of the UK COVID-19 lockdown on the screening, diagnostics and incidence of breast, colorectal, lung and prostate cancer in the UK: A population-based cohort study. *Front. Oncol.* **14**, 1370862 (2024).
37. Gallagher, A. M., Dedman, D., Padmanabhan, S., Leufkens, H. G. & de Vries, F. The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol. Drug Saf.* **28**, 563–569 (2019).
38. Strongman, H., Williams, R. & Bhaskaran, K. What are the implications of using individual and combined sources of routinely collected data to identify and characterise incident site-specific cancers? A concordance and validation study using linked English electronic health records data. *BMJ Open* **10**, e037719 (2020).
39. Huszno, J., Kolosza, Z. & Grzybowska, E. BRCA1 mutation in breast cancer patients: Analysis of prognostic factors and survival. *Oncol. Lett.* **17**, 1986–1995 (2019).
40. Denu, R. A. *et al.* Racial and socioeconomic disparities are more pronounced in inflammatory breast cancer than other breast cancers. *J. Cancer Epidemiol.* **2017**, 1 (2017).
41. Taheri, M., Tavakol, M., Akbari, M. E., Almasi-Hashiani, A. & Abbasi, M. Relationship of socio economic status, income, and education with the survival rate of breast cancer: a meta-analysis. *Iran. J. Public Health* **48**, 1428 (2019).
42. Dong, J.-Y. & Qin, L.-Q. Education level and breast cancer incidence: a meta-analysis of cohort studies. *Menopause* **27**, 113–118 (2020).
43. Herrett, E. *et al.* Data resource profile: clinical practice research datalink (CPRD). *Int. J. Epidemiol.* **44**, 827–836 (2015).
44. Wolf, A. *et al.* Data resource profile: Clinical practice research datalink (CPRD) aurum. *Int. J. Epidemiol.* **48**, 1740–1740g (2019).
45. Voss, E. A. *et al.* Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J. Am. Med. Inform. Assoc.* **22**, 553–564 (2015).
46. Hripesak, G. *et al.* Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers. *Stud. Health Technol. Inform.* **216**, 574–578 (2015).
47. Gilbert, J., Rao, G., Schuemie, M., Ryan, P. & Weaver, J. *CohortDiagnostics: Diagnostics for OHDSI Cohorts*, [https://github.com/OHDSI/CohortDiagnostics](https://ohdsi.github.io/CohortDiagnostics) (2023).
48. National Cancer Institute. Joinpoint Regression Program, Version 5.0.2. *Statistical Methodology and Applications Branch, Surveillance Research Program* (2023).
49. Goel, M. K., Khanna, P. & Kishore, J. Understanding survival analysis: Kaplan-Meier estimate. *Int. J. Ayurveda Res.* **1**, 274 (2010).
50. Burn, E., Raventos, B., Catala, M., Du, M., Guo, Y., Black, A., Inberg, G., Lopez, K. *IncidencePrevalence: Estimate Incidence and Prevalence using the OMOP Common Data Model. R Package Verion 0.4.0*, <https://cran.r-project.org/web/packages/IncidencePrevalence/index.html> (2023).

Author contributions

All authors were involved in the study conception and design, interpretation of the results and the preparation of the manuscript. D.N. carried out data analysis for the manuscript. A.G. and I.T. reviewed the clinical codelist used in this study. N.L.B. and D.N. wrote the initial draft of the manuscript with D.N., E.B. and D.P.A. had access to CPRD. data. All authors were involved in the interpretation of the results, critically reviewed the final manuscript and gave consent for publication.

Funding

This activity under the European Health Data & Evidence Network (EHDEN) and OPTIMA has received funding from the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement No 806968 and No. 101034347 respectively. IMI2 receives support from the European Union's Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations (EFPIA). IMI supports collaborative research projects and builds networks of industrial and academic experts in order to boost pharmaceutical innovation in Europe. The views communicated within are those of OPTIMA. Neither the IMI nor the European Union EFPIA or any Associated Partners are responsible for any use that may be made of the information contained herein. The study funders had no role in the conceptualisation, design, data collection, analysis, interpretation of data, decision to publish, or preparation of the manuscript. Additionally, there was partial support from the Oxford NIHR Biomedical Research Centre. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Competing interests

Professor Daniel Prieto-Alhambra's research group has received research grants from the European Medicines Agency from the Innovative Medicines Initiative from Amgen Chiesi and from UCB Biopharma; and consultancy

or speaker fees (paid to his department) from Astellas Amgen Astra Zeneca and UCB Biopharma. NLB is Director of Sleep Universal Limited (though this work is in no way connected to the present manuscript). All other authors declare no conflicts of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69006-1>.

Correspondence and requests for materials should be addressed to D.P.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

OPTIMA Consortium

James N'Dow⁷, Wim Witjes⁷, Emma Jane Smith⁷, Carla Bezuidenhout⁷, Sarah Collen⁷, Karin Plass⁷, Torsten Gerriet Blum⁷, Angelika Borkowetz⁷, Peter-Paul Willemse⁷, Philip Cornford⁷, Saeed Dabestani⁷, Maurice Schlieff⁷, Juan Gómez Rivas⁷, Anders Bjartell⁸, Monique Roobol⁹, Katharina Beyer⁹, Lionne Venderbos⁹, Sebastiaan Remmers⁹, Daan Nieboer⁹, Raoul Boomsma⁹, Bertrand De Meulder¹⁰, Charles Auffray¹⁰, Nesrine Taibi¹⁰, Ayman Hijazy¹⁰, Albert Saporta¹⁰, Johann Pellet¹⁰, Imran Omar¹¹, Lesley Anderson¹¹, Steven MacLennan¹¹, Sara MacLennan¹¹, Valerie Speirs¹¹, Solveiga Zibaite¹¹, Moragh Boyle¹¹, Charlotte Murray¹¹, Dianne Brown¹¹, Demi McDonald¹¹, Andres Metspalu¹², Jaak Vilo¹², Raivo Kolde¹², Sulev Reisberg¹², Elena Sügis¹², Marek Oja¹², Telve Objartel¹², Alberto Briganti¹³, Giorgio Gandaglia¹³, Martina Faticoni¹³, Greta Matteuzzi¹³, Claude Chelala¹⁴, Louise Jones¹⁴, Maryam Abdollahyan¹⁴, Emanuela Gadaleta¹⁴, Guido Juckeland¹⁵, Michael Bussmann¹⁵, Daniel Kotik¹⁵, Artur Yakimovich¹⁵, Torsten Bauer¹⁶, Jens Kollmeier¹⁶, Jessica Werchan¹⁶, Torsten Blum¹⁶, Rebecca Graebig-Rancourt¹⁶, Tobias Sjöblom¹⁷, Chatarina Larsson¹⁷, Arvid Widenlou Nordmark¹⁷, Daniel Prieto-Alhambra¹⁸, Sara Khalid¹⁸, Edward Burn¹⁸, Antonella Delmestri¹⁸, Mahkameh Mafi¹⁸, Danielle Newby¹⁸, Cheryl Tan¹⁸, Nikolaus Forgó¹⁹, Antoni Napieralski¹⁹, Martina Wimmer¹⁹, Katharina Haimbuchner¹⁹, Saskia Kaltenbrunner¹⁹, Katja Hartl¹⁹, Kseniia Guliaeva¹⁹, Giuseppe Curigliano²⁰, Carmen Criscitiello²⁰, Stefania Morganti²⁰, Chiara Corti²⁰, Elena Dal Zotto²⁰, Nadia Harbeck²¹, Julian Koch²¹, Neal Navani²², Sam Janes²², Amyn Bhamani²², Stephane Lejeune²³, Mario Campone²⁴, Jean-Sebastien Frenel²⁴, Kevin Joubel²⁴, François Bocquet²⁴, Camille Berneur²⁴, Marion Laloue²⁴, Malvina Dutot²⁴, Ludovic Jacob²⁴, Delphine Macle²⁴, Stéphanie Thauvin²⁴, Fanny Seguin²⁴, Catherine Le Manach²⁴, Philippe Lambin²⁵, Anshu Ankolekar²⁵, Talita Duarte-Salles²⁶, Laura Perez²⁶, Valérie Vaccaro²⁷, Thomy Tonia²⁷, Céline Genton²⁷, Wouter van Geffen²⁷, Ilona Tietzova²⁷, Armin Frille²⁷, Vincent Fallet²⁷, Adrien Costantini²⁷, Simone Wesselmann²⁸, Christoph Kowalski²⁸, Nora Tabea Sibert²⁸, Ellen Griesshammer²⁸, Pippa Powell²⁹, Clare Williams²⁹, Sigrid van Dorp³⁰, Nadia Honing³⁰, Javier Téllez³¹, Sandra Garrido³¹, Roberto Galán³¹, Ruben Villoria³¹, Inmaculada Perea Fernández³¹, Paloma López de Arenosa Barbeito³¹, Enric Bousoño Borrull³¹, Laura Tur Giménez³¹, Soraly Hernandez³¹, Pablo Gonzalez Fuente³¹, Juan Miguel Auñón García³¹, José Carlos Barrios González³¹, Alvaro Morandeira Galban³¹, Andreas Kremer³², Maria Quaranta³², Sebastiano La Ferla³², Loic Marc³², Nils Christian³², Christian Bauer³²,

Mariana Pina³², Sigrid Auweter³³, Julia Reichwald³³, Corinna Zur Bensen-Thomas³³, Larissa Tschetsch³³, Francisco Pinto³³, Samuel Lesuffleur³⁴, Matthieu Blottière³⁴, Louise Dufлот³⁴, David Vallas³⁴, Pierre-Olivier Chaudé³⁴, Marie Baumier³⁴, Daniele Cremonini³⁴, Patrizia Torremante³⁵, Florian Fromm³⁵, Verena von Scharfenberg³⁵, Karin Rosenits³⁵, Nuno Azevedo³⁶, Marcel Hartig³⁷, Waltraud Kantz³⁷, Frederic Kube³⁷, Amanda Matthews³⁷, Bhakti Arondekar³⁷, Bruno Gori³⁷, Hagen Krüger³⁷, Julia Ilinares³⁷, Keith Wilner³⁷, Lucile Serfass³⁷, Lynn McRoy³⁷, Robert Miller³⁷, Simon Bauer³⁷, Sofia Simon³⁷, Georgios Papanastasiou³⁷, Karen Godbold³⁷, Edwina Cahill³⁷, Stefan Langhammer³⁷, Anne Adams³⁷, Sebastian Boie³⁷, Florian Reis³⁷, Susan Evans Axelsson³⁸, John-Edward Butler-Ransohoff³⁸, Imke Meyer³⁸, Selmin Ulusu Saatci³⁸, Samu Kurki³⁸, Helene Ostojic³⁸, Abdelali Majdi³⁸, Santiago Villalba³⁸, Sai Jasti³⁸, Adrian Wolny³⁸, Lisa Schneider³⁸, Adrian Rousset³⁹, Ivo Cleuren³⁹, Sandra Eketorp Sylvan³⁹, Ellie Paintin³⁹, Monika Pokrzepa³⁹, Nicolas Pourbaix³⁹, Carolin Lorber⁴⁰, Marlene Thomas⁴⁰, Stefanie Morris⁴⁰, Joao Mouta⁴⁰, Martina von Meyenn⁴⁰, Mahesh Shivhare⁴⁰, Thomas Metcalfe⁴⁰, Camille Andre⁴⁰, Tobias Schulte in den Bäumen⁴⁰, Jason Hannon⁴⁰, Alan Mark Hochberg⁴⁰, Kartick Sukumaran⁴⁰, Jie Shen⁴¹, Nareen Katta⁴¹, Yilin Xu⁴¹, Sean Turner⁴¹, John Ossyra⁴¹, David Dellamonica⁴², Heather Moses⁴², Yiduo Zhang⁴², Christophe Dufour⁴², Marcus Simon⁴², Maria Teresa Campos⁴², Hassan Naqvi⁴², Jens Ceder⁴², Olga Alekseeva⁴², Burkhard Mueller⁴³, Tobias Flösdorf⁴³, Ruben Koch⁴³, Anastasia Goette⁴³, Gustaf Hedström⁴⁴ & Per-Henrik Edqvist⁴⁴

⁷Stichting European Urological Foundation, Arnhem, The Netherlands. ⁸Lunds Universitet, Lund, Sweden. ⁹Erasmus Universitair Medisch Centrum Rotterdam, Rotterdam, The Netherlands. ¹⁰Association EISBM, Vourles, France. ¹¹The University Court of the University of Aberdeen, Aberdeen, UK. ¹²Tartu Ülikool, Tartu, Estonia. ¹³Universita Vita-Salute San Raffaele, Milan, Italy. ¹⁴Queen Mary University of London, London, UK. ¹⁵Helmholtz-Zentrum Dresden-Rossendorf EV, Dresden, Germany. ¹⁶Helios Klinikum Emil Von Behring GMBH, Berlin, Germany. ¹⁷Uppsala Universitet, Uppsala, Sweden. ¹⁸The Chancellor, Masters and Scholars of the University of Oxford, Oxford, UK. ¹⁹Universität Wien, Vienna, Austria. ²⁰Istituto Europeo di Oncologia SRL, Milan, Italy. ²¹Ludwig-Maximiliansuniversitaet Muenchen, Munich, Germany. ²²University College London, London, UK. ²³European Organisation for Research and Treatment of Cancer AISBL, Brussels, Belgium. ²⁴Institut de Cancerologie de L'ouest, Angers, France. ²⁵Universiteit Maastricht, Maastricht, The Netherlands. ²⁶Fundacio Institut Universitari Pera la Recerca A L'atencio Primaria de Salut Jordi Gol I Gurina, Barcelona, Spain. ²⁷European Respiratory Society, Lausanne, Switzerland. ²⁸Deutsche Krebsgesellschaft EV, Berlin, Germany. ²⁹European Lung Foundation, Sheffield, UK. ³⁰ttopstart BV, Rijswijk, The Netherlands. ³¹GMV Soluciones Globales Internet SAU, Madrid, Spain. ³²Information Technology for Translational Medicine (ITTM) SA, Esch-Sur-Alzette, Luxembourg. ³³Smart Reporting GMBH, Munich, Germany. ³⁴Owkin France, Paris, France. ³⁵Artic Innovation GMBH, Munich, Germany. ³⁶Ydeal. Net Software LDA, Santa Maria Da Feira, Portugal. ³⁷Pfizer Limited, Sandwich, UK. ³⁸Bayer Aktiengesellschaft, Leverkusen, Germany. ³⁹Amegen, Diegem, Belgium. ⁴⁰F. Hoffmann-LA Roche AG, Basel, Switzerland. ⁴¹AbbVie Inc, North Chicago, USA. ⁴²AstraZeneca AB, Södertälje, Sweden. ⁴³Mutabor Technologies GMBH, Hamburg, Germany. ⁴⁴Region Uppsala, Uppsala, Sweden.