Comorbidities, medication use, and overall survival in eight cancers: a multinational cohort study of 1.7 million patients across Europe

Irene López-Sánchez,[a,b] Anna Palomar-Cros,[a] Ravinder Claire,[c] Laura Pérez-Crespo,[a] Agustina Giuliodori,[a] Ian Koblbauer,[d] Jeremy Dietz,[c] Jamie Elvidge,[c] James Koh,[c] Asieh Golozar,[e] Juan Manuel Ramirez-Anguita,[f] Angela Leis,[f] Miguel-Angel Mayer,[f,g] Nicola Symmers,[h] Mahéva Vallet,[h] Colin McLean,[h] Peter S. Hall,[h] Mees Mosseveld,[i] Katia Verhamme,[i] Espen Enerly,[j] Peter Prinsen,[k] Jelle Evers,[k] Marek Oja,[l] Raivo Kolde,[l] Rafael Marcos-Gragera,[m,n] Eric Fey,[o,p] Kimmo Porkka,[o,p] Tiago Taveira-Gomes,[q,r,s] Fernanda Estevinho,[t] Alberto Moreno Conde,[u,v] Jesus Moreno Conde,[u,v] Carlos Miguez Sanchez,[v] Evelyne Fournier,[w] Andrea Pistillo,[a] Xihang Chen,[d] George Corby,[d] Abigail Robinson,[d] Maria T. Sanchez-Santos,[d] Antonella Delmestri,[d] Wai Yi Man,[d] Martí Català,[d] Marta Alcalde-Herraiz,[d] Edward Burn,[d] Daniel Prieto-Alhambra,[d,i] Talita Duarte-Salles,[a,i,*] and Danielle Newby[d]

[a]Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
[b]Programa de Doctorat en Metodologia de la Recerca Biomèdica i Salut Pública, Universitat Autònoma de Barcelona, Barcelona, Cerdanyola del Vallès, Spain
[c]National Institute for Health and Care Excellence, London, United Kingdom
[d]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, United Kingdom
[e]Nemesis Health, New York, USA
[f]Research Programme on Biomedical Informatics, Hospital del Mar Research Institute, Doctor Aiguader 88, 08003, Barcelona, Spain
[g]Data Science Unit, Hospital del Mar, Passeig Maritim 25-29, 08003, Barcelona, Spain
[h]Edinburgh Cancer Research Centre, Institute of Genetics and Cancer, The University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh, EH4 2XR, United Kingdom
[i]Erasmus MC University Medical Center, Rotterdam, the Netherlands
[j]Department of Research, Cancer Registry of Norway, Norwegian Institute of Public Health, Oslo, Norway
[k]Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, the Netherlands
[l]Institute of Computer Science, University of Tartu, Narva mnt 18, 51009, Tartu, Estonia
[m]Epidemiology Unit and Girona Cancer Registry, Oncology Coordination Plan, Catalan Institute of Oncology (ICO), CIBER of Epidemiology and Public Health (CIBERESP), Girona, Spain
[n]Biomedical Research Institute (IDIBGI-CERCA), Josep Carreras Leukemia Research Institute, Department of Medical Sciences, Medical School, University of Girona, Girona, Spain
[o]HUS Helsinki University Hospital, Helsinki, Finland
[p]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Helsinki, Finland
[q]Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, Porto, Portugal
[r]Faculty of Health Sciences, University Fernando Pessoa, Porto, Portugal
[s]SIGIL Scientific Enterprises, Dubai, United Arab Emirates
[t]Department of Oncology, Unidade Local de Saúde de Matosinhos, Matosinhos, Portugal
[u]Institute of Biomedicine of Seville, IBiS/Virgen Macarena University Hospital/CSIC/University of Seville, Seville, Spain
[v]Hospital Universitario Virgen Macarena, Seville, Spain
[w]Université de Genève, Geneva, Switzerland

Summary

Background Real-world evidence provides valuable insights into cancer burden, presentation, and care variations. Through a large-scale federated approach, this study aims to explore patient characteristics and overall survival for eight cancers using data from 11 electronic health records and cancer registries from eight European countries, mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM).

Methods Patients aged 18 years or older with a primary cancer diagnosis between 2000 and 2019 were included. Patients were followed from cancer diagnosis until death, database exit, or study end. Mortality data was sourced from linked national or subnational death registries for most databases. Patient characteristics, including comorbidities, and medication use, were summarised. Age-standardised overall survival (OS) at one,

The Lancet Regional Health - Europe 2026;63: 101585

Published Online xxx https://doi.org/10.1016/j.lanepe.2025.101585

*Corresponding author. Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol); Gran Via Corts Catalanes, 587 àtic, 08007, Barcelona, Spain.
E-mail address: tduarte@idiapjgol.org (T. Duarte-Salles).

five, and ten years were calculated using the Kaplan–Meier method and stratified by cancer type, age group and sex.

Findings There were 1,796,278 eligible cancer patients included with most diagnoses in individuals aged 60–79 years. Top comorbidities and medications were relatively consistent across databases, with certain variations observed by cancer type, possibly indicative of early cancer signs and risk factors. For instance, anaemia was frequent in colorectal (9% [HUS]–23% [IMASIS]; 791/8395–730/3141 individuals) and stomach cancers (10% [HUS]–34% [IMASIS]; 130/1277–225/670), while chronic obstructive pulmonary disease (18% [SIDIAP]–34% [HUVM], 5310/29,009–1039/3063) and pneumonia (5% [CPRD GOLD]–33% [UTARTU], 1904/34,990–1001/3063) were common in lung cancer patients. Breast and prostate cancers had the highest one, five and ten-year overall survival, with 5-year OS ranging from 76% [ECi]–85% [IMASIS] and 75% [HUVM]–83% [SIDIAP], respectively. Pancreatic cancer showed the lowest survival ranging from 3% [NCR]–25% [IMASIS] 5-year OS. Variations in cancer survival estimates were observed across data sources and countries.

Interpretation Federated analysis of diverse European real-world databases, standardised to OMOP-CDM, offer a valuable benchmark for future cancer research, particularly in understanding prodromes and risk factors, often recorded in routinely collected healthcare data prior to cancer onset.

Funding The European Health Data & Evidence Network has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations partners.

Keywords: Cancer survival; Cancer burden; Descriptive epidemiology; Real-world evidence; Real-world data; Mortality

Research in context

Evidence before this study

We searched PubMed for English-language research articles published from january 2000 to august 2025, using a combination of the terms 'cancer', 'population-based', 'survival' and 'Europe' to identify population-based studies assessing survival for one or more of the following cancer types: breast, colorectal, head and neck, liver, lung, pancreas, prostate and stomach. Previous studies on cancer survival in Europe have predominantly relied on cancer registry data which, while valuable, often lack information on prior medical history, as linkages to this information are not possible or time-consuming.

Added value of this study

This study provides one-year, five-year and ten-year overall survival rates by age and sex for eight cancer types across eight European countries (Estonia, Finland, The Netherlands, Norway, Portugal, Spain, Switzerland and the United Kingdom) using data from 11 real-world databases. These include cancer registries, primary care and hospital databases, marking the first such comprehensive analysis conducted to our knowledge. Using a federated analysis approach based on the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) allows for efficient and standardised data analysis while ensuring coding consistency and patient privacy. Additionally, the research provides novel insights into comorbidities and medication use prior to diagnosis, information not commonly available in cancer registries.

Implications of all the available evidence

The use of diverse real-world data sources can help improve our understanding of the cancer burden in Europe, providing a complete, scalable and readily updatable view of the patient journey. From prodromes and risk factors captured in primary care records, through treatment and prognosis recorded in cancer registries, to management strategies in hospital care databases. Additionally, the study of these complementary databases, standardised into a common model like OMOP-CDM facilitate cross-country collaborative research.

Introduction

In 2022, cancer accounted for 9.7 million deaths globally, with 20 million new cases diagnosed.[1] With one in nine men and one in 12 women expected to die from cancer, and the number of cases projected to reach 35 million by 2050, the disease remains a critical global

health challenge.[1] The continuous surveillance and monitoring of cancer survival is needed for the development, implementation, and evaluation of health policies aiming to reduce the burden of disease.

Survival rates vary based on cancer type, detection stage, and treatment, and are influenced by individual health, comorbidities, and tumour-related factors. While advances in screening and treatments have improved survival rates for certain cancers, disparities can persist between different healthcare systems and countries.[2] Understanding individual factors related to cancer survival and identifying at-risk population subgroups is crucial for planning future interventions.

Recognising the complexities of researching diverse observational databases, the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) was created.[3] This standardised framework addresses structural and semantic variations in observational data, ensuring consistent and transparent analysis. Building on this foundation, the European Health Data and Evidence Network (EHDEN) aims to enhance the discovery and analysis of health data in Europe.[4] Leveraging the EHDEN network, our study aims to explore the prior medical history and the overall survival of cancer patients with eight different cancer types through a large-scale federated approach, uniquely analysing multiple primary care, hospital, and cancer registry databases from eight European countries.

Methods

Study design

We conducted an observational cohort study using routinely collected healthcare data from across Europe mapped to the OMOP CDM.[3] The OMOP CDM enabled the study to be executed in a distributed manner, allowing each site to run analytic code locally without transferring person-level data. All data partners received approval or waiver from their institutional review boards following their institutional governance guidelines.

Data sources

All databases mapped to OMOP and data partners in the EHDEN network were invited to participate. Eleven databases from eight European countries (Estonia, Finland, The Netherlands, Norway, Portugal, Spain, Switzerland, and the United Kingdom) informed the analysis. Of these, six were electronic health record (EHRs) databases (two primary care and four hospital care databases), with the remaining five being cancer registries and cancer-specific databases.

The primary care databases included the Clinical Practice Research Datalink (CPRD; UK) GOLD, and the Information System for Research in Primary Care (SIDIAP; Spain). The hospital-based databases were Helsinki University Hospital (HUS; Finland); the Virgen Macarena University Hospital (HUVM; Spain), the Hospital del Mar Institut Municipal d'Assistència Sanitària Information System (IMASIS; Spain) and the Unidade Local de Saúde de Matosinhos (ULSM; Portugal). The cancer registries were Cancer Registry of Norway (CRN; Norway), The Netherlands Cancer Registry (NCR; Netherlands) and Geneva Cancer Registry (GCR; Switzerland). The two cancer-specific databases included the University of Tartu (UTARTU; Estonia) database and the Edinburgh Cancer Informatics (ECi; Scotland). These last two included data from several sources including cancer registry and EHR data. All databases obtained date of death from national or subnational death registries, apart from CPRD GOLD. A detailed description of each database, including information on the population coverage, is available in the supplement (Table S1).

Study population and study period

The study population consisted of individuals aged 18 years or older diagnosed with a primary malignant cancer of either breast, colorectal, head and neck, liver, lung, pancreas, prostate and stomach cancers. Patients only contributed to one cancer cohort. We required a minimum of one year of prior history before cancer diagnosis, except for ECi, CRN and GCR, where this inclusion criterion was not applied, because the observation period started on the date of cancer diagnosis. Patients were excluded if they had (i) a prior history of any malignancy except for non-melanoma skin cancer, (ii) a date of death and cancer diagnosis on the same date or (iii) multiple cancer diagnoses from different sites occurred on the same date. Any database with less than 200 patients per cancer subtype was not analysed.

The study period was from the 1st of January 2000 to the 31st of December 2019. However, for some databases, the start date was the 1st of January of the year of useable data, for instance, 2003 for HUVM, 2007 for SIDIAP and 2012 from UTARTU (Table S2). Patients were followed from the date of their first recorded cancer diagnosis to whichever came first: exit from the database, date of death, or the end of the study period.

Cancer definitions

We used OMOP standard vocabularies, the Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT), and the International Classification of Diseases for Oncology, Third Edition (ICD-O-3) diagnostic codes to comprehensively identify cancer patients.[5] Codes signifying either non-malignant cancer or metastasis were excluded, as well as codes indicative of non-epithelial tumours. Cancer codelists were reviewed by clinicians with expertise in primary care and oncology and are provided in the supplement (Table S3). Additionally, computable phenotypes were further reviewed by data partners using CohortDiagnostics R package.[6] For survival analyses, mortality was

defined as all-cause mortality based on records of date of death.

Statistical methods

Population characteristics of cancer patients were summarised, with median and interquartile range (IQR) used for continuous variables, and counts and percentages used for categorical variables. For all databases except cancer registries, a range of predefined comorbidities (at any time prior to cancer diagnosis) and medication usage (one year prior to cancer diagnosis) were selected based on their availability in EHR databases and their clinical relevance, and summarised (Table S4).

For survival analysis, we used the Kaplan–Meier (KM) method to estimate the overall survival probabilities from observed survival times, including 95% confidence intervals. Individuals were followed from the date of their first recorded diagnosis to death or censoring event (end of the study period or exit from the database). We estimated survival at one, five and ten years. All results were stratified by database, sex and age groups. Results for ULSM, Portugal, were excluded from the survival analysis after review by clinical experts due to potential issues with the mortality data.

We age-standardised survival estimates using the International Cancer Survival Standard (ICSS) weights.[7] For both sexes combined, we used all age groups for age standardisation. For males and females, we used age groups 40 years and older for most cancers, except for breast and prostate cancers, where age groups 50 years and older were used to compute age-standardised results. We were unable to estimate age-standardised survival for some databases due to the small number of cases in some age and cancer strata; therefore, these age-standardised results are not presented. To avoid re-identification, for all analyses, we do not report results with less than ten cases. Survival estimates could not be adjusted for stage or biology of the cancer, due to limitations on data availability.

The statistical software R (version => 4.2.3) was used for all analyses. Analytical code for the study, with version control, is available at: https://github.com/oxford-pharmacoepi/CancerSurvivalWp2Analysis.

Ethics approval

The use of Clinical Practice Research Datalink (CPRD) data for this study was approved via the Research Data Governance (RDG) Process of the UK Medicines and Healthcare Products Regulatory Agency (protocol 22_001843).

The use of Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària data (SIDIAP) for this study was approved by the Clinical Research Ethics Committee of the IDIAP Jordi Gol (project code: 24/001-P).

The use of IMIM-Hospital del Mar Barcelona (IMASIS) data for this study was approved by the Parc de Salut Mar Research Ethics Committee CEIm-Parc de Salut Mar (2023/11262).

The use of Hospital District of Helsinki and Uusimaa (HUS) data for this study was approved under data permit HUS/325/2023.

The use of Netherlands Comprehensive Cancer Organisation (NCR) data for this study was approved under data permit K23.198.

The use of Hospital Universitario Virgen Macarena (HUVM) data for this study was approved under data permit 1651-N-23.

The Cancer Registry of Norway provided statistical data for the study. The CRN has permission to collect, process and report statistical data without the need to seek consent.

The use of University of Tartu (UTARTU) data for this study was approved by Estonian Committee on Bioethics and Human Research (1.1–12/159).

The use of Geneva Cancer Registry (GCR) data for this study was conducted in accordance with Article 32 of the Swiss Law on the Registration of Oncological Diseases (LEMO, RS 818.33), which permits cancer registries to provide aggregated data where data shared cannot trace back to individual patients. This data excludes patients who have expressed their opposition to cancer registration.

The use of South East Scotland Cancer Database (ECi) data for this study was approved by the Integrated Research Application System NHS Research Ethics Service (Dataloch service for research, IRAS ID 317626, reference number 22/NS/0093).

The use of Unidade Local de Saúde de Matosinhos (ULSM) data for this study was obtained from the Comissão de Ética para a Saúde da ULS Matosinhos (code 103/CES/JAS).

Role of funding source

The funding sources played no part in the design, analysis, interpretation of the findings, writing of the manuscript, or the decision to submit for publication. The corresponding author had the final responsibility for the decision to submit for publication.

Results

Study population

Our results are presented in a publicly available Shiny app enabling an interactive exploration of the study results: https://dpa-pde-oxford.shinyapps.io/EHDENCancerSurvivalStudyShiny/. We identified 3,165,081 patients with a diagnosis of primary malignant breast, pancreatic, prostate, colorectal, lung, stomach, liver, or head and neck cancers from participating databases. After applying exclusion criteria, 1,796,278 patients were eligible for inclusion. Study attrition was largely due to patients not being observed in the database during the study period (Table S5).

Database name	CPRD GOLD	SIDIAP	HUS	HUVM	IMASIS	ULSM	CRN	GCR	NCR	ECi[a]	UTARTU[b]
Country	UK	Spain	Finland	Spain	Spain	Portugal	Norway	Switzerland	Netherlands	Scotland	Estonia
Database type	Primary	Primary	Hospital	Hospital	Hospital	Hospital	Registry	Registry	Registry	Cancer Specific[c]	Cancer Specific[c]
Number of patients	239,876	213,590	58,233	12,275	13,315	5817	259,301	23,774	943,196	9386	17,515
Median follow-up	1186	2082	2283	2135	1704	2006	1303	1427	1799	2409	1794
Study period	2000–2019	2007–2019	2000–2019	2003–2019	2000–2019	2000–2019	2000–2019	2000–2019	2000–2019	2000–2019	2000–2019
Any comorbidity	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes
Any medication	Yes	Yes	Yes	No	Yes	Yes	No	No	No	No	Yes
Sex											
Female	122,835 (51%)	93,982 (44%)	31,076 (53%)	5971 (49%)	6058 (45%)	2729 (47%)	112,519 (43%)	11,678 (49%)	460,949 (49%)	9386 (100%)	7215 (41%)
Male	117,041 (49%)	119,608 (56%)	27,157 (47%)	6304 (51%)	7257 (55%)	3088 (53%)	146,782 (57%)	12,096 (51%)	482,247 (51%)	–	10,300 (59%)
Age [median (IQR)]	69 (60–77)	67 (58–77)	67 (59–75)	66 (56–75)	69 (59–78)	67 (57–76)	68 (60–77)	67 (57–75)	68 (59–76)	60 (50–69)	68 (60–76)
Age group											
18–39	3980 (2%)	5537 (3%)	1210 (2%)	350 (3%)	298 (2%)	206 (4%)	4410 (2%)	569 (2%)	17,867 (2%)	511 (5%)	285 (2%)
40–49	14,963 (6%)	17,504 (8%)	3717 (6%)	1280 (10%)	910 (7%)	560 (10%)	15,124 (6%)	2083 (9%)	64,890 (7%)	1619 (17%)	952 (5%)
50–59	38,777 (16%)	38,272 (18%)	9912 (17%)	2444 (20%)	2260 (17%)	1071 (18%)	41,907 (16%)	4560 (19%)	161,874 (17%)	2519 (27%)	2807 (16%)
60–69	66,264 (28%)	57,522 (27%)	18,589 (32%)	3357 (27%)	3390 (25%)	1548 (27%)	77,338 (30%)	6845 (29%)	275,071 (29%)	2415 (26%)	5717 (33%)
70–79	69,257 (29%)	55,580 (26%)	16,144 (28%)	3210 (26%)	3775 (28%)	1480 (25%)	72,635 (28%)	5965 (25%)	273,206 (29%)	1559 (17%)	5287 (30%)
80+	46,635 (19%)	39,175 (18%)	8661 (15%)	1634 (13%)	2682 (20%)	952 (16%)	47,887 (18%)	3752 (16%)	150,288 (16%)	763 (8%)	2467 (14%)
Cancer types											
Breast	77,384 (32%)	54,248 (25%)	21,248 (36%)	4137 (34%)	3423 (26%)	1588 (27%)	55,816 (22%)	7435 (31%)	257,562 (27%)	9386 (100%)	6377 (36%)
Colorectal	44,155 (18%)	50,862 (24%)	8395 (14%)	3141 (26%)	3141 (24%)	1352 (23%)	61,356 (24%)	3493 (15%)	215,109 (23%)	–	–
Head and neck	10,487 (4%)	12,639 (6%)	3756 (6%)	747 (6%)	914 (7%)	537 (9%)	10,759 (4%)	1491 (6%)	47,306 (5%)	–	–
Liver	2913 (1%)	7622 (4%)	1203 (2%)	346 (3%)	811 (6%)	0 (0%)	1894 (1%)	887 (4%)	7524 (1%)	–	–
Lung	34,990 (15%)	29,009 (14%)	3996 (7%)	510 (4%)	1247 (9%)	329 (6%)	34,358 (13%)	3066 (13%)	160,940 (17%)	–	3063 (17%)
Pancreatic	7558 (3%)	7989 (4%)	2749 (5%)	309 (3%)	589 (4%)	206 (4%)	10,257 (4%)	1064 (4%)	35,604 (4%)	–	–
Prostate	56,634 (24%)	42,771 (20%)	15,609 (27%)	2779 (23%)	2520 (19%)	1034 (18%)	77,156 (30%)	5664 (24%)	187,851 (20%)	–	8075 (46%)
Stomach	5755 (2%)	8450 (4%)	1277 (2%)	306 (2%)	670 (5%)	771 (13%)	7705 (3%)	674 (3%)	31,300 (3%)	–	–

Values are no (%) except as indicated. [a]Only breast cancer patients were included for this database and due to ethical governance for this database only results for females are reported. [b]Only breast, lung and prostate cancer patients were included in this database. [c]Cancer-specific databases included data from several sources, including cancer registry and EHR data.

Table 1: **Baseline patient characteristics at the time of cancer diagnosis across databases.**

Baseline characteristics

Table 1 shows a summary of the baseline patient characteristics of eligible patients with a diagnosis of any of the eight cancers for each database. According to Table 1, NCR from the Netherlands, contributed the largest number of patients (n = 943,196) whereas ULSM from Portugal, contributed the smallest (n = 5817). Across all databases, the most prevalent cancers were breast, prostate, colorectal, and lung. Median age ranged from 60 years in ECi in Scotland (breast cancer only), to 69 years in CPRD GOLD, UK and IMASIS, Spain, with other databases reporting a median age at diagnosis between 66 and 68 years.

Stratification by age and sex showed that breast cancer was most common among females across most databases with decreasing prevalence with increasing age. Alternatively, among males, colorectal and head and neck cancers were more prevalent in younger age groups with prostate cancer higher in older males (Tables S6 and S7).

Aside from sex-specific cancers, more males were diagnosed with lung (54% [n = 18,843, CPRD GOLD]–83% [n = 424, HUVM]), liver (66% [n = 534 IMASIS]–81% [n = 718, GCR]) and head and neck cancers (58% [n = 2160, HUS]–88% [n = 473, ULSM]) (Table S8). In contrast, pancreatic (47% [n = 1293, HUS]–55% [n = 169, HUVM] males), colorectal (50% [n = 4198, HUS]–59% [n = 1863, HUVM] males) and stomach cancers (54% [n = 693, HUS]–64% [n = 20,105, NCR] males) exhibited a more balanced distribution between sexes. Some regional differences were observed, with Spanish, Portuguese and Estonian databases (HUVM, IMASIS, SIDIAP, ULSM and UTARTU) reporting a higher proportion of males with lung and head and neck cancers.

Common comorbidities any time prior to cancer diagnosis were relatively consistent across databases, with variations observed by cancer type (Tables S9 and S10). Age-related comorbidities, such as hypertension (11% [HUS]–63% [UTARTU]; 6277/58,233–11,026/17,515 individuals), osteoarthritis (8% [HUS]–28% [UTARTU]; 4786/58,233–4863/17,515) and hyperlipidaemia (3% [HUS]–21% [UTARTU]; 1513/58,233–3614/17,515) were among the top conditions across most cancer types and databases, with ischaemic heart disease particularly common among prostate cancer patients (7% [HUS]–20% [UTARTU]; 1050/15,609–1652/8075).

Some cancer types showed increased prevalence of certain comorbidities potentially indicative of risk factors and even prodromic conditions (Table 2 and Table S10). For liver cancer patients, chronic liver disease was the most common (24% [CPRD GOLD/HUS]–77% [IMASIS]; 709/2913 and 292/1203–622/811) and, for some databases, viral hepatitis (19% [SIDIAP]–58% [IMASIS]; 1455/7622–467/811) was also prevalent. Anaemia was more prevalent among colorectal (9% [HUS]–23% [IMASIS]; 791/8395–730/3141), and stomach cancer patients (10% [HUS]–34% [IMASIS]; 130/1277–225/670) compared to other cancers. In lung cancer patients, chronic obstructive pulmonary disease (COPD) (18% [SIDIAP]–34% [HUVM/UTARTU]; 5310/29,009–171/510 and 1039/3063) was prevalent, with some databases also showing high proportions of pneumonia (15% [IMASIS]–33% [UTARTU]; 183/1247–1001/3063). COPD was also common in head and neck cancers (4% [HUS]–15% [IMASIS]; 138/3756–134/914). Type 2 Diabetes mellitus (T2D) diagnoses were higher in pancreatic (9% [HUS]–35% [HUVM]; 260/2749–107/309) and liver cancer patients (17% [HUS]–31% [IMASIS]; 200/1203–252/811). The prevalence of obesity varied depending on whether it was defined only by diagnostic codes or combined with BMI measurements, with the latter being the most common comorbidity in primary care databases across all cancer types.

Medications prescribed the year preceding diagnosis also showed relative consistency across databases (Table S11). Common medications included those for acid-related disorders (15% [HUS/IMASIS]–53% [SIDIAP]; 8945/58,233 and 1994/13,315–113,271/213,590), systemic antibacterials (12% [IMASIS]–55% [UTARTU]; 1597/13,315–9599/17,515), anti-inflammatory/antirheumatic treatments (10% [IMASIS]–51% [SIDIAP]; 1383/13,315–108,067/213,590), and psycholeptics (2% [UTARTU]–37% [SIDIAP]; 431/17,515–78,759/213,590). Diuretics were commonly prescribed among liver cancer patients (15% [IMASIS]–44% [SIDIAP]; 120/811–3324/7622), and medications for obstructive airway diseases among lung cancer patients (12% [IMASIS]–49% [SIDIAP]; 147/1247–14,294/29,009) (Table 3 and Table S12). Prostate cancer patients were more likely prescribed renin-angiotensin system agents (6% [IMASIS]–47% [UTARTU]; 142/2520–3809/8075) and lipid-modifying agents (5% [IMASIS]–38% [CPRD GOLD/SIDIAP]; 128/2520–21,776/56,634 and 16,403/42,771). Overall, the prevalence of comorbidities and medications was higher and remained more consistent between primary care databases than between hospitals.

Overall survival across databases and cancer types

Age-standardised survival trends showed differences between data sources, with the largest variations observed for pancreatic and stomach cancers (Fig. 1). Spanish databases consistently showed higher survival curves for most cancers, with confidence intervals overlapping with those from Switzerland, Finland and Estonia. Conversely, the nationwide databases from the UK, Netherlands and Norway tended to display lower rates, often overlapping with databases from Switzerland and Scotland (Figure S1).

Survival curves according to database type showed several differences (Figure S2). Primary care databases consistently displayed non-overlapping intervals, with SIDIAP (Spain) performing better than CPRD GOLD (UK). Conversely, cancer registries and cancer-specific

		CPRD GOLD (UK)	SIDIAP (Spain)	HUS (Finland)	HUVM (Spain)	IMASIS (Spain)	ULSM (Portugal)	UTARTU (Estonia)
Breast								
Cardiovascular diseases	Hypertension	15,733 (20%)	8744 (16%)	1523 (7%)	874 (21%)	565 (17%)	391 (25%)	3571 (56%)
Endocrine and metabolic disorders	Hyperlipidaemia	5225 (7%)	5600 (10%)	317 (1%)	500 (12%)	240 (7%)	230 (14%)	1362 (21%)
Musculoskeletal disorders	Osteoarthritis	13,010 (17%)	9257 (17%)	1764 (8%)	445 (11%)	268 (8%)	136 (9%)	1946 (31%)
Mental health disorders	Anxiety	11,663 (15%)	9256 (17%)	246 (1%)	115 (3%)	93 (3%)	85 (5%)	666 (10%)
	Depressive disorder	11,372 (15%)	5070 (9%)	480 (2%)	216 (5%)	205 (6%)	129 (8%)	900 (14%)
Colorectal								
Cardiovascular diseases	Hypertension	11,877 (27%)	11,963 (24%)	1086 (13%)	1172 (37%)	1226 (39%)	473 (35%)	–
Endocrine and metabolic disorders	Hyperlipidaemia	3936 (9%)	5943 (12%)	252 (3%)	607 (19%)	617 (20%)	287 (21%)	–
	Type 2 Diabetes	4564 (10%)	6480 (13%)	464 (6%)	522 (17%)	492 (16%)	180 (13%)	–
Haematologic disorders	Anaemia	7826 (18%)	10,868 (21%)	791 (9%)	315 (10%)	730 (23%)	251 (19%)	–
Musculoskeletal disorders	Osteoarthritis	8316 (19%)	9073 (18%)	691 (8%)	402 (13%)	384 (12%)	115 (9%)	–
Head and Neck								
Cardiovascular diseases	Hypertension	2383 (23%)	2496 (20%)	356 (9%)	219 (29%)	239 (26%)	167 (31%)	–
Endocrine and metabolic disorders	Hyperlipidaemia	813 (8%)	1334 (11%)	74 (2%)	119 (16%)	116 (13%)	111 (21%)	–
	Type 2 Diabetes	764 (7%)	1323 (10%)	163 (4%)	98 (13%)	96 (11%)	57 (11%)	–
Musculoskeletal Disorders	Osteoarthritis	1514 (14%)	1590 (13%)	248 (7%)	68 (9%)	51 (6%)	37 (7%)	–
Respiratory diseases	COPD	945 (9%)	1084 (9%)	138 (4%)	88 (12%)	134 (15%)	74 (14%)	–
Liver								–
Cardiovascular diseases	Hypertension	840 (29%)	1760 (23%)	229 (19%)	137 (40%)	360 (44%)	–	–
Digestive and hepatobilary diseases	Chronic liver disease	709 (24%)	2473 (32%)	292 (24%)	168 (49%)	622 (77%)	–	–
	Viral hepatitis	130 (4%)	1455 (19%)	86 (7%)	86 (25%)	467 (58%)	–	–
Endocrine and metabolic disorders	Type 2 Diabetes	768 (26%)	1468 (19%)	200 (17%)	99 (29%)	252 (31%)	–	–
Musculoskeletal disorders	Osteoarthritis	635 (22%)	1251 (16%)	105 (9%)	39 (11%)	115 (14%)	–	–
Lung								
Cardiovascular diseases	Hypertension	9316 (27%)	6784 (23%)	751 (19%)	271 (53%)	412 (33%)	134 (41%)	1950 (64%)
Endocrine and Metabolic Disorders	Hyperlipidaemia	3406 (10%)	3417 (12%)	202 (5%)	152 (30%)	259 (21%)	99 (30%)	619 (20%)
Infectious diseases	Pneumonia	1904 (5%)	3198 (11%)	640 (16%)	35 (7%)	183 (15%)	80 (24%)	1001 (33%)
Musculoskeletal Disorders	Osteoarthritis	7175 (21%)	4426 (15%)	357 (9%)	75 (15%)	105 (8%)	28 (9%)	814 (27%)
Respiratory diseases	COPD	8624 (25%)	5310 (18%)	931 (23%)	171 (34%)	370 (30%)	77 (23%)	1039 (34%)
Pancreatic								
Cardiovascular diseases	Hypertension	2115 (28%)	1955 (24%)	446 (16%)	140 (45%)	310 (53%)	84 (41%)	–
Endocrine and metabolic disorders	Hyperlipidaemia	719 (10%)	957 (12%)	105 (4%)	88 (28%)	165 (28%)	60 (29%)	–
	Type 2 Diabetes	1538 (20%)	1900 (24%)	260 (9%)	107 (35%)	187 (32%)	48 (23%)	–
Haematologic disorders	Anaemia	553 (7%)	881 (11%)	93 (3%)	33 (11%)	128 (22%)	35 (17%)	–
Musculoskeletal disorders	Osteoarthritis	1657 (22%)	1686 (21%)	296 (11%)	44 (14%)	117 (20%)	25 (12%)	–
Prostate								
Cardiovascular diseases	Hypertension	16,422 (29%)	11,161 (26%)	1722 (11%)	962 (35%)	844 (33%)	385 (37%)	5505 (68%)
	Ischaemic heart disease	6596 (12%)	2404 (6%)	1050 (7%)	219 (8%)	226 (9%)	104 (10%)	1652 (20%)
Endocrine and metabolic disorders	Hyperlipidaemia	5768 (10%)	5169 (12%)	476 (3%)	502 (18%)	465 (18%)	232 (22%)	1633 (20%)
	Type 2 Diabetes	5148 (9%)	4925 (12%)	707 (5%)	405 (15%)	279 (11%)	145 (14%)	1121 (14%)
Musculoskeletal Disorders	Osteoarthritis	11,132 (20%)	6606 (15%)	1224 (8%)	264 (9%)	215 (9%)	87 (8%)	2103 (26%)
Stomach								
Cardiovascular diseases	Hypertension	1567 (27%)	1810 (21%)	164 (13%)	119 (39%)	286 (43%)	250 (32%)	–
Endocrine and metabolic disorders	Hyperlipidaemia	472 (8%)	883 (10%)	43 (3%)	61 (20%)	141 (21%)	156 (20%)	–
	Type 2 Diabetes	680 (12%)	1053 (12%)	73 (6%)	59 (19%)	110 (16%)	96 (12%)	–
Haematologic disorders	Anaemia	1338 (23%)	2305 (27%)	130 (10%)	54 (18%)	225 (34%)	171 (22%)	–
Musculoskeletal disorders	Osteoarthritis	1268 (22%)	1578 (19%)	101 (8%)	47 (15%)	97 (14%)	68 (9%)	–

Values are no (%). Denominators used for percentage calculations are provided in Table 1 (total cases per each cancer site and specific database).CPRD GOLD, Clinical Practice Research Datalink; SIDIAP, The Information System for Research on Primary Care; UTARTU, University of Tartu; HUS, Hospital District of Helsinki and Uusimaa; HUVM, Hospital Universitario Virgen Macarena; IMASIS, Institut Municipal Assistència Sanitària Information System; ULSM, Unidade Local de Saúde de Matosinhos; COPD, Chronic Obstructive Pulmonary Disease.

***Table 2*: Top 5 baseline comorbidities any time prior to cancer diagnosis per cancer type and database.**

databases showed similar lower survival curves across most cancer types. GCR displayed the highest survival curves and widest confidence intervals, non-overlapping for breast, colorectal and prostate cancers. Lastly, hospital databases showed similar survival curves, with broad overlapping confidence intervals.

		CPRD GOLD (UK)	SIDIAP (Spain)	HUS (Finland)	IMASIS (Spain)	ULSM (Portugal)	UTARTU (Estonia)
Breast							
Anti-infective agents	Antibacterials for systemic use	30,729 (40%)	18,729 (35%)	1427 (7%)	163 (5%)	624 (39%)	1980 (31%)
Cardiovascular agents	Agents acting on the Renin-Angiotensin System	15,687 (20%)	14,331 (26%)	2347 (11%)	79 (2%)	267 (17%)	2405 (38%)
Gastrointestinal agents	Drugs Acid-Related Disorders	20,199 (26%)	23,296 (43%)	2209 (10%)	192 (6%)	543 (34%)	1168 (18%)
Immune system modulators	Anti-inflammatory and Antirheumatic agents	18,902 (24%)	27,631 (51%)	6078 (29%)	238 (7%)	746 (47%)	1824 (29%)
Nervous system agents	Psycholeptics	14,296 (18%)	21,670 (40%)	4120 (19%)	277 (8%)	532 (34%)	187 (3%)
Colorectal							
Anti-infective agents	Antibacterials for systemic use	17,558 (40%)	18,726 (37%)	1372 (16%)	380 (12%)	498 (37%)	–
Cardiovascular agents	Antithrombotic agents	5385 (12%)	11,482 (23%)	2008 (24%)	341 (11%)	371 (27%)	–
Gastrointestinal agents	Drugs Acid-Related Disorders	16,847 (38%)	27,538 (54%)	1716 (20%)	579 (18%)	540 (40%)	–
Immune system modulators	Anti-inflammatory and Antirheumatic agents	9627 (22%)	23,211 (46%)	2739 (33%)	295 (9%)	487 (36%)	–
Nervous system agents	Psycholeptics	7577 (17%)	18,724 (37%)	1657 (20%)	392 (12%)	433 (32%)	–
Head & Neck							
Anti-infective agents	Antibacterials for systemic use	5789 (55%)	6509 (51%)	551 (15%)	84 (9%)	198 (37%)	–
Gastrointestinal agents	Drugs Acid-Related Disorders	3671 (35%)	7055 (56%)	606 (16%)	140 (15%)	156 (29%)	–
Immune system modulators	Anti-inflammatory and Antirheumatic agents	3031 (29%)	7812 (62%)	1234 (33%)	206 (23%)	244 (45%)	–
Nervous system agents	Opioids	3893 (37%)	2794 (22%)	747 (20%)	52 (6%)	110 (20%)	–
	Psycholeptics	1893 (18%)	4279 (34%)	686 (18%)	197 (22%)	118 (22%)	–
Liver							
Anti-infective agents	Antibacterials for systemic use	1453 (50%)	3295 (43%)	326 (27%)	115 (14%)	–	–
Cardiovascular agents	Agents acting on the Renin-Angiotensin System	1163 (40%)	3229 (42%)	300 (25%)	59 (7%)	–	–
Gastrointestinal agents	Drugs Acid-Related Disorders	1582 (54%)	4828 (63%)	380 (32%)	192 (24%)	–	–
Immune system modulators	Anti-inflammatory and Antirheumatic agents	756 (26%)	3536 (46%)	477 (40%)	66 (8%)	–	–
Nervous system agents	Opioids	1252 (43%)	2003 (26%)	400 (33%)	66 (8%)	–	–
Lung							
Anti-infective agents	Antibacterials for systemic use	24,022 (69%)	15,416 (53%)	1102 (28%)	192 (15%)	146 (44%)	1458 (48%)
Gastrointestinal agents	Drugs Acid-Related Disorders	15,961 (46%)	18,267 (63%)	1080 (27%)	239 (19%)	152 (46%)	785 (26%)
Immune system modulators	Anti-inflammatory and Antirheumatic agents	15,243 (44%)	17,829 (61%)	1984 (50%)	201 (16%)	139 (42%)	945 (31%)
Nervous system agents	Psycholeptics	9078 (26%)	11,991 (41%)	1554 (39%)	170 (14%)	139 (42%)	86 (3%)
Respiratory system	Drugs for Obstructive Airway Diseases	15,899 (45%)	14,294 (49%)	1430 (36%)	147 (12%)	120 (36%)	976 (32%)
Pancreatic							
Anti-infective agents	Antibacterials for systemic use	3523 (47%)	3283 (41%)	684 (25%)	84 (14%)	67 (33%)	–
Gastrointestinal agents	Drugs Acid-Related Disorders	4674 (62%)	5635 (71%)	850 (31%)	148 (25%)	100 (49%)	–
Immune system modulators	Anti-inflammatory and Antirheumatic agents	2121 (28%)	4345 (54%)	1180 (43%)	77 (13%)	71 (34%)	–
Nervous system agents	Opioids	3665 (48%)	2742 (34%)	1062 (39%)	52 (9%)	64 (31%)	–
	Psycholeptics	2056 (27%)	3873 (48%)	873 (32%)	103 (17%)	83 (40%)	–
Prostate							
Anti-infective agents	Antibacterials for systemic use	25,223 (45%)	22,724 (53%)	5640 (36%)	508 (20%)	589 (57%)	6161 (76%)
Cardiovascular agents	Agents acting on the Renin-Angiotensin System	19,133 (34%)	19,348 (45%)	2589 (17%)	142 (6%)	252 (24%)	3809 (47%)
	Lipid Modifying agents	21,776 (38%)	16,403 (38%)	2147 (14%)	128 (5%)	336 (32%)	1703 (21%)
Gastrointestinal agents	Drugs Acid-Related Disorders	18,050 (32%)	20,001 (47%)	1642 (11%)	369 (15%)	368 (36%)	1337 (17%)
Immune system modulators	Anti-inflammatory and Antirheumatic agents	15,412 (27%)	19,548 (46%)	3497 (22%)	254 (10%)	409 (40%)	2299 (28%)
Stomach							
Anti-infective agents	Antibacterials for systemic use	2637 (46%)	3225 (38%)	151 (12%)	71 (11%)	242 (31%)	–
Cardiovascular agents	Antithrombotic agents	928 (16%)	1666 (20%)	250 (20%)	57 (9%)	193 (25%)	–
Gastrointestinal agents	Drugs Acid-Related Disorders	4291 (75%)	6651 (79%)	462 (36%)	135 (20%)	362 (47%)	–
Immune system modulators	Anti-inflammatory and Antirheumatic agents	1352 (23%)	4155 (49%)	380 (30%)	46 (7%)	243 (32%)	–
Nervous system agents	Psycholeptics	1211 (21%)	3452 (41%)	216 (17%)	66 (10%)	215 (28%)	–

Values are no. (%). Denominators used for percentage calculations are provided in Table 1 (total cases per each cancer site and specific database). CPRD GOLD, Clinical Practice Research Datalink; SIDIAP, The Information System for Research on Primary Care; UTARTU, University of Tartu; HUS, Hospital District of Helsinki and Uusimaa; HUVM, Hospital Universitario Virgen Macarena; IMASIS, Institut Municipal Assistència Sanitària Information System; ULSM, Unidade Local de Saúde de Matosinhos. [a]Due to incomplete prescription data, HUVM database was excluded from this part of patient characterisation analysis.

***Table 3*: Top 5 baseline medication one year prior to cancer diagnosis per cancer type and database.[a]**

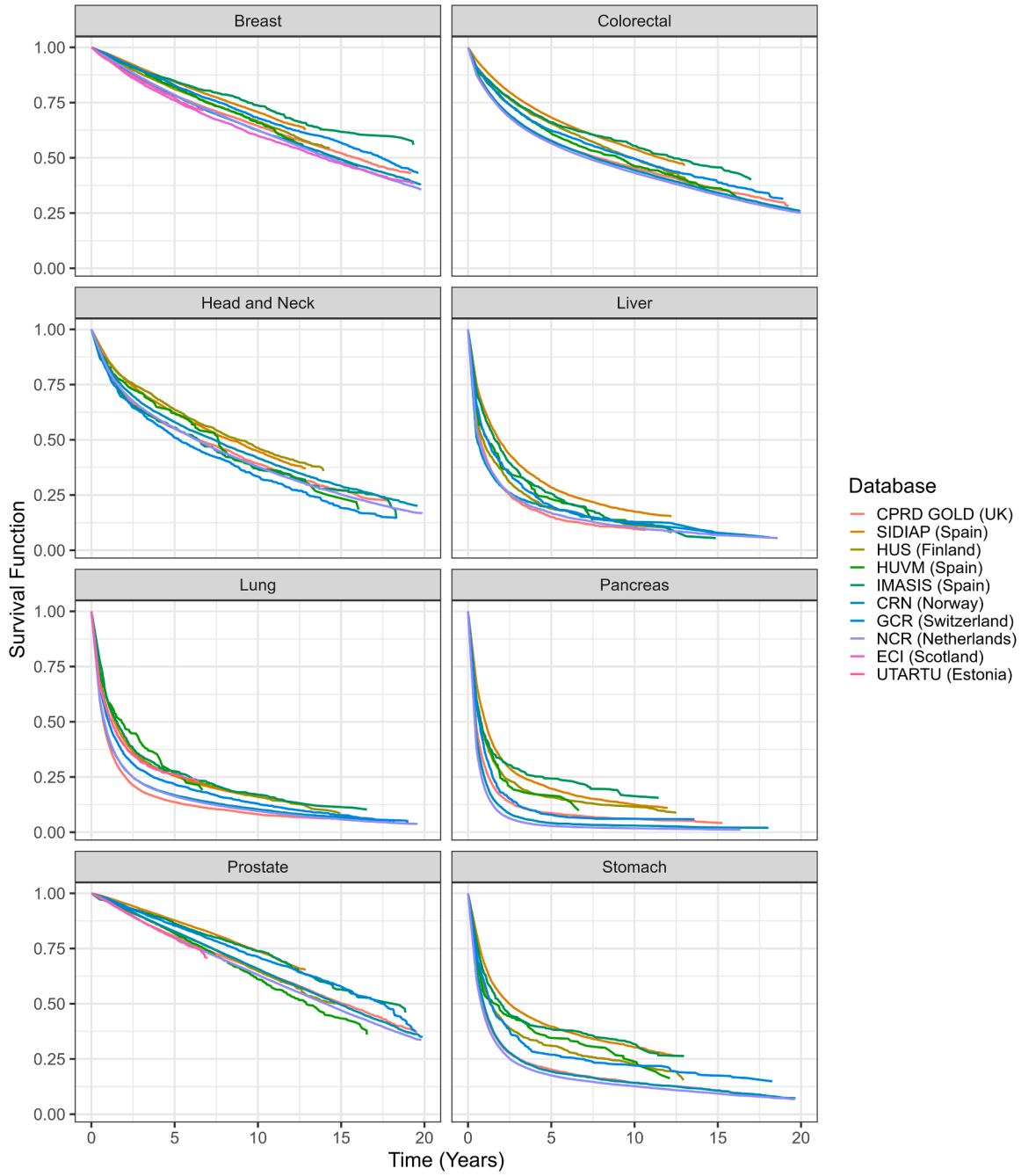

***Fig. 1:* Age-standardised Kaplan–Meier survival curves by database and cancer type.** Patients at risk can be found in the Table S13.

Sex-stratified results showed higher variability for females (Figure S3), and, for most cancers, survival curves were lower for males, except for liver and pancreatic cancers. Crude survival curves and numbers at risk are provided in the supplement (Figure S4 and Table S11), with crude KM curves showing similar but larger sex differences (Figure S5).

Breast cancer had the highest survival for one (94–97%), five (76–85%), and ten years (60–74%), apart from CPRD GOLD, SIDIAP, UTARTU, and CRN, where prostate cancer had higher one-year (95–97%) and five-year survival (77–83%) (Table 4). Pancreatic cancer showed the worst prognosis, with one-year survival ranging from 19.8% (95% CI 18.7–21.1) in

	Cases (N)	Events (N)	One-year survival (95% CI)	Five-year survival (95% CI)	Ten-year survival (95% CI)	Median survival (95% CI)	Median follow-up
Breast							
CPRD GOLD (UK)	77,384	16,870	94.7 (94.4–95.1)	78.2 (77.5–79)	64.3 (63.3–65.4)	4 (3.9–4.2)	1932
SIDIAP (Spain)	54,248	8539	96.9 (96.6–97.2)	84.3 (83.5–85)	70.6 (69.4–71.8)	0.8 (0.8–0.8)	2893
HUS (Finland)	21,248	4278	95.6 (94.9–96.2)	81 (79.7–82.3)	66.5 (64.7–68.3)	0.7 (0.7–0.8)	2986
HUVM (Spain)	4137	728	95.8 (94.3–97.2)	82 (79–85.3)	65.9 (61.5–70.9)	3.9 (3.6–4.5)	2879
IMASIS (Spain)	3423	553	95.3 (93.7–96.9)	84.7 (81.7–87.8)	73.6 (69.2–78.5)	1.3 (1–1.6)	2788
CRN (Norway)	55,816	15,121	94.7 (94.3–95.1)	78.2 (77.3–79)	62.3 (61.2–63.4)	3.5 (3.4–3.7)	2248
GCR (Switzerland)	7435	1741	96.6 (95.7–97.5)	82.7 (80.7–84.8)	67.9 (65.3–70.8)	4.6 (4.2–5.1)	2505
NCR (Netherlands)	257,562	75,574	95.1 (94.9–95.3)	78.4 (78–78.8)	62.4 (61.9–62.9)	9.1 (8.9–9.2)	3082
ECi (Scotland)	9386	2776	94.3 (93.2–95.4)	75.8 (73.7–78)	59.9 (57.3–62.6)	3.5 (3.2–3.8)	2409
UTARTU (Estonia)	6377	1046	94.1 (92.9–95.3)	76.9 (74.1–79.9)	–	0.7 (0.6–0.8)	2013
Colorectal							
CPRD GOLD (UK)	44,155	20,337	82.6 (81.7–83.5)	57.4 (56.1–58.7)	45.3 (43.8–46.8)	5.5 (5.2–5.8)	1103
SIDIAP (Spain)	50,862	18,457	90 (89.4–90.6)	68.2 (67.1–69.2)	53.8 (52.5–55.3)	3.1 (3–3.2)	2110
HUS (Finland)	8395	3351	86.8 (85.1–88.5)	65.4 (62.8–68.2)	49.5 (46.3–53)	2.8 (2.6–3)	1937
HUVM (Spain)	3141	1246	85.6 (82.8–88.5)	61.1 (56.8–65.8)	46.3 (41.2–52.2)	2.2 (2–2.8)	1846
IMASIS (Spain)	3141	1159	86.5 (83.7–89.4)	66 (61.5–70.9)	55.5 (50–61.9)	3.4 (3–3.8)	1683
CRN (Norway)	61,356	33,621	82.2 (81.5–82.9)	57.6 (56.6–58.7)	44.3 (43.2–45.4)	5.4 (5.2–5.7)	1175
GCR (Switzerland)	3493	1763	85.7 (83.1–88.4)	62.3 (58.4–66.6)	49.6 (45.3–54.5)	6.4 (5.6–7.6)	1391
NCR (Netherlands)	215,109	114,802	81.3 (80.9–81.7)	56.4 (55.9–57)	43.1 (42.5–43.7)	7.9 (7.6–8.1)	1749
Head & Neck							
CPRD GOLD (UK)	10,487	4531	80.4 (78.6–82.2)	54.9 (52.5–57.4)	39.1 (36.3–42.2)	5.9 (5.3–6.9)	1209
SIDIAP (Spain)	12,639	4903	86.1 (84.7–87.4)	62.1 (60.1–64.2)	44.8 (42.4–47.5)	5.2 (4.8–5.6)	2138
HUS (Finland)	3756	1526	85.3 (82.8–87.8)	63.6 (60–67.5)	46.3 (42.3–51)	5.5 (4.9–6.3)	2016
HUVM (Spain)	747	263	83.3 (77.2–89.7)	62.5 (53.4–73.6)	37.3 (27.3–54.8)	2.4 (1.6–3.4)	1867
IMASIS (Spain)	914	407	80.6 (74.9–86.8)	55.7 (47.9–64.7)	37.4 (28.9–49.1)	5.3 (3.7–7.9)	1340
CRN (Norway)	10,759	5336	83.1 (81.6–84.7)	58 (55.8–60.3)	41.7 (39.4–44.3)	6.8 (6.3–7.4)	1317
GCR (Switzerland)	1491	845	80.2 (75.9–84.9)	51.1 (45.6–57.4)	33.8 (28.1–41.2)	4.7 (3.8–5.9)	1271
NCR (Netherlands)	47,306	25,863	81.6 (80.8–82.4)	55.2 (54.2–56.3)	37.8 (36.6–39)	7 (6.7–7.4)	1891
Liver							
CPRD GOLD (UK)	2913	2225	43.9 (39.5–49)	15.3 (11.7–20.1)	9.8 (6.6–14.6)	0.8 (0.6–1.1)	225
SIDIAP (Spain)	7622	5333	63.2 (60.7–65.8)	28.4 (25.9–31.2)	17.5 (15–20.6)	1.9 (1.6–2.2)	610
HUS (Finland)	1203	924	48 (41.6–55.4)	19.1 (13.9–26.6)	11.2 (6.8–19.6)	0.8 (0.6–1.5)	298
HUVM (Spain)	346	234	52.8 (42.2–66.9)	23.6 (14.7–41)	12.9 (8.5–23.7)	1.1 (0.6–2.1)	389
IMASIS (Spain)	811	530	61 (52.8–70.3)	26.1 (18.7–38)	11.8 (5.6–26.8)	1.5 (1.1–2.4)	391
CRN (Norway)	1894	1499	40 (35–45.7)	19.7 (15.6–25)	12.9 (9–19.1)	0.6 (0.4–0.8)	165
GCR (Switzerland)	887	697	52.7 (45.5–61.4)	20.3 (14.5–28.9)	12.3 (7.4–21)	1.2 (0.9–1.7)	357
NCR (Netherlands)	7524	5972	42 (39.3–44.8)	16.9 (14.8–19.4)	10.4 (8.4–13)	0.7 (0.6–1)	238
Lung							
CPRD GOLD (UK)	34,990	28,066	40.9 (39.4–42.5)	13.6 (12.3–15)	8.1 (6.9–9.7)	0.8 (0.7–0.8)	216
SIDIAP (Spain)	29,009	20,592	58.5 (57.1–59.9)	25.4 (24.1–26.9)	16 (14.6–17.6)	1.4 (1.3–1.4)	478
HUS (Finland)	3996	2953	54 (50.1–58)	26 (22.1–30.7)	16.2 (12.6–21.2)	1.2 (1–1.5)	416
HUVM (Spain)	510	250	60.2 (49.8–72.3)	29.8 (17.2–56.8)	19.1 (6.7–56.9)	0.6 (0.4–2.1)	716
IMASIS (Spain)	1247	804	58.9 (52.5–66.6)	27.7 (21.7–36.1)	17.6 (11.6–26.9)	1.6 (1.3–2)	420
CRN (Norway)	34,358	28,358	43.3 (41.8–44.8)	16.7 (15.6–18)	10.4 (9.3–11.6)	0.8 (0.7–0.8)	240
GCR (Switzerland)	3066	2470	50.7 (46.3–55.8)	21.8 (18.1–26.8)	13 (9.7–18)	1 (0.9–1.2)	319
NCR (Netherlands)	160,940	133,829	43.7 (43.1–44.4)	16.3 (15.8–16.9)	9.6 (9.1–10)	0.8 (0.8–0.8)	282
UTARTU (Estonia)	3063	2063	54 (49.6–58.7)	26.2 (21.4–32.1)	–	1 (0.8–1.2)	385
Pancreas							
CPRD GOLD (UK)	7558	6456	29.9 (27.1–33.1)	8.7 (6.7–11.4)	6.2 (4.3–8.9)	0.5 (0.4–0.5)	130
SIDIAP (Spain)	7989	6042	50.7 (48.1–53.5)	19.8 (17.4–22.5)	12.6 (10.1–15.7)	1 (0.9–1.1)	319
HUS (Finland)	2749	2276	42.3 (37.7–47.6)	16 (12.3–21)	11.7 (8.2–16.8)	0.8 (0.6–1)	235
HUVM (Spain)	309	224	44.6 (32.9–60.2)	17.2 (8.4–34.8)	–	0.6 (0.5–1.2)	276
IMASIS (Spain)	589	388	44.3 (34.5–56.1)	24.9 (16–37.9)	16.8 (8.3–36.6)	0.6 (0.4–1)	167
CRN (Norway)	10,257	9591	24.4 (22.3–26.9)	4.2 (3.1–5.8)	3 (2–4.5)	0.4 (0.4–0.5)	118
GCR (Switzerland)	1064	961	36.9 (30.1–46.5)	8.2 (4.4–17)	6.1 (2.7–14.7)	0.7 (0.6–0.9)	198
NCR (Netherlands)	35,604	33,504	19.8 (18.7–21.1)	2.9 (2.3–3.5)	1.8 (1.3–2.4)	0.3 (0.3–0.4)	111

(Table 4 continues on next page)

	Cases (N)	Events (N)	One-year survival (95% CI)	Five-year survival (95% CI)	Ten-year survival (95% CI)	Median survival (95% CI)	Median follow-up
(Continued from previous page)							
Prostate							
CPRD GOLD (UK)	56,634	17,198	94.8 (94.4–95.2)	77.3 (76.3–78.2)	61.5 (60.1–63)	8.3 (7.9–8.6)	1607
SIDIAP (Spain)	42,771	10,155	97 (96.7–97.3)	83.4 (81.8–85.3)	67.5 (65.3–70.6)	4.2 (4.1–4.3)	2782
HUS (Finland)	15,609	4892	95.4 (94.5–96.2)	78.7 (76.6–80.7)	61.2 (58.7–63.7)	3.5 (3.4–3.6)	2614
HUVM (Spain)	2779	821	90.9 (88.9–92.8)	74.9 (71.3–78.8)	53.9 (48.1–60.9)	3.5 (3.2–4)	2429
IMASIS (Spain)	2520	660	94.8 (93.4–96.2)	83.1 (79.3–86.6)	68.2 (61.7–75.2)	4.4 (4–4.8)	2338
CRN (Norway)	77,156	27,565	95.5 (95.2–95.8)	78.8 (78.2–79.5)	60.5 (58.5–62.2)	8.1 (7.9–8.3)	2007
GCR (Switzerland)	5664	1699	96.2 (95.4–97)	77.8 (75.6–80.1)	63.2 (60.3–66.4)	4 (3.7–4.4)	2048
NCR (Netherlands)	187,851	72,576	95 (94.8–95.2)	75.9 (74.9–76.7)	58.3 (57.1–59.3)	7.9 (7.8–8)	2584
UTARTU (Estonia)	8075	1462	95 (94–96)	77.1 (74.7–79.7)	–	0.6 (0.6–0.7)	1964
Stomach							
CPRD GOLD (UK)	5755	4501	47.1 (43.6–50.8)	20.3 (17.3–23.9)	14.2 (11.3–18)	0.9 (0.8–1.1)	255
SIDIAP (Spain)	8450	5225	67.7 (65.4–70.1)	39.7 (37–42.6)	30.3 (27.4–33.5)	2.8 (2.3–3.5)	722
HUS (Finland)	1277	907	56.8 (50.7–63.7)	31 (25.2–38.3)	23.2 (17.6–30.8)	1.5 (1–2.6)	457
HUVM (Spain)	306	188	54 (42.4–69.1)	34.7 (23.6–51.9)	25.2 (13.6–48)	0.3 (0.2–0.9)	466
IMASIS (Spain)	670	382	63.5 (54.9–73.3)	38.7 (29.8–50.8)	31.8 (22.5–45.5)	1.3 (0.9–3.1)	427
CRN (Norway)	7705	6520	48.3 (45.5–51.2)	19.3 (17.1–21.8)	14.2 (12.2–16.6)	1 (0.9–1.1)	278
GCR (Switzerland)	674	493	59.5 (51.6–68.6)	27.2 (20.2–36.9)	22.5 (16–32.3)	1.6 (1–3.5)	428
NCR (Netherlands)	31,300	26,655	43.9 (42.6–45.3)	17.5 (16.4–18.7)	12.7 (11.7–13.8)	0.8 (0.8–0.9)	262

CPRD GOLD, Clinical Practice Research Datalink; SIDIAP, The Information System for Research on Primary Care; UTARTU, University of Tartu; HUS, Hospital District of Helsinki and Uusimaa; HUVM, Hospital Universitario Virgen Macarena; IMASIS, Institut Municipal Assistència Sanitària Information System; ULSM, Unidade Local de Saúde de Matosinhos; CRN, Cancer Registry of Norway; GCR, Geneva Cancer Registry Data-Base; NCR, Netherlands Cancer Registry; ECi, South East Scotland Cancer Database.

Table 4: **Age-standardised survival estimates per cancer type and database.**

NCR–50.7% (95% CI 48.1–53.5) in SIDIAP, and ten-year survival ranging from 1.8% (1.3–2.4) in NCR–16.8 (8.3–36.6) in IMASIS.

Stratification by sex showed across most databases, males had poorer survival at five and ten years for colorectal, lung, and head and neck cancers compared to females, with lung cancer also showing better survival in females at one year (Table S12). No sex differences were observed for liver, pancreatic nor stomach cancers across most databases with few exceptions. Crude survival estimates showed similar results, but with larger differences between sexes (Tables S13 and S14).

Discussion

In this study, we present a detailed analysis of baseline characteristics and overall survival for eight cancers recorded from EHRs and cancer registries from 11 databases across Europe that were mapped to the OMOP-CDM. Our findings indicate similarities in patient demographics across databases, and higher prevalence of comorbidities and prescriptions among primary care databases. Survival trends varied between cancer types and databases, especially among pancreatic and stomach cancers. Across databases, breast and prostate cancers were the most prevalent and had the highest survival compared to pancreatic cancer, with the poorest. Excluding breast cancer, more males were diagnosed with cancer and in general, had poorer long-term survival for specific cancers. Medication use and comorbidity profiles for patients with lung, gastrointestinal, pancreatic and liver cancers were consistent with known risk factors and early cancer indicators.

Breast, prostate, colorectal, and lung cancers were the most prevalent cancers in this study, consistent with findings across Europe.[1] Males constituted a higher proportion of cases, apart from breast cancer, aligning with existing literature.[1,8] While higher behavioural and environmental risk factors contribute to higher male cancer predominance, other studies also point to sex-related biologic and genetic factors as important contributors to increasing cancer susceptibility in males.[9,10] Survival advantages for females have been widely reported, though these differences may also diminish with increasing age.[10]

Although cancer is more prevalent in older populations, diagnoses among younger adults have increased in recent decades.[11] Screening programs, reproductive factors and public awareness campaigns, particularly for breast cancer have likely led to a lower age at diagnosis.[10] Our findings also show that other cancers, such as colorectal, are becoming increasingly common among younger adults, likely due to birth cohort effects.[12] Furthermore, the rising prevalence of Human Papillomavirus infections may explain the higher prevalence of head and neck cancer in younger age groups.[13]

The differences observed in the overall survival across databases can be attributed to several factors,

with the database source likely being the most significant. Cancer registries are the gold standard, and although they did show the lowest survival estimates for most studied cancers, this is due to their high level of comprehension in documenting all cancer events in the population in comparison to EHR databases. Whereas, for EHR, it captures health-seeking individuals and can lack information on the cause of death. This might result in a delay or oversight of those who do not seek health care due to the lack of early symptoms of more aggressive cancers, such as pancreatic and stomach cancers. Notably, two prior studies in both primary care databases in this study demonstrated relatively high sensitivity for cancer diagnoses when validated against cancer registries, which contributes to the reliability of our results based on primary care databases.[14,15]

Regional disparities in cancer mortality might also explain survival differences. According to the 2023 Spanish Cancer Profile Report by the OECD, Spain has one of the lowest rates of cancer incidence and mortality in the European Union (EU).[16] Furthermore, the country has large cancer mortality variability depending on the region, with some regions at the level of best-performing country in EU, and others close to the average. Three Spanish databases informed our study, which belonged to the regions with lowest mortality in Spain, which likely contribute to the observed results, and might explain differences with studies using data from other Spanish regions.[1,2,16] Variations in data source quality, allocation of health care provisions, including screenings and treatment protocols, or their absence for cancers with poor survival, and lifestyle and socioeconomic factors can also explain variations in cancer survival across countries.[8,17] Furthermore, differences in the prevalence of cancer subsites for specific cancers across countries can also explain the variations in survival. For example, gastric cardia cancers have poorer prognosis compared to gastric distal cancers, mainly due to differences in stage at diagnosis, tumour aggressiveness and treatment complexity, which could explain variation in survival across different countries.[18]

Specific cancer types have distinct natural histories with certain risk factor profiles, early signs and symptoms with several examples evident in this study.[1] T2D has been associated with an increased risk of pancreatic cancer and may even be early symptom of undiagnosed cases.[19] Other notable findings include chronic liver disease and viral hepatitis among liver cancer patients, both of which are established risk factors due to inflammation, cirrhosis and cellular damage.[20] Additionally, COPD and pneumonia were more prevalent in lung cancer patients, the first one likely due to shared risk factors such as smoking, while the second may reflect an early symptom of undiagnosed disease or risk factor.[21,22] Finally, anaemia, a well-recognised early indicator of gastrointestinal cancers, was commonly observed prior to diagnosis of both colorectal and stomach cancers.[23,24]

Prior medication usage before cancer diagnosis also provides valuable insights into various aspects of patient health, potential risk factors and underlying disease progression. Proton pump inhibitors and antibiotics were commonly prescribed prior to cancer diagnosis aligning with other studies.[25] While these medications are frequently prescribed in general medical care, their use could suggest the treatment of early symptoms of underlying disease.[26] Variations in specific medications by cancer type appeared as expected in line with known risk factors, for instance, higher prevalence of COPD-related medications among lung cancer patients and higher prevalence of treatment of common complications of chronic liver disease and cirrhosis among liver cancer patients.[27]

The federated approach taken in this study was possible by the prior mapping of all participating databases to a common data model which enables for a more harmonised, transparent and efficient process compared to other strategies for multi-database analyses.[28] Data partners retained full governance of their data and actively participated throughout the study, from refining outcome variables, to conducting the analyses and discussing the results. Collaboration was facilitated though a study-a-thon, during which data partners shared findings and contributed to a deeper understanding of the results. The federated approach required complex coordination, for example managing the processing of obtaining protocol approvals, addressing any technical issues remotely, and analysing the large volume of aggregated results generated. However, these challenges are not unique to the federated approach but to any large-scale multi-database study. In this context, the federated approach provides a practical and effective strategy for assessing diverse databases within a single study in a standardised and reproducible way.

The main strength of this study lies in the detailed characterisation of over 1.7 million patients, from multiple databases across several European countries, providing comprehensive insights into their demographics, comorbidities, and medication use. This level of granularity, available in certain databases, offered a valuable perspective essential for understanding cancer in real-world settings. This diversity of databases allowed for a thorough assessment across countries and database types, enhancing the generalisability of our results.

Despite these strengths, the study has some limitations. Integrating and comparing diverse health data sources introduces challenges too, due to differences in population coverage and data comprehensiveness. Cancer registries typically achieve full population representativeness within their countries or regions,

while primary care and hospital databases often cover smaller catchment areas. This variation should be considered when interpreting the results, as it might lead to selection bias. It is also important to note a potential overlap between the CPRD and ECi databases, as CPRD includes primary care records covering 28.6% of the Scottish population as of 2024.[29] However, ECi contains data exclusively from Scotland and from varied sources: paper-based patient case notes, EHRs, secondary healthcare databases and morbidity registers (Table S1). Therefore, the inclusion of both databases contributes to a more comprehensive coverage of the Scottish population.

The lack of potential prognostic factors across databases also likely influences the observed survival outcomes. Many EHR databases do not routinely capture cancer-specific details like stage at diagnosis or progression. Conversely, cancer-registries generally provide staging data but lack medical history and socioeconomic status data, the latter not having a standard indicator in OMOP CDM yet, though standarisation efforts are ongoing.[30] Additionally, variation in diagnostic and treatment guidelines across healthcare centres, countries and calendar time might also impact survival.

EHR databases currently contain missing data for certain key risk factors such as smoking status or alcohol intake, preventing their inclusion in the analysis. Addressing missingness in EHR data is particularly challenging, as it is often difficult to differentiate between missing information and true negative records, and might lead to bias if data is not missing completely at random. Also, primary care databases might incompletely capture patients with severe cancers as these patients are less likely to initially attend these settings. This could result in selection bias and overestimation of survival as well as index date misclassification and delays of cancer diagnoses. Compared to EHR which only included SNOMED CT codes, registries additionally included ICD-O-3 codes, which may have contributed to differences in phenotype precision across data sources.

Furthermore, sparse survival data in younger age groups for some cancers may skew age-standardised rates towards older age groups with more abundant data, meaning survival might be underestimated. However, we included both crude and age-standardised survival estimates in this study for comparison. Finally, for most databases, the study period started in 2000, however for some data sources the study started after this date. This variation in start date for some databases may introduce bias, as more recent data could reflect advances in treatment and survival, affecting longitudinal comparisons across data sources.

In summary, this study demonstrates the use of federated analysis that enables collaboration across multiple databases and countries, while preserving data-privacy. The variations in overall cancer survival were observed across different databases and caution is warranted when interpreting the results as they likely reflect differences in data comprehensiveness between cancer registries and EHRs, and regional variations. Importantly, the inclusion of comorbidity profiles and medication use prior to diagnosis, ascertained in EHR, but often not present in cancer registries, provides novel insights into the complexity of real-world cancer populations, paving the way for deeper investigation into their impact on cancer prognosis and patient survival outcomes. The overall results of this study confirm established epidemiological patterns and demonstrate them through a large-scale integration of EHRs and cancer registry data from multiple countries using a standardised common data model and federated analysis to preserve data privacy. This novel approach establishes a methodological benchmark for future research on real-world cancer populations in Europe.

Contributors

DN, RC, TDS, and AG were involved in the study conception and design. DN led and developed the standardised analytical code for the study. RC obtained funding for the study. ILS and DN wrote the first draft of the manuscript. Because the study followed a federated approach using the OMOP CDM, data verification and access to raw patient-level data were conducted locally by investigators at each participating site. IL-S, AP-C, RC, LP-C, AG, IK, JD, JE, JK, AG, JMR-A, AL, M-AM, NS, MV, CM, PSH, MM, KV, EE, PP, JE, MO, RK, RM-G, EF, KP, TT-G, FE, AMC, JMC, CMS, EF, AP, XC, GC, AR, MTS-S, AD, WYM, MC, MA-H, EB, DP-A, TD-S, DN were involved in the appraisal of the study design, interpretation of the results, critically reviewed the final manuscript and gave consent for publication.

Data sharing statement

We analysed aggregated results from pseudoanonymised data collected across 11 real-world databases, following the approval of each database's committee. We are not permitted to share individual data, however, results are publicly available in a shiny application for interactive exploration.

Declaration of interests

Laura Pérez-Crespo has received a grant from the Spanish Institute of Health Carlos III (Sara Borrell fellowship (CD23/00223).

Ian Koblbauer is the Director of Atlin Analytics Ltd, which provides health economics consultancy services for clients in the pharmaceutical industry, outside of the submitted work.

Maheva Vallet and Peter S Hall have received institutional research funding for various unrelated clinical studies from Lilly, Eisai, Novartis, Gilead, Sanofi, Roche, AstraZeneca, DxCover, AbbVie, MSD, and SeaGen.

Mees Mossevelt has received unconditional research grants from Chiesi, UCB, Amgen, Johnson & Johnson, Innovative Medicines Initiative, and the European Medicines Agency.

Eric Fey has received research grants from the University of Helsinki, HUS Helsinki University Hospital, and the Clinical Research Institute HUS; consulting fees from Roche and the European Commission; and travel support from the University of Helsinki. He serves as National Node Lead for OHDSI Finland.

Fernanda Estrevinho has received payments or honoraria for lectures, presentations, or educational events from AstraZeneca, Boehringer Ingelheim, Sanofi, Bristol Myers Squibb, Pierre Fabre, Roche, Daiichi Sankyo, Janssen-Cilag, MSD, Pfizer, Takeda, Amgen, Merck, and Novartis. She has also received support for attending meetings and/or travel from these same companies. Additionally, she has served on advisory boards for Janssen-Cilag, MSD, Merck, Roche, Daiichi

Sankyo, Novartis, AstraZeneca, Boehringer Ingelheim, and BMS. All unrelated to the submitted work.

Maria T. Sanchez-Santos received consultancy fees from Theramex HQ UK Ltd, unrelated to this work.

Antonella Delmestri is a consultant for the Saudi FDA.

Carlos Míguez has received payments for lectures, presentations, participation in speaker bureaus, and educational or manuscript-related events from Johnson & Johnson, Boehringer, and AbbVie. He has also received support for meeting attendance and/or travel from Johnson & Johnson, AbbVie, and Novartis. All unrelated to the submitted work.

Xihang Chen has received grants or contracts from University of Oxford.

Professor Daniel Prieto-Alhambra's research group from the University of Oxford has received research grants from the European Medicines Agency, from the Innovative Medicines Initiative, from Amgen, Chiesi-Taylor, Gilead, Lilly, Janssen, Novartis, and UCB Biopharma. He also conducted consultancy, with consultancy fees paid to the University from UCB Biopharma. Finally, Janssen has funded or supported training programmes organised by DPA's department. DPA sits in the Board of the EHDEN Foundation.

All other authors declare no conflicts of interest. The views expressed are those of the authors and not their employers or funders.

Acknowledgements

This work was supported by the European Health Data & Evidence Network, which received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and European Federation of Pharmaceutical Industries and Associations partners.

DN and Oxford team thanks Dr Ilona Tietzova (Department of Tuberculosis and Respiratory Diseases First Faculty of Medicine Charles University), Dr Andreas Weinberger Rosen (Centre for Surgical Science, Zealand University Hospital), Andrea Miquel Dominguez, (Otorrinolaringology department, Hospital Joan XXIII de Tarragona), Francesc Xavier Avilés-Jurado (Head Neck Tumors Unit, Hospital Clínic de Barcelona), Àlvar Roselló Serrano (Institut Català d'Oncologia, Hospital Universitari Dr Josep Trueta, Girona, Spain), Patricia Pedregal-Pascual and Carlos Guarner-Argente (Gastroenterology Department, Hospital de la Santa Creu i Sant Pau), for help with reviewing preliminary diagnostics codelists used for this study.

MO and RK acknowledge the Innovative Medicines Agency for supporting this work with a grant to their institution for data processing, salaries and travel expenses.

EE acknowledges Dr Tor Åge Myklebust and Dr Yngvar Nilssen (Department of Research, Cancer Registry of Norway (CRN), Norwegian Institute of Public Health) for checking the data and supporting the study analytics at CRN.

RC thanks Dr Páll Jónsson (National Institute for Health and Care Excellence), Dr Jacoline Bouvy (National Institute for Health and Care Excellence), and Dr Seamus Kent (Erasmus School of Health Policy & Management) for contributing to the study conception and initial protocol development.

MAM and IMASIS team thanks Dr. Maria Sala, Dr. Andrea Burón and Adrià Moncusí (Cancer Registry Programme, Epidemiology and Evaluation Department, Hospital del Mar Barcelona) for providing additional data from the cancer registry.

ILS thanks the Doctoral Programme in Biomedical Research Methodology and Public Health at the Autonomous University of Barcelona.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanepe.2025.101585.

References

1 Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74:229–263.
2 Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513,025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet*. 2018;391:1023–1075.
3 Observational Health Data Sciences and Informatics. *Observational medical outcomes partnership*. The Book of OHDSI; 2021. https://ohdsi.github.io/TheBookOfOhdsi/OhdsiCommunity.html#observational-medical-outcomes-partnership. Accessed December 4, 2024.
4 European Health Data Evidence Network. ehden.eu. https://www.ehden.eu/. Accessed November 16, 2023.
5 Belenkaya R, Gurley MJ, Golozar A, et al. Extending the OMOP common data model and standardized vocabularies to support observational cancer research. *JCO Clin Cancer Inform*. 2021;5:12–20.
6 Gilbert J, Rao G, Schuemie M, Ryan P, Weaver J. CohortDiagnostics: diagnostics for OHDSI cohorts. https://ohdsi.github.io/CohortDiagnostics; 2024.
7 Corazziari I, Quinn M, Capocaccia R. Standard cancer patient population for age standardising survival ratios. *Eur J Cancer*. 1990 2004;40:2307–2316.
8 OECD. Beating cancer inequalities in the EU. https://www.oecd.org/en/publications/beating-cancer-inequalities-in-the-eu_14fdc89a-en.html; 2024.
9 Jackson SS, Marks MA, Katki HA, et al. Sex disparities in the incidence of 21 cancer types: quantification of the contribution of risk factors. *Cancer*. 2022;128:3531–3540.
10 GBD 2019 Cancer Risk Factors Collaborators, Lang JJ, Compton K, et al. The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2022;400:563–591.
11 Zhao J, Xu L, Sun J, et al. Global trends in incidence, death, burden and risk factors of early-onset cancer from 1990 to 2019. *BMJ Oncol*. 2023;2:e000049. https://doi.org/10.1136/bmjonc-2023-000049.
12 Spaander MCW, Zauber AG, Syngal S, et al. Young-onset colorectal cancer. *Nat Rev Dis Primer*. 2023;9:21.
13 Barsouk A, Aluru JS, Rawla P, Saginala K, Barsouk A. Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma. *Med Sci*. 2023;11:42.
14 Recalde M, Manzano-Salgado CB, Díaz Y, et al. Validation of cancer diagnoses in electronic health records: results from the information system for research in primary care (SIDIAP) in Northeast Spain. *Clin Epidemiol*. 2019;11:1015–1024.
15 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*. 2020;10:e037719.
16 OCDE. EU country cancer profile: Spain 2023. https://doi.org/10.1787/260cd4e0-en; 2023. Accessed February 5, 2025.
17 Gheorghe G, Bungau S, Ilie M, et al. Early diagnosis of pancreatic cancer: the key for survival. *Diagnostics*. 2020;10:869.
18 Xue J, Yang H, Huang S, Zhou T, Zhang X, Zu G. Comparison of the overall survival of proximal and distal gastric cancer after gastrectomy: a systematic review and meta-analysis. *World J Surg Oncol*. 2021;19:17.
19 Ruze R, Song J, Yin X, et al. Mechanisms of obesity- and diabetes mellitus-related pancreatic carcinogenesis: a comprehensive and systematic review. *Signal Transduct Target Ther*. 2023;8:139.
20 Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. *Nat Rev Dis Primer*. 2021;7:6–28.
21 Song L, Wu D, Wu J, Zhang J, Li W, Wang C. Investigating causal associations between pneumonia and lung cancer using a bidirectional mendelian randomization framework. *BMC Cancer*. 2024;24:721.
22 Qi C, Sun S-W, Xiong X-Z. From COPD to lung cancer: mechanisms linking, diagnosis, treatment, and prognosis. *Int J Chron Obstruct Pulmon Dis*. 2022;17:2603–2621.
23 Krieg S, Loosen S, Krieg A, Luedde T, Roderburg C, Kostev K. Association between iron deficiency anemia and subsequent stomach and colorectal cancer diagnosis in Germany. *J Cancer Res Clin Oncol*. 2024;150:53.
24 Demb J, Kolb JM, Dounel J, et al. Red flag signs and symptoms for patients with early-onset colorectal cancer: a systematic review and meta-analysis. *JAMA Netw Open*. 2024;7:e2413157.

25 Pottegård A, Hallas J. New use of prescription drugs prior to a cancer diagnosis. *Pharmacoepidemiol Drug Saf*. 2017;26:223–227.
26 National Institute for Health and Care Excellence. Suspected cancer: recognition and referral. NICE guideline [NG12] https://www.nice.org.uk/guidance/ng12/chapter/Recommendations-organised-by-symptom-and-findings-of-primary-care-investigations; 2015. Accessed December 4, 2024.
27 Sharma A, Nagalli S. Chronic liver disease. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2024. http://www.ncbi.nlm.nih.gov/books/NBK554597/. Accessed December 4, 2024.
28 Gini R, Sturkenboom MCJ, Sultana J, et al. Different strategies to execute multi-database studies for medicines surveillance in real-world setting: a reflection on the European model. *Clin Pharmacol Ther*. 2020;108:228–235.
29 Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A. Data resource profile update: CPRD GOLD. *Int J Epidemiol*. 2025;54:dyaf077.
30 Jiang X, Beaton MA, Gillberg J, Williams A, Natarajan K. Feasibility of linking area deprivation index data to the OMOP common data model. *AMIA Annu Symp Proc*. 2022;2022:587–595.