Study Report

P4-C2-005

DARWIN EU® - Prevalence of selected cancers

28/01/2026

Version 3.0

Authors: Berta Raventós, Talita Duarte-Salles, Julieta Politi, Cesar Barboza, Ioanna Nika, Anton Barchuk, Yuqing Hu.

Public

CONTENTS

P4-C2-005 Study Report
Version: V3.0
Dissemination level: Public

Study title[1]	DARWIN EU® - Prevalence of selected cancers
Study report version	V3.0
Date	28/01/2026
EUPAS number	EUPAS1000000715
Active substance	N/A
Medicinal product	N/A
Research question and objectives	Research question What is the prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma in Europe? Study objectives The primary aim of this study was to estimate the point prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma across selected data sources in Europe. The secondary aim of this study was to compare the prevalence of selected cancers included in the primary aim of the study, using denominator data available within the data source versus denominator data obtained from external sources.
Countries of study	Croatia, Denmark, Germany, The Netherlands, Norway
Authors	Berta Raventós (b.raventos@darwin-eu.org) Talita Duarte-Salles (t.duarte@darwin-eu.org) Julieta Politi (j.politi@darwin-eu.org) Cesar Barboza (c.barboza@darwin-eu.org) Ioanna Nika (i.nika@darwin-eu.org) Anton Barchuk (a.barchuk@darwin-eu.org) Yuqing Hu (y.hu@darwin-eu.org)

[1]This is a routinely repeated study from P1-C1-001 with EUPAS50800.

LIST OF ABBREVIATIONS

Acronyms/term	Description
ALL	Acute lymphocytic leukaemia
AML	Acute myeloid leukaemia
CDM	Common Data Model
CC	Coordination Centre
CIPH	Croatian Institute of Public Health
CLL	Chronic lymphocytic leukaemia
CRN	Cancer Registry Norway
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DK-DHR	Danish Data Health Registries
DKMA	Danish Medicines Agency
DLBCL	Diffuse large B-Cell Lymphoma
DNCR	Danish National Cancer Registry
DNPR	Danish National Patient Registry
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Authorisation Studies Register
GDPR	General Data Protection Regulation
IARC	International Agency for Research on Cancer
ICD-O-3	International Classification of Diseases for Oncology, 3rd Edition
IKNL	Netherlands Comprehensive Cancer Organisation
InGef	Institut für angewandte Gesundheitsforschung Berlin GmbH
InGef RDB	InGef Research Database
IPCI	Integrated Primary Care Information
IQVIA DA	IQVIA Disease Analyzer
IRB	Institutional Review Board
N/A	Not Applicable
NAJS	Croatian National Public Health Information System
NCI	National Cancer Institute
NCR	Netherlands Cancer Registry
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
SEER	Surveillance, Epidemiology, and End Results
SHI	Statutory health insurance

Acronyms/term	Description
SNOMED	Systematised Nomenclature of Medicine
US	The United States
UK	The United Kingdom

1. TITLE

DARWIN EU® - Prevalence of selected cancers

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Berta Raventós Talita Duarte-Salles	Erasmus MC
Epidemiologists	Julieta Politi Yuqing Hu*	Erasmus MC
Data Scientist	Cesar Barboza Ioanna Nika Ger Inberg Maarten van Kessel Ross Williams	Erasmus MC
Clinical Domain Expert	Anton Barchuk	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation****
NAJS	Mario Sekerija Jakov Vukovic Anamaria Jurcevic Ivan Pristas Antea Jezdzic Marko Cavlina	Croatian Institute of Public Health (CIPH)
DK-DHR	Elvira Bräuner Susanne Bruun	Danish Medicines Agency (DKMA)
InGef RDB	Raeleesha Norris Annika Vivirito Alexander Harms Josephine Jacob	InGef - Institut für angewandte Gesundheitsforschung Berlin GmbH
NCR	Jelle Evers Maaike van Swieten	Netherlands Comprehensive Cancer Organisation (IKNL)
CRN	Espen Enerly Anna Skog	Norwegian Institute of Public Health

*Added to the study team on 10 October 2025.
**Data partners do not have an investigator role. Data partners execute code at their data source, review, and approve their results.

3. ABSTRACT

Title

DARWIN EU® - Prevalence of selected cancers

Rationale and background

Substantial uncertainty surrounds the prevalence of rare blood cancers. Estimates derived from real-world data can be influenced by various factors, including how population denominators are defined. In some cases, data sources may only provide information on disease cases, requiring the use of national demographic statistics to define population denominators. This study aimed to estimate prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma in registries and claims data, using denominator information obtained from external sources when not available in the data.

Research question and objectives

Research questions

What is the prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma in Europe?

Objectives

The primary aim of this study was to estimate the point prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma across selected data sources in Europe. The secondary aim of this study was to compare the prevalence of selected cancers included in the primary aim of the study, using denominator data available within the data source versus denominator data obtained from external sources.

Methods

Study design

Population-level descriptive epidemiology.

Population

Individuals present in the data source from 2010 (or start of accurate data if later) to the last year with complete data.

Variables

Outcome:

Outcomes included rare blood cancers, pancreatic cancer, and soft tissue sarcoma. Rare blood cancers were assessed separately and included acute lymphocytic leukaemia (ALL), acute myeloid leukaemia (AML), chronic lymphocytic leukaemia (CLL), diffuse large B-Cell lymphoma (DLBCL), follicular lymphoma, and multiple myeloma. Study outcomes were identified based on the presence of a relevant diagnosis or observation. Individuals were considered as a prevalent case if they have had such a record in the prior 5 years.

Relevant covariates:

Age and sex.

Data sources

1. Croatia: Croatian National Public Health Information System (NAJS)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Germany: InGef Research Database (InGef RDB)

4. The Netherlands: Netherlands Cancer Registry (NCR)
5. Norway: Cancer Registry Norway (CRN)

Study size

No sample size was calculated, as this was an exploratory study which did not test a specific hypothesis.

Statistical analysis

We estimated 5-year partial prevalence of each outcome of interest. Estimates were calculated annually as of the 1st of January. For the primary analysis, denominators were directly derived from the data sources, except in cancer registries (i.e., NCR and CRN), where denominators were obtained from national statistics available from Eurostat. For the secondary objective, prevalence estimates based on denominators directly derived from the data sources were compared with prevalence estimates based on denominators from external sources. The secondary analysis was performed in data sources with population-level denominators and national coverage (i.e., NAJS, DK-DHR).

We performed two sensitivity analyses: 1) One year of prior history was required for the analyses using denominators directly derived from the data (i.e., NAJS, DK-DHR, InGef RDB); 2) In DK-DHR and NAJS, we restricted cancer cases to those originating from national cancer registries.

All results were stratified by age and sex. When using internal denominators, age was calculated at denominator entry, allowing individuals to transition between age groups, while for external denominators, age was fixed at diagnosis.

A minimum cell count of 5 was used when reporting results, with any smaller counts reported as "<5".

Results

Estimated 5-year partial prevalence per 10,000 population in 2023 ranged from 0.51 to 1.00 for ALL, from 0.71 to 1.50 for AML, from 2.50 to 4.90 for CLL, from 2.66 to 3.80 for DLBCL, from 1.80 to 3.70 for follicular lymphoma, and from 2.80 to 4.60 for multiple myeloma. Lastly, estimates for pancreatic cancer ranged from 1.72 to 3.80, and for soft tissue sarcoma from 2.83 to 4.30. NAJS and DK-DHR obtained the highest estimates, while estimates from NCR and CRN were consistent and obtained the lowest figures.

The highest prevalence estimates for all included cancers were observed in the 60–69, 70–79, and 80–89 age groups and decreased as age decreased, except for ALL, where this pattern was reversed. Results were consistent when applying a one-year prior history requirement. Estimates obtained in DK-DHR restricting to cases mapped from the national cancer registry were consistent with those obtained in NCR and CRN, except for CLL and follicular lymphoma, where estimates remained higher. In NAJS, this restriction led to comparable estimates for ALL, AML, and pancreatic cancer, while other outcomes assessed obtained lower estimates than those reported for other data sources.

Estimated prevalence using internal and external denominators yielded consistent results, with some minor discrepancies. In NAJS, the use of denominators derived from the data source resulted in marginally lower estimates across outcomes studied. This is likely attributed to an overestimation of the population in the data source, which exceeds the Croatian population reported in Eurostat by approximately 400,000 individuals as of 2023, primarily due to the lack of data on emigration. Small discrepancies in age-stratified results were also noted, likely reflecting different approaches to age calculations applied in the analysis.

Discussion

Results from this study were in line with estimates from the previous DARWIN EU® study (EUPAS50800) and were consistent with the literature, with the most notable differences observed in NAJS and DK-DHR, likely due to the inclusion of unconfirmed cancer diagnoses, which may have led to an overestimation of

prevalence. When restricting to cases derived from the cancer registries, estimates in DK-DHR and NAJS (for some cancer types only) were comparable to those obtained in other cancer registries.

5-year partial prevalence estimates based on denominators from external sources were generally consistent with those derived directly from the data source. Minor variations were observed due to discrepancies between the data source's denominator and national statistics in NAJS and different methods for handling time-varying stratifications.

Findings from this study highlight the complementarity of different data sources to improve validity and suggest that data sources with population-level denominators affected by missing or incomplete population data may benefit from the use of external denominators for prevalence estimation.

4. AMENDMENTS AND UPDATES

Amendment number	Previous approved version of the report	Date	Section of the study report	Amendment or update	Reason
1	V2.0	30/06/2025	Not applicable.	Update from initial study protocol P1-C1-001 (EUPAS50800)	This is a routine repeated study.

Comparison with previous protocols:

Study deliverables	C1-001 (EUPAS50800)	P4-C2-005 (EUPAS1000000715)
Study period	2010–end of available data	2010–end of available data
Data partners (Country):[1]		
- IQVIA LPD Belgium (Belgium)	X	
- NAJS (Croatia)		X
- DK-DHR (Denmark)		X
- InGef RDB (Germany)		X (primary aim only)
- IQVIA DA Germany (Germany)	X	
- IPCI (Netherlands)	X	
- NCR (Netherlands)		X (primary aim only)
- CRN (Norway)		X (primary aim only)
- SIDIAP (Spain)	X	
- CPRD GOLD (UK)	X	
Reference study protocol	N/A	C1-001 (EUPAS50800)

[1] Some data partners have been included to participate in the primary aim only (noted in brackets).

Changes from reference study protocol:

- Objectives: The reference protocol included one specific objective per cancer type. In this protocol, these have been summarised into a single primary objective. A secondary aim comparing prevalence results derived from denominators identified from the data or external sources has been added.
- Study population: In the reference study protocol, participants were required to have one year of prior observation before contributing to the main analysis. In sensitivity analyses, this requirement was first removed and then extended to three years. In the current study, the one-year prior observation requirement was excluded from the main analysis and included as a sensitivity analysis for data sources with defined denominator populations (i.e., NAJS, DK-DHR, InGef RDB).
- Outcomes: The reference study protocol only included rare blood cancers. In the current study, additional outcomes (pancreatic cancer and soft tissue sarcoma) were included. Concept sets for rare blood cancers were revised and refined. Additionally, concepts were updated to incorporate ICD-O-3 codes, which were available in three of the data sources used in the current analysis (i.e., DK-DHR, NCR, CRN).

- Covariates: In the current study, the age category '>80' was intended to be used to represent the oldest age group (instead of 80–89, 90–99, and >100 in the reference study protocol). This change was made for analyses involving external denominators to align with data from national statistics, which do not provide information at this level of granularity.
- Analyses: The reference study protocol included 26 different analyses, exploring different design choices in terms of: 1) prevalence type (main analysis: 5-year partial prevalence, sensitivity: 2-year partial prevalence, and complete prevalence); 2) prevalence estimate (main analysis: point prevalence, sensitivity: period prevalence); 3) prior history requirement (main analysis: 365 days, sensitivity: 0 and 1,095 days); and 4) observable time requirement for period prevalence (main analysis: not applicable; sensitivity: one day and full period year). For the current study, we replicated the main analysis (i.e., 5-year partial point prevalence), changing the prior history to 0 days, with a sensitivity analysis of 365 days for some of the data sources included. The current study involved the use of national statistics to derive denominators for data sources where these are not available (i.e., CRN, NCR) and for comparison purposes (secondary aim) in data sources where both internal and external denominators were available (i.e., NAJS, DK-DHR). This approach was not used in the reference study protocol, where all prevalence estimates were calculated using denominators directly derived from the data sources included.

5. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Final Study Protocol	11th August 2025	18th August 2025
Creation of Analytical code	September 2025	September/October 2025
Execution of Analytical Code on the data	October 2025	November 2025
Draft Study Report	5th December 2025	12th December 2025
Final Study Report	To be confirmed with EMA	To be confirmed with EMA

6. RATIONALE AND BACKGROUND

Substantial uncertainty surrounds the prevalence of rare blood cancers. Estimates generated from real-world data can be influenced by data source type, setting, and quality. This present study builds on a previous DARWIN EU® study (EUPAS50800), which estimated the prevalence of rare blood cancers using data from five sources in Europe, primarily reflecting primary care settings. Across the data sources included in the study, acute lymphocytic leukaemia (ALL) and acute myeloid leukaemia (AML) were the least prevalent diseases, with the highest estimates for their 5-year partial point prevalence being 0.65 and 1.03 per 10,000, respectively. The highest estimate of prevalence of diffuse large B-Cell lymphoma (DLBCL) was 1.73 per 10,000, while the highest prevalence of follicular lymphoma was 2.83 per 10,000. Lastly, the highest estimates for prevalence of chronic lymphocytic leukaemia (CLL) and multiple myeloma were 4.13 and 4.27 per 10,000, respectively. Importantly, the study assessed the impact of different design choices on the obtained results, including different windows for prevalence assessment (i.e., point vs. period prevalence) and outcome duration (i.e., complete vs. partial prevalence, using 2 and 5 years to define outcome duration).

Building on the methodological work of the previous study, the current study has now been repeated to incorporate additional outcomes (i.e., pancreatic cancer and soft tissue sarcoma) and data sources, including registry data and claims data. Importantly, this study incorporated two national cancer registries in Europe which only captured cancer patient data. Consequently, denominators for prevalence estimation could not be derived directly and had to be sourced externally from national estimates.

7. RESEARCH QUESTION AND OBJECTIVES

Research question

What is the prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma in Europe?

Objectives

The primary aim of this study was to estimate the point prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma across selected data sources in Europe.

The secondary aim of this study was to compare the prevalence of selected cancers included in the primary aim of the study, using denominator data available within the data source versus denominator data obtained from external sources.

The study was built upon a previous DARWIN EU® study (EUPAS50800), which estimated the prevalence of rare blood cancers in 5 primary care data sources in Europe. We replicated the main analysis of the previous study (i.e. 5-year partial point prevalence), with additional outcomes, data sources, and follow-up. Importantly, we calculated prevalence using denominators obtained from external sources in cases where they could not be directly derived and compared these estimates with those based on internal denominators in data sources with national coverage.

8. RESEARCH METHOD

8.1. Study design

Retrospective cohort study based on routinely collected health data from five data sources across five European countries. This study consisted of a descriptive epidemiological study and provided point prevalence estimates of selected cancers in each participating data source.

The analytical approach used to estimate prevalence depended on the source used to identify denominators. For the primary analysis, denominators were derived directly from the data. In cancer registries, these were not available in the data and had to be obtained from external sources. Both strategies were compared, where possible, in the secondary objective.

The source of denominators impacted the definition of index date (**Figure 1**). For analyses based on denominators derived directly from the data, the cohort entry date was defined as the date on which each individual began contributing observation time to the denominator. Similarly, the end of follow-up was outlined as the date on which they ceased contributing to the denominator. This logic is supported in the current analytical pipeline for descriptive epidemiological studies and is described in more detail in section "**8.2. Follow-up**". For analyses based on denominators derived from external sources, the identification and follow-up of patients was defined relative to the first outcome occurrence (i.e., index date). Follow-up was established as the end of observation or the end of the study period.

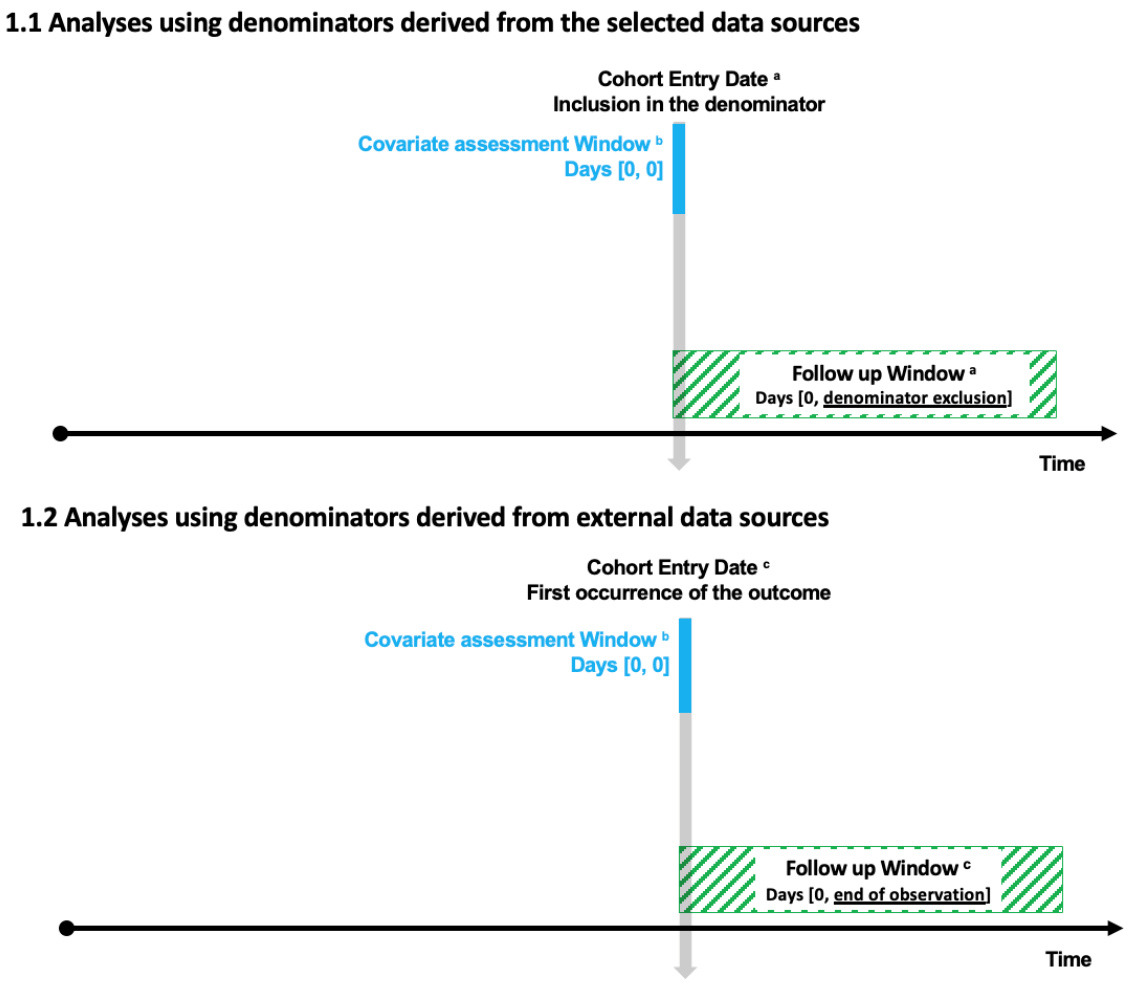

Figure 1. Graphical depiction of the study design.

a. For analyses using denominators from the selected data sources, index date was defined as the earliest date in which an individual fulfilled pre-specified inclusion criteria (e.g., study period, age for stratified results, prior history requirements for sensitivity analyses). Follow-up ended at the earliest day in which an individual did no longer fulfil pre-specified criteria.
b. Variables assessed at index date included age, sex, and prior history requirement (where applicable).
c. For analyses using external denominators, index date was defined as the date in which an individual was diagnosed with an outcome of interest. End of follow up was defined as the first of end of observation or study period end.

8.2. Follow-up

Follow-up differed depending on the strategy used to identify denominators. When derived from the data source itself, the strategy to define follow-up was based on the logic defined to identify denominators in the current analytical pipeline for disease epidemiology studies (see "**8.9. Statistical methods**").

Participants began contributing person time on the respective date of the latest of the following:

- Study start date (1st January 2010)
- Date at which they had sufficient prior history (365 days, or, if there was no requirement for prior history, this date would coincide with the date at which their observation period started)
- Date at which they reached a minimum age (where age strata was considered)

Participants ceased contributing person time at the earliest date of the following:

- Study end date (most recent year with data up to 31st December)
- Date at which their observation period ended
- The last day in which they have the maximum age (where age strata was considered)

Where there are multiple age strata, study participants contributed to each strata as long as they satisfied the age conditions relevant for that strata. When they reached the limit of one age strata, they began contributing to the next applicable strata.

When the denominator was derived from external sources, longitudinal follow-up was not conducted. Instead, a cross-sectional analysis was conducted to identify individuals in the data source who met the criteria for the condition of interest. This approach differed from the first strategy, where data was assessed longitudinally, and patients were included or excluded based on when they satisfied specific criteria (e.g., strata age thresholds or prior history requirements).

An overview of study entry and exit as defined in the current analytical pipeline is shown below in **Figure 2** for the overall analysis, with entry of exit for a given age strata shown in **Figure 3**.

In the example shown in **Figure 3**:

- **Person ID 1 and 3** enter the study on the day they reach the minimum age and exit at the maximum age.
- **Person ID 2** enters the study on the study start date and exits at the maximum age.
- **Person ID 4** enters the study on the day they reach the minimum age and leaves when they exit the data source (the end of their observation period).
- **Person ID 5** enters the study on the day they have sufficient prior history and exits at the maximum age.
- **Person ID 6** has two observation periods in the data source. For the first, they enter at study start date and contribute time until their exit, for the second, they start contributing time again within the observation period once they have sufficient prior history and exit at the maximum age.

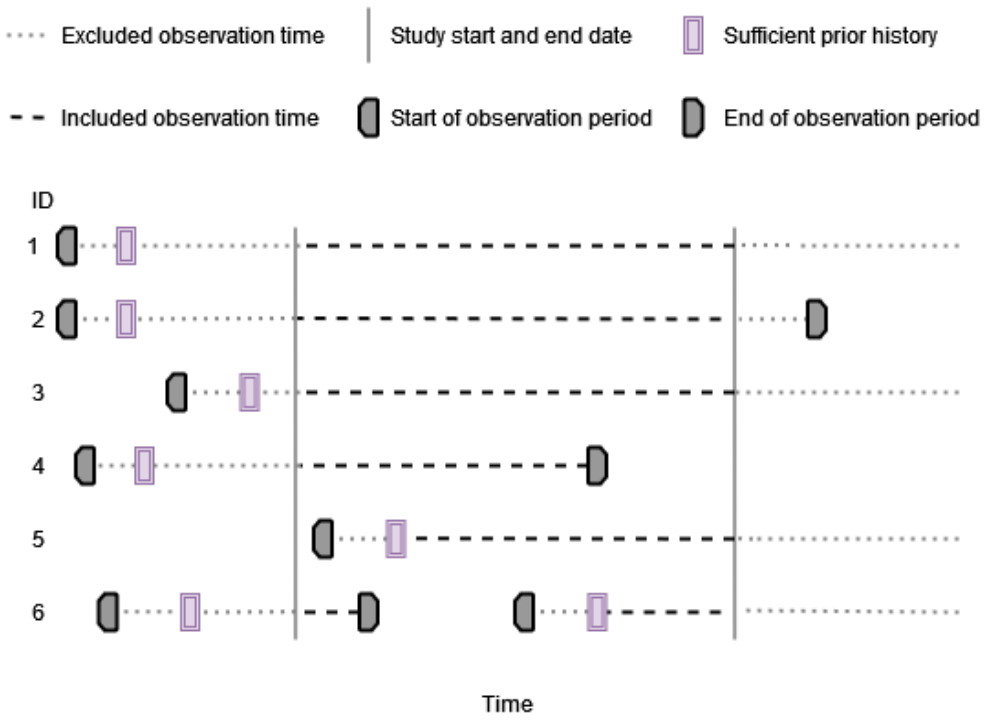

Figure 2. Included observation time for study participants overall.

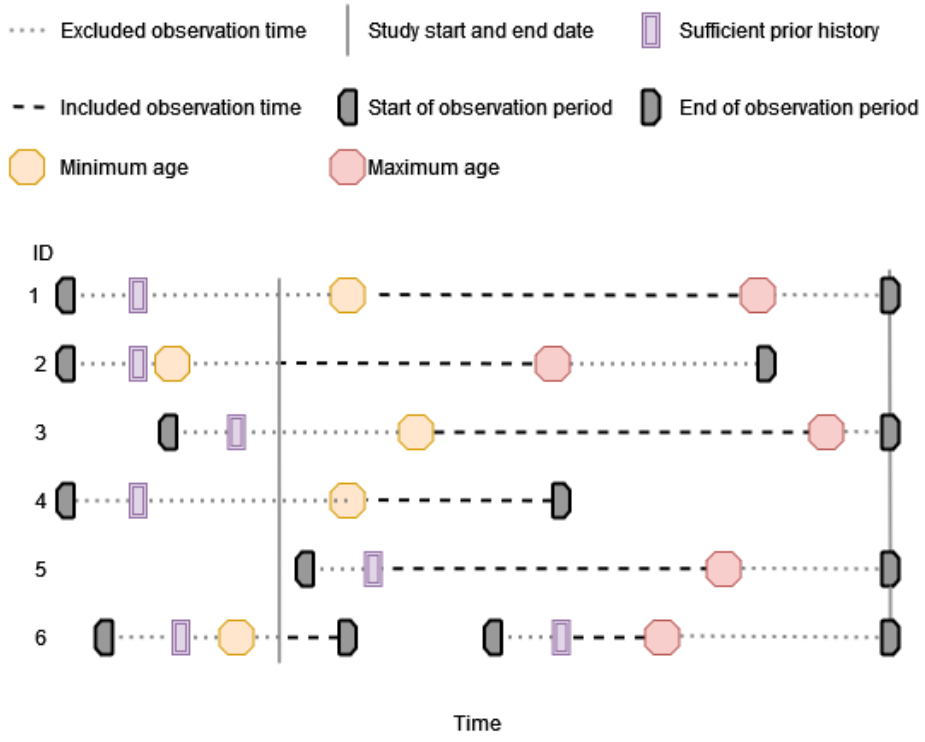

Figure 3. Included observation time for study participants for a given age strata.

8.3. Study population with inclusion and exclusion criteria

All individuals present in the selected data sources during the study period were eligible for inclusion. No prior history requirement was applied in the primary analysis. This decision is based on the logic implemented in the current analytical pipeline for disease epidemiology studies (see "**8.9.2. Main statistical methods**"), where the fulfilment of pre-specified criteria (e.g., age, prior history) determines when participants can be included in the denominator used for prevalence calculations. This approach leverages the longitudinal nature of the data and cannot be applied in data sources such as cancer registries, where inclusion and follow-up are based on the occurrence of the outcome. In the selected cancer registries, the start of observation of patients is defined as either the start date of the registry or the individual's date of birth or immigration (whichever is later), and therefore, prior observation time does not reflect a measure of prior patient observability in the data.

In the sensitivity analysis, we added a one year of prior history for data sources with population-level denominators (i.e., NAJS, DK-DHR, InGef RDB). This is described in more detail in section "**8.9.4 Sensitivity analyses**".

8.4. Study setting and data sources

This study was conducted using routinely collected data from data sources in the DARWIN EU® network. The study was informed by 5 data sources in the DARWIN EU® network of data partners from 5 European countries, of which 4 European Union (EU) member states (**Table 1**). All data sources were a priori mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM).

Data sources

1. Croatia: Croatian National Public Health Information System (NAJS)
2. Denmark: Danish Data Health Registries (DK-DHR)
3. Germany: InGef Research Database (InGef RDB)
4. The Netherlands: Netherlands Cancer Registry (NCR)
5. Norway: Cancer Registry Norway (CRN)

All data sources were used to inform the primary analysis. For this analysis, the use of external denominators was restricted to NCR and CRN, which were the only data sources from which denominators could not be directly derived (**Table 1**). The secondary objective, aimed at comparing prevalence estimates when using different sources to identify denominators, was restricted to NAJS and DK-DHR, as these were the only data sources with denominator populations that have national-level coverage.

Data sources selection

These data sources fulfilled the criteria required in terms of data quality, completeness, timeliness, and representativeness to conduct studies on cancer epidemiology while covering different regions of Europe. In terms of relevance, the selection of data sources was based on the availability of data on the selected outcomes to perform the described analyses.

Additional sensitivity analyses restricting cases to those originating from cancer registries were incorporated after the approval of the protocol in NAJS and DK-DHR, and new data releases became available for CRN and NCR. Additional details are described in "**8.9.4 Sensitivity analyses**" and "**8.10. Deviations from the protocol**".

Data source justification and key characteristics

Croatian National Public Health Information System (NAJS):

NAJS was included in this study to contribute to the diversity of data sources in terms of data type and health settings. It includes data from electronic health records and health registries on approximately 4.3 million active patients (please see "**10.2. Strengths and limitations**" for further details), which allowed for the estimation of prevalence using population-level denominators to be directly derived from the data.

Data from the different registries that comprise NAJS have different periods, with primary care data beginning in 2015 and hospital inpatient records starting in 2017 (**Annex I**). NAJS also includes information from the Cancer Registry at the Croatian Institute of Public Health, a population registry that includes all newly diagnosed cases of cancer in people living in the Republic of Croatia. The Cancer Registry has been operating since 1959 and has been part of NAJS since 2015.[1]

Based on a preliminary feasibility assessment, the number of person counts for outcomes of interest was estimated to be in the order of thousands, except for soft tissue sarcoma, where counts did not reach 1,000, and pancreatic cancer, where counts exceeded 10,000. Data availability for NAJS extended up to January 2025. Data prior to 2017 was not included, due to the unavailability of hospital records before that year.

In the present study, we used data from the whole NAJS data source, with mapped data from different registries. For the sensitivity analysis, we restricted cases to those recorded in the Cancer Registry.

Danish Data Health Registries (DK-DHR):

DK-DHR was included in this study because it comprises data from health registries in Denmark, covering the entire population of Denmark since 1995 (approximately 5.9 million active patients). DK-DHR includes mapped data from the Danish National Cancer Registry (DNCR). DNCR is considered the gold standard for identifying incident cancer cases in Denmark. Its comprehensive coverage and long-standing history are why it is utilised by international consortia, such as NORDCAN, for comparable Nordic cancer statistics. The DNCR captures the incidence, type, and location of malignancies with the highest degree of validity and completeness, especially for solid tumours. Over 90% of cancers in DNCR are histologically confirmed.

The DK-DHR also includes mappings of cancers from the Danish National Patient Registry (DNPR). While the Danish National Patient Registry (DNPR) is a powerful resource and often exhibits a good positive predictive value for various cancer diagnoses, its primary role in the DK-DHR is not for cancer incidence calculation. The DNPR was mapped and integrated in the DK-DHR for cancers specifically to document all procedures, treatments, and other patient contacts following the initial diagnosis of cancer. This allows for construction of robust, longitudinal history of the patient's care pathway. DK-DHR captures information on all people in Denmark, which allows the estimation of prevalence using population-level denominators derived directly from the data. Given its national coverage, it enabled the comparison of prevalence estimates based on both internal and external denominators.

Based on a preliminary feasibility assessment, the number of person counts for outcomes of interest was estimated to exceed 10,000 across the entire data source, except for ALL and soft tissue sarcoma, where counts were on the order of thousands. Data availability extended up to 2024, but there was a lag in data for histologically confirmed cancers of one year, with data for DNCR available up to December 2022.

In the present study, we used data from the whole DK-DHR data source, with mapped data from DNPR (up to December 2023) and DNCR (up to December 2022). For the sensitivity analysis, we restricted cases to those recorded in the DNCR.

InGef Research Database (InGef RDB):

InGef RDB was included in this study to contribute to the diversity of data sources in terms of data type and health settings, as it includes claims data from primary and hospital care settings on approximately 10 million individuals across more than 50 statutory health insurance providers (SHIs) throughout Germany. InGef RDB includes data on all patients insured through specific SHIs, which allowed the estimation of prevalence using population-level denominators derived directly from the data.

Based on a preliminary feasibility assessment, the number of person counts for the outcomes of interest was estimated to be in the order of thousands across the entire data source, except for soft tissue sarcoma, where counts did not reach 1,000, and pancreatic cancer, where counts exceeded 10,000. InGef RDB contributed data from 2015 to 2024, as legal regulations only allow the use of data from the most recent 10 years (see "**10.2. Strengths and limitations**" for further details).

Netherlands Cancer Registry (NCR):

NCR was included in this study because it is the national cancer registry in the Netherlands. It includes data from all patients newly diagnosed with cancer in the country since 1989. The number of patients included is approximately 2.7 million. NCR only captures information on patients diagnosed with cancer and is therefore not suitable for estimates that require population-level denominators unless those denominators are derived from external sources.

Based on a preliminary feasibility assessment, the number of person counts for the outcomes was estimated to exceed 10,000 across the entire data source, except for ALL and multiple myeloma, with counts below 8,000. Data availability extended up to January 2025. However, due to lack of completeness of most recent data, the study was restricted up to December 2023.

Cancer Registry of Norway (CRN):

CRN was included in this study, as it is a national cancer registry, including information on all patients diagnosed with cancer in Norway since 1951 (approximately 1.2 million people included). CRN only captures information on patients diagnosed with cancer, and therefore, this data source is not suitable for estimates that require population-level denominators unless those denominators are derived from external sources.

Based on a preliminary feasibility assessment, the number of person counts for the outcomes were expected to exceed 10,000 across the entire data source, except for ALL, where counts did not achieve 5,000. Data availability extends up to January 2025.

Information on the data sources used in this study is provided in **Annex I**.

Table 1. Data sources and their participation in study objectives.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals[1]	Calendar period covered by each data source[2]	Contributing to	Denominator source, primary aim[3]	Denominator source, secondary aim[3]
Croatia	NAJS	- Primary care - Hospital care (IP, OP)	- Registries, EHRs	4.30 M	2010 to 2025	Primary and secondary aim	Internal	Internal vs. External
Denmark	DK-DHR	- Hospital care (IP, OP)	- EHRs	5.98 M	1995 to 2024	Primary and secondary aim	Internal	Internal vs. External
Germany	InGef RDB	- Primary care - Hospital care (IP, OP)	- Claims	7.67 M	2015 to 2024	Primary aim	Internal	N/A
The Netherlands	NCR	- Hospital care (IP) - Others	- Registries	2.43 M	1993 to 2025	Primary aim	External	N/A
Norway	CRN	- Hospital care (IP, OP)	- Registries	0.36 M	1953 to 2025	Primary aim	External	N/A

CRN=Cancer Registry Norway, DK-DHR=Danish Data Health Registries, EHR=Electronic health record, InGef RDB=InGef Research Database, IP=inpatient, NAJS= Croatian National Public Health Information System, NCR=Netherlands Cancer Registry, N/A= Not applicable, OP=outpatient.

[1]Defined as the maximum number of individuals in observation in the last 6 months of data. Regarding NAJS, it is important to note that information on individuals who have migrated out of the country was not available. As a result, the number of patients included in the data source is higher than the actual number of residents (i.e., 4.3 million active patients in current data release vs. 3.8 million individuals in 2024 based on Eurostat data).

[2]Period covered in the data release used for this study. NAJS contributed data from 2017 onwards, as hospital data is available from that year onwards. Data from 2025 was not used, as data for 2025 was not complete. For InGef RDB, legal regulations only allow for the use of the most recent 10 years of data. DK-DHR and NCR provided data up to end of 2023, due to lack of completeness of most recent data. Data from CRN was complete up to December 2024.

[3]External denominators were obtained from publicly available demographic data from Eurostat. External denominators were used for the primary analysis in the Netherlands (NCR) and Norway (CRN), as they only captured information on patients diagnosed with cancer.

8.5. Study period

The study period spanned from 1st January 2010 to the most recent year with complete data. In some data sources, the availability of accurate data varied and did not cover the entire study period. The study started later in NAJS (2017) and InGef RDB (2015). Data sources provided data up to 2024, except for DK-DHR and NCR, which provided data up to December 2023. For the sensitivity analysis restricting cases to those mapped from the cancer registry, DK-DHR provided data up to December 2022 (see "**8.9.4. Sensitivity analyses**").

8.6. Variables

8.6.1. Exposure

None.

8.6.2. Outcome

All Objectives:

- Rare blood cancers:
    - Acute lymphocytic leukaemia (ALL)
    - Acute myeloid leukaemia (AML)
    - Chronic lymphocytic leukaemia (CLL)
    - Diffuse large B-Cell Lymphoma (DLBCL)
    - Follicular lymphoma
    - Multiple myeloma
- Pancreatic cancer
- Soft tissue sarcoma

Outcomes were assessed based on the presence of records indicating the specified diseases. Each outcome, including the different rare blood cancer types, was assessed separately. Concept sets for rare blood cancers were based on those included in the DARWIN EU® study (EUPAS50800), which were based on the Systematised Nomenclature of Medicine (SNOMED) terminology. These codes were refined during the study execution following the DARWIN EU® phenotyping standard processes, which involved the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources. Codes based on the International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) were included, as these were available in some of the data sources participating in the study (i.e., DK-DHR, NCR, and CRN). Concept sets used for the identification of outcomes are described in **Annex III**, where we also outline the concepts that differed from the ones used in the previous study (**Table S1** to **Table S8**).

Outcome cohort definitions:

For this study, we focused on the primary analysis performed as part of the prior DARWIN EU® study (EUPAS50800), which estimated the 5-year partial point prevalence of outcomes of interest.

Point prevalence represents the proportion of a population that is in observation and affected by the specific condition at a specific point in time. Partial prevalence limits the number of patients to those alive and diagnosed during a fixed time in the past. It differs from complete prevalence (also referred to as full prevalence or life prevalence), which includes all individuals previously diagnosed with the cancer who are still alive at the specific point in time, regardless of how long ago the diagnosis was, or whether the patients are considered cured or still require treatment at that specific time.[2] Both types of estimates were

reported in the prior DARWIN EU® study (EUPAS50800), where complete and point prevalence (using 2 and 5 years to define outcome duration) were compared.

In this present study, individuals were considered as a prevalent case if they had a relevant record observed in the prior 5-years (i.e., 1,825 days). This is shown below in **Figure 4** for a set of hypothetical patients, where the outcome duration (i.e. one time unit) is shorter than the one used in this study. To note, as an individual may have multiple records that would lead to cohort entry, partial prevalence outcome durations may vary, and individuals may have more than one outcome period (e.g., a recurrence). This is shown in the figure where person ID 1 and 3 have one instance of the outcome recorded (with the outcome duration being the same for both), whereas ID 4 has two instances and so has a longer outcome duration.

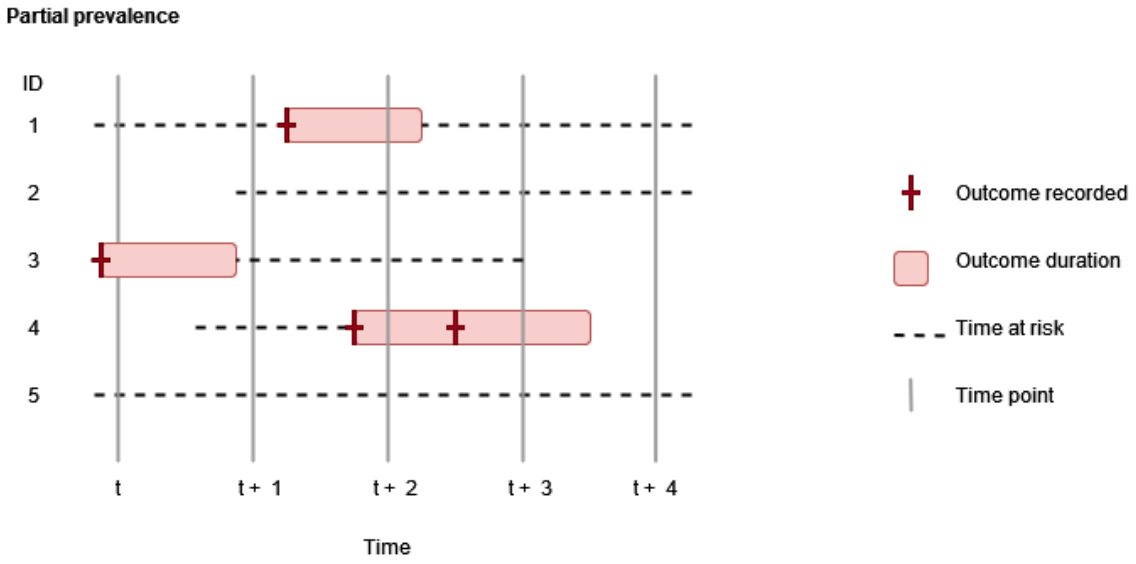

Figure 4. Outcome duration for partial prevalence.

8.6.3. Covariates, including confounders, effect modifiers, intercurrent events, and other variables

Two age groupings were used: 1) 0–9; 10–19; 20–29; 30–39; 40–49; 50–59; 60–69; 70–79; 80–89, 90–99, and 100–150 and 2) 0–44; 45–64; >65. The latter, broader age groups were used where low counts preclude reporting results for the 10-year age bands. For analyses involving the use of external denominators, the age category '>80' was used to represent the oldest age group to align with national statistics (see "**8.10. Deviations from the protocol**"). The sex (male/female) of study participants was also identified.

8.7. Study size

No sample size was calculated, as this was a descriptive disease epidemiology study which did not test a specific hypothesis. In addition, we used already collected available data to estimate prevalence. Thus, the sample size was driven by the availability of data for patients with outcomes of interest. Based on a preliminary feasibility assessment, the expected number of persons counts exceeded 10,000 in DK-DHR, NCR, and CRN across most outcomes studied. In NAJS and InGef RDB, counts were expected to be in the order of thousands in most cases. In the feasibility assessment, the lowest expected number of cases was observed for ALL (ranging from 2,000 to 4,500) and soft tissue sarcoma (ranging from 300 to 15,000).

8.8. Data transformation

Analyses were conducted separately for each data source. Before study initiation, test runs of the analyses were performed on a subset of the data sources and quality control checks were performed. Once all the

tests passed (see **Annex II**), the final study codes package was released in the version-controlled Study Repository for execution against all the participating data sources.

The data partners locally executed the analytics against the OMOP CDM in R Studio and reviewed and approved the, by default, aggregated results.

The study results of all data sources were checked, after which they were made available to the team, and the dissemination phase started. All results were locked and timestamped for reproducibility and transparency.

8.9. Statistical methods

8.9.1. Main summary measures

The measure of effect of selected cancers reported in this study was point prevalence, defined as the proportion of individuals in a population who have the outcome of interest at a given point in time. For an example of calculations of point prevalence see **Figure 5**. At time t+2, two of the five study participants are in an outcome cohort, giving a point prevalence of 40%.

More specifically, we calculated 5-year partial point prevalence. When calculating partial prevalence, the outcome end date was determined by a fixed number of days since the last record of the outcome events, which was defined as 5 years in this case, provided that the patient was not lost to follow-up or died before this time window elapsed.

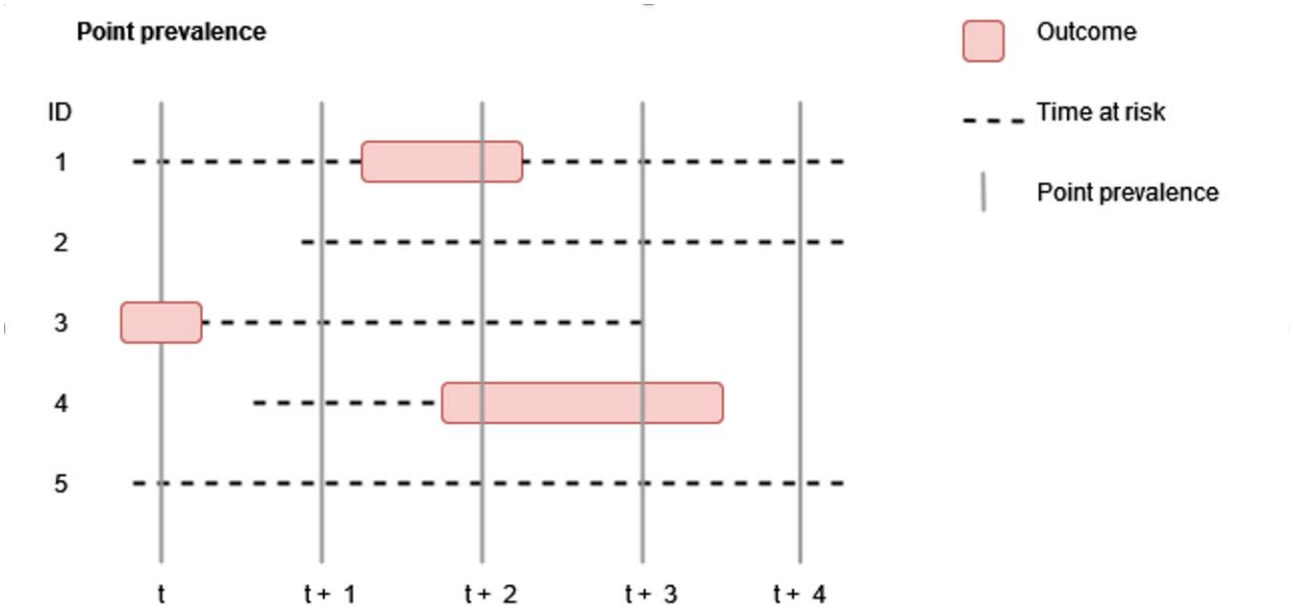

Figure 5. Point prevalence.

8.9.2. Main statistical methods

Point prevalence was estimated as of the 1st January of each calendar year. For the primary analysis, denominators were derived directly from the data in NAJS, DK-DHR, and InGef RDB. In NCR and CRN, these could not be derived from the data and were obtained from external sources. In the secondary analysis, the same approach was applied to NAJS and DK-DHR, and the results were compared with those from the primary analysis.

National statistics were obtained from publicly available demographic data in Eurostat, which collects information on the number of people having their usual residence in an EU Member State on 1st January of each calendar year. Population data collected in Eurostat is transmitted by national statistical institutes of

the Member States and can also be either based on data from the most recent census adjusted by the components of the population change produced since the last census or based on population registries. In the case of Croatia, the reported reference population corresponds to the "Usually Resident Population", as stated in Article 2 of Regulation 1260/2013. For Denmark, this corresponds to the "Registered Residence Population", which is composed of those persons who are listed in one or more registers owned by national authorities at the reference date.[3] The study used aggregate statistics by sex and five-year age groups,[4] which were further combined when required to align with this study's age group stratifications (see "**8.6.3. Covariates, including confounders, effect modifiers, intercurrent events, and other variables**").

The *IncidencePrevalence* R package, developed by DARWIN EU®, was used to calculate prevalence when using denominators derived directly from the data.[5] Custom code was created for analyses requiring the use of external sources to identify denominators, as this is not currently implemented in the current DARWIN EU® analytical pipelines.

Cell counts <5 were suppressed when reporting results to comply with the data source's privacy protection regulations.

8.9.3. Missing values

Outcomes assessed in the study were based on the recorded diagnoses available in the data. The absence of a record was considered as the study participant not having been diagnosed with the condition of interest.

8.9.4. Sensitivity analyses

We included a sensitivity analysis adding one year of prior observation to ensure a minimum prior observation time (**Table 2**). This analysis was applied in data sources that allow the identification of population-level denominators (i.e., NAJS, DK-DHR, and InGef RDB). Please see section "**8.3. Study population**" for more details on prior observation requirements.

We conducted a second sensitivity analysis, deviating from the original protocol, by restricting outcome cases to those mapped from national cancer registries in DK-DHR and NAJS. This analysis was incorporated to minimise potential outcome misclassification as, in these data sources, cases derived from sources other than cancer registries are not histologically confirmed and may involve suspected cases (see "**8.10. Deviations from the protocol**").

Table 2. Sensitivity analyses – rationale, strengths, and limitations.

	What is being varied? How?	Why?	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
Prior history requirement (365 days)	Participants were required to have one year of prior history before contributing observation time to the study. This was only applied in data sources with population-level denominators (i.e., NAJS, DK-DHR, InGef RDB).	To assess the impact of this requirement on prevalence estimates.	To identify any ongoing outcomes at the time at which a participant entered the study.	This was not applied in cancer registries due to how observability is defined in these data sources.
Restriction of cases originating from	Outcome cases were restricted to those	NAJS and DK-DHR comprise multiple registries and data	Allow for a more accurate capture of cases, reducing potential	Data availability from DNCR ends in December 2022.

	What is being varied? How?	Why?	Strengths of the sensitivity analysis compared to the primary	Limitations of the sensitivity analysis compared to the primary
cancer registries in DK-DHR and NAJS[1]	recorded in national cancer registries.	sources, including national cancer registries. In these data sources, cancer cases originating from sources other than cancer registries are not histologically confirmed and may represent suspected cases. In contrast, cancer registries are regarded as the gold standard for accurately identifying cancers within these datasets.	outcome misclassification.	

CRN=Cancer Registry Norway, DK-DHR=Danish Data Health Registries, DNCR=Danish National Cancer Registry, InGef RDB=InGef Research Database, NAJS= Croatian National Public Health Information System, NCR=Netherlands Cancer Registry.
[1]Cancer registries: the Danish National Cancer Registry (DNCR) in Denmark and the Cancer Registry at the Croatian Institute of Public Health in Croatia.

8.10. Deviations from the protocol

Since the last publication of the protocol, this has been amended to incorporate the following changes:

Deviation number	Protocol version	Date	Section of study protocol	Deviation	Reason
1	3.0	16/10/2025	8.9.4. Sensitivity analyses	Addition of a sensitivity analysis in DK-DHR, restricting cases to those originating from DNCR. In addition, the data source description has been improved by providing additional information regarding the different sources of cancer diagnoses.	Cancer diagnoses in DK-DHR are mapped from the Danish National Patient Registry (DNPR) and the Danish National Cancer Registry (DNCR). The DNPR is expected to have a high predictive value for most cancers but may also include suspected cases that are not histologically confirmed. DNCR is considered the definitive gold standard for identifying cancers in Denmark, with a high degree of validity, completeness, and over 90% histologically confirmed cancers. By restricting cases to those from the DNCR (available in DK-DHR up to December 2022), we expected more accurate case capture and reduced potential outcome misclassification.
2	3.0	03/11/2025	8.6.3. Covariates, including confounders, effect modifiers, intercurrent events, and other variables	According to the approved protocol, the age category ">80" was intended to represent the oldest age group, rather than using separate categories of 80–89, 90–99, and >100 as specified in the previous DARWIN EU® study. This adjustment was planned to align with national statistics, which do not provide data at the more granular level. However, this modification was not implemented in the study code, and the results are therefore presented using the age categories 80–89, 90–99, and 100–150 in analyses where internal denominators were used for NAJS (main analysis) and InGef RDB. Consequently, results using internal denominators stratified by age cannot be directly compared with those obtained from external sources for individuals aged >80 in NAJS. Due to the low number of outcomes observed in the 100–150 age group, this group has been excluded from figures presenting results stratified by age group. However, all	The intended age grouping (“>80”) was inadvertently not implemented in the study code. This was subsequently corrected for data sources where the study code had to be executed a second time, and the correction was implemented in DK-DHR and, in NAJS, only as part of the sensitivity analysis.

Deviation number	Protocol version	Date	Section of study protocol	Deviation	Reason
				results are available in the ShinyApp.	
3	3.0	06/11/2025	8.4. Study settings and data sources	NCR and CRN updated data releases after the study protocol was approved. For NCR, individuals aged <18 were not included in previous releases onboarded in DARWIN EU®. In the latest release, individuals <18 years were included and therefore contributed to the study. For CRN, complete cancer data became available for 2024, which allowed us to estimate prevalence for that year (whereas the protocol had originally specified 2023).	New data releases became available and were onboarded in DARWIN EU®.
4	3.0	04/12/2025	8.9.4. Sensitivity analyses	Addition of a sensitivity analysis in NAJS, restricting cases to those originating from the national cancer registry. In addition, the data source description has been improved by providing additional information regarding the Cancer Registry in Croatia. The method of constructing outcome cohorts in the sensitivity analysis has also changed. In the main analysis, we applied the study period criteria (study period start: 2017 for NAJS) when creating the cohort and when estimating prevalence. For the sensitivity analysis, we applied the study period restriction only in the calculation of prevalence. As a result, cases identified before 2017 were considered prevalent for 5 years, even if diagnosed before the study start date (e.g., a patient diagnosed in 2015 in NAJS was considered prevalent up to 2020, provided that the patient was not lost of follow-up or died before this time window elapsed).	Cancer diagnoses in NAJS originating from sources other than the cancer registry can include “working diagnoses”, which might involve diagnoses that are not confirmed through histologic or radiologic diagnostics. The restriction of cases originating from the cancer registry was incorporated to reduce potential outcome misclassification. The use of data prior to 2017 to derive outcome cohorts was decided in light of the results of the main analysis, where we observed a peak in prevalence estimates in data sources with data availability starting after 2010. This was not possible to explore in InGef RDB, as data availability started on 2015. However, NAJS had data from other sources prior to 2017 (primary care data was available from 2015 onwards).

CRN=Cancer Registry Norway, DK-DHR=Danish Data Health Registries, DNCR=Danish National Cancer Registry, DNPR=Danish National Patient Registry, InGef RDB=InGef Research Database, NAJS= Croatian National Public Health Information System, NCR=Netherlands Cancer Registry.

9. RESULTS

The full set of results for this study is available through an interactive web-application ShinyApp at: EUPAS1000000715.

The main results derive from the analysis conducted as originally planned in the protocol, which utilised various health registries contained within NAJS and DK-DHR. Results for DK-DHR and NAJS should be interpreted in light of the potential inclusion of unconfirmed diagnoses (see "**8.10. Deviations from the protocol**"). Results from the sensitivity analyses restricting cases to those originating from cancer registries (which were available for both data sources) are presented for comparison in "**9.5. Other analyses**".

9.1. Participants

9.1.1. Denominator populations

For data sources with internal denominators, a total of 22,801,202 study participants were included in the denominator for prevalence calculations during the entire study period (n=4,721,070 in NAJS; n=7,567,850 in DK-DHR; n=10,512,282 in InGef RDB). The attrition process to identify these denominators, and the number of individuals excluded in each step, is detailed in **Annex IV** (**Table S1**).

For data sources where denominators had to be derived from external data, estimates for the entire study period were not available. However, we have information on the study population by calendar year. Based on the last year of available data in each data source, a total of 41,208,774 participants were included in the denominators for the primary aim of the study, 17,847,266 of which were included data sources with population-level denominators, and 23,361,508 from external sources with aggregated data (**Table 3**).

Counts of participants over time are shown in **Annex IV**, along with the comparison between internal and external denominators for DK-DHR and NAJS (**Figure S1** and **Figure S2**). In most cases, the number of participants remained similar or increased over the years. In DK-DHR, denominators derived from internal and external sources aligned, with some minor differences. While the number of subjects included in the denominator for 2010 was similar in NAJS, the number of subjects in the denominator increased when using internal denominators but decreased when using national statistics data, with a difference of approximately 400,000 subjects (**Table 3**). Reasons behind this discrepancy are described in "**10.2. Strengths and limitations"**.

Table 3. Participants included in internal and external denominators used for prevalence estimation.

Data source name	Denominator source[1]	Number of subjects		
		Entire study period[2]	2010 or first year with available data[3]	Last year with available data[3]
Primary analysis				
NAJS	Internal	4,721,070	4,041,486	4,260,887
DK-DHR	Internal	7,567,850	5,519,435	5,919,391
InGef RDB	Internal	10,512,282	7,224,937	7,666,988
NCR	External	N/A	16,574,989	17,811,291
CRN	External	N/A	4,858,199	5,550,217
Secondary analysis				
NAJS	External	N/A	4,078,769	3,861,967
DK-DHR	External	N/A	5,534,738	5,932,654

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; N/A= Not applicable; NCR=Netherlands Cancer Registry.
[1] External denominators were obtained from publicly available demographic data from Eurostat. External denominators were used for the primary analysis in NCR and CRN, as they only capture information on cancer patients.
[2] Only available for data sources with population-level denominators. The study period spanned from 2010 to 2023.
[3] First year of available data: 2010 for all data sources except from NAJS (2017) and InGef RDB (2015). Last year of available data: 2023 for DK-DHR and NCR, 2024 for NAJS, InGef RDB and CRN.

9.1.2. Outcome cohorts

Outcome cohorts were identified for each cancer type and data source. **Figure 6** provides an example of the flowchart illustrating the outcome cohort population for ALL in DK-DHR. For all cancer types, we defined an outcome duration of 5 years (i.e., 1,825 days) and ensured that cases occurred within the study period. We allowed cases that began within 5 years of the study period start to be considered prevalent from the study start onwards (e.g., a case was diagnosed in 2009 was considered prevalent from 2010 to 2014, as long as the patient remained under observation during this period).

The number of subjects identified in each outcome cohort is described in "**9.2 Descriptive data**". Further details on cohort attrition for each cancer type and data source are provided in **Table S10**, with all flowcharts available in the ShinyApp.

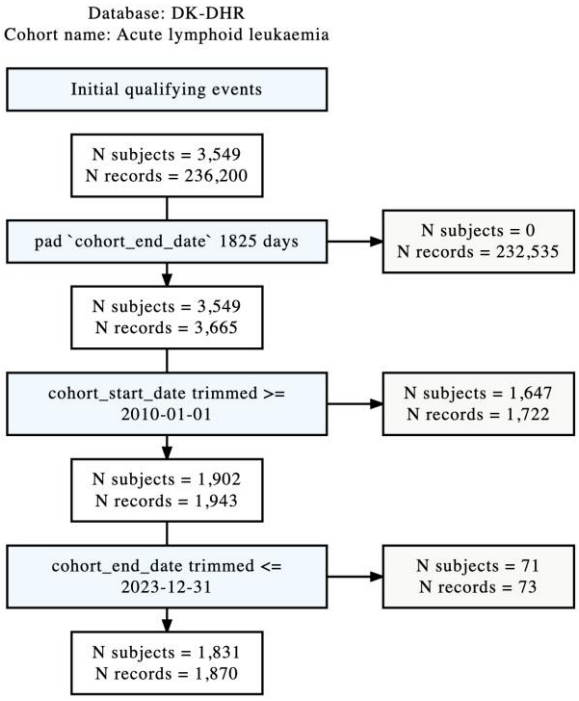

Figure 6. Flowchart depicting individuals with acute lymphoid leukaemia in DK-DHR.

DK-DHR=Danish Data Health Registries

9.2. Descriptive data

We identified a total of 188,935 individuals with rare blood cancers, 99,446 with pancreatic cancer, and 53,621 with soft tissue sarcoma. **Table 4** includes the number of subjects identified by cancer type in each data source and their main demographic characteristics. Additional demographic characteristics can be found in the ShinyApp.

In general, the proportion of males was slightly higher than females across all cancer types studied. Median age at diagnosis ranged from approximately 65 to 74 across rare blood cancers, with the exception of ALL, where it ranged from 13 to 19 years in DK-DHR, InGef RDB, NCR, and CRN. In NAJS, median age at ALL diagnosis was estimated at 49 [Q25–Q75 13–70] years (see "**9.5. Other analyses**"). For pancreatic cancer, the median age at diagnosis was 70 to 73 years, and for soft tissue sarcoma, 61 to 66 years.

The amount of data visibility after the first diagnosis varied across cancer types. The median future observation time exceeded 700 days, with many cancer types surpassing 1,500 days. For AML and pancreatic cancer, the median future observation time was substantially lower, ranging from 259 to 348 days for AML and from 132 to 206 days for pancreatic cancer across data sources. Prior and future observation time for each cancer cohort and data source can be explored in the ShinyApp.

Table 4. Number of subjects included, and main demographic characteristics.

Variable name	Estimate name	Outcome cohort: NAJS 2010–2024	Data source name & Study period: DK-DHR 2010–2023	InGef RDB 2015–2024	NCR 2010–2023	CRN 2010–2024
Acute lymphoid leukaemia						
Subjects	N	1,608	1,831	2,446	4,074	1,383
Males	N (%)	911 (56.7%)	1,073 (58.6%)	1,430 (58.5%)	2,354 (57.8%)	766 (55.4%)

Variable name	Estimate name	Outcome cohort / Data source name & Study period: NAJS 2010–2024	DK-DHR 2010–2023	InGef RDB 2015–2024	NCR 2010–2023	CRN 2010–2024
Age at diagnosis	Median [Q25–Q75]	49 [13–70]	19 [6–56]	19 [9–53]	15 [5–46]	13 [5–40]
Acute myeloid leukaemia						
Subjects	N	3,389	5,371	5,691	11,109	3,089
Males	N (%)	1,859 (54.9%)	3,000 (55.9%)	3,287 (57.8%)	6,226 (56.0%)	1,707 (55.3%)
Age at diagnosis	Median [Q25–Q75]	68 [57–77]	70 [59–78]	67 [54–77]	68 [55–77]	69 [54–78]
Chronic lymphocytic leukaemia						
Subjects	N	6,344	10,570	6,255	19,759	6,342
Males	N (%)	3,487 (55.0%)	6,357 (60.1%)	4,062 (64.9%)	12,276 (62.1%)	3,843 (60.6%)
Age at diagnosis	Median [Q25–Q75]	71 [62–79]	71 [64–79]	74 [63–81]	70 [62–77]	70 [63–78]
Diffuse large B-cell lymphoma						
Subjects	N	4,147	10,172	7,053	23,920	7,015
Males	N (%)	2,072 (50.0%)	5,798 (57.0%)	4,150 (58.8%)	13,484 (56.4%)	3,947 (56.3%)
Age at diagnosis	Median [Q25–Q75]	67 [57–75]	70 [60–77]	68 [57–78]	69 [59–77]	70 [59–78]
Follicular lymphoma						
Subjects	N	5,185	7,235	3,909	11,783	4,221
Males	N (%)	2,486 (48.0%)	3,667 (50.7%)	2,071 (53.0%)	6,146 (52.2%)	2,112 (50.0%)
Age at diagnosis	Median [Q25–Q75]	65 [55–73]	68 [59–76]	65 [55–74]	65 [56–73]	66 [58–74]
Multiple myeloma						
Subjects	N	6,444	9,525	8,064	22,408	8,584
Males	N (%)	3,157 (49.0%)	5,373 (56.4%)	4,785 (59.3%)	13,066 (58.3%)	4,911 (57.2%)
Age at diagnosis	Median [Q25–Q75]	70 [62–78]	71 [63–78]	69 [60–78]	70 [62–77]	72 [63 –79]
Pancreatic cancer						
Subjects	N	10,883	18,082	17,124	39,813	13,544
Males	N (%)	5,522 (50.7%)	9,351 (51.7%)	9,491 (55.4%)	20,192 (50.7%)	6,904 (51.0%)
Age at diagnosis	Median [Q25–Q75]	70 [63–79]	71 [64–78]	70 [61–79]	71 [63–78]	73 [64–80]
Soft tissue sarcoma						
Number subjects	N	6,326	9,728	7,131	23,337	7,099
Males	N (%)	3,460 (54.7%)	4,986 (51.3%)	3,945 (55.3%)	12,507 (53.6%)	3,598 (50.7%)
Age at diagnosis	Median [Q25–Q75]	66 [55–76]	64 [50–74]	61 [47–74]	63 [49–74]	63 [47–75]

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.

9.3. Main results

Table 5 presents the estimated prevalence of the study outcomes for the primary aim of the study (5-year partial point prevalence with no prior history requirement) in 2023, the most recent calendar year for which data was available across all included data sources. Estimates per 10,000 in 2023 ranged from 0.51 to 1.00 for ALL, from 0.71 to 1.50 for AML, from 2.50 to 4.90 for CLL, from 2.66 to 3.80 for DLBCL, from 1.80 to 3.70 for follicular lymphoma, and from 2.80 to 4.60 for multiple myeloma. Lastly, estimates for pancreatic cancer ranged from 1.72 to 3.80, and for soft tissue sarcoma from 2.83 to 4.30.

The highest prevalence was observed for NAJS across all selected cancers, except for DLBCL, where the highest estimates were observed for DK-DHR. DK-DHR ranked second for most outcomes studied, except for ALL and AML, where results for InGef RDB aligned with those obtained for NAJS. In general, estimates obtained for NCR and CRN were consistent across outcomes studied and obtained the lowest figures. Differences between NCR and CRN were limited to multiple myeloma and pancreatic cancer and were driven by differing temporal trends, as described later in this report (see **Figure 7**). Results for InGef RDB in 2023 were similar to those obtained for NCR and CRN, except for ALL, AML, and pancreatic cancer, where estimates for InGef RDB were higher (**Table 5**).

Table 5. Estimated 5-year partial point prevalence of selected cancers in 2023.

Data source name	Estimate name		
	Denominator (N)[1]	Outcome (N)	Prevalence per 10,000 [95% CI]
Acute lymphoid leukaemia			
NAJS	4,249,572	421	1.00 (0.90–1.10)
DK-DHR	5,919,391	392	0.70 (0.60–0.70)
InGef RDB	7,641,957	732	1.00 (0.90–1.00)
NCR	17,811,291	907	0.51 (0.48 –0.54)
CRN	5,488,984	334	0.61 (0.54–0.68)
Acute myeloid leukaemia			
NAJS	4,249,572	644	1.50 (1.40–1.60)
DK-DHR	5,919,391	583	1.00 (0.90–1.10)
InGef RDB	7,641,957	992	1.30 (1.20–1.40)
NCR	17,811,291	1,389	0.78 (0.74–0.82)
CRN	5,488,984	391	0.71 (0.64–0.78)
Chronic lymphocytic leukaemia			
NAJS	4,249,572	2,071	4.90 (4.70–5.10)
DK-DHR	5,919,391	2,705	4.60 (4.40–4.70)
InGef RDB	7,641,957	1,928	2.50 (2.40–2.60)
NCR	17,811,291	4,875	2.74 (2.66–2.81)
CRN	5,488,984	1,629	2.97 (2.83–3.11)
Diffuse large B-cell lymphoma			
NAJS	4,249,572	1,292	3.00 (2.90–3.20)
DK-DHR	5,919,391	2,222	3.80 (3.60–3.90)
InGef RDB	7,641,957	2,061	2.70 (2.60–2.80)
NCR	17,811,291	5,153	2.89 (2.81–2.97)
CRN	5,488,984	1,461	2.66 (2.53 –2.80)
Follicular lymphoma			
NAJS	4,249,572	1,582	3.70 (3.50–3.90)
DK-DHR	5,919,391	2,012	3.40 (3.30–3.60)
InGef RDB	7,641,957	1,398	1.80 (1.70–1.90)
NCR	17,811,291	3,245	1.82 (1.76–1.89)
CRN	5,488,984	1,125	2.05 (1.93–2.17)
Multiple myeloma			
NAJS	4,249,572	1,964	4.60 (4.40–4.80)
DK-DHR	5,919,391	2,561	4.30 (4.20–4.50)
InGef RDB	7,641,957	2,149	2.80 (2.70–2.90)

Data source name	Estimate name		
	Denominator (N)[1]	Outcome (N)	Prevalence per 10,000 [95% CI]
NCR	17,811,291	5,125	2.88 (2.80–2.96)
CRN	5,488,984	2,018	3.68 (3.52–3.84)
Pancreatic cancer			
NAJS	4,249,572	1,600	3.80 (3.60–4.00)
DK-DHR	5,919,391	1,654	2.80 (2.70–2.90)
InGef RDB	7,641,957	2,360	3.10 (3.00–3.20)
NCR	17,811,291	3,071	1.72 (1.66–1.79)
CRN	5,488,984	1,170	2.13 (2.01–2.26)
Soft-tissue sarcoma			
NAJS	4,249,572	1,843	4.30 (4.10–4.50)
DK-DHR	5,919,391	2,441	4.10 (4.00–4.30)
InGef RDB	7,641,957	2,215	2.90 (2.80–3.00)
NCR	17,811,291	5,343	3.00 (2.92–3.08)
CRN	5,488,984	1,553	2.83 (2.69–2.97)

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.
[1]External denominators were used for the primary analysis in NCR and CRN, as they only capture information on cancer patients. External denominators were obtained from publicly available demographic data from Eurostat.

In general, results were higher in males across cancer types and data sources. Sex differences were particularly pronounced for CLL, DLBCL, and multiple myeloma (**Figure 7**). For these cancer types, males had higher estimates than females in all data sources, except for NAJS, where sex-differences were subtle and were only pronounced for CLL. The highest estimates were observed in the 60–69, 70–79, and 80–89 age groups and decreased as age decreased. This pattern was seen for all outcomes, except for ALL, where the pattern was reversed obtaining the highest estimates for 0–9 and 10–19 years and declined with age (**Figure 8**).

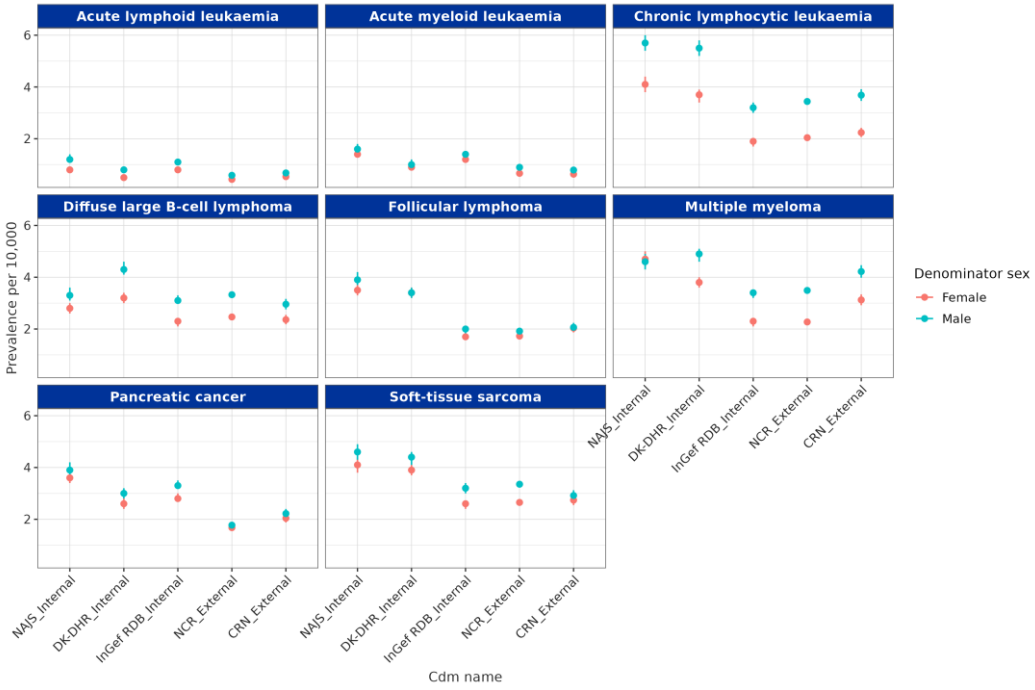

Figure 7. Estimated 5-year partial point prevalence of selected cancers in 2023 by sex.

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.

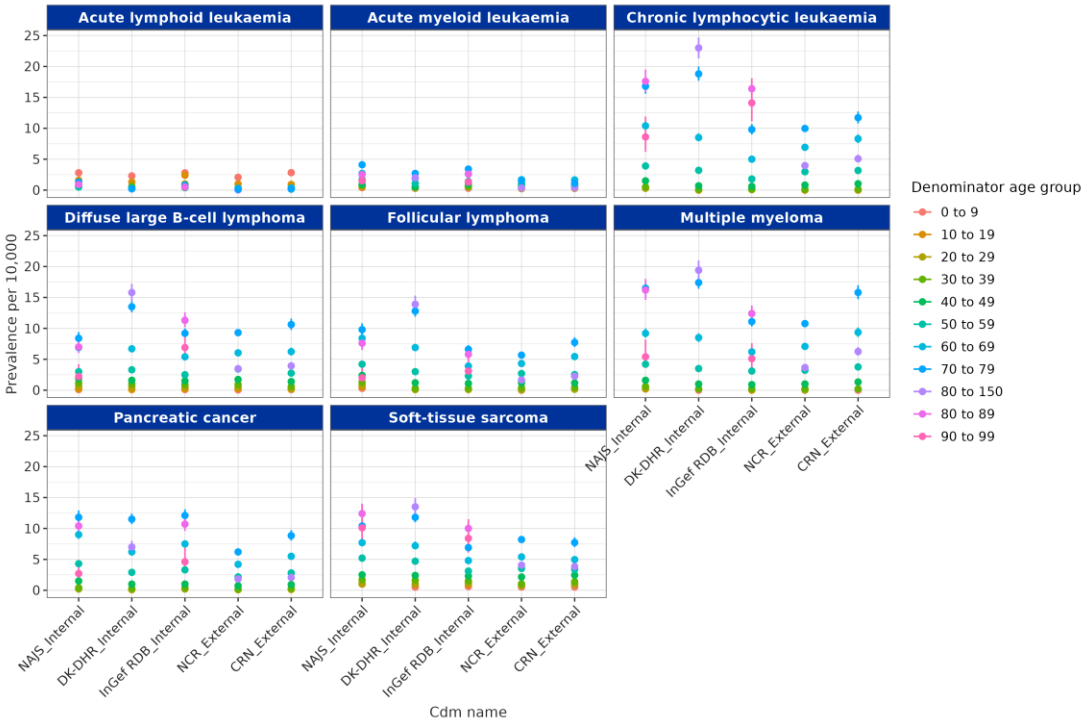

Figure 8. Estimated 5-year partial point prevalence of selected cancers in 2023 by age group.

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.

Prevalence over time showed similar figures for both NCR and CRN, with trends remaining stable throughout the period for these data sources (**Figure 9**). Trends showed a modest upward trend in CRN for multiple myeloma and, to a lesser extent, pancreatic cancer, which resulted in estimates being slightly higher in CRN compared to NCR from 2020 onwards. Results also showed an upward trend in DK-DHR across selected cancer types, except for ALL and AML, where trends were stable throughout the study period. Importantly, we observed a particular trend in NAJS (which began contributing data in 2017) and InGef RDB (which began in 2015). For these data sources, we observed an increasing trend during the first points of data availability, reaching a peak in 2020 in most cases (see "**10.2 Strengths and limitations**"). From 2021 onwards, results for InGef RDB were comparable to those for NCR and CRN across all cancer types, except for pancreatic cancer, and to a lesser extent ALL and AML, for which estimates remained higher (**Figure 9**).

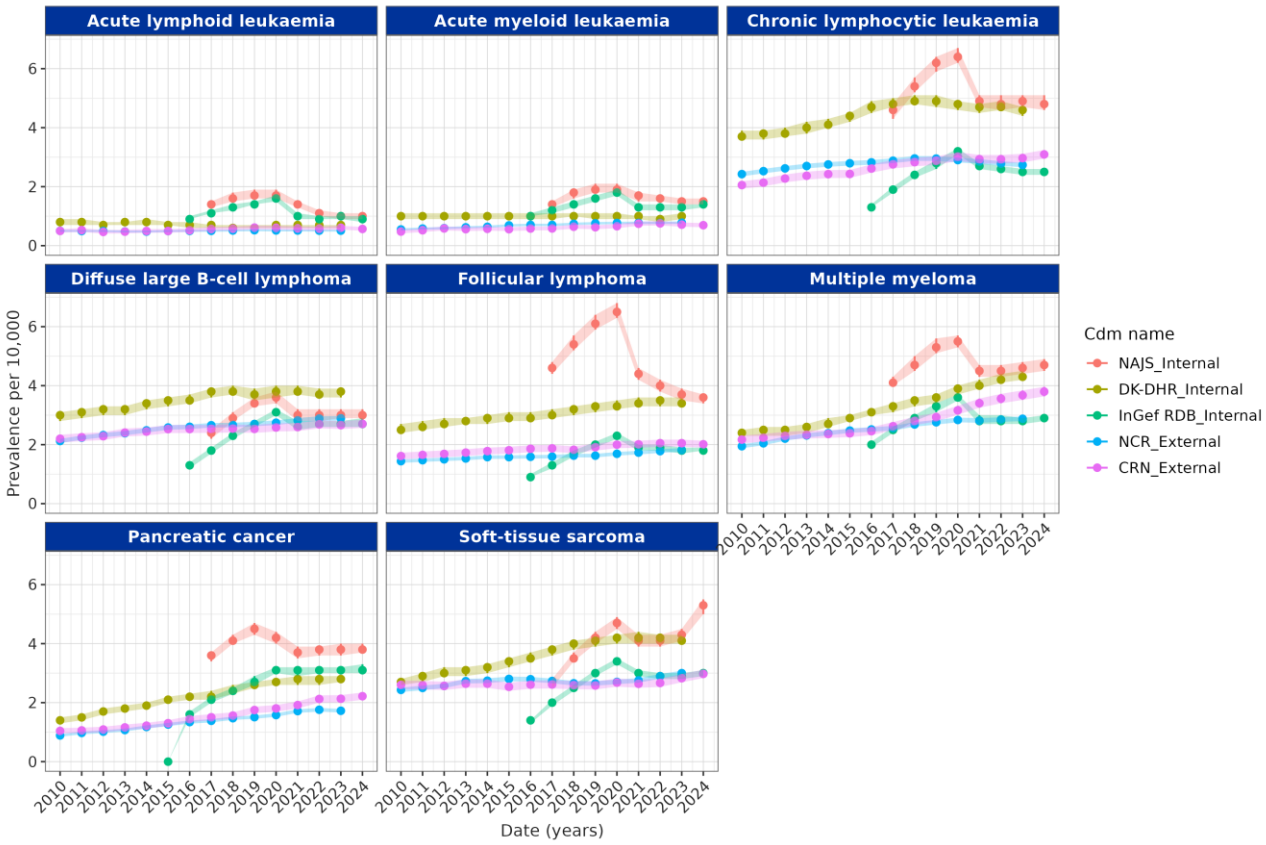

Figure 9. Estimated 5-year partial point prevalence per 10,000 of selected cancers by calendar year.

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.

9.4. Denominator comparison

The secondary aim of this study was to compare the prevalence of selected cancers included in the primary aim of the study, using denominator data available within the data source versus denominator data obtained from external sources. This analysis was performed in NAJS and DK-DHR only.

Estimated prevalence using internal and external denominators showed consistent results for both data sources (**Figure 10**). For DK-DHR, results were closely aligned, while for NAJS, estimates were slightly higher when using external denominators. The number of outcomes in both analyses (internal vs. external denominators) was consistent in both data sources. However, the number of individuals included in the denominator population differed in NAJS, exceeding the Eurostat-reported population by 400,000

individuals as of 2023 (see "**9.4 Denominators**"). Discrepancies in the number of individuals included did not affect all calendar years equally, with the differences becoming more pronounced from 2019 onwards (**Figure S2**), aligning with the time points where prevalence showed more divergence (**Figure 10**).

Results were consistent by sex and age groups. Results by sex and age groups are presented in **Annex IV** (**Figure S3** to **Figure S6**). While the number of outcomes in both analyses (internal vs. external denominators) was consistent overall and when stratified by sex, some variations in outcome counts were observed when stratifying by age groups (see "**10.2 Strengths and limitations**"). **Figure S7** and **Figure S8** exemplify this for follicular lymphoma, with results for all outcomes and age groups studied available in the ShinyApp.

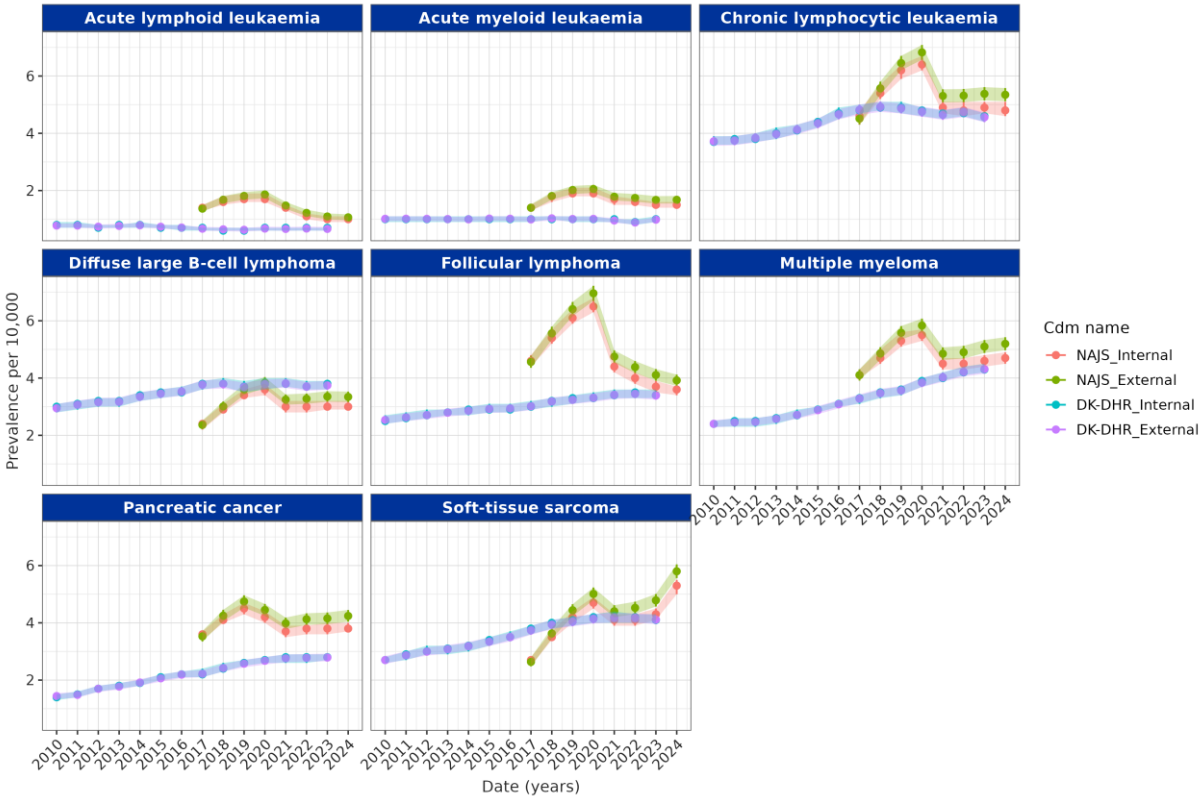

Figure 10. Estimated 5-year partial point prevalence per 10,000 using internal and external denominators in DK-DHR and NAJS.

DK-DHR=Danish Data Health Registries; NAJS=Croatian National Public Health Information System.

9.5. Other analyses

9.5.1. Prior history requirement

We conducted a sensitivity analysis by incorporating a 365-day prior history requirement, which was applied to data sources with population-level denominators (i.e., NAJS, DK-DHR, and InGef RDB). Results from this sensitivity analysis aligned with those observed in the primary aim of the study. Findings remained aligned throughout the study period and were consistent overall, and within sex and age groups. **Figure 10** illustrates overall results over time, with results by sex and age groups available in the ShinyApp.

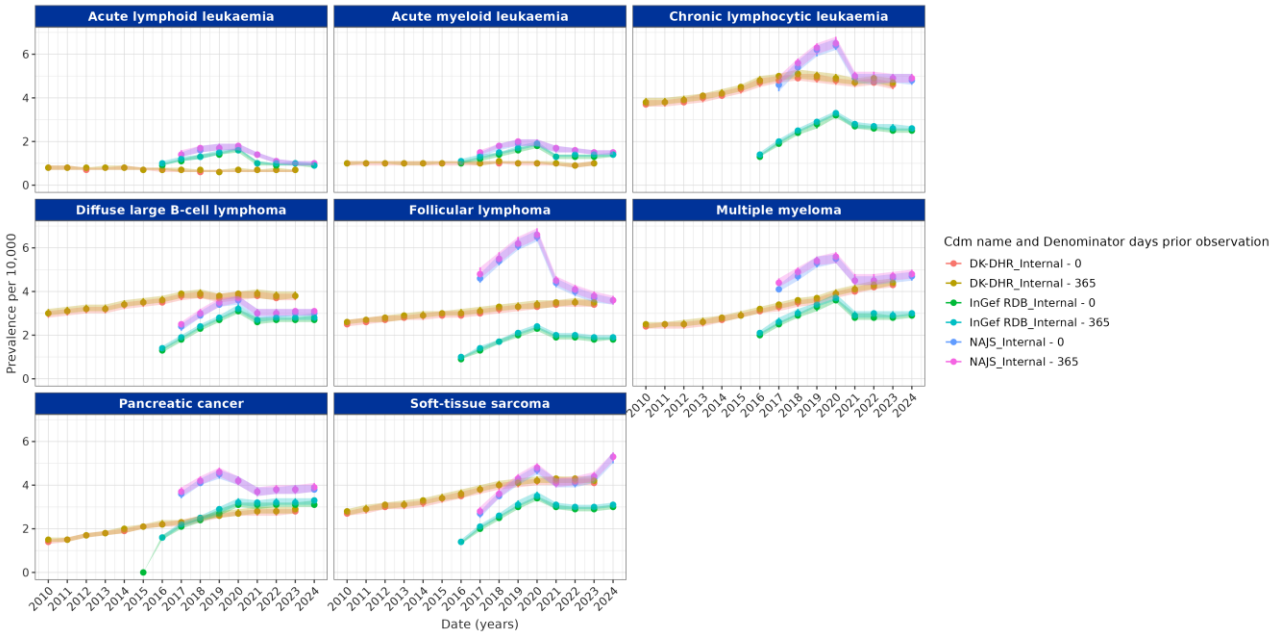

Figure 11. Estimated 5-year partial point prevalence per 10,000 using 0- vs. 365-day prior history requirement in data sources with population-level denominators.

DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System.

9.5.2. Restriction to cancer cases recorded in national cancer registries

In a second sensitivity analysis, we restricted cancer cases to those originating from the cancer registry in DK-DHR and NAJS. In DK-DHR, results from the sensitivity analysis showed lower 5-year partial point prevalence estimates for all cancer types in the sensitivity analysis (DK-DHR, DNCR only) compared to the main analysis (DK-DHR, use of DNPR and DNCR) (**Figure 12**). The highest point prevalence reductions were observed for DLBCL and pancreatic cancer, in which the sensitivity analysis captured approximately 70% and 78% of the prevalent cases included in the main analysis in 2022. The proportion of cases 0captured in the sensitivity analysis compared to the main analysis in 2022 was over 82% for the rest of cancers studied (**Table S11** in **Annex IV**).

For NAJS, results showed a greater reduction in the number of cases captured and, consequently, in the prevalence estimates obtained. Based on estimates for 2022, the proportion of cases captured in the sensitivity analysis (compared to the main) were approximately 60% for AML and DLBCL. For other cancer types, these reductions were more pronounced, with the proportion of cases captured being approximately 45–50%. The highest reductions were observed for follicular lymphoma and soft tissue sarcoma, with only 27% of cases captured, and CLL, with 37% (**Table S12** in **Annex IV**). In addition, the increasing trend observed during the first years of available data, reaching a peak in 2020 in the main analysis, was not reflected in the sensitivity analysis (**Figure 12**). In the sensitivity analysis, cases diagnosed before 2017 were not excluded and could contribute as a prevalent case during the first years of the study, as cases were considered prevalent up to 5 years after diagnosis. An example of a flowchart depicting cohort creation in both main and sensitivity analyses is illustrated in **Figure S9**, with additional details in "**8.10. Deviations from the protocol**" and "**10.2. Strengths and limitations"**. Attrition for all cohorts and demographics are available in the Shiny App. For ALL, median age at diagnosis changed from 49 [Q25–Q75 13–70] years in the main analysis to 22 [5–62] in the sensitivity analysis, closer to that obtained for other data sources, where it ranged from 13 to 19 years.

When comparing sensitivity results to those obtained from other data sources, results from this sensitivity analysis restricting to cancer registry cases in DK-DHR were generally consistent with those from NCR and CRN (**Figure 13**). While prevalence estimates in DK-DHR decreased when restricting to cancer registry cases, results obtained for CLL and follicular lymphoma remained higher than those observed in other data sources. Estimates were comparable with those from NCR and CRN for multiple myeloma (before 2019) and soft tissue sarcoma (before 2017), while an upward trend was observed during the last points of available data. This trend was also observed in CRN for multiple myeloma. Regarding NAJS, results were consistent for ALL, AML, and pancreatic cancer when compared to those from other data sources, except for InGef RDB, where estimates were higher. However, estimates for the other cancers studied were notably lower than those observed in other data sources (**Figure 13**).

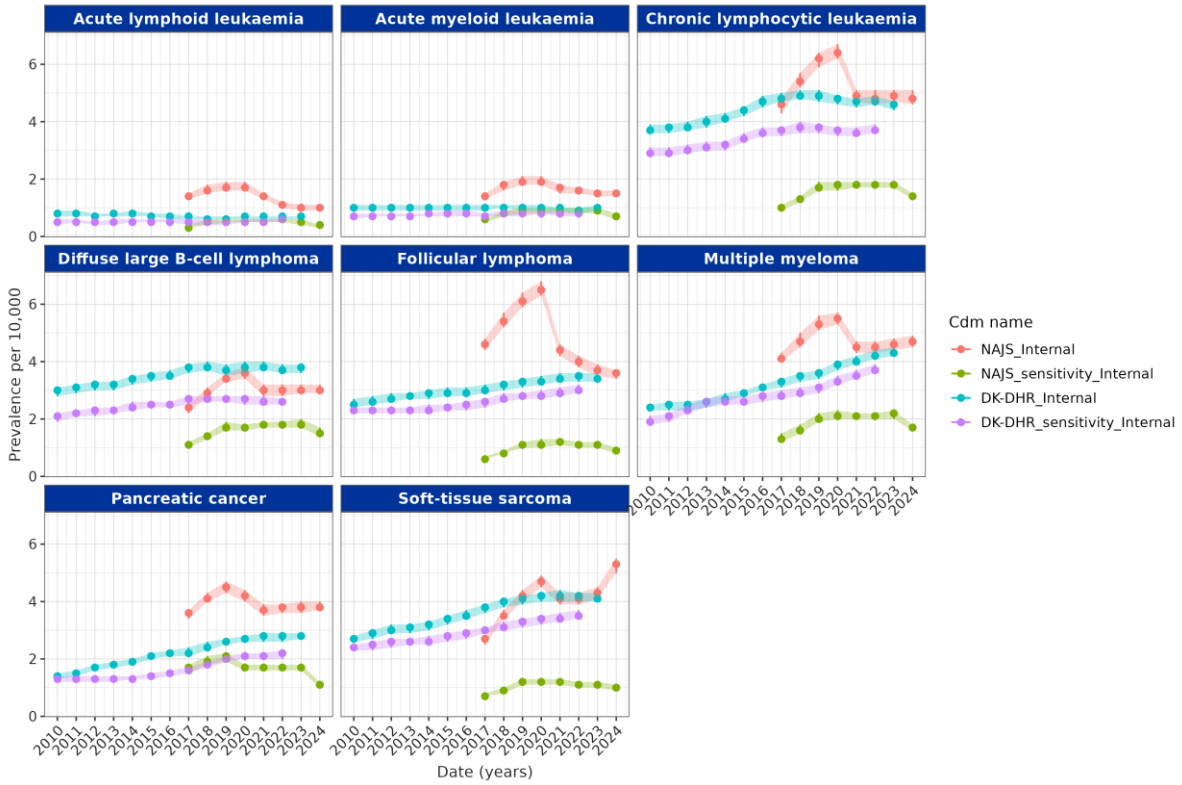

Figure 12. Estimated 5-year partial point prevalence per 10,000 considering all recorded cases vs. cases recorded from national cancer registries in NAJS and DK-DHR.

DK-DHR=Danish Data Health Registries; NAJS=Croatian National Public Health Information System.

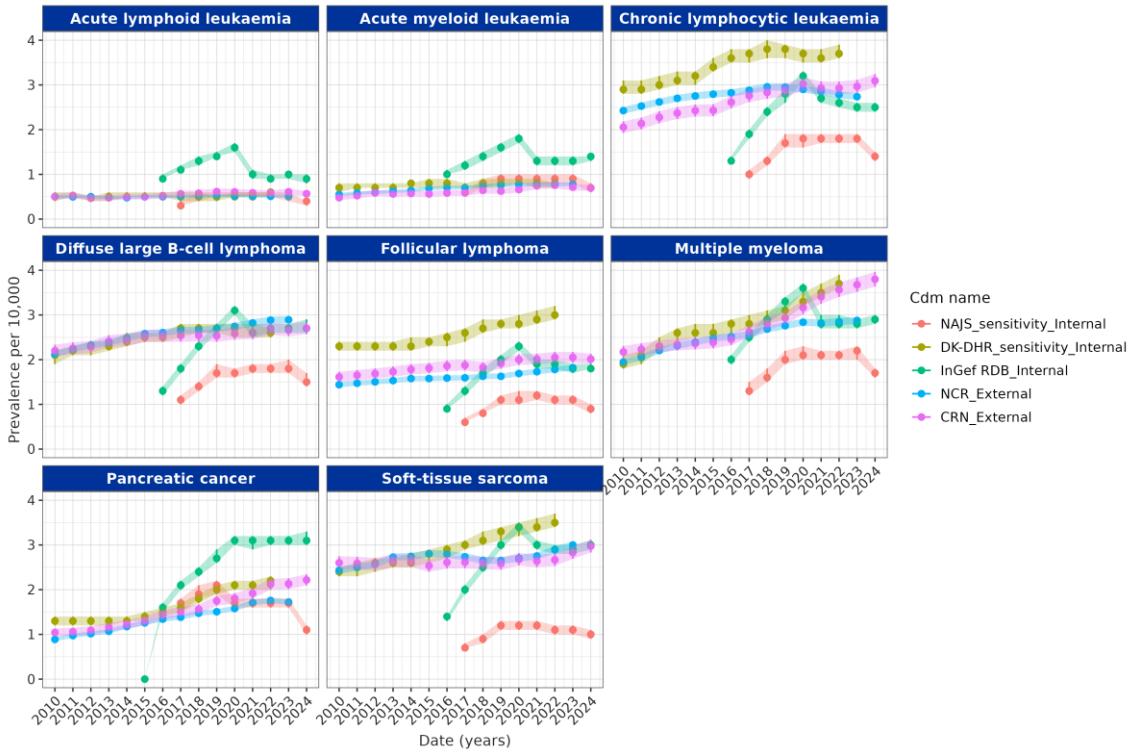

Figure 13. Estimated 5-year partial point prevalence per 10,000 of selected cancers for all data sources, considering cases recorded from national cancer registries in NAJS and DK-DHR.

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.

10. DISCUSSION

10.1. Key results

In this study, which was built upon a previous DARWIN EU® study (EUPAS50800), we estimated point prevalence of rare blood cancers, pancreatic cancer, and soft tissue sarcoma. Estimates were derived from 5 data sources from 5 European countries, three of which had population-level denominators (Croatia, NAJS; Denmark, DK-DHR; Germany, InGef RDB) and two of which were cancer registries (the Netherlands, NCR; Norway, CRN), for which denominators had to be obtained from external sources. In a secondary aim, conducted in NAJS and DK-DHR only, we compared prevalence estimates obtained from population-level denominators with estimates obtained with denominators from external sources.

Prevalence (estimated as 5-year partial point prevalence with no prior history requirement) per 10,000 in 2023 ranged from 0.51 to 1.00 for ALL, from 0.71 to 1.50 for AML, from 2.50 to 4.90 for CLL, from 2.66 to 3.80 for DLBCL, from 1.80 to 3.70 for follicular lymphoma, and from 2.80 to 4.60 for multiple myeloma. Lastly, estimates for pancreatic cancer ranged from 1.72 to 3.80, and for soft tissue sarcoma from 2.83 to 4.30. In general, prevalence estimates were similar by sex or slightly higher in males compared to females. Sex differences were particularly pronounced for CLL, DLBCL, and multiple myeloma. The highest estimates were observed in the 60–69, 70–79, and 80–89 age groups and decreased as age decreased, except for ALL, where the pattern was reversed obtaining the highest estimates among age groups of 0–9 and 10–19 years. Results were consistent when applying a one-year prior history requirement in the data sources with population-level denominators (i.e., NAJS, DK-DHR, InGef RDB).

Trends in prevalence were stable throughout the study period, albeit an upward trend was observed in some cases. We observed an increase in prevalence during the first points of data availability in NAJS and InGef RDB, reaching a peak in 2020 in most cases, after which the trend generally stabilised (see "**10.2. Strengths and limitations"**). NAJS yielded the highest estimates across studied cancers, while NCR and CRN maintained consistently similar estimates throughout the study period. Results for InGef RDB were similar to those in NCR and CRN from 2020 onwards, except for ALL, AML, and pancreatic cancer, where estimates in InGef RDB were higher. When restricting DK-DHR cases to those from the Danish National Cancer Registry (DNCR), results aligned with those observed in NCR and CRN for all cancer types, except for CLL and follicular lymphoma, where estimates in DK-DHR were still higher than those reported in other data sources. In NAJS, restricting cases to those originating from the Cancer Registry from the Croatian Institute of Public Health showed comparable results to those reported for other data sources (except for InGef RDB) for ALL, AML, and pancreatic cancer. This restriction led to marked reductions in the other cancer types assessed, with sensitivity results in NAJS presenting the lowest values across the data sources examined for CLL, DLBCL, follicular lymphoma, multiple myeloma, and soft tissue sarcoma.

Compared to the previous DARWIN EU® study (EUPAS50800), which estimated prevalence of rare blood cancers up to 2020, this study yielded broadly consistent results when restricting cases to those recorded in cancer registries in DK-DHR and NAJS (**Table S13**). Considering this restriction, the prevalence range obtained in the prior study was consistent with that observed in this study for follicular lymphoma (prevalence range per 10,000 in 2020 in previous study vs. current study: 0.90 to 2.83 vs. 1.10 to 2.80) and multiple myeloma (2.15 to 4.27 vs. 2.10 to 3.60). Results for ALL and AML were also consistent for all data sources, except for InGef RDB (prevalence range per 10,000 in 2020 in previous study vs. current study without InGef RDB: 0.44 to 0.65 vs. 0.50 to 0.61 for ALL; 0.72 to 1.03 vs. 0.66 to 0.90 for AML). Differences in prevalence estimates were observed for DLBCL, with estimates in this study exceeding those from the previous study (prevalence range per 10,000; previous study vs. current study considering restriction in DK-DHR and NAJS: 0.47 to 1.73 vs. 1.70 to 3.10).

Estimated prevalence using internal and external denominators showed consistent results in data sources where this comparison was examined (i.e., DK-DHR, NAJS). For DK-DHR, results were closely aligned, while for NAJS, estimates were slightly higher when using external denominators (see "**10.2 Strengths and limitations"**). Results were consistent when stratified by age and sex. Aside from expected variations in the number of participants included in the denominator depending on the source used to derive denominators (i.e., internal vs. external), we observed some differences in the number of outcomes observed when stratifying by age group, likely due to different strategies to handle time-varying covariates (see "**10.2 Strengths and limitations"**). These variations did not impact the results and were not observed in the overall analyses or in those stratified by sex.

10.2. Strengths and limitations of the research methods

The study was informed by routinely collected health care data and so, data quality issues must be considered. Although we expected relatively few false positives, not all cases identified in DK-DHR and NAJS in the main analysis were confirmed through histologic or radiologic diagnostics. This may have led to misclassification of the outcome in the main analysis, with individuals having unconfirmed diagnoses being classified as having the disease, which may overestimate prevalence. Therefore, the results of the main analysis for these data sources should be interpreted with caution. The sensitivity analysis strengthens our findings by limiting cases to those originating from national cancer registries only and draw attention to the impact of the potential outcome misclassification, which vary by outcome and seem to be affect NAJS disproportionally. While results of this analysis are more in line with those seen for other data sources, some findings differ and warrant further investigation. Importantly, two additional considerations related to the cancer registry in NAJS were identified upon reviewing obtained results. First, cancer cases recorded in the cancer registry represent only incident cases (the first time an individual is diagnosed with a specific cancer type), which may lead to an underestimation of prevalence. Second, there is an 18–24 month lag in

the reporting of cases to the registry,[1] which affects estimates for 2024 and, to a lesser extent, 2023. This reporting delay may partially explain the lower prevalence estimates observed in the sensitivity analysis for certain cancers in NAJS, as well as the decrease in prevalence observed in the most recent data (**Figure 12** and **Figure 13**).

Another more general limitation is around uncertainty in outcome definitions, in particular the duration over which a person was considered as a prevalent case following their diagnosis. Previous studies have demonstrated the substantial differences between partial and complete prevalence,[6] with the difference between these estimates in part driven by the severity of disease and typical age at diagnosis. Different strategies to define outcome duration were explored in the prior DARWIN EU® study on rare blood cancers and were found to have a substantial impact on the obtained results (EUPAS50800).

Additionally, the results estimated from this study only reflected the populations from the included data sources and cannot be generalised to other countries or healthcare settings. For InGef RDB, we focused on data from hospital settings (inpatient and outpatient). The impact of this was expected to be limited, as most cancers diagnoses were expected to be recorded in secondary care settings. In addition, InGef RDB contributed data from 2015 to 2024, as legal regulations only allow the use of data from the most recent 10 years. This implies that prevalent cases during the first years of available data were likely to be underestimated, as information on cancer diagnoses prior to 2015 was not available. For example, a patient diagnosed with cancer in 2013 should have been considered prevalent for five years according to our definition. However, since the diagnosis occurred before 2015, this case was not captured, and the individual was incorrectly considered disease-free. This likely caused the progressive observed increase in prevalence observed for InGef RDB during the first 5 years of available data, consistent with the duration of our outcome definition (i.e., 5 years) and estimates starting from 2020 may be more valid. This was also observed for NAJS in the main analysis, which contributed data from 2017 onwards (the start of data availability for hospital records, with primary care data available from 2015 onwards). Both data sources exhibited a peak in prevalence in 2020 (January 1st), followed by a decrease in 2021 and stabilisation thereafter. This peak was not observed in the sensitivity analysis in NAJS, where we restricted cases to those originating from the cancer registry and allowed cases diagnosed before 2017 to be considered prevalent for 5 years, even if diagnosed before the study start date. This aligned with the strategy used for other data sources in the study, where the study start was defined in 2010, albeit data availability started earlier. The extent to which this pattern is due to limited historical data should be further explored, as we cannot rule out changes in coding practices over time.

Regarding the results obtained using internal vs. external denominators, two important considerations should be noted. First, information on individuals who have migrated out of the country was not available in NAJS. As a result, the number of patients included in the data source is higher than the actual number of residents (i.e., 4.30 million active patients in current data release vs. 3.86 million individuals in 2024 based on Eurostat data). This led to an inflated denominator, potentially causing the prevalence to be underestimated when using internal denominators, which explains the modest discrepancies observed in NAJS in the secondary analysis. Second, strategies for handling time-varying covariates differed between the two approaches compared (**Figure 1**). When using internal denominators, age was calculated at the denominator entry, and individuals were able to move between age groups as they aged. For example, a participant who experienced the outcome at age 43 in November 2011 contributed as a case to the 0–44 age group in 2012 (January 1st), and to the 45–64 age group in the subsequent years, up to five years from the start of diagnosis, provided the individual remained under observation during this period. In contrast, when using external denominators, age was calculated at the date of cancer diagnosis, and individuals were assigned to age groups based on that date only (see "**8.2. Follow up**"). Consequently, they did not move between age groups as they aged during the five-year period in which the disease was considered prevalent, which likely explains the minor variations observed in outcome counts by age group between the two strategies. Future studies using external denominators could compute age of prevalent cases on an

annual basis, which would likely mitigate the age discrepancies observed. Such an approach might be particularly relevant in cases where partial prevalence is defined over a long period of time (e.g., 10-year partial prevalence) or when complete prevalence is calculated (i.e., once diagnosed, the individual is considered to always have the disease).

10.3. Interpretation

In this section, we compared our results with data reported in global cancer statistics from various sources and settings, as well as those from the previous DARWIN EU® study (EUPAS50800). The rationale for including multiple sources, including one from the United States (US), is to provide a broader context and additional estimates for comparison, as not all outcomes are covered in every source. However, US estimates should be compared with caution due to differences in population characteristics such as race and ethnicity, which plays a role in the susceptibility of developing some of the outcomes studied.[7, 8]

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) in the US estimated the 5-year partial prevalence per 10,000 to be 0.68 for ALL, 0.86 for AML, 2.65 for CLL, and 3.12 for multiple myeloma, as of January 1, 2022.[9] Figures per 10,000 were estimated 2.12 per 10,000 for pancreatic cancer, and 1.3 per 10,000 for soft tissue, including heart. In general, findings from NCR, CRN, and InGef RDB were in line with these findings, except for soft tissue sarcoma, where estimates obtained in this study were higher (prevalence range from 2.83 to 4.30 per 10,000 in the main analysis; from 1.10 to 3.50, when restricting cases to those from the cancer registry in DK-DHR and NAJS) than estimates from SEER. These differences are likely to be attributed to discrepancies in the phenotypes used to identify cases.

Based on the latest GLOBOCAN estimates produced for 2022 by the International Agency for Research on Cancer (IARC), estimates of 5-year partial prevalence for countries included in our study ranged from 2.45 to 3.54 per 10,000 for multiple myeloma and from 1.52 to 1.98 per 10,000 for pancreatic cancer.[10] Prevalence estimates for multiple myeloma obtained in this study were broadly consistent with those reported in GLOBOCAN. Prevalence per 10,000 for 2022 was estimated at 2.80 in Croatia (vs. 2.10 [95% CI 3.60–3.90] in NAJS restricting cases to those from the cancer registry), 3.54 in Denmark (vs. 3.70 [3.60–3.90] in DK-DHR restricting cases to those from the cancer registry), 2.45 in Germany (vs. 2.80 [2.70–3.00] in InGef RDB), 2.97 in the Netherlands (vs. 2.85 [2.77–2.93] in NCR), and 3.18 in Norway (vs. 3.56 [3.40–3.72] in CRN). Prevalence estimates for pancreatic cancer obtained in this study for 2022 aligned with those reported in GLOBOCAN in Croatia (1.63 vs. 1.70 [1.60–1.80] in NAJS in the sensitivity analysis) and the Netherlands (1.77 vs. 1.76 [1.70–1.82]), albeit the number of cases appears to be higher in our study than those reported by the NCR in their source data (cases of pancreatic cancer in 2023 in the current study vs. those identified by NCR: 3,071 vs. 2,303). For other countries, our study yielded higher results than those observed in GLOBOCAN.[11]

Estimates for other specific cancers were not available in the SEER Program or GLOBOCAN. For Non-Hodgkin Lymphoma, of which DLBCL and follicular lymphoma are the most common subtypes, 5-year partial prevalence was estimated as 8.2 per 10,000 according to SEER results in the US [9] and ranged from 7.68 to 10.29 in countries included in the current study according to GLOBOCAN.[10] In our study, prevalence per 10,000 in 2022 ranged from 2.69 to 3.70 for DLBCL and 1.78 to 4.00 for follicular lymphoma. When restricting cases to those from the cancer registry in DK-DHR and NAJS, estimates ranged from 1.80 to 2.88 and 1.78 to 3.00, respectively.

The 5-year prevalence per 10,000 in Denmark as of 2023 were 0.49 for ALL, 0.73 for AML, 3.5 for CLL, 4.0 for multiple myeloma, and 2.29 for pancreatic cancer according to NORDCAN, which includes detailed cancer statistics from national cancer registries in the Nordic countries.[12] When using data from DK-DHR, our study produced higher estimates than those reported in NORDCAN in the main analysis. However, estimates were consistent when restricting cases to those originating from the cancer registry (i.e., DNCR) (prevalence per 10,000 for 2022: 0.60 [95%CI 0.50–0.60] for ALL; 0.80 [0.70–0.80] for AML; 3.70 [3.60–3.90]

for CLL; 3.70 [3.60–3.90] for multiple myeloma; and 2.20 [2.20–2.30] for pancreatic cancer). Discrepancies between both analysis in DK-DHR (based on cases from DNPR and DNCR, or DNCR only) may point to a potential misclassification of cases in DNPR, affecting 12–22% of cases in the cancer types assessed in the current study. In NAJS, restricting cases to those originating from the cancer registry led to greater reduction of cases, dropping approximately 40–60% of cases of cancer types included in the main analysis as of 2022, and up to 73% of cases for follicular lymphoma and soft tissue sarcoma. This discrepancy is likely to be explained by coding practices in NAJS, where "working diagnoses" are recorded as conditions even if not confirmed, which could result in misclassification and could explain the higher estimates observed in the main analysis. However, it should also be noted that the cancer registry in NAJS only includes incident diagnoses, which might also contribute to the differences observed when restricting cases to those originating from the cancer registry (see "**10.2. Strengths and limitations**"). The discrepancies observed between both analysis (main vs. sensitivity) in DK-DHR and NAJS also reflect the different characteristics of the registries that conform these data sources, with some registries (like DNPR in Denmark) providing information on procedures, treatments, and other patient contacts for cancer patients, and cancer registries representing the gold standard for identifying cancers, with detailed information on tumour characteristics and histopathological factors. This also highlights the potential for complementarity between cancer registries and other data sources, which has also been described for InGef RDB elsewhere.[13]

Estimates from the previous DARWIN EU® study (EUPAS50800), which were primarily derived from primary care data, broadly aligned with those obtained in the current study for NCR, CRN, and when restricting cases to those recorded in cancer registries, for DK-DHR and NAJS (see "**10.1. Key results**"). The main differences in prevalence estimates were observed for DLBCL, with estimates in this study exceeding those from the prior study. Differences in outcome definitions, health settings covered, types of data, and granularity in the source coding system are likely to have contributed to the discrepancies observed.

All data sources included in this study consisted of or included data from cancer registries, except for InGef RDB (claims data). A previous study comparing data from InGef RDB and the Cancer Registry of Rhineland-Palatinate showed good agreement between the two sources in age at diagnosis, tumour location, and treatments for patients with breast, prostate, or lung cancer in the same area.[13] The extent to which this agreement can be applied to other cancer types others than those included in that study is unclear. In the current study, estimates from InGef RDB (from 2020 onwards) were largely consistent with estimates obtained in other data sources, except for pancreatic cancer, and to a lesser extent ALL and AML, for which estimates were higher. The early increase observed in trends observed during the first years of available data was also observed in NAJS in the main analysis and is likely caused to the lack of prior data availability, although we cannot exclude the possibility of changes in coding practices over time (see "**10.2. Strengths and limitations**"). Prior evidence has highlighted key design choices for estimating prevalence using real-world data that can lead to substantial variations, with many of them stemming from assumptions of observable person-time and lookback time to identify for prevalent cases.[14] As illustrated for InGef RDB and NAJS (main analysis), the lack of prior data availability can likely introduce artefacts in prevalence trends during the first years of available data. Future studies assessing partial prevalence should consider delaying the start of the study period to ensure sufficient time to capture prevalent cases based on the amount of time defined for partial prevalence estimation.

Importantly, estimates obtained using denominators from external sources were consistent with those derived from the denominator in the data source in DK-DHR. In NAJS, estimates were marginally higher when external denominators were applied, likely due to inflated denominators in the data source, which does not have information on individuals who have migrated outside the country. In addition, we identified some variations in age-stratified results resulting from differences in how time-varying covariates are handled in the analytical code(see "**10.2. Strengths and limitations**"). Such age-differences should not be viewed as a limitation of using external denominators, but rather as a reflection of different approaches for

handling time-varying covariates. Despite some minor variability, the results were consistent across both approaches, which reinforces the validity of employing both methods to estimate prevalence in data sources where this comparison was assessed.

In this study, external denominators were obtained from Eurostat, which collects data on the "usually resident population" from the EU Member States. This population can be derived either from the most recent census adjusted by the components of the population change produced since the last census (the most common approach, and the one used by Croatia) or from population registries owned by national authorities (which is the case of 10 EU Member States, including Denmark, Norway, and Sweden).[3] For data sources with population-level denominators, such as national health registries, the use of external denominators is unlikely to provide additional benefit, as these sources typically offer more detailed, individual-level information and more comprehensive longitudinal follow-up. In addition, the denominators used by these data sources may originate from the same sources that are reported to Eurostat. However, the use of external denominators may still be relevant for data sources in which denominators are not accurately captured due to missing or incomplete data on births, mortality, or migration. This is the case of NAJS in this study, where the lack of data on emigration resulted in inflated denominators and might have led to an underestimation of prevalence. Lastly, it should be noted that, in this study, the use of external denominators was limited to data sources with national-level coverage. Applying this approach to other data sources may require additional assumptions and could be particularly challenging for data sources where the geographical coverage is not well-defined.

10.4. Generalisability

This study included data from 5 data sources from 5 European countries (Croatia, Denmark, Germany, the Netherlands, Norway). We consider our findings to largely reflect prevalence of selected cancers in the respective countries, within the healthcare settings covered by these data sources. However, results should not be generalised to Europe, as differences in population characteristics and coding practices might vary by country.

Considering both the current and the previous study (EUPAS50800), evidence provided in the context of DARWIN EU® was informed by a total of 10 data sources from 8 European countries. It should be noted that data sources included in the previous study mostly reflected primary care settings, while those included in the current study primarily reflect cases from hospital settings or, in some of the data sources included, those recorded in national cancer registries. Countries included in the first study included Belgium, Germany, the Netherlands, Spain, and the United Kingdom (UK). However, only data sources from Germany, Spain, and the UK provided estimates for all rare blood cancers studied. DLBCL and AML were not captured in the Belgian data source, and data from the Netherlands only assessed prevalence for multiple myeloma due to a lack of granularity in their source coding system. Data sources from Germany and the Netherlands were also included in the current study, but they reflected different healthcare settings compared to the previous study. In the first study, both countries relied on primary care data (IQVIA DA Germany and IPCI, respectively), while in the current study, Germany's data was from claims data (InGef RDB), and the Netherlands' data came from a national cancer registry (NCR).

Importantly, differences in healthcare settings, geographical variations, outcome definitions, and types of data used may limit the comparability of results across data sources.

11. CONCLUSION

Across the included data sources, ALL and AML exhibited the lowest prevalence, with a 5-year partial point prevalence of less than 1.5 per 10,000. For other cancer types, estimated prevalence in the included data sources was below 5 per 10,000. Estimated prevalence in the included data sources also varied

substantially across age groups, with older age groups obtaining the highest estimates across all cancer types, except for ALL.

Prevalence estimates obtained from cancer registries, and for some cancer types in InGef RDB from 2020, were broadly consistent across data sources and aligned with the estimates reported in the prior DARWIN EU® study. NAJS and DK-DHR obtained the highest estimates, likely because of the inclusion of unconfirmed cancer diagnoses, which likely resulted in an overestimation of prevalence. In DK-DHR, restricting the analysis to cases originating from the national cancer registry produced estimates comparable to other data sources for most cancers studied. In NAJS, this restriction led to comparable estimates for ALL, AML, and pancreatic cancer, while other outcomes assessed obtained lower estimates than those reported for other data sources and warrant further investigation.

Prevalence estimates based on denominators from Eurostat were generally consistent with those derived directly from data sources with population-level denominators and national coverage. In NAJS, the use of denominators derived from the data source resulted in marginally lower prevalence estimates across outcomes studied. This is likely attributed to an overestimation of the population in the data source, which exceeds the Croatian population reported in Eurostat by 400,000 individuals as of 2023. Small discrepancies in age-stratified results were also noted, likely reflecting the analytical choices applied in the analysis to handle time-varying stratifications.

Findings from this study highlight the complementarity of different data sources to improve validity and suggest that data sources with population-level denominators affected by missing or incomplete population data may benefit from the use of external denominators for prevalence estimation.

12. REFERENCES

1.Incidencija raka u Hrvatskoj u 2022. godini: Središnja medicinska ustanova javnog zdravstva u Hrvatskoj; 2025 [updated 24-04-2025. Available from: https://www.hzjz.hr/periodicne-publikacije/incidencija-raka-u-hrvatskoj-u-2022-godini/.
2.Points to consider on the estimation and reporting on the prevalence of a condition for the purpose of orphan designation.: Committee for Orphan Medicinal Products (COMP); 2019 [updated 20-06-2019; cited 2025. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/points-consider-estimation-and-reporting-prevalence-condition-orphan-designation_en.pdf.
3.Reference metadata: Population (national level): Eurostat; 2025 [updated 14/02/2025; cited 2025 01/08/2025]. Available from: https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm#shortsource_typeDisseminated.
4.Population on 1 January by age group and sex Eurostat2025 [updated 23/05/2024. Available from: https://ec.europa.eu/eurostat/databrowser/view/demo_pjangroup/default/table?lang=en&category=demo.demo_pop.
5.Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. Pharmacoepidemiology and Drug Safety. 2024;33(1).
6.Li J, Smith A, Crouch S, Oliver S, Roman E. Estimating the prevalence of hematological malignancies and precursor conditions using data from Haematological Malignancy Research Network (HMRN). Cancer Causes & Control. 2016;27(8):1019-26.
7.Kirtane K, Lee SJ. Racial and ethnic disparities in hematologic malignancies. Blood. 2017;130(15):1699-705.
8.Waxman AJ, Mink PJ, Devesa SS, Anderson WF, Weiss BM, Kristinsson SY, et al. Racial disparities in incidence and outcome in multiple myeloma: a population-based study. Blood. 2010;116(25):5501-6.
9.SEER*Explorer: An interactive website for SEER cancer statistics: Surveillance Research Program, National Cancer Institute; 2025 [Available from: https://seer.cancer.gov/statistics-network/explorer.
10. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al. Global Cancer Observatory: Cancer Today: International Agency for Research on Cancer; 2025 [Available from: https://gco.iarc.who.int/today/.
11. NCR data: Netherlands Comprehensive Cancer Organisation (IKNL); 2025 [updated 27-01-2025.
12. Larønningen S, Arvidsson G, Bray F, Dahl-Olsen E, Engholm G, Ervik M, et al. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 9.5 (19.06.2025). 2025 [Available from: https://nordcan.iarc.fr/.
13. Lang LM, Behr C, Ludwig M, Walker J, Christian Lange H, Basedow F, et al. Routine practice data of three cancer entities: Comparison among cancer registry and health insurance data. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2023;177:65-72.
14. Rassen JA, Bartels DB, Schneeweiss S, Patrick AR, Murk W. Measuring prevalence and incidence of chronic conditions in claims and electronic health record databases. Clinical Epidemiology. 2018;Volume 11:1-15.

Committee for Orphan Medicinal Products (COMP); 2019 [updated 20-06-2019; cited 2025. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedulation (national level): Eurostat; 2025 [updated 14/02/2025; cited 2025 01/08/2025]. Availabled.4.Population on 1 January by age group and sex Eurostat2025 [updated 23/05/2024. Available from: httpfile:///ER cancer statistics/ Surveillance Research Programfile:/// Cancer Observatory/ Cancer Todfile:///vik M, et al. NORDCAN/ C

13. ANNEXES

ANNEX I. Description of data sources

Croatian National Public Health Information System (NAJS), Croatia

#	Section	Description
1	Data source identification and country	NAJS (Croatian National Public Health Information System) Croatia
2	Data partner information section	Croatian Institute of Public Health Department of Data Science and Analytics
3	Coverage and timespan	Data collection since: 1998 Extent: Nation-wide. Geographic coverage covers whole Croatia, with various levels of resolution for different registries. Current estimates for the population in Croatia will be available at: https://podaci.dzs.hr/hr/podaci/stanovnistvo/procjena-stanovnistva/ for each year.
4	Healthcare setting / type of data	Primary care – General Practioner, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. Primary care – gps, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. For both inpatient and outpatient setting diagnoses, medication, procedures, and measurements are captured. The year of availability of information depends on the setting • 2014-2025 for biochemical lab tests in primary care from EHR patients records (measurements with results) • 2015-2025 for primary care data from EHR patient records (conditions, procedures, and drug prescriptions) • 2015- 2024 for inpatient hospital data from EHR administrative records (conditions, procedures, measurements without results and drug administrations) • 2016-2025 for health risk assessment data entered by GPs (measurements with results - height, weight…) • 2016-2024 for secondary conciliatory care data from EHR administrative records (conditions, procedures, measurements without results and drug administrations) • 2016-2022 for emergency care data from EHR patient records (conditions) • 2017-2025 for hospital records from registry data (conditions and procedures) • 2020-2025 for vaccination data from EHR patient records
5	Data collection process	Inpatient hospital billing systems, and Other. Data is entered by clinicians at healthcare contact, then combined by CIPH into the NAJS database and integrated with registries for public health purposes.
6	General representativeness	The data is collected from the evidence of public health records collected for public health purposes, as the majority of health care in Croatia is public and under single health insurance provider. Personal details are collected to a better extent for insured individuals compared to uninsured patients, who are excluded in the ETL process.
7	Data content /source coding	Medication prescriptions are recorded with ATC codes with an additional 3 digit code denoting the package. Diagnoses with ICD10 codes (Australian modification). Procedures with local source codes. Lab results with local source codes.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. Records from 2017 include insured patients with reliable IDs. Uninsured patients do not have reliable IDs. For example, if a patient changed her status from insured to uninsured, or vice versa, she could be counted several times, as could tracking records from before 2017 and after. By using the unique personal identifier for Croatian citizens, it can be checked and verified.
9	Quality control (data source specific)	There is a network of registry personnel (leaders, administrators, coders, sources) working on data coverage and other quality dimensions. An analytical team routinely checks for erroneous entries in hospital records, removing double entries, false dates, and overlapping stays. Entries

#	Section	Description
		without enough data or with obviously erroneous dates from primary care analysis are being excluded.
10	Linkage	The national death registry is updated yearly, with one year lag, but the fact of someone's death (just the date) is updated daily, without the cause of death or any other additional details. Primary care is updated weekly and hospital level care monthly. Specific registries are included in NAJS (e.g. diabetes registry), where inclusion criteria vary across these registries.
11	Vital status	NAJS is linked to the national death registry.
12	Limitations	Hospital data is available from 2017 onwards. This is often used as start of data collection, while laboratory and GP data is captured before that (since 2014 and 2015 respectively). The total and active person count in the NAJS data is larger than the current population of Croatia. This explained by: a) the person table included deceased and all previously insured people and b) there is no information about insurance ending, c) healthcare is also used by people with dual citizenship from neighbouring countries It is known that a lot of people emigrated (300k-400k) and weren't included in the last population census but still are in the NAJS database. There is also an influx of immigrant workers that are insured and registered but weren't included in the census. In-hospital administrations are managed via paper drug charts and hospital discharge summaries are currently not captured into NAJS. Reliability of hospital drug administration data is less reliable than prescription data from primary care, with some drugs (monoclonal antibodies / precision medicine drugs) that require additional approval not being recorded at all. In-hospital administrations are managed via paper drug charts and hospital discharge summaries are currently not captured into NAJS.
13	Main references	No main reference provided.
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111155 Website: https://www.hzjz.hr/nacionalni-javnozdravstveni-informacijski-sustav-najs/

Danish Health Registries (DK-DHR), Denmark

#	Section	Description
1	Data source Identification and country	DK-DHR (Danish Data Health Registries) Denmark
2	Data partner information section	Danish Medicines Agency (DKMA) Data Analytics Centre (DAC)
3	Coverage and timespan	Data collection since: 1995 Extent: Nation-wide. The data is representative of the entire Danish population.
4	Healthcare setting / type of data	Community pharmacists, and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnosis (including rare diseases and pregnancy data), hospital admissions, discharge and ICU data, Cause of death, Drug prescriptions, dispensing, vaccination and contraception, Procedures, Devices, and Sociodemographic information.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. All causes of deaths, all retrieved drug prescriptions, all records of vaccinations, all hospital inpatient and outpatients contacts including disease diagnoses and hospital surgical and non-surgical procedures, cancers, laboratory test results for the entire Danish population from 1/1/1995 onwards.
6	General representativeness	The data is representative of the entire Danish population. Healthcare is free in Denmark, so we do not expect any bias in data collection based on socio-economic status.

#	Section	Description
7	Data content /source coding	Diagnoses and causes of death are collected using the ICD-10 vocabulary. ATC and RxNorm are used for Drugs. SNOMED codes are used for Procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No.
9	Quality control (data source specific)	The data we have received relating to nationwide Danish Health Data registries offer an opportunity for large-scale, population-based studies with several advantages 1) Their large size improves the precision of estimates and enables the study of rare exposures and outcomes with long-term latency, 2) Inclusion of nearly all individuals in the target population ensures that the data reflect routine clinical care and all clinical segments of the source population, 3) Data are collected independently of each research study, thus minimising certain types of bias, e.g., non-response, and the influence from attention to the research question on the diagnostic process. Before the source data is sent to us, the Danish Health Data Authority runs and does comprehensive checks of the registry table data validity of the variables, breaks in data, changes in variable coding, missingness, etc. We perform checks of missingness/completeness in relation to requested variables. In essence, we are receiving a dump of a mirror of the data that is controlled by the SDS. The documentation performed by SDS is available online, in Danish primarily https://www.esundhed.dk/Dokumentation (all variables), but also in English https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/national_health_registers
10	Linkage	There is no linkage in this data source.
11	Vital status	The Cause of Death registry (DAR) is used, the cause of death is collected using ICD-10 codes.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Schmidt M, Schmidt SAJ,Adelborg K,Sundbøll J,Laugesen K,Ehrenstein V,Sørensen HT "The Danish health care system and epidemiological research: from health care contacts to database records." Clinical epidemiology (2019): 31372058
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111217 Website: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/healthdatadenmark

Institute for Applied Health Research Berlin GmbH (InGef RDB), Germany

#	Section	Description
1	Data source Identification and country	InGef RDB (InGef Research Database) Germany
2	Data partner information section	Institut für angewandte Gesundheitsforschung Berlin GmbH
3	Coverage and timespan	Data collection since: 2014 Extent: Nation-wide. The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs. The data in the database depicts all health care use which has been reimbursed by the SHI.
4	Healthcare setting / type of data	Primary care – gps, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The following data elements are collected: pregnancy data, hospital admission and/or discharge also with ICU admission. Prescription, dispensing drugs and Advanced therapy medicinal products. Contraception, medical devices, vaccinations, procedures, diagnoses and demographic information.

#	Section	Description
5	Data collection process	Insurance/administrative claims. The data in the database depicts all health care use which has been reimbursed by the SHI (statutory health insurances).
6	General representativeness	The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs.
7	Data content /source coding	The ATC and OPS (Operationen- und Prozedurenschlüssel) are used for prescription and dispensing drugs. For Procedures the EBM (Einheitlicher Bewertungsmaßstab - doctor's fee scale) and for ambulatory procedures; OPS (Operationen- und Prozedurenschlüssel) for operations conducted at the hospital are used. Medical events are coded in ICD-10-GM and another vocabulary used is PZN (Pharmazentralzummer -pharmaceutical reference number).
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No. In the German statutory health system, a person can only be enrolled in one health insurance at a time. However, if a person changes from one contributing insurer to another, a new ID number will be generated.
9	Quality control (data source specific)	Before entering the InGef database, the data elements are checked with respect to data format, completeness, and plausibility. After each data update, data are compared with the previous data update in regard to number of records, number of data providers, etc. Due to the anonymized nature of the database, no direct validation of the data (e.g. using medical charts as the gold standard) is possible. Data delivery by health care providers is generally based upon standardized data requirements and formats provided by the National Association of Statutory Health Insurance Funds (compare: https://www.gkv-datenaustausch.de/leistungserbringer/leistungserbringer.jsp)
10	Linkage	No
11	Vital status	The Cause of Death is not captured, the date of death is captured.
12	Limitations	Ambulatory diagnosis are received from the source on a quarterly basis. These diagnoses are mapped to the observation table with the concept_id History of event within 3 months (1340222), with the actual diagnosis concept_id recorded in the field value_as_concept_id, and the date as the last day of the respective quarter (i.e. 30/31st of Mar/Jun/Sept/Dec). Ambulatory prescriptions are available with exact dates. The cause of death is not captured and there is no linkage with other data sources. Approx. 10.5 Million insurees are included in the database, 7.8 Million of these actively insured in 2024. This corresponds to 7% of the total German population. Data are longitudinally linked over a period of currently ten years.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): 26530279
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

Netherlands Cancer Registry (NCR), The Netherlands

#	Section	Description
1	Data source identification and country	NCR (Netherlands Cancer Registry) The Netherlands
2	Data partner information section	IKNL (Netherlands Comprehensive Cancer Organisation) Department of Research and Development
3	Coverage and timespan	Data collection since: 1989 Extent: Nation-wide. The NCR compiles clinical data of all individuals newly diagnosed with cancer in the Netherlands. Data managers register newly diagnosed cancer patients, starting in 1989, on a national basis, with 3 million patients included.
4	Healthcare setting / type of data	Hospital inpatient care, and other (specify). The NCR is a disease-specific registry that contains characteristics of patients, details about the cancer diagnosis and the primary treatment, and information on the vital status which is updated once a year.
5	Data collection process	Registries. Data managers employed by the Netherlands Comprehensive Cancer Organisation (IKNL) enter and process the data.
6	General representativeness	The data has nationwide coverage in The Netherlands of people having a cancer diagnosis since 1989. Data since 1993 is OMOP standardised.
7	Data content /source coding	ATC codes, ICD-O-3. The diagnosis contains the type of cancer (ICD-O-3) and the stage (TNM). In addition, internally developed coding for diagnostic/therapeutic procedures.
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. No, we use identifiable data for linkage and identification across hospitals.
9	Quality control (data source specific)	All data is collected and checked by data managers. These are some of the quality control procedures: - All data managers receive a training when they start, and they get yearly refresher courses. - The registration application limits what can be registered, depending on certain disease characteristics. This prevents a lot of errors. - A team of data managers is responsible for quality (werkgroep kwaliteit). - Random samples of registered data are checked. - Automatic checks are done. - Researchers, who use the data, can request quality checks if they suspect issues.
10	Linkage	Established linkages: - Pathology data from PALGA - Vital status by Personal Records Database
11	Vital status	Yealy updated linkage with Personal Records Database
12	Limitations	The database only has patients with a cancer diagnosis. The primary treatment is registered. There are some lab results, mainly around the diagnosis and relevant for primary treatment. There is no sociodemographic information. Only ATC codes are available for some drugs, and this can only be mapped to ingredient level in OMOP.
13	Main references	Gijs Geleijnse, RuRu Chun-Ju Chiang, Melle Sieswerda, Melinda Schuurman, K. C. Lee, Johan van Soest, Andre Dekker, Wen-Chung Lee & Xander A. A. M. Verbeek "Prognostic factors analysis for oral cavity cancer survival in the Netherlands and Taiwan using a privacy-preserving federated infrastructure" Scientific Reports (2020):
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111117 Website: https://iknl.nl/international

Cancer Registry Norway (CRN), Norway

#	Section	Description
1	Data source identification and country	CRN (Cancer Registry Norway) Norway
2	Data partner information section	Norwegian Institute of Public Health Cancer Registry Norway
3	Coverage and timespan	Data collection since: 1953 Extent: Nation-wide. The catchment area is all of Norway. The approximate population is around 5.6 Million. The cancer-specific clinical registries have collected nationwide data since the mid-2010s and are continuously updated.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The following data elements are collected: diagnoses, pathology, lab reports, clinical notifications, deaths, procedures, measurements, drug prescription, and dispensing. The CRN also includes data from various clinical cancer registries for specific cancer types, providing extended information on diagnosis, treatment and follow-up. For example, there are clinical registries for lung, breast, gynecological (ovarian, cervical), prostate, melanoma, sarcoma, and other cancers, which supplement the main incidence registry.
5	Data collection process	Registries. Sources of information for the cancer registry are: clinical notifications from hospitals and specialists (public and private), pathology reports from pathology departments (public and private), radiation therapy and SACT-information from hospitals, data from the Norwegian Patient Registry, the Population Registry and the Cause of Death Registry, and PREM and PROM from patients. The data sources send data to the Cancer Registry, mainly in electronic form. Data is received and pre-processed before they are further enriched and processed by the data controllers, following a standard SOP for data quality checks. In addition, strict quality assessment procedures are carried out to manually evaluate the quality of the information regarding general statistics and cancer-specific domain data.
6	General representativeness	The Cancer Registry of Norway is a population-based cancer registry covering the entire Norwegian population. Registration of cancer patients in the CRN is mandatory in Norway, with no possibility for opt-out. The registry therefore includes all cancer patients (including premalignant cases and some benign tumors of the CNS) and provides a comprehensive representation of the national cancer population.
7	Data content /source coding	ICD03 and ICD10 are used for topography/morphology/disease of the cancer diagnosis. Some examples of other used classifications are ISUP, ECOG, TNM and SNOMED. The cancer registry also has extensive in-house classifications developed through over 70 years of cancer registration. Drugs prescriptions and dispensing is recorded with ATC codes, indications for the prescriptions are recorded with ICD10 codes. All information on the content of the CRN can be found in our online metadata-catalogue at https://metadata.kreftregisteret.no
8	Data Harmonisation	The data has been mapped to the OMOP CDM v5.4 and the OMOP standard vocabularies (SNOMED, RxNorm, LOINC). The format, structural and semantic conformance has been verified upon onboarding into the DARWIN EU® data network. All patients have 11-digit unique national identifiers for identification and linkage.
9	Quality control (data source specific)	The CRN follows a strict quality assurance regime, including validation of all incoming data and a rule engine that enforces international, national, and internal standards uniformly. Manual checks are performed for specific cancer types to identify unlikely data, such as melanoma cases with tumour diameters exceeding 2 cm. Data quality is documented in scientific articles and

#	Section	Description
		annual statistics, including metrics like morphological verification rate, estimated completeness, proportion of cases identified only via death certificates, and male-to-female ratios.
10	Linkage	Every citizen has a unique personal identifier, and via this identifier, patients can be linked to other national data sources.
11	Vital status	Vital status is regularly updated in the Cancer Registry of Norway through linkage with the National Population Register. Information on cause for death is obtained from the Cause of death Registry through linkage.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Larsen IK, Småstuen M,Johannesen TB,Langmark F,Parkin DM,Bray F,Møller B "Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness." European journal of cancer (Oxford, England : 1990) (2009): 19091545
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111128 Website: https://www.kreftregisteret.no/en/

ANNEX II. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enabled the use of standardised analytics and using DARWIN EU® tools across the network since the structure of the data and the terminology system was harmonised. The OMOP CDM was developed and maintained by the Observational Health Data Sciences and Informatics (OHDSI) initiative and is described in detail on the wiki page of the CDM: https://ohdsi.github.io/CommonDataModel and in The Book of OHDSI: http://book.ohdsi.org

The analytic code for this study was written in R and used standardised analytics wherever possible. Each data partner executed the study code against their data source containing individual data and then returned the results (csv files) which only contained aggregated data. The results from each of the contributing data sites were then combined in tables and figures for the study report.

Data storage and protection

For this study, personal data from individuals in various EU member states were processed, using information collected from national/regional electronic health record data sources. Due to the sensitive nature of this personal medical data, it is important to be fully aware of ethical and regulatory aspects and to strive to take all reasonable measures to ensure compliance with ethical and regulatory issues on privacy.

All data sources used in this study were already used for pharmaco-epidemiological research and have a well-developed mechanism to ensure that European and local regulations dealing with ethical use of the data and adequate privacy control were adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses were run, which generate non-identifiable aggregate summary results.

QUALITY CONTROL

Data source quality control

When defining cohorts, a systematic search of possible codes for inclusion were identified using the *CodelistGenerator* R package (https://github.com/darwin-eu/CodelistGenerator). This package allows the user to define a search strategy and will use this to query the vocabulary tables of the OMOP common data model so as to find potentially relevant codes. In addition, the *CohortDiagnostics* (https://github.com/OHDSI/CohortDiagnostics) were executed to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error.

The study code to calculate point prevalence was based on DARWIN EU® R package *IncidencePrevalence*, which includes numerous automated unit tests to ensure the validity of the codes, alongside software peer review and user testing. The R package will be made publicly available via GitHub. *IncidencePrevalence* was used to calculate prevalence when using denominators derived directly from the data. Custom code was created for analyses requiring the use of external sources to identify denominators, as this is not currently implemented in the current DARWIN EU® analytical pipelines.

ANNEX III: List of conditions definitions

Table S1. Acute Lymphocytic Leukaemia.

SNOMED codes:

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4081867	Acute biphenotypic leukemia	X	
4153344	Acute lymphoblastic leukemia, transitional pre-B-cell	X	X
134305	Acute lymphoid leukemia	X	X
36712834	Acute lymphoid leukemia relapse	X	X
4094550	Adult T-cell leukemia	X	
4003188	Adult T-cell leukemia/lymphoma	X	
4300174	Adult T-cell leukemia/lymphoma	X	
36717161	Aggressive natural killer-cell leukemia	X	
37207573	ALL (acute lymphoblastic leukaemia) tumour and germline WGS (whole genome sequencing)		X
37156124	B lymphoblastic leukemia lymphoma		X
42872958	B lymphoblastic leukemia lymphoma with hyperdiploidy	X	X
42872959	B lymphoblastic leukemia lymphoma with hypodiploidy (Hypodiploid ALL)	X	X
37206196	B lymphoblastic leukemia lymphoma with iAMP21	X	X
42872961	B lymphoblastic leukemia lymphoma with t(1;19)(q23;p13.3); E2A-PBX1 (TCF3-PBX1)	X	X
42872957	B lymphoblastic leukemia lymphoma with t(12;21)(p13;q22); TEL-AML1 (ETV6-RUNX1)	X	X
42872960	B lymphoblastic leukemia lymphoma with t(5;14)(q31;q32); IL3-IGH	X	X
37109936	B lymphoblastic leukemia lymphoma with t(9:22) (q34;q11.2); BCR-ABL 1	X	X
42872955	B lymphoblastic leukemia lymphoma with t(9;22)(q34;q11.2); BCR-ABL1	X	X
42872956	B lymphoblastic leukemia lymphoma with t(v;11q23); MLL rearranged	X	X
37204838	B-lymphoblastic leukemia lymphoma BCR-ABL1-like	X	X
37204662	NK-lymphoblastic leukemia/lymphoma	X	X
37017893	Philadelphia chromosome-negative precursor B-cell acute lymphoblastic leukemia	X	X
36712835	Refractory acute lymphoid leukemia	X	X
4153344	Acute lymphoblastic leukemia, transitional pre-B-cell	X	X
134305	Acute lymphoid leukemia	X	X

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
3654653	B lymphoblastic leukemia lymphoma with hyperdiploidy	X	X
3654651	B lymphoblastic leukemia lymphoma with hypodiploidy	X	X
3654650	B lymphoblastic leukemia lymphoma with t(1;19)(Q23;P13.3); E2A-PBX1 (TCF3/PBX1)	X	X
3654649	B lymphoblastic leukemia lymphoma with t(12;21)(p13;q22); TEL/AML1 (ETV6-RUNX1)	X	X
3654647	B lymphoblastic leukemia lymphoma with t(5;14)(q31;q32); IL3-IGH	X	X
3654648	B lymphoblastic leukemia lymphoma with t(v;11q23); MLL rearranged	X	X
4173963	B-cell acute lymphoblastic leukemia	X	X
4079280	Common acute lymphoblastic leukemia	X	X
4079281	Null cell acute lymphoblastic leukemia	X	X
4138008	Philadelphia chromosome-positive acute lymphoblastic leukemia	X	X
4143821	Philadelphia chromosome-positive acute lymphoblastic leukemia	X	X
4082461	Precursor B-cell acute lymphoblastic leukemia	X	X
4029662	Precursor B-cell lymphoblastic leukemia	X	X
4264448	Precursor B-lymphoblastic leukemia/lymphoblastic lymphoma	X	X
4030260	Precursor cell lymphoblastic leukemia	X	X
4030261	T-lymphoblastic leukemia	X	X
4082462	T-lymphoblastic leukemia	X	X
4221907	T-lymphoblastic leukemia/lymphoblastic lymphoma	X	X
45766617	T-lymphoblastic leukemia/lymphoma	X	
4227963	T-lymphoblastic lymphoma	X	X
4288091	T-lymphoblastic lymphoma/leukemia	X	X
4189936	Acute lymphoblastic leukemia - category	X	(*)
42872925	B lymphoblastic leukemia/lymphoma- catergory	X	(*)
42872954	B lymphoblastic leukemia lymphoma- no ICD-O subtype	X	(*)
45766617	T-lymphoblastic leukemia/lymphoma	X	(*)

SNOMED= Systematised Nomenclature of Medicine.
*Not standard in the vocabulary version used in the current study (27-AUG-2025).

ICD-O-3 Codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
9727/3, 9728/3, 9729/3, 9811/3, 9812/3, 9813/3, 9814/3, 9815/3, 9816/3, 9817/3, 9818/3, 9819/3, 9826/3, 9836/3, 9837/3, 9819/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

Table S2. Acute Myeloid Leukaemia.

SNOMED codes:

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4289318	Acute basophilic leukemia	X	X
4031360	Acute erythroid leukemia	X	X
4079686	Acute megakaryoblastic leukemia	X	X
4180418	Acute megakaryoblastic leukemia	X	X
135768	Acute monocytic leukemia	X	X
37151872	Acute myeloid leukemia	X	X
35623630	Acute myeloid leukemia and myelodysplastic syndrome related to alkylating agent	X	X
35623633	Acute myeloid leukemia and myelodysplastic syndrome related to radiation	X	X
35623631	Acute myeloid leukemia and myelodysplastic syndrome related to topoisomerase type 2 inhibitor	X	X
4029177	Acute myeloid leukemia myelodysplasia related	X	X
608277	Acute myeloid leukemia with 11q23 abnormality		X
37165057	Acute myeloid leukemia with BCR-ABL1	X	X
37204375	Acute myeloid leukemia with BCR::ABL1 fusion	X	X
37312067	Acute myeloid leukemia with biallelic CEBPA mutation	X	X
3654662	Acute myeloid leukemia with CBFB::MYH11 fusion	X	X
4304355	Acute myeloid leukemia with CBFB::MYH11 fusion	X	X
45771384	Acute myeloid leukemia with CEBPA mutation	X	X
35622696	Acute myeloid leukemia with CEBPA somatic mutations	X	X
42872933	Acute myeloid leukemia with DEK::NUP214 fusion	X	X
36683269	Acute myeloid leukemia with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); RPN1-EVI1	X	X
4029663	Acute myeloid leukemia with KMT2A rearrangement	X	X
4304051	Acute myeloid leukemia with maturation	X	X

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4234749	Acute myeloid leukemia with maturation, FAB M2	X	X
42872934	Acute myeloid leukemia with MECOM rearrangement	X	X
4304199	Acute myeloid leukemia with minimal differentiation	X	X
45766616	Acute myeloid leukemia with mutated NPM1	X	X
40483761	Acute myeloid leukemia with myelodysplasia-related changes	X	X
35622003	Acute myeloid leukemia with NPM1 somatic mutation	X	X
42872942	Acute myeloid leukemia with RBM15::MKL1 fusion	X	X
4028713	Acute myeloid leukemia with RUNX1::RUNX1T1 fusion	X	X
37116722	Acute myeloid leukemia with t(6;9)(p23;q34) translocation	X	X
37110870	Acute myeloid leukemia with t(8;16)(p11;p13) translocation	X	X
37110871	Acute myeloid leukemia with t(8;16)(p11;p13) translocation	X	X
40481524	Acute myeloid leukemia with t(9:11)(p22;q23); MLLT3-MLL	X	X
4304356	Acute myeloid leukemia without maturation	X	X
4230253	Acute myeloid leukemia without maturation, FAB M1	X	X
140352	Acute myeloid leukemia, disease	X	X
4233531	Acute myeloid leukemia, minimal differentiation, FAB M0	X	X
4003187	Acute myelomonocytic leukemia, FAB M4	X	X
4003184	Acute panmyelosis with myelofibrosis	X	X
4112803	Acute promyelocytic leukemia - hypogranular variant	X	X
4002497	Acute promyelocytic leukemia, FAB M3	X	X
37207565	AML (acute myeloid leukaemia) tumour and germline WGS (whole genome sequencing)		X
4144191	Basophilic leukemia	X	X
45765495	Core binding factor acute myeloid leukemia	X	X
45766268	Cytogenetically normal acute myeloid leukemia	X	X
138099	Erythroleukemia, FAB M6	X	X
4175688	Hypergranular promyelocytic leukemia	X	X
35622760	Inherited acute myeloid leukemia	X	X
4300784	Leukemic infiltration of skin in acute myeloid leukemia	X	X
4173970	Malignant myeloid/lymphoid neoplasm	X	X
35607963	Megakaryoblastic acute myeloid leukemia with t(1;22)(p13;q13)	X	X
313159	Megakaryocytic leukemia	X	X

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
42872921	Mixed phenotype acute leukemia B/myeloid	X	
42872922	Mixed phenotype acute leukemia T/myeloid	X	
3572256	Refractory acute myeloid leukemia	X	X
3572249	Relapsing acute myeloid leukemia	X	X
4326339	Smoldering chronic lymphocytic leukemia	X	
36717461	Therapy related acute myeloid leukemia and myelodysplastic syndrome	X	X
42538579	Therapy related acute myeloid leukemia due to and following administration of antineoplastic agent	X	X
36715587	Acute myeloid leukemia due to recurrent genetic abnormality	X	(*)
42535969	Acute myeloid leukemia with FMS-like tyrosine kinase-3 mutation	X	(*)
42539431	Acute myeloid leukemia with FMS-like tyrosine kinase-3 mutation	X	(*)
4213196	Acute panmyelosis with myelofibrosi	X	(*)

SNOMED= Systematised Nomenclature of Medicine.
*Not standard in the vocabulary version used in the current study (27-AUG-2025).

ICD-O-3 Codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
9840/3, 9861/3, 9865/3, 9866/3, 9867/3, 9869/3, 9870/3, 9871/3, 9872/3, 9873/3, 9874/3, 9877/3, 9878/3, 9879/3, 9891/3, 9895/3, 9897/3, 9910/3, 9911/3, 9912/3, 9920/3, 9931/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

Table S3. Chronic Lymphocytic Leukaemia.

SNOMED codes:

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4173824	B-cell chronic lymphocytic leukemia variant	X	X
4180093	B-cell chronic lymphocytic leukemia/small lymphocytic lymphoma	X	X
37110902	Chronic lymphocytic leukemia genetic mutation variant	X	X
4082338	Chronic lymphocytic prolymphocytic leukemia syndrome	X	X
760935	Chronic lymphoid leukemia in relapse	X	X
138379	Chronic lymphoid leukemia, disease	X	X
44811227	Clinical stage A chronic lymphocytic leukaemia	X	X
44811228	Clinical stage B chronic lymphocytic leukaemia	X	X
44814026	Clinical stage C chronic lymphocytic leukaemia	X	X
4173962	Lymphoma with spill	X	X
4082311	B-cell chronic lymphocytic leukemia	X	(*)
4186899	Chronic lymphoid leukemia - category	X	(*)
4139554	Atypical hairy cell leukemia	X	
4173974	B-cell prolymphocytic leukemia	X	
4182216	Hairy cell leukemia	X	
4038845	Hairy cell leukemia (clinical)	X	
4245460	Hairy cell leukemia of spleen	X	
4082459	Hairy cell leukemia variant	X	
4082460	Large granular lymphocytic leukemia	X	
4299151	Leukemic infiltration of skin in hairy-cell leukemia	X	
193429	Leukemic reticuloendotheliosis of intra-abdominal lymph nodes	X	
132570	Leukemic reticuloendotheliosis of lymph nodes of head, face and neck	X	
37206728	Monoclonal B-cell lymphocytosis chronic lymphocytic leukemia-type	X	
37312112	Monoclonal B-cell lymphocytosis chronic lymphocytic leukemia-type	X	
37312109	Monoclonal B-cell lymphocytosis non-chronic lymphocytic leukemia type	X	
37204530	Non-chronic lymphocytic leukemia monoclonal B-cell lymphocytosis	X	
4029800	Prolymphocytic leukemia	X	
4001331	Prolymphocytic leukemia (clinical)	X	

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4173956	Richter's syndrome	X	
4326339	Smoldering chronic lymphocytic leukemia	X	
4173957	Splenic lymphoma with villous lymphocytes	X	
44783718	T-cell large granular lymphocytic leukemia	X	
4079683	T-cell prolymphocytic leukemia	X	

SNOMED= Systematised Nomenclature of Medicine.
*Not standard in the vocabulary version used in the current study (27-AUG-2025).

ICD-O-3 Codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
9823/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

Table S4. Diffuse Large B-cell Lymphoma.

SNOMED codes:

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
37151287	Diffuse large B cell lymphoma	X	X
432574	Diffuse large B-cell lymphoma	X	X
37206914	Diffuse large B-cell lymphoma activated B-cell subtype	X	X
37116992	Diffuse large B-cell lymphoma associated with chronic inflammation	X	X
37110401	Diffuse large B-cell lymphoma co-occurrent with chronic inflammation caused by Epstein-Barr virus	X	X
37206933	Diffuse large B-cell lymphoma germinal center B-cell subtype	X	X
37116982	Diffuse large B-cell lymphoma of central nervous system	X	X
3654886	Diffuse large B-cell lymphoma of small intestine	X	X
3654887	Diffuse large B-cell lymphoma of stomach	X	X
4121330	Diffuse malignant lymphoma - centroblastic	X	X
4079293	Diffuse malignant lymphoma - centroblastic polymorphic	X	X
4173978	Diffuse malignant lymphoma - large cleaved cell	X	X
4079291	Diffuse malignant lymphoma - large non-cleaved cell	X	X
764839	Diffuse non-Hodgkin's lymphoma Lugano stage I	X	X

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
37396838	Epstein-Barr virus positive diffuse large B-cell lymphoma of elderly	X	X
37399015	Epstein-Barr virus positive diffuse large B-cell lymphoma of elderly	X	X
40481357	Large cell lymphoma of intrapelvic lymph nodes	X	X
440058	Malignant lymphoma of extranodal AND/OR solid organ site	X	X
42872963	Malignant lymphoma, diffuse large B-cell, immunoblastic	X	X
4178883	Malignant lymphoma, mixed small and large cell, diffuse	X	X
4028703	Mediastinal large B-cell lymphoma	X	X
4297358	Primary cutaneous diffuse large cell B-cell lymphoma	X	X
42539527	Primary cutaneous diffuse large cell B-cell lymphoma of lower extremity	X	X
4028702	Primary effusion lymphoma	X	X
44808122	Diffuse large B-cell lymphoma		X (**)
4300704	Diffuse large B-cell lymphoma (nodal/systemic with skin involvement)		X (**)
36716774	B-cell lymphoma unclassifiable with features intermediate between Burkitt lymphoma and diffuse large B-cell lymphoma	X	
36716775	B-cell lymphoma unclassifiable with features intermediate between classical Hodgkin lymphoma and diffuse large B-cell lymphoma	X	
36712836	B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma	X	
36712837	B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and Hodgkin lymphoma	X	
37117038	B-cell lymphoma with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma	X	
4144199	Diffuse malignant lymphoma - centroblastic-centrocytic	X	
4205271	Follicular lymphoma	X	
4230587	Follicular lymphoma, grade 3	X	
44808118	Follicular lymphoma grade 3b	X	
4146630	Malignant lymphoma, centroblastic-centrocytic, follicular	X	
4141258	Malignant lymphoma, centroblastic type, follicular	X	

SNOMED= Systematised Nomenclature of Medicine.
**Not standard in the vocabulary version used in the current study (27-AUG-2025). Included as they were standard in some of the vocabulary versions used in included data sources.

ICD-O-3 codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
9675/3, 9678/3, 9679/3, 9680/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

Table S5. Follicular lymphoma.

SNOMED codes:

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
40486171	Diffuse follicle center lymphoma	X	X
4144199	Diffuse malignant lymphoma - centroblastic-centrocytic	X	X
4079288	Follicular low grade B-cell lymphoma	X	X
4200880	Follicular low grade B-cell lymphoma morphology	X	X
4205271	Follicular lymphoma	X	X
4147411	Follicular lymphoma	X	X
44808015	Follicular lymphoma grade 1	X	X
44814156	Follicular lymphoma grade 2	X	X
44808028	Follicular lymphoma grade 3	X	X
606919	Follicular lymphoma grade 3a		X
607791	Follicular lymphoma grade 3a		X
606914	Follicular lymphoma grade 3b		X
36715795	Follicular lymphoma of small intestine	X	X
4265301	Follicular lymphoma, cutaneous follicle center sub-type	X	X
35610325	Follicular lymphoma, cutaneous follicle centre	X	X
4288232	Follicular lymphoma, diffuse follicle center cell sub-type, grade 2	X	X
4265735	Follicular lymphoma, diffuse follicle center sub-type, grade 1	X	X
4188203	Follicular lymphoma, grade 1	X	X
4204524	Follicular lymphoma, grade 2	X	X
4230587	Follicular lymphoma, grade 3	X	X
4173977	Follicular malignant lymphoma - mixed cell type	X	X
4079289	Follicular malignant lymphoma - small cleaved cell	X	X
45765919	Follicular non-Hodgkin's lymphoma diffuse follicle center cell sub-type grade 2	X	X

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
45765770	Follicular non-Hodgkin's lymphoma diffuse follicle center sub-type grade 1	X	X
40490991	Follicular non-Hodgkin's lymphoma of bone	X	X
40492940	Follicular non-Hodgkin's lymphoma of central nervous system	X	X
40490467	Follicular non-Hodgkin's lymphoma of extranodal site	X	X
40487142	Follicular non-Hodgkin's lymphoma of intestine	X	X
40490998	Follicular non-Hodgkin's lymphoma of lung	X	X
36684826	Follicular non-Hodgkin's lymphoma of lymph nodes of multiple sites	X	X
40486169	Follicular non-Hodgkin's lymphoma of nasopharynx	X	X
40488917	Follicular non-Hodgkin's lymphoma of nose	X	X
40493017	Follicular non-Hodgkin's lymphoma of oral cavity	X	X
40486654	Follicular non-Hodgkin's lymphoma of ovary	X	X
40488901	Follicular non-Hodgkin's lymphoma of prostate	X	X
40492018	Follicular non-Hodgkin's lymphoma of skin	X	X
40489407	Follicular non-Hodgkin's lymphoma of soft tissue	X	X
40486173	Follicular non-Hodgkin's lymphoma of stomach	X	X
40487141	Follicular non-Hodgkin's lymphoma of testis	X	X
40493011	Follicular non-Hodgkin's lymphoma of tonsil	X	X
40493012	Follicular non-Hodgkin's lymphoma of uterine cervix	X	X
4003833	Follicular non-Hodgkin's lymphoma, large cell (clinical)	X	X
4001329	Follicular non-Hodgkin's lymphoma, mixed small cleaved cell and large cell (clinical)	X	X
4002357	Follicular non-Hodgkin's lymphoma, small cleaved cell (clinical)	X	X
4097572	Follicular non-Hodgkin's mixed small cleaved and large cell lymphoma	X	X
440058	Malignant lymphoma of extranodal AND/OR solid organ site	X	X
435753	Malignant lymphoma of intrathoracic lymph nodes	X	X
37165833	Malignant lymphoma of lymph nodes of head and neck	X	X
4141258	Malignant lymphoma, centroblastic type, follicular	X	X
4146630	Malignant lymphoma, centroblastic-centrocytic, follicular		X
4017277	Malignant lymphoma, follicular AND/OR nodular	X	X
4123298	Malignant lymphoma, follicular center cell	X	X
4121853	Malignant lymphoma, follicular center cell, cleaved	X	X
4121970	Malignant lymphoma, follicular center cell, non-cleaved	X	X

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4146031	Malignant lymphoma, lymphocytic, poorly differentiated, nodular	X	X
4146027	Malignant lymphoma, mixed lymphocytic-histiocytic, nodular	X	X
4149840	Nodular lymphoma		X
198088	Nodular lymphoma of intrapelvic lymph nodes	X	X
320347	Nodular lymphoma of lymph nodes of multiple sites	X	X
4119131	Nodular malignant lymphoma, lymphocytic - well differentiated	X	X
42537145	Pediatric follicular lymphoma	X	X
37116892	Pediatric follicular lymphoma	X	X
4301668	Primary cutaneous follicular center B-cell lymphoma	X	X
37205181	Follicular T-cell lymphoma	X	
4146024	Malignant lymphoma - centrocytic	X	
37203899	Nodal peripheral T-cell lymphoma with T follicular helper phenotype	X	
4121331	Nodular malignant lymphoma, lymphocytic - intermediate differentiation	X	
4299152	Follicular center B-cell lymphoma (nodal/systemic with skin involvement)	X	(*)
438698	Malignant lymphoma of lymph nodes of head, face AND/OR neck	X	(*)
194878	Nodular lymphoma of extranodal AND/OR solid organ site	X	(*)
44808117	Follicular lymphoma grade 3a	X	(*)
44806990	Follicular lymphoma grade 3a	X	(*)
44808118	Follicular lymphoma grade 3b	X	(*)

SNOMED= Systematised Nomenclature of Medicine.
*Not standard in the vocabulary version used in the current study (27-AUG-2025).

ICD-O-3 codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
9591/3, 9597/3, 9673/3, 9690/3, 9695/3, 9698/3, 9709/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

Table S6. Multiple Myeloma.

SNOMED codes:

Concept ID	Concept name	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
4258135	Asymptomatic multiple myeloma	X	X
4094548	Extramedullary plasmacytoma	X	X
4111355	IgA myeloma	X	X
4111356	IgG myeloma	X	X
4112310	Immunoglobulin D myeloma	X	X
4259972	Indolent multiple myeloma	X	X
4188299	Kappa light chain myeloma	X	X
4197600	Lambda light chain myeloma	X	X
4082464	Light chain myeloma	X	X
437233	Multiple myeloma	X	X
4214660	Multiple solitary plasmacytomas	X	X
4019477	Myeloma-associated amyloidosis	X	X
4079684	Non-secretory myeloma	X	X
4137510	Osteosclerotic myeloma	X	X
4028859	Plasma cell leukemia	X	X
133154	Plasma cell leukemia	X	X
760936	Plasma cell leukemia in relapse	X	X
37151852	Plasma cell myeloma		X
4163558	Plasma cell myeloma/plasmacytoma	X	X
4190642	Plasma cell myeloma/plasmacytoma	X	X
4216139	Plasmacytoma	X	X
37156887	Plasmacytoma		X
4300702	Primary cutaneous plasmacytoma	X	X
764229	Relapse multiple myeloma	X	X
4145040	Solitary osseous myeloma		X
4184985	Smoldering myeloma	X	
4190641	Plasma cell myeloma - category	X	(*)
4024874	Plasmacytoma	X	(*)
4210177	Multiple myeloma	X	(*)

SNOMED= Systematised Nomenclature of Medicine.
*Not standard in the vocabulary version used in the current study (27-AUG-2025).

ICD-O-3 codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
9731/3, 9732/3, 9733/3, 9734/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

Table S7. Pancreatic cancer.

SNOMED codes:

Concept ID	Concept name
37204852	Acinar cell carcinoma of pancreas
45763891	Adenocarcinoma of pancreas
605820	Adenocarcinoma of pancreas with NRG1 fusion
4181331	Carcinoma of body of pancreas
4110585	Carcinoma of endocrine pancreas
4209933	Carcinoma of head of pancreas
4157459	Carcinoma of pancreas
4178960	Carcinoma of tail of pancreas
4340498	Cystadenocarcinoma of pancreas
37395837	Familial malignant neoplasm of pancreas
1244601	Gastric inhibitory peptide-secreting tumor of pancreas
4282771	Glucagonoma, malignant
1244600	Growth hormone releasing factor-secreting tumor of pancreas
606747	Infiltrating duct carcinoma of pancreas
4051445	Insulinoma, malignant
37165043	Intraductal tubulopapillary malignant neoplasm of pancreas
36683250	Invasive intraductal papillary-mucinous carcinoma of pancreas
765387	Malignant carcinoid tumor of pancreas
37162352	Malignant gastrinoma of pancreas
37162354	Malignant glucagonoma of pancreas
37017016	Malignant insulinoma
4092074	Malignant neoplasm of ectopic pancreatic tissue
4092072	Malignant tumor of body of pancreas
4112734	Malignant tumor of endocrine pancreas
4111024	Malignant tumor of exocrine pancreas
4178967	Malignant tumor of head of pancreas
4095437	Malignant tumor of Islets of Langerhans

Concept ID	Concept name
4180793	Malignant tumor of pancreas
4094866	Malignant tumor of pancreatic duct
4095436	Malignant tumor of tail of pancreas
3655588	Mixed ductal-neuroendocrine carcinoma of pancreas
37162106	Mixed neuroendocrine-non neuroendocrine neoplasm of pancreas
37206235	Mucinous cystadenocarcinoma of pancreas
3655584	Mucinous cystic neoplasm with invasive carcinoma of pancreas
37162169	Neuroendocrine carcinoma of pancreas
1244583	Neuroendocrine neoplasm of pancreas
37163229	Non-functioning neuroendocrine neoplasm of pancreas
192261	Overlapping malignant neoplasm of pancreas
37311469	Pancreatic ductal adenocarcinoma
4114220	Pancreatic polypeptidoma
4205868	Pancreatoblastoma
40391739	Pancreatoblastoma
1244602	Parathyroid hormone-related peptide-secreting tumor of pancreas
37167610	Primary acinar cell carcinoma of pancreas
36713362	Primary adenocarcinoma of body of pancreas
36713363	Primary adenocarcinoma of head of pancreas
602008	Primary adenocarcinoma of neck of pancreas
42872399	Primary adenocarcinoma of pancreas
602009	Primary adenocarcinoma of pancreatic duct
602011	Primary adenocarcinoma of tail of pancreas
37166180	Primary carcinoma of body of pancreas
37166228	Primary carcinoma of endocrine pancreas
37166282	Primary carcinoma of head of pancreas
37166400	Primary carcinoma of pancreas
37166181	Primary carcinoma of tail of pancreas
37166182	Primary cystadenocarcinoma of pancreas
37166533	Primary infiltrating duct carcinoma of pancreas
37167596	Primary invasive intraductal papillary-mucinous carcinoma of pancreas
37166648	Primary malignant gastrinoma of pancreas
609180	Primary malignant gastrointestinal stromal neoplasm of pancreas
37166647	Primary malignant glucagonoma of pancreas
37166594	Primary malignant insulinoma
434293	Primary malignant neoplasm of body of pancreas
440649	Primary malignant neoplasm of head of pancreas

Concept ID	Concept name
25486	Primary malignant neoplasm of islets of Langerhans
601133	Primary malignant neoplasm of neck of pancreas
199754	Primary malignant neoplasm of pancreas
433423	Primary malignant neoplasm of pancreatic duct
432843	Primary malignant neoplasm of tail of pancreas
42536743	Primary malignant neuroendocrine neoplasm of pancreas
37162707	Primary malignant solid pseudopapillary neoplasm of pancreas
37162713	Primary mixed acinar ductal carcinoma of pancreas
37162711	Primary mixed acinar endocrine-ductal carcinoma of pancreas
37167600	Primary mucinous cystadenocarcinoma of pancreas
37166549	Primary mucinous cystic neoplasm with invasive carcinoma of pancreas
37167745	Primary pancreatoblastoma
37167609	Primary serous cystadenocarcinoma of pancreas
37167650	Primary solid pseudopapillary carcinoma of pancreas
1244451	Primary well-differentiated neuroendocrine tumor of pancreas
37163230	Serotonin-producing neuroendocrine neoplasm of pancreas
37204808	Serous cystadenocarcinoma of pancreas
37204187	Solid pseudopapillary carcinoma of pancreas
36674768	Squamous cell carcinoma of exocrine pancreas
37168354	Undifferentiated carcinoma with osteoclast-like giant cells of pancreas
4098951	Vasoactive intestinal peptide-secreting tumor
36713547	Well-differentiated neuroendocrine tumor of pancreas

SNOMED= Systematised Nomenclature of Medicine.

ICD-O-3 Codes:

Concept ID	ICD-O-3 Codes[1]	Name
36532486	8550/3-C25.1	Acinar cell carcinoma of body of pancreas
44500463	8550/3-C25.0	Acinar cell carcinoma of head of pancreas
36556489	8550/3-C25.4	Acinar cell carcinoma of islets of Langerhans
44501959	8550/3-C25.7	Acinar cell carcinoma of other specified parts of pancreas
36525442	8550/3-C25.8	Acinar cell carcinoma of overlapping lesion of pancreas
36551058	8550/3-C25.3	Acinar cell carcinoma of pancreatic duct
36548164	8550/3-C25.2	Acinar cell carcinoma of tail of pancreas
36549500	8551/3-C25.1	Acinar cell cystadenocarcinoma of body of pancreas
36560796	8551/3-C25.0	Acinar cell cystadenocarcinoma of head of pancreas
36543853	8551/3-C25.4	Acinar cell cystadenocarcinoma of islets of Langerhans
36554611	8551/3-C25.7	Acinar cell cystadenocarcinoma of other specified parts of pancreas
36519153	8551/3-C25.8	Acinar cell cystadenocarcinoma of overlapping lesion of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36532005	8551/3-C25.9	Acinar cell cystadenocarcinoma of pancreas, NOS
36540162	8551/3-C25.3	Acinar cell cystadenocarcinoma of pancreatic duct
36564157	8551/3-C25.2	Acinar cell cystadenocarcinoma of tail of pancreas
36524267	8245/3-C25.1	Adenocarcinoid tumor of body of pancreas
36524362	8245/3-C25.0	Adenocarcinoid tumor of head of pancreas
36530881	8245/3-C25.4	Adenocarcinoid tumor of islets of Langerhans
36544725	8245/3-C25.7	Adenocarcinoid tumor of other specified parts of pancreas
36543801	8245/3-C25.8	Adenocarcinoid tumor of overlapping lesion of pancreas
36533450	8245/3-C25.9	Adenocarcinoid tumor of pancreas, NOS
36542041	8245/3-C25.3	Adenocarcinoid tumor of pancreatic duct
36521830	8245/3-C25.2	Adenocarcinoid tumor of tail of pancreas
36533908	8210/3-C25.1	Adenocarcinoma in adenomatous polyp of body of pancreas
36553227	8210/3-C25.0	Adenocarcinoma in adenomatous polyp of head of pancreas
36526586	8210/3-C25.4	Adenocarcinoma in adenomatous polyp of islets of Langerhans
36561813	8210/3-C25.7	Adenocarcinoma in adenomatous polyp of other specified parts of pancreas
36531468	8210/3-C25.8	Adenocarcinoma in adenomatous polyp of overlapping lesion of pancreas
36546281	8210/3-C25.9	Adenocarcinoma in adenomatous polyp of pancreas, NOS
36527068	8210/3-C25.3	Adenocarcinoma in adenomatous polyp of pancreatic duct
36527900	8210/3-C25.2	Adenocarcinoma in adenomatous polyp of tail of pancreas
36560940	8263/3-C25.1	Adenocarcinoma in tubulovillous adenoma of body of pancreas
36524288	8263/3-C25.0	Adenocarcinoma in tubulovillous adenoma of head of pancreas
36523537	8263/3-C25.4	Adenocarcinoma in tubulovillous adenoma of islets of Langerhans
36562260	8263/3-C25.7	Adenocarcinoma in tubulovillous adenoma of other specified parts of pancreas
36548594	8263/3-C25.8	Adenocarcinoma in tubulovillous adenoma of overlapping lesion of pancreas
36524859	8263/3-C25.9	Adenocarcinoma in tubulovillous adenoma of pancreas, NOS
36543167	8263/3-C25.3	Adenocarcinoma in tubulovillous adenoma of pancreatic duct
36543527	8263/3-C25.2	Adenocarcinoma in tubulovillous adenoma of tail of pancreas
36525324	8261/3-C25.1	Adenocarcinoma in villous adenoma of body of pancreas
36526839	8261/3-C25.0	Adenocarcinoma in villous adenoma of head of pancreas
36529419	8261/3-C25.4	Adenocarcinoma in villous adenoma of islets of Langerhans
36520322	8261/3-C25.7	Adenocarcinoma in villous adenoma of other specified parts of pancreas
36558301	8261/3-C25.8	Adenocarcinoma in villous adenoma of overlapping lesion of pancreas
36541498	8261/3-C25.9	Adenocarcinoma in villous adenoma of pancreas, NOS
36526967	8261/3-C25.3	Adenocarcinoma in villous adenoma of pancreatic duct
36527537	8261/3-C25.2	Adenocarcinoma in villous adenoma of tail of pancreas
36553219	8573/3-C25.1	Adenocarcinoma with apocrine metaplasia of body of pancreas
36527126	8573/3-C25.0	Adenocarcinoma with apocrine metaplasia of head of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36529348	8573/3-C25.4	Adenocarcinoma with apocrine metaplasia of islets of Langerhans
36549960	8573/3-C25.7	Adenocarcinoma with apocrine metaplasia of other specified parts of pancreas
36553040	8573/3-C25.8	Adenocarcinoma with apocrine metaplasia of overlapping lesion of pancreas
36561577	8573/3-C25.9	Adenocarcinoma with apocrine metaplasia of pancreas, NOS
36548294	8573/3-C25.3	Adenocarcinoma with apocrine metaplasia of pancreatic duct
36535779	8573/3-C25.2	Adenocarcinoma with apocrine metaplasia of tail of pancreas
36551494	8571/3-C25.1	Adenocarcinoma with cartilaginous and osseous metaplasia of body of pancreas
36529147	8571/3-C25.0	Adenocarcinoma with cartilaginous and osseous metaplasia of head of pancreas
36517883	8571/3-C25.4	Adenocarcinoma with cartilaginous and osseous metaplasia of islets of Langerhans
36528094	8571/3-C25.7	Adenocarcinoma with cartilaginous and osseous metaplasia of other specified parts of
36563485	8571/3-C25.8	Adenocarcinoma with cartilaginous and osseous metaplasia of overlapping lesion of
36566171	8571/3-C25.9	Adenocarcinoma with cartilaginous and osseous metaplasia of pancreas, NOS
36542649	8571/3-C25.3	Adenocarcinoma with cartilaginous and osseous metaplasia of pancreatic duct
36518986	8571/3-C25.2	Adenocarcinoma with cartilaginous and osseous metaplasia of tail of pancreas
44499965	8255/3-C25.1	Adenocarcinoma with mixed subtypes of body of pancreas
44502941	8255/3-C25.0	Adenocarcinoma with mixed subtypes of head of pancreas
36522699	8255/3-C25.4	Adenocarcinoma with mixed subtypes of islets of Langerhans
36519920	8255/3-C25.7	Adenocarcinoma with mixed subtypes of other specified parts of pancreas
36545921	8255/3-C25.8	Adenocarcinoma with mixed subtypes of overlapping lesion of pancreas
44501082	8255/3-C25.9	Adenocarcinoma with mixed subtypes of pancreas, NOS
36536754	8255/3-C25.3	Adenocarcinoma with mixed subtypes of pancreatic duct
44500024	8255/3-C25.2	Adenocarcinoma with mixed subtypes of tail of pancreas
44501586	8574/3-C25.1	Adenocarcinoma with neuroendocrine differentiation of body of pancreas
36564417	8574/3-C25.0	Adenocarcinoma with neuroendocrine differentiation of head of pancreas
36557393	8574/3-C25.4	Adenocarcinoma with neuroendocrine differentiation of islets of Langerhans
36542034	8574/3-C25.7	Adenocarcinoma with neuroendocrine differentiation of other specified parts of pancreas
36567271	8574/3-C25.8	Adenocarcinoma with neuroendocrine differentiation of overlapping lesion of pancreas
36563586	8574/3-C25.9	Adenocarcinoma with neuroendocrine differentiation of pancreas, NOS
36566559	8574/3-C25.3	Adenocarcinoma with neuroendocrine differentiation of pancreatic duct
36552225	8574/3-C25.2	Adenocarcinoma with neuroendocrine differentiation of tail of pancreas
36529493	8572/3-C25.1	Adenocarcinoma with spindle cell metaplasia of body of pancreas
36536510	8572/3-C25.0	Adenocarcinoma with spindle cell metaplasia of head of pancreas
36525089	8572/3-C25.4	Adenocarcinoma with spindle cell metaplasia of islets of Langerhans
36544775	8572/3-C25.7	Adenocarcinoma with spindle cell metaplasia of other specified parts of pancreas
36536231	8572/3-C25.8	Adenocarcinoma with spindle cell metaplasia of overlapping lesion of pancreas
36558642	8572/3-C25.9	Adenocarcinoma with spindle cell metaplasia of pancreas, NOS
36561682	8572/3-C25.3	Adenocarcinoma with spindle cell metaplasia of pancreatic duct

Concept ID	ICD-O-3 Codes[1]	Name
36553634	8572/3-C25.2	Adenocarcinoma with spindle cell metaplasia of tail of pancreas
36518298	8570/3-C25.1	Adenocarcinoma with squamous metaplasia of body of pancreas
36566684	8570/3-C25.0	Adenocarcinoma with squamous metaplasia of head of pancreas
36567248	8570/3-C25.4	Adenocarcinoma with squamous metaplasia of islets of Langerhans
36550922	8570/3-C25.7	Adenocarcinoma with squamous metaplasia of other specified parts of pancreas
36552940	8570/3-C25.8	Adenocarcinoma with squamous metaplasia of overlapping lesion of pancreas
36521427	8570/3-C25.9	Adenocarcinoma with squamous metaplasia of pancreas, NOS
36552462	8570/3-C25.3	Adenocarcinoma with squamous metaplasia of pancreatic duct
36534876	8570/3-C25.2	Adenocarcinoma with squamous metaplasia of tail of pancreas
42512017	8144/3-C25.0	Adenocarcinoma, intestinal type of head of pancreas
1553231	8144/3-C25.9	Adenocarcinoma, intestinal type of pancreas, NOS
36554902	8140/3-C25.4	Adenocarcinoma, NOS, of islets of Langerhans
44500872	8140/3-C25.7	Adenocarcinoma, NOS, of other specified parts of pancreas
36565700	8140/3-C25.8	Adenocarcinoma, NOS, of overlapping lesion of pancreas
44500011	8140/3-C25.2	Adenocarcinoma, NOS, of tail of pancreas
44503160	8560/3-C25.1	Adenosquamous carcinoma of body of pancreas
44499977	8560/3-C25.0	Adenosquamous carcinoma of head of pancreas
36555211	8560/3-C25.4	Adenosquamous carcinoma of islets of Langerhans
36527718	8560/3-C25.7	Adenosquamous carcinoma of other specified parts of pancreas
36545449	8560/3-C25.8	Adenosquamous carcinoma of overlapping lesion of pancreas
36545325	8560/3-C25.9	Adenosquamous carcinoma of pancreas, NOS
36532856	8560/3-C25.3	Adenosquamous carcinoma of pancreatic duct
44502960	8560/3-C25.2	Adenosquamous carcinoma of tail of pancreas
36566423	8147/3-C25.1	Basal cell adenocarcinoma of body of pancreas
36555095	8147/3-C25.0	Basal cell adenocarcinoma of head of pancreas
36532279	8147/3-C25.4	Basal cell adenocarcinoma of islets of Langerhans
36548809	8147/3-C25.7	Basal cell adenocarcinoma of other specified parts of pancreas
36566513	8147/3-C25.8	Basal cell adenocarcinoma of overlapping lesion of pancreas
36561541	8147/3-C25.9	Basal cell adenocarcinoma of pancreas, NOS
36517949	8147/3-C25.3	Basal cell adenocarcinoma of pancreatic duct
36538230	8147/3-C25.2	Basal cell adenocarcinoma of tail of pancreas
42512295	8083/3-C25.2	Basaloid squamous cell carcinoma of tail of pancreas
36538219	8231/3-C25.1	Carcinoma simplex of body of pancreas
36559274	8231/3-C25.0	Carcinoma simplex of head of pancreas
36525530	8231/3-C25.4	Carcinoma simplex of islets of Langerhans
36523970	8231/3-C25.7	Carcinoma simplex of other specified parts of pancreas
36545252	8231/3-C25.8	Carcinoma simplex of overlapping lesion of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36556407	8231/3-C25.9	Carcinoma simplex of pancreas, NOS
36551212	8231/3-C25.3	Carcinoma simplex of pancreatic duct
36536273	8231/3-C25.2	Carcinoma simplex of tail of pancreas
36531801	8035/3-C25.1	Carcinoma with osteoclast-like giant cells of body of pancreas
36527971	8035/3-C25.0	Carcinoma with osteoclast-like giant cells of head of pancreas
36547808	8035/3-C25.4	Carcinoma with osteoclast-like giant cells of islets of Langerhans
36556757	8035/3-C25.7	Carcinoma with osteoclast-like giant cells of other specified parts of pancreas
36551123	8035/3-C25.8	Carcinoma with osteoclast-like giant cells of overlapping lesion of pancreas
36538901	8035/3-C25.9	Carcinoma with osteoclast-like giant cells of pancreas, NOS
36561209	8035/3-C25.3	Carcinoma with osteoclast-like giant cells of pancreatic duct
36533104	8035/3-C25.2	Carcinoma with osteoclast-like giant cells of tail of pancreas
36554976	8021/3-C25.1	Carcinoma, anaplastic, NOS, of body of pancreas
44499789	8021/3-C25.0	Carcinoma, anaplastic, NOS, of head of pancreas
36532385	8021/3-C25.4	Carcinoma, anaplastic, NOS, of islets of Langerhans
36520068	8021/3-C25.7	Carcinoma, anaplastic, NOS, of other specified parts of pancreas
36557310	8021/3-C25.8	Carcinoma, anaplastic, NOS, of overlapping lesion of pancreas
36536431	8021/3-C25.9	Carcinoma, anaplastic, NOS, of pancreas, NOS
36524764	8021/3-C25.3	Carcinoma, anaplastic, NOS, of pancreatic duct
44502569	8021/3-C25.2	Carcinoma, anaplastic, NOS, of tail of pancreas
44501505	8010/3-C25.7	Carcinoma, NOS, of other specified parts of pancreas
36531610	8010/3-C25.8	Carcinoma, NOS, of overlapping lesion of pancreas
44500916	8010/3-C25.3	Carcinoma, NOS, of pancreatic duct
36530712	8020/3-C25.1	Carcinoma, undifferentiated, NOS, of body of pancreas
36567126	8020/3-C25.0	Carcinoma, undifferentiated, NOS, of head of pancreas
36521773	8020/3-C25.4	Carcinoma, undifferentiated, NOS, of islets of Langerhans
36565565	8020/3-C25.7	Carcinoma, undifferentiated, NOS, of other specified parts of pancreas
36540610	8020/3-C25.8	Carcinoma, undifferentiated, NOS, of overlapping lesion of pancreas
36560164	8020/3-C25.9	Carcinoma, undifferentiated, NOS, of pancreas, NOS
36554556	8020/3-C25.3	Carcinoma, undifferentiated, NOS, of pancreatic duct
36553250	8020/3-C25.2	Carcinoma, undifferentiated, NOS, of tail of pancreas
36553705	8981/3-C25.1	Carcinosarcoma, embryonal of body of pancreas
36563549	8981/3-C25.0	Carcinosarcoma, embryonal of head of pancreas
36559004	8981/3-C25.4	Carcinosarcoma, embryonal of islets of Langerhans
36535785	8981/3-C25.7	Carcinosarcoma, embryonal of other specified parts of pancreas
36563378	8981/3-C25.8	Carcinosarcoma, embryonal of overlapping lesion of pancreas
36528939	8981/3-C25.9	Carcinosarcoma, embryonal of pancreas, NOS
36530251	8981/3-C25.3	Carcinosarcoma, embryonal of pancreatic duct

Concept ID	ICD-O-3 Codes[1]	Name
36526267	8981/3-C25.2	Carcinosarcoma, embryonal of tail of pancreas
36523087	8980/3-C25.1	Carcinosarcoma, NOS, of body of pancreas
44502650	8980/3-C25.0	Carcinosarcoma, NOS, of head of pancreas
36532048	8980/3-C25.4	Carcinosarcoma, NOS, of islets of Langerhans
36543305	8980/3-C25.7	Carcinosarcoma, NOS, of other specified parts of pancreas
36529459	8980/3-C25.8	Carcinosarcoma, NOS, of overlapping lesion of pancreas
44502406	8980/3-C25.9	Carcinosarcoma, NOS, of pancreas, NOS
36535914	8980/3-C25.3	Carcinosarcoma, NOS, of pancreatic duct
36539894	8980/3-C25.2	Carcinosarcoma, NOS, of tail of pancreas
36565326	8160/3-C25.8	Cholangiocarcinoma of overlapping lesion of pancreas
1553535	8160/3-C25.9	Cholangiocarcinoma of pancreas, NOS
36553522	8310/3-C25.1	Clear cell adenocarcinoma, NOS, of body of pancreas
44502777	8310/3-C25.0	Clear cell adenocarcinoma, NOS, of head of pancreas
36563828	8310/3-C25.4	Clear cell adenocarcinoma, NOS, of islets of Langerhans
36536710	8310/3-C25.7	Clear cell adenocarcinoma, NOS, of other specified parts of pancreas
36526485	8310/3-C25.8	Clear cell adenocarcinoma, NOS, of overlapping lesion of pancreas
36547730	8310/3-C25.9	Clear cell adenocarcinoma, NOS, of pancreas, NOS
36534500	8310/3-C25.3	Clear cell adenocarcinoma, NOS, of pancreatic duct
44502234	8310/3-C25.2	Clear cell adenocarcinoma, NOS, of tail of pancreas
42512776	8045/3-C25.0	Combined small cell carcinoma of head of pancreas
36536729	8501/3-C25.1	Comedocarcinoma, NOS, of body of pancreas
36531402	8501/3-C25.0	Comedocarcinoma, NOS, of head of pancreas
36517168	8501/3-C25.4	Comedocarcinoma, NOS, of islets of Langerhans
36554867	8501/3-C25.7	Comedocarcinoma, NOS, of other specified parts of pancreas
36554164	8501/3-C25.8	Comedocarcinoma, NOS, of overlapping lesion of pancreas
36550558	8501/3-C25.9	Comedocarcinoma, NOS, of pancreas, NOS
36556628	8501/3-C25.3	Comedocarcinoma, NOS, of pancreatic duct
36537539	8501/3-C25.2	Comedocarcinoma, NOS, of tail of pancreas
36547251	8440/3-C25.1	Cystadenocarcinoma, NOS, of body of pancreas
44499646	8440/3-C25.0	Cystadenocarcinoma, NOS, of head of pancreas
36555571	8440/3-C25.4	Cystadenocarcinoma, NOS, of islets of Langerhans
36558135	8440/3-C25.7	Cystadenocarcinoma, NOS, of other specified parts of pancreas
36532276	8440/3-C25.8	Cystadenocarcinoma, NOS, of overlapping lesion of pancreas
44504793	8440/3-C25.9	Cystadenocarcinoma, NOS, of pancreas, NOS
36534376	8440/3-C25.3	Cystadenocarcinoma, NOS, of pancreatic duct
36537113	8440/3-C25.2	Cystadenocarcinoma, NOS, of tail of pancreas
36558441	8514/3-C25.1	Duct carcinoma, desmoplastic type of body of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36567065	8514/3-C25.0	Duct carcinoma, desmoplastic type of head of pancreas
36529584	8514/3-C25.4	Duct carcinoma, desmoplastic type of islets of Langerhans
36531945	8514/3-C25.7	Duct carcinoma, desmoplastic type of other specified parts of pancreas
36553472	8514/3-C25.8	Duct carcinoma, desmoplastic type of overlapping lesion of pancreas
36526312	8514/3-C25.9	Duct carcinoma, desmoplastic type of pancreas, NOS
36552933	8514/3-C25.3	Duct carcinoma, desmoplastic type of pancreatic duct
36567058	8514/3-C25.2	Duct carcinoma, desmoplastic type of tail of pancreas
36525356	8504/3-C25.1	Encapsulated papillary carcinoma with invasion of body of pancreas
36533741	8504/3-C25.0	Encapsulated papillary carcinoma with invasion of head of pancreas
36566906	8504/3-C25.4	Encapsulated papillary carcinoma with invasion of islets of Langerhans
36530071	8504/3-C25.7	Encapsulated papillary carcinoma with invasion of other specified parts of pancreas
36541762	8504/3-C25.8	Encapsulated papillary carcinoma with invasion of overlapping lesion of pancreas
36566950	8504/3-C25.9	Encapsulated papillary carcinoma with invasion of pancreas, NOS
36547425	8504/3-C25.3	Encapsulated papillary carcinoma with invasion of pancreatic duct
36520577	8504/3-C25.2	Encapsulated papillary carcinoma with invasion of tail of pancreas
36529591	8241/3-C25.1	Enterochromaffin cell carcinoid of body of pancreas
36538651	8241/3-C25.0	Enterochromaffin cell carcinoid of head of pancreas
36562690	8241/3-C25.7	Enterochromaffin cell carcinoid of other specified parts of pancreas
36531534	8241/3-C25.8	Enterochromaffin cell carcinoid of overlapping lesion of pancreas
36558895	8241/3-C25.9	Enterochromaffin cell carcinoid of pancreas, NOS
36551560	8241/3-C25.3	Enterochromaffin cell carcinoid of pancreatic duct
36532194	8241/3-C25.2	Enterochromaffin cell carcinoid of tail of pancreas
36552544	8242/3-C25.1	Enterochromaffin-like cell tumor of body of pancreas
36535845	8242/3-C25.0	Enterochromaffin-like cell tumor of head of pancreas
36557721	8242/3-C25.4	Enterochromaffin-like cell tumor of islets of Langerhans
36541129	8242/3-C25.7	Enterochromaffin-like cell tumor of other specified parts of pancreas
36544547	8242/3-C25.8	Enterochromaffin-like cell tumor of overlapping lesion of pancreas
36542417	8242/3-C25.9	Enterochromaffin-like cell tumor of pancreas, NOS
36522778	8242/3-C25.3	Enterochromaffin-like cell tumor of pancreatic duct
36539294	8242/3-C25.2	Enterochromaffin-like cell tumor of tail of pancreas
36558838	8157/3-C25.1	Enteroglucagonoma, malignant of body of pancreas
36536923	8157/3-C25.0	Enteroglucagonoma, malignant of head of pancreas
36558214	8157/3-C25.4	Enteroglucagonoma, malignant of islets of Langerhans
36551149	8157/3-C25.7	Enteroglucagonoma, malignant of other specified parts of pancreas
36517216	8157/3-C25.8	Enteroglucagonoma, malignant of overlapping lesion of pancreas
36561476	8157/3-C25.9	Enteroglucagonoma, malignant of pancreas, NOS
36527283	8157/3-C25.3	Enteroglucagonoma, malignant of pancreatic duct

Concept ID	ICD-O-3 Codes[1]	Name
36563701	8157/3-C25.2	Enteroglucagonoma, malignant of tail of pancreas
36524839	8562/3-C25.1	Epithelial-myoepithelial carcinoma of body of pancreas
36551913	8562/3-C25.0	Epithelial-myoepithelial carcinoma of head of pancreas
36555061	8562/3-C25.4	Epithelial-myoepithelial carcinoma of islets of Langerhans
36529871	8562/3-C25.7	Epithelial-myoepithelial carcinoma of other specified parts of pancreas
36563335	8562/3-C25.8	Epithelial-myoepithelial carcinoma of overlapping lesion of pancreas
36518464	8562/3-C25.9	Epithelial-myoepithelial carcinoma of pancreas, NOS
36546377	8562/3-C25.3	Epithelial-myoepithelial carcinoma of pancreatic duct
36561074	8562/3-C25.2	Epithelial-myoepithelial carcinoma of tail of pancreas
36552706	8011/3-C25.1	Epithelioma, malignant of body of pancreas
36558885	8011/3-C25.0	Epithelioma, malignant of head of pancreas
36521022	8011/3-C25.4	Epithelioma, malignant of islets of Langerhans
36525443	8011/3-C25.7	Epithelioma, malignant of other specified parts of pancreas
36567199	8011/3-C25.8	Epithelioma, malignant of overlapping lesion of pancreas
36538656	8011/3-C25.9	Epithelioma, malignant of pancreas, NOS
36562610	8011/3-C25.3	Epithelioma, malignant of pancreatic duct
36543098	8011/3-C25.2	Epithelioma, malignant of tail of pancreas
36523618	8153/3-C25.1	Gastrinoma of body of pancreas
44501631	8153/3-C25.0	Gastrinoma of head of pancreas
36564706	8153/3-C25.4	Gastrinoma of islets of Langerhans
36566583	8153/3-C25.7	Gastrinoma of other specified parts of pancreas
36560505	8153/3-C25.8	Gastrinoma of overlapping lesion of pancreas
36531406	8153/3-C25.9	Gastrinoma of pancreas, NOS
36524458	8153/3-C25.3	Gastrinoma of pancreatic duct
36522355	8153/3-C25.2	Gastrinoma of tail of pancreas
36555467	8030/3-C25.1	Giant cell and spindle cell carcinoma of body of pancreas
36543905	8030/3-C25.0	Giant cell and spindle cell carcinoma of head of pancreas
36527253	8030/3-C25.4	Giant cell and spindle cell carcinoma of islets of Langerhans
36523824	8030/3-C25.7	Giant cell and spindle cell carcinoma of other specified parts of pancreas
36565004	8030/3-C25.8	Giant cell and spindle cell carcinoma of overlapping lesion of pancreas
36525538	8030/3-C25.9	Giant cell and spindle cell carcinoma of pancreas, NOS
36555128	8030/3-C25.3	Giant cell and spindle cell carcinoma of pancreatic duct
36544574	8030/3-C25.2	Giant cell and spindle cell carcinoma of tail of pancreas
36535958	8031/3-C25.1	Giant cell carcinoma of body of pancreas
36548854	8031/3-C25.0	Giant cell carcinoma of head of pancreas
36525451	8031/3-C25.4	Giant cell carcinoma of islets of Langerhans
36562416	8031/3-C25.7	Giant cell carcinoma of other specified parts of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36546073	8031/3-C25.8	Giant cell carcinoma of overlapping lesion of pancreas
36529887	8031/3-C25.9	Giant cell carcinoma of pancreas, NOS
36547156	8031/3-C25.3	Giant cell carcinoma of pancreatic duct
36563878	8031/3-C25.2	Giant cell carcinoma of tail of pancreas
36536265	8015/3-C25.1	Glassy cell carcinoma of body of pancreas
36543462	8015/3-C25.0	Glassy cell carcinoma of head of pancreas
36558531	8015/3-C25.4	Glassy cell carcinoma of islets of Langerhans
36529918	8015/3-C25.7	Glassy cell carcinoma of other specified parts of pancreas
36530115	8015/3-C25.8	Glassy cell carcinoma of overlapping lesion of pancreas
36558013	8015/3-C25.9	Glassy cell carcinoma of pancreas, NOS
36551900	8015/3-C25.3	Glassy cell carcinoma of pancreatic duct
36545422	8015/3-C25.2	Glassy cell carcinoma of tail of pancreas
36553478	8152/3-C25.1	Glucagonoma of body of pancreas
36541840	8152/3-C25.0	Glucagonoma of head of pancreas
44502016	8152/3-C25.4	Glucagonoma of islets of Langerhans
36529138	8152/3-C25.7	Glucagonoma of other specified parts of pancreas
36546772	8152/3-C25.8	Glucagonoma of overlapping lesion of pancreas
44501342	8152/3-C25.9	Glucagonoma of pancreas, NOS
36543195	8152/3-C25.3	Glucagonoma of pancreatic duct
36543505	8152/3-C25.2	Glucagonoma of tail of pancreas
36560997	8243/3-C25.1	Goblet cell carcinoid of body of pancreas
44502465	8243/3-C25.0	Goblet cell carcinoid of head of pancreas
36518714	8243/3-C25.4	Goblet cell carcinoid of islets of Langerhans
36556096	8243/3-C25.7	Goblet cell carcinoid of other specified parts of pancreas
36553763	8243/3-C25.8	Goblet cell carcinoid of overlapping lesion of pancreas
36541589	8243/3-C25.9	Goblet cell carcinoid of pancreas, NOS
36522828	8243/3-C25.3	Goblet cell carcinoid of pancreatic duct
36557786	8243/3-C25.2	Goblet cell carcinoid of tail of pancreas
36536865	8576/3-C25.1	Hepatoid adenocarcinoma of body of pancreas
36520813	8576/3-C25.0	Hepatoid adenocarcinoma of head of pancreas
36534946	8576/3-C25.4	Hepatoid adenocarcinoma of islets of Langerhans
36550666	8576/3-C25.7	Hepatoid adenocarcinoma of other specified parts of pancreas
36557059	8576/3-C25.8	Hepatoid adenocarcinoma of overlapping lesion of pancreas
36550800	8576/3-C25.9	Hepatoid adenocarcinoma of pancreas, NOS
36565567	8576/3-C25.3	Hepatoid adenocarcinoma of pancreatic duct
36545660	8576/3-C25.2	Hepatoid adenocarcinoma of tail of pancreas
36558867	8522/3-C25.0	Infiltrating duct and lobular carcinoma of head of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
44500418	8500/3-C25.1	Infiltrating duct carcinoma, NOS, of body of pancreas
44500460	8500/3-C25.0	Infiltrating duct carcinoma, NOS, of head of pancreas
36559933	8500/3-C25.4	Infiltrating duct carcinoma, NOS, of islets of Langerhans
44500076	8500/3-C25.7	Infiltrating duct carcinoma, NOS, of other specified parts of pancreas
36517321	8500/3-C25.8	Infiltrating duct carcinoma, NOS, of overlapping lesion of pancreas
44502780	8500/3-C25.9	Infiltrating duct carcinoma, NOS, of pancreas, NOS
44502113	8500/3-C25.3	Infiltrating duct carcinoma, NOS, of pancreatic duct
44503029	8500/3-C25.2	Infiltrating duct carcinoma, NOS, of tail of pancreas
44500078	8523/3-C25.0	Infiltrating duct mixed with other types of carcinoma of head of pancreas
36518941	8523/3-C25.9	Infiltrating duct mixed with other types of carcinoma of pancreas, NOS
42512697	8523/3-C25.2	Infiltrating duct mixed with other types of carcinoma of tail of pancreas
36554904	8521/3-C25.1	Infiltrating ductular carcinoma of body of pancreas
36559974	8521/3-C25.0	Infiltrating ductular carcinoma of head of pancreas
36529704	8521/3-C25.4	Infiltrating ductular carcinoma of islets of Langerhans
36552650	8521/3-C25.7	Infiltrating ductular carcinoma of other specified parts of pancreas
36525930	8521/3-C25.8	Infiltrating ductular carcinoma of overlapping lesion of pancreas
36532697	8521/3-C25.9	Infiltrating ductular carcinoma of pancreas, NOS
36546666	8521/3-C25.3	Infiltrating ductular carcinoma of pancreatic duct
36542231	8521/3-C25.2	Infiltrating ductular carcinoma of tail of pancreas
36552136	8151/3-C25.1	Insulinoma, NOS, of body of pancreas
44499637	8151/3-C25.0	Insulinoma, NOS, of head of pancreas
44499738	8151/3-C25.4	Insulinoma, NOS, of islets of Langerhans
36568319	8151/3-C25.7	Insulinoma, NOS, of other specified parts of pancreas
36533766	8151/3-C25.8	Insulinoma, NOS, of overlapping lesion of pancreas
36524669	8151/3-C25.3	Insulinoma, NOS, of pancreatic duct
36566698	8151/3-C25.2	Insulinoma, NOS, of tail of pancreas
36546008	8503/3-C25.1	Intraductal papillary adenocarcinoma with invasion of body of pancreas
36556819	8503/3-C25.0	Intraductal papillary adenocarcinoma with invasion of head of pancreas
36529474	8503/3-C25.4	Intraductal papillary adenocarcinoma with invasion of islets of Langerhans
36541054	8503/3-C25.7	Intraductal papillary adenocarcinoma with invasion of other specified parts of pancreas
36530527	8503/3-C25.8	Intraductal papillary adenocarcinoma with invasion of overlapping lesion of pancreas
36525859	8503/3-C25.9	Intraductal papillary adenocarcinoma with invasion of pancreas, NOS
36557420	8503/3-C25.3	Intraductal papillary adenocarcinoma with invasion of pancreatic duct
36554693	8503/3-C25.2	Intraductal papillary adenocarcinoma with invasion of tail of pancreas
36524030	8453/3-C25.1	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of body of
44501087	8453/3-C25.0	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of head of
36565136	8453/3-C25.4	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of islets of

Concept ID	ICD-O-3 Codes[1]	Name
36541803	8453/3-C25.7	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of other
36531307	8453/3-C25.8	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of
36537141	8453/3-C25.3	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of
36567666	8453/3-C25.2	Intraductal papillary mucinous neoplasm with an associated invasive carcinoma of tail of
36403005	8507/3-C25.0	Invasive micropapillary carcinoma of breast of head of pancreas
36547307	8014/3-C25.1	Large cell carcinoma with rhabdoid phenotype of body of pancreas
36527242	8014/3-C25.0	Large cell carcinoma with rhabdoid phenotype of head of pancreas
36557240	8014/3-C25.4	Large cell carcinoma with rhabdoid phenotype of islets of Langerhans
36549675	8014/3-C25.7	Large cell carcinoma with rhabdoid phenotype of other specified parts of pancreas
36552000	8014/3-C25.8	Large cell carcinoma with rhabdoid phenotype of overlapping lesion of pancreas
36559679	8014/3-C25.9	Large cell carcinoma with rhabdoid phenotype of pancreas, NOS
36526035	8014/3-C25.3	Large cell carcinoma with rhabdoid phenotype of pancreatic duct
36535551	8014/3-C25.2	Large cell carcinoma with rhabdoid phenotype of tail of pancreas
36559719	8012/3-C25.1	Large cell carcinoma, NOS, of body of pancreas
44499998	8012/3-C25.0	Large cell carcinoma, NOS, of head of pancreas
36525740	8012/3-C25.4	Large cell carcinoma, NOS, of islets of Langerhans
36547636	8012/3-C25.7	Large cell carcinoma, NOS, of other specified parts of pancreas
36563560	8012/3-C25.8	Large cell carcinoma, NOS, of overlapping lesion of pancreas
36521136	8012/3-C25.9	Large cell carcinoma, NOS, of pancreas, NOS
36538299	8012/3-C25.3	Large cell carcinoma, NOS, of pancreatic duct
36562646	8012/3-C25.2	Large cell carcinoma, NOS, of tail of pancreas
36534612	8013/3-C25.1	Large cell neuroendocrine carcinoma of body of pancreas
44499999	8013/3-C25.0	Large cell neuroendocrine carcinoma of head of pancreas
36519036	8013/3-C25.4	Large cell neuroendocrine carcinoma of islets of Langerhans
36552024	8013/3-C25.7	Large cell neuroendocrine carcinoma of other specified parts of pancreas
36552759	8013/3-C25.8	Large cell neuroendocrine carcinoma of overlapping lesion of pancreas
36539779	8013/3-C25.9	Large cell neuroendocrine carcinoma of pancreas, NOS
36523139	8013/3-C25.3	Large cell neuroendocrine carcinoma of pancreatic duct
36545961	8013/3-C25.2	Large cell neuroendocrine carcinoma of tail of pancreas
42512445	9251/3-C25.0	Malignant giant cell tumor of soft parts of head of pancreas
36526835	8005/3-C25.1	Malignant tumor, clear cell type of body of pancreas
36536542	8005/3-C25.0	Malignant tumor, clear cell type of head of pancreas
36539072	8005/3-C25.4	Malignant tumor, clear cell type of islets of Langerhans
36539008	8005/3-C25.7	Malignant tumor, clear cell type of other specified parts of pancreas
36539320	8005/3-C25.8	Malignant tumor, clear cell type of overlapping lesion of pancreas
36553562	8005/3-C25.9	Malignant tumor, clear cell type of pancreas, NOS
36560075	8005/3-C25.3	Malignant tumor, clear cell type of pancreatic duct

Concept ID	ICD-O-3 Codes[1]	Name
36531020	8005/3-C25.2	Malignant tumor, clear cell type of tail of pancreas
36561298	8003/3-C25.1	Malignant tumor, giant cell type of body of pancreas
36565019	8003/3-C25.0	Malignant tumor, giant cell type of head of pancreas
36550315	8003/3-C25.4	Malignant tumor, giant cell type of islets of Langerhans
36558964	8003/3-C25.7	Malignant tumor, giant cell type of other specified parts of pancreas
36526566	8003/3-C25.8	Malignant tumor, giant cell type of overlapping lesion of pancreas
36543227	8003/3-C25.9	Malignant tumor, giant cell type of pancreas, NOS
36559091	8003/3-C25.3	Malignant tumor, giant cell type of pancreatic duct
36536046	8003/3-C25.2	Malignant tumor, giant cell type of tail of pancreas
36527033	8002/3-C25.1	Malignant tumor, small cell type of body of pancreas
36548296	8002/3-C25.0	Malignant tumor, small cell type of head of pancreas
36561072	8002/3-C25.4	Malignant tumor, small cell type of islets of Langerhans
36520510	8002/3-C25.7	Malignant tumor, small cell type of other specified parts of pancreas
36532137	8002/3-C25.8	Malignant tumor, small cell type of overlapping lesion of pancreas
36526425	8002/3-C25.9	Malignant tumor, small cell type of pancreas, NOS
36527104	8002/3-C25.3	Malignant tumor, small cell type of pancreatic duct
36543245	8002/3-C25.2	Malignant tumor, small cell type of tail of pancreas
36548578	8004/3-C25.1	Malignant tumor, spindle cell type of body of pancreas
36541237	8004/3-C25.0	Malignant tumor, spindle cell type of head of pancreas
36543340	8004/3-C25.4	Malignant tumor, spindle cell type of islets of Langerhans
36529808	8004/3-C25.7	Malignant tumor, spindle cell type of other specified parts of pancreas
36531678	8004/3-C25.8	Malignant tumor, spindle cell type of overlapping lesion of pancreas
36540475	8004/3-C25.9	Malignant tumor, spindle cell type of pancreas, NOS
36549994	8004/3-C25.3	Malignant tumor, spindle cell type of pancreatic duct
36531790	8004/3-C25.2	Malignant tumor, spindle cell type of tail of pancreas
36551522	8510/3-C25.1	Medullary carcinoma, NOS, of body of pancreas
36564218	8510/3-C25.0	Medullary carcinoma, NOS, of head of pancreas
36558945	8510/3-C25.4	Medullary carcinoma, NOS, of islets of Langerhans
36562185	8510/3-C25.7	Medullary carcinoma, NOS, of other specified parts of pancreas
36563151	8510/3-C25.8	Medullary carcinoma, NOS, of overlapping lesion of pancreas
36549995	8510/3-C25.9	Medullary carcinoma, NOS, of pancreas, NOS
36557144	8510/3-C25.3	Medullary carcinoma, NOS, of pancreatic duct
36520646	8510/3-C25.2	Medullary carcinoma, NOS, of tail of pancreas
36546695	8575/3-C25.1	Metaplastic carcinoma, NOS, of body of pancreas
36532549	8575/3-C25.0	Metaplastic carcinoma, NOS, of head of pancreas
36547361	8575/3-C25.4	Metaplastic carcinoma, NOS, of islets of Langerhans
36546789	8575/3-C25.7	Metaplastic carcinoma, NOS, of other specified parts of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36556998	8575/3-C25.8	Metaplastic carcinoma, NOS, of overlapping lesion of pancreas
36555052	8575/3-C25.9	Metaplastic carcinoma, NOS, of pancreas, NOS
36565307	8575/3-C25.3	Metaplastic carcinoma, NOS, of pancreatic duct
36540811	8575/3-C25.2	Metaplastic carcinoma, NOS, of tail of pancreas
42512815	8552/3-C25.1	Mixed acinar-ductal carcinoma of body of pancreas
42511661	8552/3-C25.0	Mixed acinar-ductal carcinoma of head of pancreas
42512319	8552/3-C25.4	Mixed acinar-ductal carcinoma of islets of Langerhans
42511728	8552/3-C25.7	Mixed acinar-ductal carcinoma of other specified parts of pancreas
42512937	8552/3-C25.8	Mixed acinar-ductal carcinoma of overlapping lesion of pancreas
42512190	8552/3-C25.9	Mixed acinar-ductal carcinoma of pancreas, NOS
42512915	8552/3-C25.3	Mixed acinar-ductal carcinoma of pancreatic duct
42512022	8552/3-C25.2	Mixed acinar-ductal carcinoma of tail of pancreas
44500249	8244/3-C25.1	Mixed adenoneuroendocrine carcinoma of body of pancreas
36560106	8244/3-C25.0	Mixed adenoneuroendocrine carcinoma of head of pancreas
36534315	8244/3-C25.4	Mixed adenoneuroendocrine carcinoma of islets of Langerhans
36552177	8244/3-C25.7	Mixed adenoneuroendocrine carcinoma of other specified parts of pancreas
36519101	8244/3-C25.8	Mixed adenoneuroendocrine carcinoma of overlapping lesion of pancreas
36528490	8244/3-C25.9	Mixed adenoneuroendocrine carcinoma of pancreas, NOS
36544439	8244/3-C25.3	Mixed adenoneuroendocrine carcinoma of pancreatic duct
36546697	8244/3-C25.2	Mixed adenoneuroendocrine carcinoma of tail of pancreas
36557606	8323/3-C25.1	Mixed cell adenocarcinoma of body of pancreas
44502538	8323/3-C25.0	Mixed cell adenocarcinoma of head of pancreas
36556636	8323/3-C25.4	Mixed cell adenocarcinoma of islets of Langerhans
36557209	8323/3-C25.7	Mixed cell adenocarcinoma of other specified parts of pancreas
36533132	8323/3-C25.8	Mixed cell adenocarcinoma of overlapping lesion of pancreas
36524270	8323/3-C25.9	Mixed cell adenocarcinoma of pancreas, NOS
36521669	8323/3-C25.3	Mixed cell adenocarcinoma of pancreatic duct
36544534	8323/3-C25.2	Mixed cell adenocarcinoma of tail of pancreas
36402508	8154/3-C25.1	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of body of pancreas
36402568	8154/3-C25.0	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of head of pancreas
36402631	8154/3-C25.4	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of islets of Langerhans
36402688	8154/3-C25.7	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of other specified parts of
36402526	8154/3-C25.8	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of overlapping lesion of
36402481	8154/3-C25.9	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of pancreas, NOS
36402372	8154/3-C25.3	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of pancreatic duct
36402494	8154/3-C25.2	Mixed neuroendocrine non-neuroendocrine neoplasm (MiNEN) of tail of pancreas
44502031	8481/3-C25.1	Mucin-producing adenocarcinoma of body of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
44503092	8481/3-C25.0	Mucin-producing adenocarcinoma of head of pancreas
36537866	8481/3-C25.4	Mucin-producing adenocarcinoma of islets of Langerhans
36521252	8481/3-C25.7	Mucin-producing adenocarcinoma of other specified parts of pancreas
36519053	8481/3-C25.8	Mucin-producing adenocarcinoma of overlapping lesion of pancreas
44501650	8481/3-C25.9	Mucin-producing adenocarcinoma of pancreas, NOS
36558561	8481/3-C25.3	Mucin-producing adenocarcinoma of pancreatic duct
44499649	8481/3-C25.2	Mucin-producing adenocarcinoma of tail of pancreas
44500145	8480/3-C25.1	Mucinous adenocarcinoma of body of pancreas
44501853	8480/3-C25.0	Mucinous adenocarcinoma of head of pancreas
36528147	8480/3-C25.4	Mucinous adenocarcinoma of islets of Langerhans
44502708	8480/3-C25.7	Mucinous adenocarcinoma of other specified parts of pancreas
36517631	8480/3-C25.8	Mucinous adenocarcinoma of overlapping lesion of pancreas
44500034	8480/3-C25.9	Mucinous adenocarcinoma of pancreas, NOS
36526404	8480/3-C25.3	Mucinous adenocarcinoma of pancreatic duct
44499444	8480/3-C25.2	Mucinous adenocarcinoma of tail of pancreas
44500651	8470/3-C25.1	Mucinous cystadenocarcinoma, NOS, of body of pancreas
44500200	8470/3-C25.0	Mucinous cystadenocarcinoma, NOS, of head of pancreas
36550808	8470/3-C25.4	Mucinous cystadenocarcinoma, NOS, of islets of Langerhans
36556285	8470/3-C25.7	Mucinous cystadenocarcinoma, NOS, of other specified parts of pancreas
36552424	8470/3-C25.8	Mucinous cystadenocarcinoma, NOS, of overlapping lesion of pancreas
36547750	8470/3-C25.3	Mucinous cystadenocarcinoma, NOS, of pancreatic duct
44500798	8470/3-C25.2	Mucinous cystadenocarcinoma, NOS, of tail of pancreas
36539751	8430/3-C25.1	Mucoepidermoid carcinoma of body of pancreas
36525621	8430/3-C25.0	Mucoepidermoid carcinoma of head of pancreas
36524647	8430/3-C25.4	Mucoepidermoid carcinoma of islets of Langerhans
36558552	8430/3-C25.7	Mucoepidermoid carcinoma of other specified parts of pancreas
36562424	8430/3-C25.8	Mucoepidermoid carcinoma of overlapping lesion of pancreas
36550711	8430/3-C25.9	Mucoepidermoid carcinoma of pancreas, NOS
36521630	8430/3-C25.3	Mucoepidermoid carcinoma of pancreatic duct
36525602	8430/3-C25.2	Mucoepidermoid carcinoma of tail of pancreas
44500046	8000/3-C25.7	Neoplasm, malignant of other specified parts of pancreas
36529877	8000/3-C25.8	Neoplasm, malignant of overlapping lesion of pancreas
42512910	9500/3-C25.0	Neuroblastoma, NOS, of head of pancreas
42512575	9500/3-C25.8	Neuroblastoma, NOS, of overlapping lesion of pancreas
44502530	8246/3-C25.1	Neuroendocrine carcinoma, NOS, of body of pancreas
44499542	8246/3-C25.0	Neuroendocrine carcinoma, NOS, of head of pancreas
44500064	8246/3-C25.4	Neuroendocrine carcinoma, NOS, of islets of Langerhans

Concept ID	ICD-O-3 Codes[1]	Name
44501725	8246/3-C25.7	Neuroendocrine carcinoma, NOS, of other specified parts of pancreas
36566617	8246/3-C25.8	Neuroendocrine carcinoma, NOS, of overlapping lesion of pancreas
36555360	8246/3-C25.3	Neuroendocrine carcinoma, NOS, of pancreatic duct
44500727	8246/3-C25.2	Neuroendocrine carcinoma, NOS, of tail of pancreas
36543944	8249/3-C25.1	Neuroendocrine tumor, grade 2 of body of pancreas
44501727	8249/3-C25.0	Neuroendocrine tumor, grade 2 of head of pancreas
36545045	8249/3-C25.4	Neuroendocrine tumor, grade 2 of islets of Langerhans
36519604	8249/3-C25.7	Neuroendocrine tumor, grade 2 of other specified parts of pancreas
36548610	8249/3-C25.8	Neuroendocrine tumor, grade 2 of overlapping lesion of pancreas
44502774	8249/3-C25.9	Neuroendocrine tumor, grade 2 of pancreas, NOS
36532703	8249/3-C25.3	Neuroendocrine tumor, grade 2 of pancreatic duct
44502279	8249/3-C25.2	Neuroendocrine tumor, grade 2 of tail of pancreas
44501723	8240/3-C25.1	Neuroendocrine tumor, NOS, of body of pancreas
44499437	8240/3-C25.0	Neuroendocrine tumor, NOS, of head of pancreas
36546782	8240/3-C25.4	Neuroendocrine tumor, NOS, of islets of Langerhans
44500864	8240/3-C25.7	Neuroendocrine tumor, NOS, of other specified parts of pancreas
36551684	8240/3-C25.8	Neuroendocrine tumor, NOS, of overlapping lesion of pancreas
44500593	8240/3-C25.9	Neuroendocrine tumor, NOS, of pancreas, NOS
36528010	8240/3-C25.3	Neuroendocrine tumor, NOS, of pancreatic duct
44500022	8240/3-C25.2	Neuroendocrine tumor, NOS, of tail of pancreas
44502511	8046/3-C25.0	Non-small cell carcinoma of head of pancreas
44502767	8046/3-C25.9	Non-small cell carcinoma of pancreas, NOS
36552457	8046/3-C25.2	Non-small cell carcinoma of tail of pancreas
44501714	8150/3-C25.1	Pancreatic neuroendocrine tumor, nonfunctioning of body of pancreas
44498951	8150/3-C25.0	Pancreatic neuroendocrine tumor, nonfunctioning of head of pancreas
44502097	8150/3-C25.4	Pancreatic neuroendocrine tumor, nonfunctioning of islets of Langerhans
44500014	8150/3-C25.7	Pancreatic neuroendocrine tumor, nonfunctioning of other specified parts of pancreas
36518642	8150/3-C25.8	Pancreatic neuroendocrine tumor, nonfunctioning of overlapping lesion of pancreas
44502172	8150/3-C25.9	Pancreatic neuroendocrine tumor, nonfunctioning of pancreas, NOS
36564262	8150/3-C25.3	Pancreatic neuroendocrine tumor, nonfunctioning of pancreatic duct
44499885	8150/3-C25.2	Pancreatic neuroendocrine tumor, nonfunctioning of tail of pancreas
1553359	8163/3-C25.0	Pancreatobiliary type carcinoma of head of pancreas
1553353	8163/3-C25.9	Pancreatobiliary type carcinoma of pancreas, NOS
36565815	8971/3-C25.1	Pancreatoblastoma of body of pancreas
44501981	8971/3-C25.0	Pancreatoblastoma of head of pancreas
36563101	8971/3-C25.4	Pancreatoblastoma of islets of Langerhans
36533780	8971/3-C25.7	Pancreatoblastoma of other specified parts of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36547408	8971/3-C25.8	Pancreatoblastoma of overlapping lesion of pancreas
36563917	8971/3-C25.9	Pancreatoblastoma of pancreas, NOS
36546996	8971/3-C25.3	Pancreatoblastoma of pancreatic duct
36540366	8971/3-C25.2	Pancreatoblastoma of tail of pancreas
44500791	8260/3-C25.1	Papillary adenocarcinoma, NOS, of body of pancreas
44502822	8260/3-C25.0	Papillary adenocarcinoma, NOS, of head of pancreas
36551544	8260/3-C25.4	Papillary adenocarcinoma, NOS, of islets of Langerhans
36561013	8260/3-C25.7	Papillary adenocarcinoma, NOS, of other specified parts of pancreas
36559463	8260/3-C25.8	Papillary adenocarcinoma, NOS, of overlapping lesion of pancreas
36554498	8260/3-C25.9	Papillary adenocarcinoma, NOS, of pancreas, NOS
36522112	8260/3-C25.3	Papillary adenocarcinoma, NOS, of pancreatic duct
36529253	8260/3-C25.2	Papillary adenocarcinoma, NOS, of tail of pancreas
36527334	8050/3-C25.1	Papillary carcinoma, NOS, of body of pancreas
36538825	8050/3-C25.0	Papillary carcinoma, NOS, of head of pancreas
36538132	8050/3-C25.4	Papillary carcinoma, NOS, of islets of Langerhans
36562198	8050/3-C25.7	Papillary carcinoma, NOS, of other specified parts of pancreas
36523744	8050/3-C25.8	Papillary carcinoma, NOS, of overlapping lesion of pancreas
36521761	8050/3-C25.9	Papillary carcinoma, NOS, of pancreas, NOS
36559572	8050/3-C25.3	Papillary carcinoma, NOS, of pancreatic duct
36537718	8050/3-C25.2	Papillary carcinoma, NOS, of tail of pancreas
36544923	8450/3-C25.1	Papillary cystadenocarcinoma, NOS, of body of pancreas
36541519	8450/3-C25.0	Papillary cystadenocarcinoma, NOS, of head of pancreas
36545679	8450/3-C25.4	Papillary cystadenocarcinoma, NOS, of islets of Langerhans
36547867	8450/3-C25.7	Papillary cystadenocarcinoma, NOS, of other specified parts of pancreas
36534321	8450/3-C25.8	Papillary cystadenocarcinoma, NOS, of overlapping lesion of pancreas
36518993	8450/3-C25.9	Papillary cystadenocarcinoma, NOS, of pancreas, NOS
36531800	8450/3-C25.3	Papillary cystadenocarcinoma, NOS, of pancreatic duct
36526653	8450/3-C25.2	Papillary cystadenocarcinoma, NOS, of tail of pancreas
36554557	8052/3-C25.1	Papillary squamous cell carcinoma of body of pancreas
36557533	8052/3-C25.0	Papillary squamous cell carcinoma of head of pancreas
36550079	8052/3-C25.4	Papillary squamous cell carcinoma of islets of Langerhans
36542603	8052/3-C25.7	Papillary squamous cell carcinoma of other specified parts of pancreas
36544209	8052/3-C25.8	Papillary squamous cell carcinoma of overlapping lesion of pancreas
36567687	8052/3-C25.9	Papillary squamous cell carcinoma of pancreas, NOS
36551753	8052/3-C25.3	Papillary squamous cell carcinoma of pancreatic duct
36524944	8052/3-C25.2	Papillary squamous cell carcinoma of tail of pancreas
42512113	8680/3-C25.0	Paraganglioma, NOS, of head of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
44502292	8680/3-C25.4	Paraganglioma, NOS, of islets of Langerhans
44500260	8680/3-C25.9	Paraganglioma, NOS, of pancreas, NOS
44502017	8162/3-C25.0	Perihilar cholangiocarcinoma of head of pancreas
44500116	8022/3-C25.1	Pleomorphic carcinoma of body of pancreas
44502214	8022/3-C25.0	Pleomorphic carcinoma of head of pancreas
36524612	8022/3-C25.4	Pleomorphic carcinoma of islets of Langerhans
36560386	8022/3-C25.7	Pleomorphic carcinoma of other specified parts of pancreas
36518705	8022/3-C25.8	Pleomorphic carcinoma of overlapping lesion of pancreas
44501700	8022/3-C25.9	Pleomorphic carcinoma of pancreas, NOS
36524055	8022/3-C25.3	Pleomorphic carcinoma of pancreatic duct
44501467	8022/3-C25.2	Pleomorphic carcinoma of tail of pancreas
36556735	8034/3-C25.1	Polygonal cell carcinoma of body of pancreas
36519114	8034/3-C25.0	Polygonal cell carcinoma of head of pancreas
36540657	8034/3-C25.4	Polygonal cell carcinoma of islets of Langerhans
36555148	8034/3-C25.7	Polygonal cell carcinoma of other specified parts of pancreas
36522133	8034/3-C25.8	Polygonal cell carcinoma of overlapping lesion of pancreas
36537838	8034/3-C25.9	Polygonal cell carcinoma of pancreas, NOS
36558073	8034/3-C25.3	Polygonal cell carcinoma of pancreatic duct
36525780	8034/3-C25.2	Polygonal cell carcinoma of tail of pancreas
44503193	8033/3-C25.1	Pseudosarcomatous carcinoma of body of pancreas
44500338	8033/3-C25.0	Pseudosarcomatous carcinoma of head of pancreas
36533481	8033/3-C25.4	Pseudosarcomatous carcinoma of islets of Langerhans
36521685	8033/3-C25.7	Pseudosarcomatous carcinoma of other specified parts of pancreas
36553026	8033/3-C25.8	Pseudosarcomatous carcinoma of overlapping lesion of pancreas
36538136	8033/3-C25.9	Pseudosarcomatous carcinoma of pancreas, NOS
36562518	8033/3-C25.3	Pseudosarcomatous carcinoma of pancreatic duct
36532823	8033/3-C25.2	Pseudosarcomatous carcinoma of tail of pancreas
44503143	8141/3-C25.1	Scirrhous adenocarcinoma of body of pancreas
36531436	8141/3-C25.0	Scirrhous adenocarcinoma of head of pancreas
36538908	8141/3-C25.4	Scirrhous adenocarcinoma of islets of Langerhans
36533396	8141/3-C25.7	Scirrhous adenocarcinoma of other specified parts of pancreas
36564986	8141/3-C25.8	Scirrhous adenocarcinoma of overlapping lesion of pancreas
36526179	8141/3-C25.9	Scirrhous adenocarcinoma of pancreas, NOS
36526955	8141/3-C25.3	Scirrhous adenocarcinoma of pancreatic duct
44499883	8141/3-C25.2	Scirrhous adenocarcinoma of tail of pancreas
36561639	8441/3-C25.1	Serous carcinoma, NOS, of body of pancreas
36537054	8441/3-C25.0	Serous carcinoma, NOS, of head of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36551786	8441/3-C25.4	Serous carcinoma, NOS, of islets of Langerhans
36527497	8441/3-C25.7	Serous carcinoma, NOS, of other specified parts of pancreas
36540547	8441/3-C25.8	Serous carcinoma, NOS, of overlapping lesion of pancreas
36533910	8441/3-C25.3	Serous carcinoma, NOS, of pancreatic duct
36539442	8441/3-C25.2	Serous carcinoma, NOS, of tail of pancreas
44501445	8490/3-C25.1	Signet ring cell carcinoma of body of pancreas
44501651	8490/3-C25.0	Signet ring cell carcinoma of head of pancreas
36536777	8490/3-C25.4	Signet ring cell carcinoma of islets of Langerhans
36561348	8490/3-C25.7	Signet ring cell carcinoma of other specified parts of pancreas
36522381	8490/3-C25.8	Signet ring cell carcinoma of overlapping lesion of pancreas
44502710	8490/3-C25.9	Signet ring cell carcinoma of pancreas, NOS
44502241	8490/3-C25.3	Signet ring cell carcinoma of pancreatic duct
44503157	8490/3-C25.2	Signet ring cell carcinoma of tail of pancreas
36532980	8043/3-C25.1	Small cell carcinoma, fusiform cell of body of pancreas
36538255	8043/3-C25.0	Small cell carcinoma, fusiform cell of head of pancreas
36549488	8043/3-C25.4	Small cell carcinoma, fusiform cell of islets of Langerhans
36545608	8043/3-C25.7	Small cell carcinoma, fusiform cell of other specified parts of pancreas
36524381	8043/3-C25.8	Small cell carcinoma, fusiform cell of overlapping lesion of pancreas
36531915	8043/3-C25.9	Small cell carcinoma, fusiform cell of pancreas, NOS
36534149	8043/3-C25.3	Small cell carcinoma, fusiform cell of pancreatic duct
36519938	8043/3-C25.2	Small cell carcinoma, fusiform cell of tail of pancreas
44500186	8041/3-C25.1	Small cell carcinoma, NOS, of body of pancreas
44500711	8041/3-C25.0	Small cell carcinoma, NOS, of head of pancreas
36549084	8041/3-C25.4	Small cell carcinoma, NOS, of islets of Langerhans
36549603	8041/3-C25.7	Small cell carcinoma, NOS, of other specified parts of pancreas
36563075	8041/3-C25.8	Small cell carcinoma, NOS, of overlapping lesion of pancreas
44499868	8041/3-C25.9	Small cell carcinoma, NOS, of pancreas, NOS
36517301	8041/3-C25.3	Small cell carcinoma, NOS, of pancreatic duct
36533022	8041/3-C25.2	Small cell carcinoma, NOS, of tail of pancreas
36537607	8230/3-C25.1	Solid carcinoma, NOS, of body of pancreas
36526062	8230/3-C25.0	Solid carcinoma, NOS, of head of pancreas
36523566	8230/3-C25.4	Solid carcinoma, NOS, of islets of Langerhans
36537548	8230/3-C25.7	Solid carcinoma, NOS, of other specified parts of pancreas
36555814	8230/3-C25.8	Solid carcinoma, NOS, of overlapping lesion of pancreas
36546659	8230/3-C25.9	Solid carcinoma, NOS, of pancreas, NOS
36539281	8230/3-C25.3	Solid carcinoma, NOS, of pancreatic duct
36554066	8230/3-C25.2	Solid carcinoma, NOS, of tail of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36520941	8452/3-C25.1	Solid pseudopapillary neoplasm of pancreas of body of pancreas
36567644	8452/3-C25.0	Solid pseudopapillary neoplasm of pancreas of head of pancreas
36543806	8452/3-C25.4	Solid pseudopapillary neoplasm of pancreas of islets of Langerhans
36532986	8452/3-C25.7	Solid pseudopapillary neoplasm of pancreas of other specified parts of pancreas
36539200	8452/3-C25.8	Solid pseudopapillary neoplasm of pancreas of overlapping lesion of pancreas
36565792	8452/3-C25.3	Solid pseudopapillary neoplasm of pancreas of pancreatic duct
44501527	8452/3-C25.2	Solid pseudopapillary neoplasm of pancreas of tail of pancreas
36535572	8156/3-C25.1	Somatostatinoma of body of pancreas
36552308	8156/3-C25.0	Somatostatinoma of head of pancreas
36565423	8156/3-C25.4	Somatostatinoma of islets of Langerhans
36547203	8156/3-C25.7	Somatostatinoma of other specified parts of pancreas
36533510	8156/3-C25.8	Somatostatinoma of overlapping lesion of pancreas
36567520	8156/3-C25.9	Somatostatinoma of pancreas, NOS
36526473	8156/3-C25.3	Somatostatinoma of pancreatic duct
36532277	8156/3-C25.2	Somatostatinoma of tail of pancreas
36542553	8032/3-C25.1	Spindle cell carcinoma, NOS, of body of pancreas
44501766	8032/3-C25.0	Spindle cell carcinoma, NOS, of head of pancreas
36537573	8032/3-C25.4	Spindle cell carcinoma, NOS, of islets of Langerhans
36537418	8032/3-C25.7	Spindle cell carcinoma, NOS, of other specified parts of pancreas
36557909	8032/3-C25.8	Spindle cell carcinoma, NOS, of overlapping lesion of pancreas
36551708	8032/3-C25.9	Spindle cell carcinoma, NOS, of pancreas, NOS
36528423	8032/3-C25.3	Spindle cell carcinoma, NOS, of pancreatic duct
36546106	8032/3-C25.2	Spindle cell carcinoma, NOS, of tail of pancreas
36530531	8078/3-C25.1	Squamous cell carcinoma with horn formation of body of pancreas
36564137	8078/3-C25.0	Squamous cell carcinoma with horn formation of head of pancreas
36544472	8078/3-C25.4	Squamous cell carcinoma with horn formation of islets of Langerhans
36554581	8078/3-C25.7	Squamous cell carcinoma with horn formation of other specified parts of pancreas
36550757	8078/3-C25.8	Squamous cell carcinoma with horn formation of overlapping lesion of pancreas
36532039	8078/3-C25.9	Squamous cell carcinoma with horn formation of pancreas, NOS
36545005	8078/3-C25.3	Squamous cell carcinoma with horn formation of pancreatic duct
36559937	8078/3-C25.2	Squamous cell carcinoma with horn formation of tail of pancreas
36517192	8075/3-C25.1	Squamous cell carcinoma, adenoid of body of pancreas
36525449	8075/3-C25.0	Squamous cell carcinoma, adenoid of head of pancreas
36538810	8075/3-C25.4	Squamous cell carcinoma, adenoid of islets of Langerhans
36530183	8075/3-C25.7	Squamous cell carcinoma, adenoid of other specified parts of pancreas
36535671	8075/3-C25.8	Squamous cell carcinoma, adenoid of overlapping lesion of pancreas
36560330	8075/3-C25.9	Squamous cell carcinoma, adenoid of pancreas, NOS

Concept ID	ICD-O-3 Codes[1]	Name
36544241	8075/3-C25.3	Squamous cell carcinoma, adenoid of pancreatic duct
36527557	8075/3-C25.2	Squamous cell carcinoma, adenoid of tail of pancreas
36549407	8071/3-C25.1	Squamous cell carcinoma, keratinizing, NOS, of body of pancreas
36560434	8071/3-C25.0	Squamous cell carcinoma, keratinizing, NOS, of head of pancreas
36551330	8071/3-C25.4	Squamous cell carcinoma, keratinizing, NOS, of islets of Langerhans
36554323	8071/3-C25.7	Squamous cell carcinoma, keratinizing, NOS, of other specified parts of pancreas
36543444	8071/3-C25.8	Squamous cell carcinoma, keratinizing, NOS, of overlapping lesion of pancreas
36522559	8071/3-C25.9	Squamous cell carcinoma, keratinizing, NOS, of pancreas, NOS
36564146	8071/3-C25.3	Squamous cell carcinoma, keratinizing, NOS, of pancreatic duct
36547093	8071/3-C25.2	Squamous cell carcinoma, keratinizing, NOS, of tail of pancreas
36542543	8072/3-C25.1	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of body of pancreas
36530594	8072/3-C25.0	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of head of pancreas
36566604	8072/3-C25.4	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of islets of Langerhans
36558590	8072/3-C25.7	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of other specified parts of
36525206	8072/3-C25.8	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of overlapping lesion of pancreas
36533480	8072/3-C25.9	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of pancreas, NOS
36559450	8072/3-C25.3	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of pancreatic duct
36545655	8072/3-C25.2	Squamous cell carcinoma, large cell, nonkeratinizing, NOS, of tail of pancreas
36546351	8076/3-C25.1	Squamous cell carcinoma, microinvasive of body of pancreas
36527912	8076/3-C25.0	Squamous cell carcinoma, microinvasive of head of pancreas
36531981	8076/3-C25.4	Squamous cell carcinoma, microinvasive of islets of Langerhans
36531863	8076/3-C25.7	Squamous cell carcinoma, microinvasive of other specified parts of pancreas
36559559	8076/3-C25.8	Squamous cell carcinoma, microinvasive of overlapping lesion of pancreas
36542609	8076/3-C25.9	Squamous cell carcinoma, microinvasive of pancreas, NOS
36518571	8076/3-C25.3	Squamous cell carcinoma, microinvasive of pancreatic duct
36537172	8076/3-C25.2	Squamous cell carcinoma, microinvasive of tail of pancreas
44500345	8070/3-C25.1	Squamous cell carcinoma, NOS, of body of pancreas
44500847	8070/3-C25.0	Squamous cell carcinoma, NOS, of head of pancreas
36556853	8070/3-C25.4	Squamous cell carcinoma, NOS, of islets of Langerhans
36562373	8070/3-C25.7	Squamous cell carcinoma, NOS, of other specified parts of pancreas
36557531	8070/3-C25.8	Squamous cell carcinoma, NOS, of overlapping lesion of pancreas
44501512	8070/3-C25.9	Squamous cell carcinoma, NOS, of pancreas, NOS
36535074	8070/3-C25.3	Squamous cell carcinoma, NOS, of pancreatic duct
36521188	8070/3-C25.2	Squamous cell carcinoma, NOS, of tail of pancreas
36533652	8073/3-C25.1	Squamous cell carcinoma, small cell, nonkeratinizing of body of pancreas
36524642	8073/3-C25.0	Squamous cell carcinoma, small cell, nonkeratinizing of head of pancreas
36564072	8073/3-C25.4	Squamous cell carcinoma, small cell, nonkeratinizing of islets of Langerhans

Concept ID	ICD-O-3 Codes[1]	Name
36549273	8073/3-C25.7	Squamous cell carcinoma, small cell, nonkeratinizing of other specified parts of pancreas
36548283	8073/3-C25.8	Squamous cell carcinoma, small cell, nonkeratinizing of overlapping lesion of pancreas
36556659	8073/3-C25.9	Squamous cell carcinoma, small cell, nonkeratinizing of pancreas, NOS
36538588	8073/3-C25.3	Squamous cell carcinoma, small cell, nonkeratinizing of pancreatic duct
36537903	8073/3-C25.2	Squamous cell carcinoma, small cell, nonkeratinizing of tail of pancreas
36524982	8074/3-C25.1	Squamous cell carcinoma, spindle cell of body of pancreas
36549944	8074/3-C25.0	Squamous cell carcinoma, spindle cell of head of pancreas
36526219	8074/3-C25.4	Squamous cell carcinoma, spindle cell of islets of Langerhans
36530804	8074/3-C25.7	Squamous cell carcinoma, spindle cell of other specified parts of pancreas
36536138	8074/3-C25.8	Squamous cell carcinoma, spindle cell of overlapping lesion of pancreas
36543966	8074/3-C25.9	Squamous cell carcinoma, spindle cell of pancreas, NOS
36547763	8074/3-C25.3	Squamous cell carcinoma, spindle cell of pancreatic duct
36526040	8074/3-C25.2	Squamous cell carcinoma, spindle cell of tail of pancreas
36541196	8143/3-C25.1	Superficial spreading adenocarcinoma of body of pancreas
36553090	8143/3-C25.0	Superficial spreading adenocarcinoma of head of pancreas
36562696	8143/3-C25.4	Superficial spreading adenocarcinoma of islets of Langerhans
36534609	8143/3-C25.7	Superficial spreading adenocarcinoma of other specified parts of pancreas
36554511	8143/3-C25.8	Superficial spreading adenocarcinoma of overlapping lesion of pancreas
36529433	8143/3-C25.9	Superficial spreading adenocarcinoma of pancreas, NOS
36519988	8143/3-C25.3	Superficial spreading adenocarcinoma of pancreatic duct
36566798	8143/3-C25.2	Superficial spreading adenocarcinoma of tail of pancreas
36561588	8211/3-C25.1	Tubular adenocarcinoma of body of pancreas
36544283	8211/3-C25.0	Tubular adenocarcinoma of head of pancreas
36533631	8211/3-C25.4	Tubular adenocarcinoma of islets of Langerhans
36558114	8211/3-C25.7	Tubular adenocarcinoma of other specified parts of pancreas
36550681	8211/3-C25.8	Tubular adenocarcinoma of overlapping lesion of pancreas
36539383	8211/3-C25.9	Tubular adenocarcinoma of pancreas, NOS
36565740	8211/3-C25.3	Tubular adenocarcinoma of pancreatic duct
36546398	8211/3-C25.2	Tubular adenocarcinoma of tail of pancreas
36568203	8001/3-C25.1	Tumor cells, malignant of body of pancreas
36568236	8001/3-C25.0	Tumor cells, malignant of head of pancreas
36568250	8001/3-C25.4	Tumor cells, malignant of islets of Langerhans
36568207	8001/3-C25.7	Tumor cells, malignant of other specified parts of pancreas
36534312	8001/3-C25.8	Tumor cells, malignant of overlapping lesion of pancreas
36568360	8001/3-C25.9	Tumor cells, malignant of pancreas, NOS
36568346	8001/3-C25.3	Tumor cells, malignant of pancreatic duct
36568332	8001/3-C25.2	Tumor cells, malignant of tail of pancreas

Concept ID	ICD-O-3 Codes[1]	Name
36521234	8051/3-C25.1	Verrucous carcinoma, NOS, of body of pancreas
36549741	8051/3-C25.0	Verrucous carcinoma, NOS, of head of pancreas
36543922	8051/3-C25.4	Verrucous carcinoma, NOS, of islets of Langerhans
36558579	8051/3-C25.7	Verrucous carcinoma, NOS, of other specified parts of pancreas
36528193	8051/3-C25.8	Verrucous carcinoma, NOS, of overlapping lesion of pancreas
36540640	8051/3-C25.9	Verrucous carcinoma, NOS, of pancreas, NOS
36541248	8051/3-C25.3	Verrucous carcinoma, NOS, of pancreatic duct
36562350	8051/3-C25.2	Verrucous carcinoma, NOS, of tail of pancreas
36540629	8262/3-C25.1	Villous adenocarcinoma of body of pancreas
36528404	8262/3-C25.0	Villous adenocarcinoma of head of pancreas
36532836	8262/3-C25.4	Villous adenocarcinoma of islets of Langerhans
36540783	8262/3-C25.7	Villous adenocarcinoma of other specified parts of pancreas
36536276	8262/3-C25.8	Villous adenocarcinoma of overlapping lesion of pancreas
36566976	8262/3-C25.9	Villous adenocarcinoma of pancreas, NOS
36565469	8262/3-C25.3	Villous adenocarcinoma of pancreatic duct
36557435	8262/3-C25.2	Villous adenocarcinoma of tail of pancreas
36520953	8155/3-C25.1	Vipoma of body of pancreas
36551750	8155/3-C25.0	Vipoma of head of pancreas
36522638	8155/3-C25.4	Vipoma of islets of Langerhans
36555301	8155/3-C25.7	Vipoma of other specified parts of pancreas
36562626	8155/3-C25.8	Vipoma of overlapping lesion of pancreas
36551288	8155/3-C25.9	Vipoma of pancreas, NOS
36555520	8155/3-C25.3	Vipoma of pancreatic duct
36544424	8155/3-C25.2	Vipoma of tail of pancreas

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour.

Table S8. Soft tissue sarcoma.

SNOMED:

Concept ID	Concept name
40486535	Angiosarcoma of cheek
37204279	Epithelioid sarcoma
40487487	Ewing's sarcoma of soft tissue
44783559	Glandular malignant peripheral nerve sheath tumor
37151903	Leiomyosarcoma
40482829	Leiomyosarcoma
40486084	Leiomyosarcoma of cardia of stomach
40493008	Leiomyosarcoma of cardioesophageal junction
40486588	Leiomyosarcoma of connective tissue
36674730	Leiomyosarcoma of corpus uteri
40488914	Leiomyosarcoma of lower esophagus
44782480	Leiomyosarcoma of orbit
40486083	Leiomyosarcoma of scalp
37396739	Leiomyosarcoma of small intestine
40486566	Leiomyosarcoma of stomach
40492385	Leiomyosarcoma of uterus
40490459	Liposarcoma of connective tissue
37166993	Malignant neoplasm of perineum
37166990	Malignant neoplasm of soft tissue of axilla
37166991	Malignant neoplasm of soft tissue of buttock
37166985	Malignant neoplasm of soft tissue of finger
37166987	Malignant neoplasm of soft tissue of hand
40490994	Malignant neoplasm of soft tissue of orbit
37166989	Malignant neoplasm of soft tissue of thumb
37166986	Malignant neoplasm of soft tissue of toe
37166992	Malignant neoplasm soft tissue of lower limb
37168060	Malignant peripheral nerve sheath neoplasm of peripheral nerve of lumbar spine
37161005	Malignant peripheral nerve sheath tumor of gingival mucosa
42539510	Malignant peripheral nerve sheath tumor with perineurial differentiation
37311919	Malignant pilonidal cyst
46271138	Metastatic leiomyosarcoma
45766450	Myxoinflammatory fibroblastic sarcoma
36717175	Primary angiosarcoma of breast

Concept ID	Concept name
36715808	Primary angiosarcoma of heart
36716489	Primary leiomyosarcoma of peritoneum
36717566	Primary leiomyosarcoma of retroperitoneum
36715810	Primary liposarcoma of male genital organ
36716491	Primary liposarcoma of peritoneum
36716490	Primary liposarcoma of retroperitoneum
36715809	Primary liposarcoma of soft tissue of limb
37167740	Primary malignant peripheral nerve sheath tumor
36716678	Primary synovial sarcoma of intrathoracic organ
36716677	Primary synovial sarcoma of respiratory organ
36715811	Primary synovial sarcoma of soft tissue of limb
37116997	Solitary fibrous tumor and hemangiopericytoma grade 3
4298238	Adipocytic liposarcoma
4004514	Angiomyoliposarcoma
4297200	Angiosarcoma
3661612	Angiosarcoma
4003021	Angiosarcoma of liver
4111948	Angiosarcoma of skin
761863	Angiosarcoma of skin of cheek
4092083	Angiosarcoma of spleen
4111943	Cutaneous leiomyosarcoma
4300681	Cutaneous leiomyosarcoma with granular cell change
4284833	Dedifferentiated liposarcoma
4301652	Dedifferentiated liposarcoma
4173148	Epithelioid leiomyosarcoma
4301644	Epithelioid leiomyosarcoma of skin
4298153	Epithelioid malignant nerve sheath tumor
4029190	Fibroblastic liposarcoma
4029477	Glomangiosarcoma
4191591	Hemangiosarcoma
607059	Histologic grade of primary malignant neoplasm of oral soft tissue
607030	Histologic type of primary malignant neoplasm of oral soft tissue
619825	Histologic type of primary malignant neoplasm of peritoneum
607039	Histologic type of primary malignant neoplasm of soft tissue
4300582	Intravascular angiosarcoma

Concept ID	Concept name
375479	Kaposi's sarcoma of soft tissue
35622339	Leiomyosarcoma of cervix uteri
608050	Leiomyosarcoma of colon
608048	Leiomyosarcoma of duodenum
608049	Leiomyosarcoma of esophagus
605745	Leiomyosarcoma of fundus of stomach
606200	Leiomyosarcoma of gastric corpus structure
606202	Leiomyosarcoma of greater curvature of stomach
606201	Leiomyosarcoma of lesser curvature of stomach
606199	Leiomyosarcoma of pyloric antrum of stomach
606198	Leiomyosarcoma of pylorus of stomach
608046	Leiomyosarcoma of rectum
605507	Leiomyosarcoma of skin of chest
4110857	Liposarcoma
4116219	Liposarcoma of orbit
761082	Liposarcoma of retroperitoneum
4102790	Liposarcoma, well differentiated
4201615	Local recurrence of malignant tumor of soft tissue
4112102	Localized malignant reticulohistiocytoma
4298141	Lymphangiosarcoma
4272004	Lymphangiosarcoma
4094870	Malignant neoplasm of cartilage of nose
35610240	Malignant neoplasm of connective and soft tissues of trunk
4095764	Malignant neoplasm of connective tissue of orbit
4089762	Malignant neoplasm of costal cartilage
4092529	Malignant neoplasm of extraocular muscle of orbit
4095309	Malignant neoplasm of glossoepiglottic fold
4094868	Malignant neoplasm of perinephric tissue
4089666	Malignant neoplasm of retrocecal tissue
4153882	Malignant neoplasm of soft tissue
35610250	Malignant neoplasm of soft tissues of lower leg
35610548	Malignant neoplasm of soft tissues of thigh
4095158	Malignant neoplasm of squamocolumnar junction of cervix
4089775	Malignant neoplasm soft tissues of cervical spine
3655604	Malignant nerve sheath neoplasm of peripheral nerve of abdomen

Concept ID	Concept name
3655627	Malignant nerve sheath neoplasm of peripheral nerve of head and neck
3655628	Malignant nerve sheath neoplasm of peripheral nerve of lower limb
3655625	Malignant nerve sheath neoplasm of peripheral nerve of pelvis
3655626	Malignant nerve sheath neoplasm of peripheral nerve of thorax
3655624	Malignant nerve sheath neoplasm of peripheral nerve of upper limb
4300679	Malignant Triton tumor
4178978	Malignant tumor of aryepiglottic fold - laryngeal aspect
4177237	Malignant tumor of false cord
4081041	Malignant tumor of fibrous tissue
4090227	Malignant tumor of lower buccal sulcus
4090229	Malignant tumor of lower labial sulcus
4112974	Malignant tumor of peripheral nerve
4179094	Malignant tumor of soft tissue of abdomen
4149450	Malignant tumor of soft tissue of back
4180902	Malignant tumor of soft tissue of face
4181350	Malignant tumor of soft tissue of head
4118989	Malignant tumor of soft tissue of head, face and neck
4181329	Malignant tumor of soft tissue of hip
4180903	Malignant tumor of soft tissue of neck
4178958	Malignant tumor of soft tissue of pelvis
4180781	Malignant tumor of soft tissue of shoulder
4181330	Malignant tumor of soft tissue of thorax
4157319	Malignant tumor of soft tissue of upper limb
4094716	Malignant tumor of upper buccal sulcus
4090228	Malignant tumor of upper labial sulcus
4301643	Melanotic malignant nerve sheath tumor
4007138	Mixed liposarcoma
4295439	Myeloliposarcoma
4059632	Myxoid leiomyosarcoma
4298233	Myxoid leiomyosarcoma of skin
4300686	Myxoid liposarcoma
4101123	Myxoid liposarcoma
4129865	Neoplasm of soft tissues of head
4010104	Pleomorphic liposarcoma
4300687	Pleomorphic liposarcoma

Concept ID	Concept name
4298312	Pleomorphic rhabdomyosarcoma
4301647	Pleomorphic rhabdomyosarcoma
602015	Primary angiosarcoma of abdomen
602020	Primary angiosarcoma of axillary region
602021	Primary angiosarcoma of face
602022	Primary angiosarcoma of head
602665	Primary angiosarcoma of left hip region
605431	Primary angiosarcoma of left lower limb
602664	Primary angiosarcoma of left upper limb
602023	Primary angiosarcoma of neck
602024	Primary angiosarcoma of pelvis
605433	Primary angiosarcoma of right hip region
605434	Primary angiosarcoma of right lower limb
605432	Primary angiosarcoma of right upper limb
609034	Primary angiosarcoma of skin of forehead
609037	Primary angiosarcoma of skin of lip
609035	Primary angiosarcoma of skin of nose
609036	Primary angiosarcoma of skin of temporal region
602025	Primary angiosarcoma of thorax
602026	Primary angiosarcoma of trunk
4313887	Primary leiomyosarcoma
602051	Primary leiomyosarcoma of abdomen
602052	Primary leiomyosarcoma of buttock
602053	Primary leiomyosarcoma of face
602054	Primary leiomyosarcoma of head
602055	Primary leiomyosarcoma of inguinal region
605449	Primary leiomyosarcoma of left hip region
605450	Primary leiomyosarcoma of left kidney
605451	Primary leiomyosarcoma of left lower limb
605452	Primary leiomyosarcoma of left shoulder region
605448	Primary leiomyosarcoma of left upper limb
602056	Primary leiomyosarcoma of neck
602057	Primary leiomyosarcoma of pelvis
602058	Primary leiomyosarcoma of perineum
605454	Primary leiomyosarcoma of right hip region

Concept ID	Concept name
605455	Primary leiomyosarcoma of right kidney
602683	Primary leiomyosarcoma of right lower limb
602684	Primary leiomyosarcoma of right shoulder region
605453	Primary leiomyosarcoma of right upper limb
602059	Primary leiomyosarcoma of thorax
602060	Primary leiomyosarcoma of trunk
602061	Primary liposarcoma of soft tissue of abdomen
602062	Primary liposarcoma of soft tissue of axilla
602063	Primary liposarcoma of soft tissue of buttock
602064	Primary liposarcoma of soft tissue of inguinal region
605457	Primary liposarcoma of soft tissue of left hip
605458	Primary liposarcoma of soft tissue of left lower limb
605459	Primary liposarcoma of soft tissue of left shoulder
605456	Primary liposarcoma of soft tissue of left upper limb
609233	Primary liposarcoma of soft tissue of neck
609234	Primary liposarcoma of soft tissue of pelvis
602065	Primary liposarcoma of soft tissue of perineum
605461	Primary liposarcoma of soft tissue of right hip
602685	Primary liposarcoma of soft tissue of right lower limb
602686	Primary liposarcoma of soft tissue of right shoulder
605460	Primary liposarcoma of soft tissue of right upper limb
602066	Primary liposarcoma of soft tissue of thorax
602067	Primary liposarcoma of soft tissue of trunk
4247720	Primary malignant neoplasm of blood vessel of abdomen
4246798	Primary malignant neoplasm of blood vessel of head
4246028	Primary malignant neoplasm of cartilage of nose
4311477	Primary malignant neoplasm of great vessels
4247838	Primary malignant neoplasm of muscle of head
4311488	Primary malignant neoplasm of muscle of lower limb
4311489	Primary malignant neoplasm of muscle of neck
4247839	Primary malignant neoplasm of muscle of trunk
4247840	Primary malignant neoplasm of muscle of upper limb
4246139	Primary malignant neoplasm of para-aortic body
4246142	Primary malignant neoplasm of periadrenal tissue
4307721	Primary malignant neoplasm of peripheral nerve

Concept ID	Concept name
601134	Primary malignant neoplasm of peripheral nerves of left lower limb
4001319	Primary malignant neoplasm of peripheral nerves of lower limb
601136	Primary malignant neoplasm of peripheral nerves of right lower limb
4246143	Primary malignant neoplasm of perirenal tissue
4246147	Primary malignant neoplasm of rectovaginal septum
4312566	Primary malignant neoplasm of rectovesical septum
4155172	Primary malignant neoplasm of retrocecal tissue
609043	Primary malignant neoplasm of soft tissue of left hip
35617893	Primary malignant neoplasm of soft tissue of left lower limb
35617803	Primary malignant neoplasm of soft tissue of left upper extremity
609044	Primary malignant neoplasm of soft tissue of right hip
35617882	Primary malignant neoplasm of soft tissue of right lower limb
35617871	Primary malignant neoplasm of soft tissue of right upper extremity
376647	Primary malignant neoplasm of soft tissues
197807	Primary malignant neoplasm of soft tissues of abdomen
4247356	Primary malignant neoplasm of soft tissues of axilla
4312689	Primary malignant neoplasm of soft tissues of buttock
4311619	Primary malignant neoplasm of soft tissues of face
4247357	Primary malignant neoplasm of soft tissues of head
442123	Primary malignant neoplasm of soft tissues of hip
4311620	Primary malignant neoplasm of soft tissues of inguinal region
609046	Primary malignant neoplasm of soft tissues of left shoulder
438094	Primary malignant neoplasm of soft tissues of lower limb
4246235	Primary malignant neoplasm of soft tissues of neck
4247358	Primary malignant neoplasm of soft tissues of pelvis
4312690	Primary malignant neoplasm of soft tissues of perineum
609047	Primary malignant neoplasm of soft tissues of right shoulder
4311621	Primary malignant neoplasm of soft tissues of shoulder
317801	Primary malignant neoplasm of soft tissues of thorax
197808	Primary malignant neoplasm of soft tissues of trunk
436357	Primary malignant neoplasm of soft tissues of upper limb
602092	Primary sarcoma of soft tissues of abdomen
605473	Primary sarcoma of soft tissues of left hip
605474	Primary sarcoma of soft tissues of left shoulder
605472	Primary sarcoma of soft tissues of left upper limb

Concept ID	Concept name
602694	Primary sarcoma of soft tissues of right hip
602158	Primary sarcoma of soft tissues of right shoulder
605475	Primary sarcoma of soft tissues of right upper limb
4110891	Rhabdomyosarcoma of orbit
4314042	Sarcoma of soft tissue
4297341	Sclerosing liposarcoma
4226398	Soft tissue of renal sinus involved by tumor
4297343	Spindle cell liposarcoma
4269784	Stewart-Treves syndrome
4268491	Synovial sarcoma
4116202	Synovial sarcoma
4085470	Synovial sarcoma, biphasic
4207078	Synovial sarcoma, epithelioid cell
4289992	Synovial sarcoma, spindle cell

SNOMED= Systematised Nomenclature of Medicine.

ICD-O-3 codes:

ICD-O-3 Codes[1]	Included in the previous study (EUPAS50800)	Included in the current study (EUPAS1000000715)
8933/3, 8821/3, 8920/3, 9581/3, 8894/3, 8934/3, 8941/3, 9181/3, 9220/3, 9044/3, 8858/3, 8832/3, 8806/3,8921/3, 8910/3, 8931/3, 8930/3, 9133/3, 8891/3, 8804/3, 9364/3, 8813/3, 9182/3, 8810/3, 8936/3, 8802/3,9580/3, 9130/3, 9120/3, 8814/3, 9140/3, 8890/3, 8850/3, 8851/3, 9170/3, 8830/3, 9561/3, 9542/3, 9540/3,9240/3, 8990/3, 8951/3, 8855/3, 8940/3, 8902/3, 8982/3, 8825/3, 8895/3, 8811/3, 9231/3, 8896/3, 8852/3,8840/3, 8822/3, 8826/3, 8841/3, 8893/3, 9186/3, 9230/3, 8823/3, 8991/3, 9125/3, 8857/3, 9194/3, 8831/3,9195/3, 9124/3, 9187/3, 8836/3, 8835/3, 8897/3, 8824/3, 8842/3, 9184/3, 9200/3, 9192/3, 9221/3, 8812/3,9193/3, 8714/3, 8833/3, 8854/3, 8963/3, 8853/3, 9185/3, 8803/3, 8815/3, 8801/3, 8935/3, 9043/3, 9042/3,9040/3, 9041/3, 9183/3, 8805/3, 9560/3, 9180/3, 9571/3, 8901/3, 8900/3, 8800/3, 8912/3, 8982/3, 8825/3, 8895/3, 8811/3, 9231/3, 8896/3 8852/3, 8840/3, 8822/3, 8826/3, 8841/3, 8893/3, 9186/3, 9230/3, 8823/3,8991/3, 9125/3, 8857/3, 9194/3, 8831/3, 9195/3, 9124/3, 9187/3, 8836/3, 8835/3, 8897/3, 8824/3, 8842/3, 9184/3, 9200/3, 9192/3, 9221/3, 8812/3, 9193/3, 8714/3, 8833/3, 8854/3, 8963/3, 8853/3, 9185/3, 8803/3, 8815/3, 8801/3, 8935/3, 9043/3,9042/3, 9040/3, 9041/3, 9183/3, 8805/3, 9560/3, 9180/3, 9571/3, 8901/3, 8900/3, 8800/3, 8912/3		X

ICD-O-3= International Classification of Diseases for Oncology, 3rd Edition.
[1] Morphology ICD-O-3 codes. The first four digits indicate the histological term and the fifth digit the behaviour code. Behaviour code of /3 indicates a malignant tumour. We included concept IDs corresponding to these morphology ICD-O-3 codes, regardless of topography (site).

ANNEX IV: Supplementary results

Table S9. Study attrition of individuals included in denominators from data sources with population-level denominators during the whole study period.

Data source name[1]	Estimate name			
	Records (N)	Subjects (N)	Excluded records (N)	Excluded subjects (N)
NAJS				
Starting population	4,853,340	4,853,340	–	–
Missing year of birth and/or sex	4,853,340	4,853,340	0	0
Cannot satisfy age criteria during the study period based on year of birth	4,852,810	4,852,810	530	530
No observation time available during study period	4,721,070	4,721,070	131,740	131,740
Does not satisfy age criteria during the study period	4,721,070	4,721,070	0	0
Prior history requirement not fulfilled during study period	4,721,070	4,721,070	0	0
No observation time available after applying age, prior observation and, if applicable, target criteria	4,721,070	4,721,070	0	0
DK-DHR				
Starting population	9,235,411	8,593,356	–	–
Missing year of birth and/or sex	9,235,411	8,593,356	0	0
Cannot satisfy age criteria during the study period based on year of birth	9,185,905	8,543,853	49,506	49,503
No observation time available during study period	7,567,850	7,306,472	1,618,055	1,237,381
Does not satisfy age criteria during the study period	7,567,850	7,306,472	0	0
Prior history requirement not fulfilled during study period	7,567,850	7,306,472	0	0
No observation time available after applying age, prior observation and, if applicable, target criteria	7,567,850	7,306,472	0	0
InGef RDB				
Starting population	10,512,283	10,512,283	–	–
Missing year of birth and/or sex	10,512,283	10,512,283	0	0
Cannot satisfy age criteria during the study period based on year of birth	10,512,282	10,512,282	<5	<5
No observation time available during study period	10,512,282	10,512,282	0	0

Data source name[1]	Estimate name			
	Records (N)	Subjects (N)	Excluded records (N)	Excluded subjects (N)
Does not satisfy age criteria during the study period	10,512,282	10,512,282	0	0
Prior history requirement not fulfilled during study period	10,512,282	10,512,282	0	0
No observation time available after applying age, prior observation and, if applicable, target criteria	10,512,282	10,512,282	0	0

DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System.
[1] Attrition corresponds to the steps used to derive the internal denominator. The study period spanned from 2010 to 2023. First year of available data: 2010 for all data sources, except from NAJS (2017) and InGef RDB (2015). Last year of available data: 2023 for DK-DHR, 2024 for NAJS and InGef RDB.

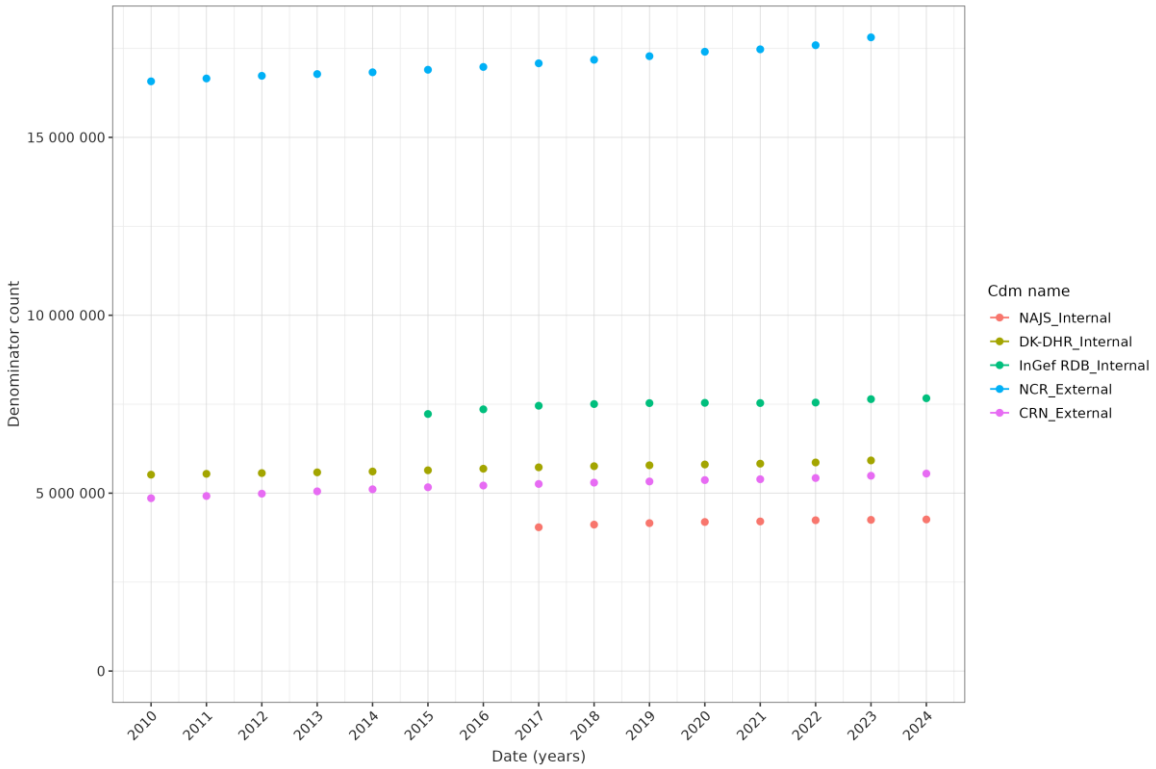

Figure S1. Number of individuals included in the denominator used for prevalence estimations.

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; N/A= Not applicable; NCR=Netherlands Cancer Registry.

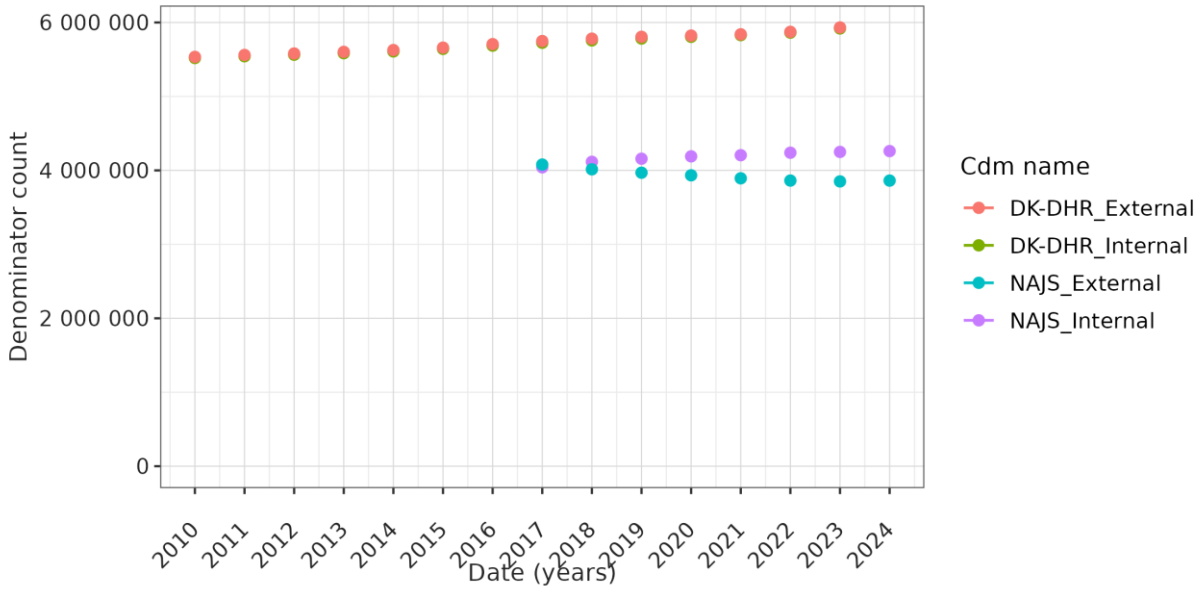

Figure S2. Number of individuals included based on internal and external denominators in DK-DHR and NAJS.

DK-DHR=Danish Data Health Registries; NAJS=Croatian National Public Health Information System.

Table S10. Study attrition of outcome cohorts during the study period.

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
Acute lymphoid leukaemia						
Initial qualifying events	number_records	63,999	236,200	36,464	7,065	3,788
	number_subjects	1,721	3,549	2,446	7,065	3,786
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	1,736	3,665	2,461	7,065	3,788
	number_subjects	1,721	3,549	2,446	7,065	3,786
	excluded_records	62,263	232,535	34,003	0	0
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	1,625	–	–	–	–
	number_subjects	1,610	–	–	–	–
	excluded_records	111	–	–	–	–
	excluded_subjects	111	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	1,623	–	2,461	–	1,383
	number_subjects	1,608	–	2,446	–	1,383
	excluded_records	<5	–	0	–	0
	excluded_subjects	<5	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	1,943	–	4,251	1,383
	number_subjects	–	1,902	–	4,251	1,383
	excluded_records	–	1,722	–	2,814	2,405
	excluded_subjects	–	1,647	–	2,814	2,403
cohort_end_date trimmed <= 2023-12-31	number_records	–	1,870	–	4,074	–
	number_subjects	–	1,831	–	4,074	–
	excluded_records	–	73	–	177	–
	excluded_subjects	–	71	–	177	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	2,461	–	–
	number_subjects	–	–	2,446	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Acute myeloid leukaemia						
Initial qualifying events	number_records	109,511	317,245	40,879	20,129	9,793
	number_subjects	3,881	10,271	5,691	20,116	9,792
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	3,897	10,381	5,720	20,123	9,792

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
	number_subjects	3,881	10,271	5,691	20,116	9,792
	excluded_records	105,614	306,864	35,159	6	<5
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	3,405	–	–	–	–
	number_subjects	3,389	–	–	–	–
	excluded_records	492	–	–	–	–
	excluded_subjects	492	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	3,405	–	5,720	–	3,089
	number_subjects	3,389	–	5,691	–	3,089
	excluded_records	0	–	0	–	0
	excluded_subjects	0	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	5,706	–	11,800	3,089
	number_subjects	–	5,671	–	11,797	3,089
	excluded_records	–	4,675	–	8,323	6,703
	excluded_subjects	–	4,600	–	8,319	6,703
cohort_end_date trimmed <= 2023-12-31	number_records	–	5,404	–	11,112	–
	number_subjects	–	5,371	–	11,109	–
	excluded_records	–	302	–	688	–
	excluded_subjects	–	300	–	688	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	5,720	–	–
	number_subjects	–	–	5,691	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Chronic lymphocytic leukaemia						
Initial qualifying events	number_records	233,256	481,814	26,456	29,330	12,479
	number_subjects	6,875	16,998	6,255	29,330	12,479
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	6,961	17,668	6,455	29,330	12,479
	number_subjects	6,875	16,998	6,255	29,330	12,479
	excluded_records	226,295	464,146	20,001	0	0
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	6,432	–	–	–	–
	number_subjects	6,346	–	–	–	–
	excluded_records	529	–	–	–	–
	excluded_subjects	529	–	–	–	–

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
cohort_end_date trimmed <= 2024-12-31	number_records	6,430	–	6,455	–	6,342
	number_subjects	6,344	–	6,255	–	6,342
	excluded_records	<5	–	0	–	0
	excluded_subjects	<5	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	11,286	–	20,833	6,342
	number_subjects	–	10,987	–	20,833	6,342
	excluded_records	–	6,382	–	8,497	6,137
	excluded_subjects	–	6,011	–	8,497	6,137
cohort_end_date trimmed <= 2023-12-31	number_records	–	10,842	–	19,759	–
	number_subjects	–	10,570	–	19,759	–
	excluded_records	–	444	–	1,074	–
	excluded_subjects	–	417	–	1,074	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	6,455	–	–
	number_subjects	–	–	6,255	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Diffuse large B-cell lymphoma						
Initial qualifying events	number_records	129,597	411,818	42,971	38,019	11,016
	number_subjects	4,489	15,587	7,053	38,002	11,007
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	4,549	15,907	7,121	38,015	11,016
	number_subjects	4,489	15,587	7,053	38,002	11,007
	excluded_records	125,048	395,911	35,850	<5	0
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	4,207	–	–	–	–
	number_subjects	4,147	–	–	–	–
	excluded_records	342	–	–	–	–
	excluded_subjects	342	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	4,207	–	7,121	–	7,015
	number_subjects	4,147	–	7,053	–	7,015
	excluded_records	0	–	0	–	0
	excluded_subjects	0	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	10,786	–	25,302	7,015
	number_subjects	–	10,642	–	25,298	7,015
	excluded_records	–	5,121	–	12,713	4,001

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
	excluded_subjects	–	4,945	–	12,704	3,992
cohort_end_date trimmed <= 2023-12-31	number_records	–	10,302	–	23,924	–
	number_subjects	–	10,172	–	23,920	–
	excluded_records	–	484	–	1,378	–
	excluded_subjects	–	470	–	1,378	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	7,121	–	–
	number_subjects	–	–	7,053	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Follicular lymphoma						
Initial qualifying events	number_records	182,987	288,943	20,183	17,714	6,214
	number_subjects	5,418	11,097	3,909	17,709	6,213
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	5,498	11,511	3,992	17,713	6,214
	number_subjects	5,418	11,097	3,909	17,709	6,213
	excluded_records	177,489	277,432	16,191	<5	0
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	5,269	–	–	–	–
	number_subjects	5,189	–	–	–	–
	excluded_records	229	–	–	–	–
	excluded_subjects	229	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	5,265	–	3,992	–	4,221
	number_subjects	5,185	–	3,909	–	4,221
	excluded_records	<5	–	0	–	0
	excluded_subjects	<5	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	7,627	–	12,483	4,221
	number_subjects	–	7,457	–	12,481	4,221
	excluded_records	–	3,884	–	5,230	1,993
	excluded_subjects	–	3,640	–	5,228	1,992
cohort_end_date trimmed <= 2023-12-31	number_records	–	7,397	–	11,784	–
	number_subjects	–	7,235	–	11,783	–
	excluded_records	–	230	–	699	–
	excluded_subjects	–	222	–	698	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	3,992	–	–
	number_subjects	–	–	3,909	–	–

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Multiple myeloma						
Initial qualifying events	number_records	431,977	852,191	63,290	35,794	21,327
	number_subjects	7,074	15,157	8,064	35,785	21,325
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	7,110	15,264	8,217	35,786	21,325
	number_subjects	7,074	15,157	8,064	35,785	21,325
	excluded_records	424,867	836,927	55,073	8	<5
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	6,488	–	–	–	–
	number_subjects	6,452	–	–	–	–
	excluded_records	622	–	–	–	–
	excluded_subjects	622	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	6,480	–	8,217	–	8,584
	number_subjects	6,444	–	8,064	–	8,584
	excluded_records	8	–	0	–	0
	excluded_subjects	8	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	10,137	–	23,655	8,584
	number_subjects	–	10,084	–	23,655	8,584
	excluded_records	–	5,127	–	12,131	12,741
	excluded_subjects	–	5,073	–	12,130	12,741
cohort_end_date trimmed <= 2023-12-31	number_records	–	9,577	–	22,408	–
	number_subjects	–	9,525	–	22,408	–
	excluded_records	–	560	–	1,247	–
	excluded_subjects	–	559	–	1,247	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	8,217	–	–
	number_subjects	–	–	8,064	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Pancreatic cancer						
Initial qualifying events	number_records	333,401	485,700	79,067	68,495	40,261
	number_subjects	13,042	32,296	17,124	68,469	40,230
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
pad `cohort_end_date` 1825 days	number_records	13,074	32,418	17,175	68,476	40,233
	number_subjects	13,042	32,296	17,124	68,469	40,230
	excluded_records	320,327	453,282	61,892	19	28
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	10,919	–	–	–	–
	number_subjects	10,887	–	–	–	–
	excluded_records	2,155	–	–	–	–
	excluded_subjects	2,155	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	10,914	–	17,175	–	13,546
	number_subjects	10,883	–	17,124	–	13,544
	excluded_records	5	–	0	–	0
	excluded_subjects	<5	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	19,232	–	42,680	13,546
	number_subjects	–	19,156	–	42,674	13,544
	excluded_records	–	13,186	–	25,796	26,687
	excluded_subjects	–	13,140	–	25,795	26,686
cohort_end_date trimmed <= 2023-12-31	number_records	–	18,151	–	39,819	–
	number_subjects	–	18,082	–	39,813	–
	excluded_records	–	1,081	–	2,861	–
	excluded_subjects	–	1,074	–	2,861	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	17,175	–	–
	number_subjects	–	–	17,124	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–
Soft-tissue sarcoma						
Initial qualifying events	number_records	124,336	251,762	35,706	38,095	17,781
	number_subjects	6,603	14,963	7,131	37,810	17,597
	excluded_records	0	0	0	0	0
	excluded_subjects	0	0	0	0	0
pad `cohort_end_date` 1825 days	number_records	6,653	15,368	7,195	37,922	17,679
	number_subjects	6,603	14,963	7,131	37,810	17,597
	excluded_records	117,683	236,394	28,511	173	102
	excluded_subjects	0	0	0	0	0
cohort_start_date trimmed >= 2017-01-01	number_records	6,378	–	–	–	–
	number_subjects	6,328	–	–	–	–
	excluded_records	275	–	–	–	–

Reason	Variable name	Data source name				
		NAJS	DK-DHR	InGef RDB	NCR	CRN
	excluded_subjects	275	–	–	–	–
cohort_end_date trimmed <= 2024-12-31	number_records	6,376	–	7,195	–	7,110
	number_subjects	6,326	–	7,131	–	7,099
	excluded_records	<5	–	0	–	0
	excluded_subjects	<5	–	0	–	0
cohort_start_date trimmed >= 2010-01-01	number_records	–	10,356	–	24,844	7,110
	number_subjects	–	10,147	–	24,800	7,099
	excluded_records	–	5,012	–	13,078	10,569
	excluded_subjects	–	4,816	–	13,010	10,498
cohort_end_date trimmed <= 2023-12-31	number_records	–	9,912	–	23,378	–
	number_subjects	–	9,728	–	23,337	–
	excluded_records	–	444	–	1,466	–
	excluded_subjects	–	419	–	1,463	–
cohort_start_date trimmed >= 2015-01-01	number_records	–	–	7,195	–	–
	number_subjects	–	–	7,131	–	–
	excluded_records	–	–	0	–	–
	excluded_subjects	–	–	0	–	–

CRN=Cancer Registry Norway; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; NAJS=Croatian National Public Health Information System; NCR=Netherlands Cancer Registry.
[1] The study period spanned from 2010 to 2023. First year of available data: 2010 for all data sources, except from NAJS (2017) and InGef RDB (2015). Last year of available data: 2023 for DK-DHR and NCR, 2024 for NAJS, InGef RDB, and CRN.

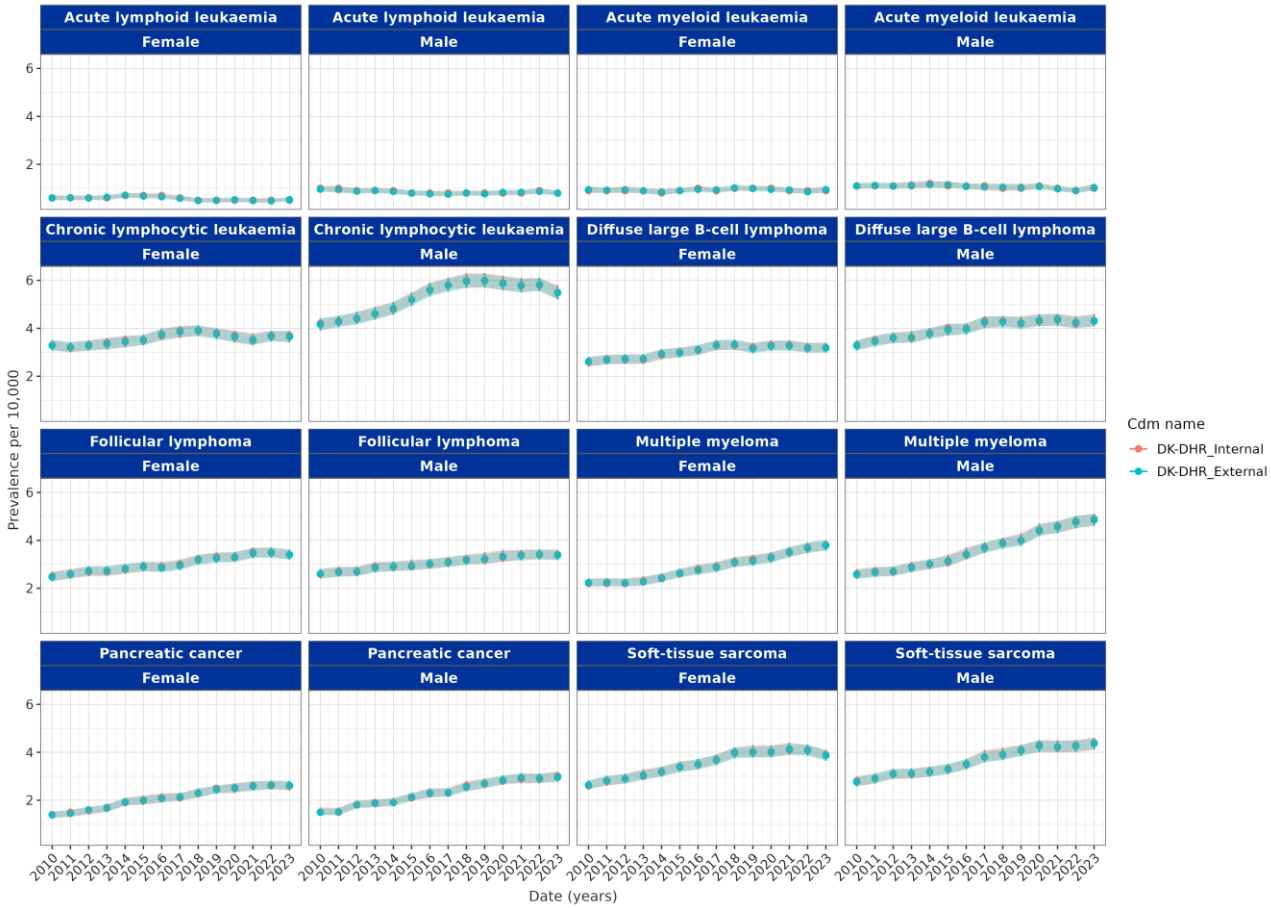

Figure S3. Estimated 5-year partial point prevalence per 10,000 using internal and external denominators in DK-DHR, stratified by sex.

DK-DHR=Danish Data Health Registries.

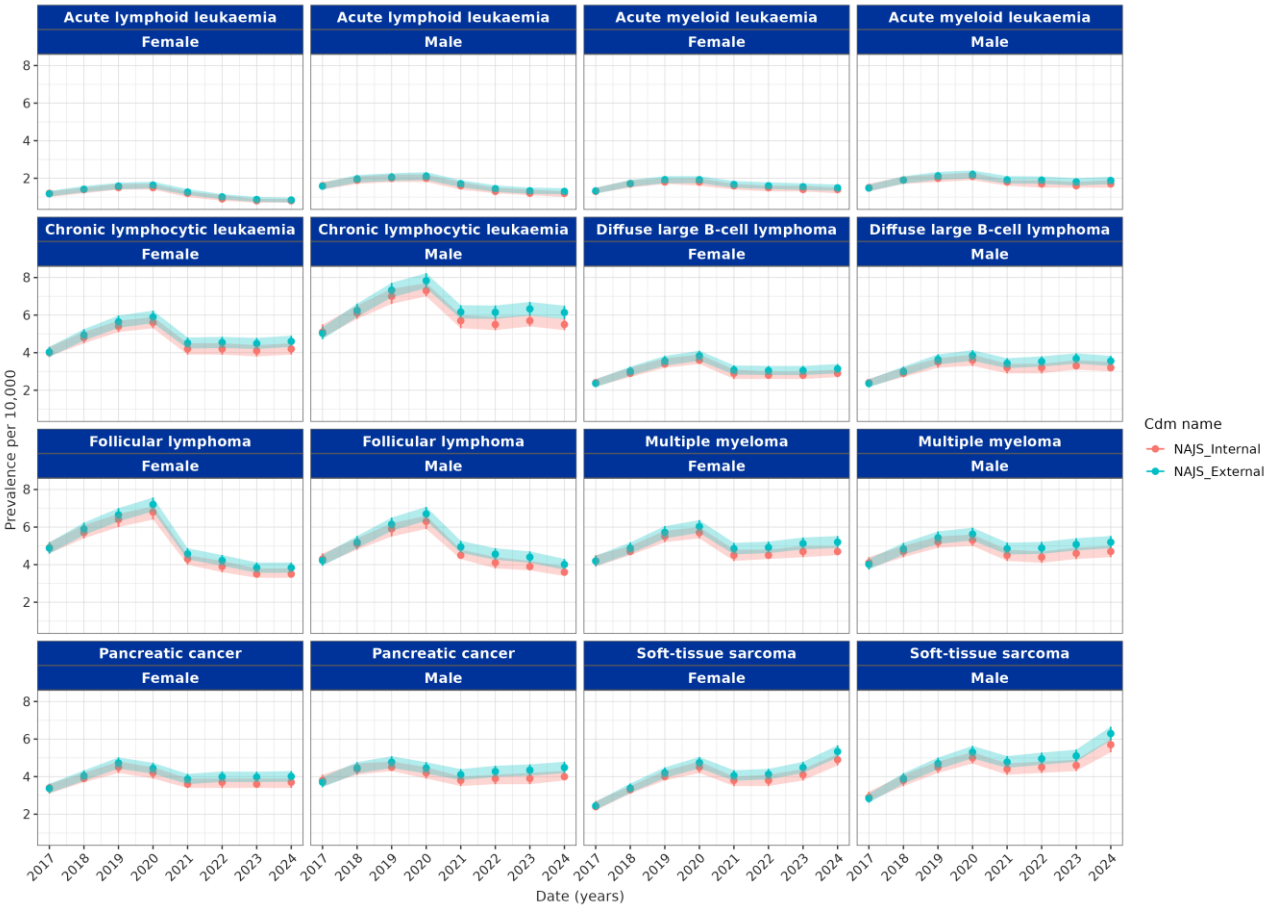

Figure S4. Estimated 5-year partial point prevalence per 10,000 using internal and external denominators in NAJS, stratified by sex.

NAJS=Croatian National Public Health Information System.

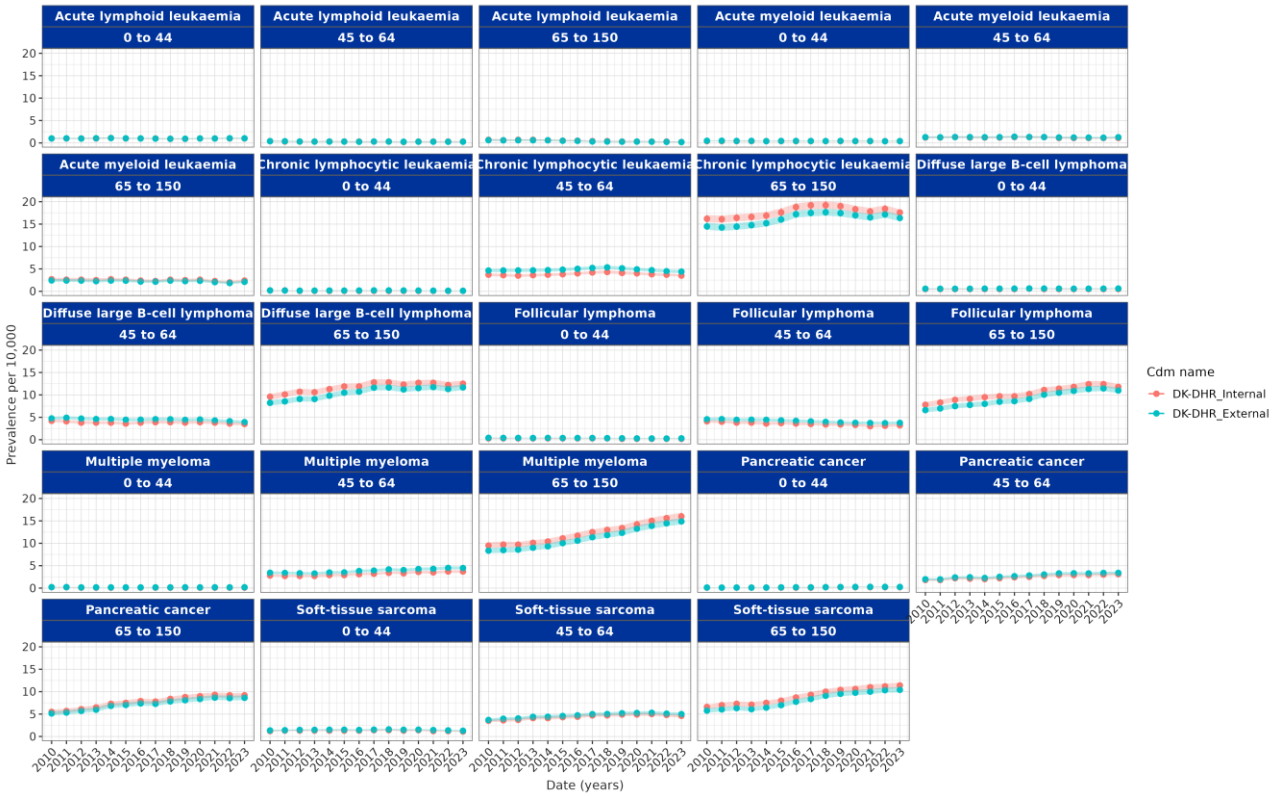

Figure S5. Estimated 5-year partial point prevalence per 10,000 using internal and external denominators in DK-DHR, stratified by age groups.

DK-DHR=Danish Data Health Registries.

Figure S6. Estimated 5-year partial point prevalence per 10,000 using internal and external denominators in NAJS, stratified by age groups.

NAJS=Croatian National Public Health Information System.

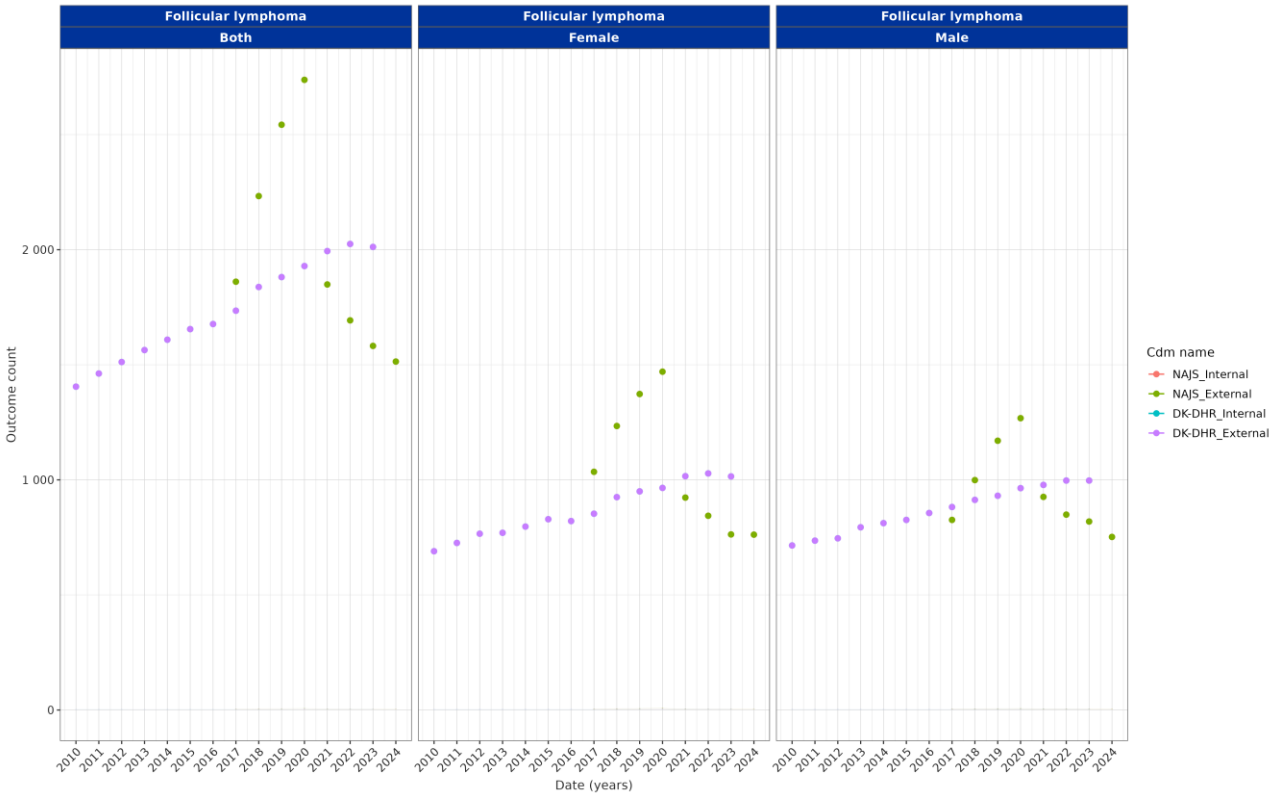

Figure S7. Number of individuals with the outcome of interest (follicular lymphoma) using internal and external denominators in NAJS and DK-DHR, stratified by sex.

DK-DHR=Danish Data Health Registries; NAJS=Croatian National Public Health Information System.

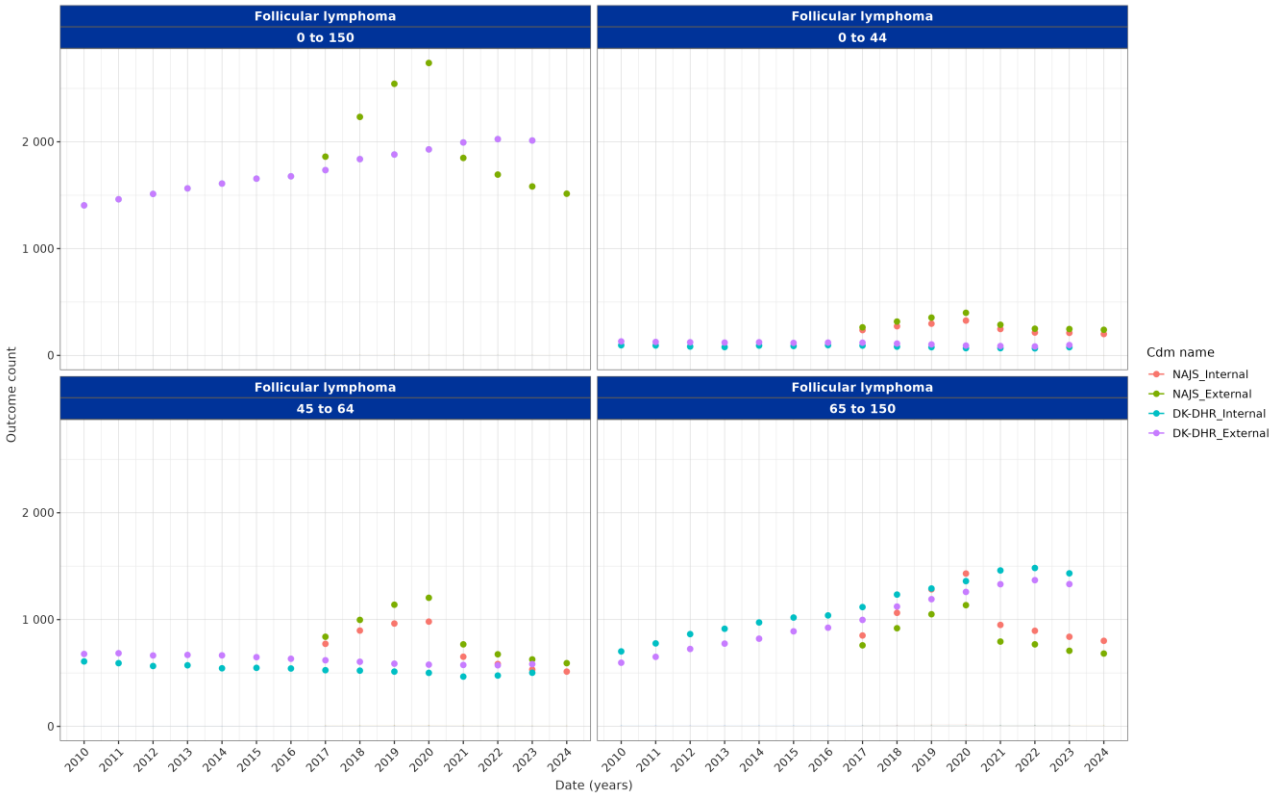

Figure S8. Number of individuals with the outcome of interest (follicular lymphoma) using internal and external denominators in NAJS and DK-DHR, stratified by age groups.

DK-DHR=Danish Data Health Registries; NAJS=Croatian National Public Health Information System.

Table S11. Estimated 5-year partial point prevalence of selected cancers in 2022, considering all recorded cases vs. cases originating from the Danish National Cancer Registry in DK-DHR.

Data source name[1]	Estimate name			
	Denominator (N)	Outcome (N)	Outcomes captured (%)[1]	Prevalence per 10,000 [95% CI]
Acute lymphoid leukaemia				
DK-DHR	5,862,553	395	N/A	0.70 (0.60–0.70)
DK-DHR (sensitivity)	5,862,553	326	82.5%	0.60 (0.50–0.60)
Acute myeloid leukaemia				
DK-DHR	5,862,553	519	N/A	0.90 (0.80–1.00)
DK-DHR (sensitivity)	5,862,553	441	85.0%	0.80 (0.70–0.80)
Chronic lymphocytic leukaemia				
DK-DHR	5,862,553	2,784	N/A	4.60 (4.40–4.70)
DK-DHR (sensitivity)	5,862,553	2,181	78.3%	3.70 (3.60 –3.90)
Diffuse large B-cell lymphoma				
DK-DHR	5,862,553	2,177	N/A	3.70 (3.60–3.90)
DK-DHR (sensitivity)	5,862,553	1,525	70.1%	2.60 (2.50–2.70)
Follicular lymphoma				
DK-DHR	5,862,553	2,025	N/A	3.50 (3.30–3.60)
DK-DHR (sensitivity)	5,862,553	1,1774	87.6%	3.00 (2.90–3.20)
Multiple myeloma				
DK-DHR	5,862,553	2,478	N/A	4.20 (4.10–4.40)
DK-DHR (sensitivity)	5,862,553	2,172	87.7%	3.70 (3.60–3.90)
Pancreatic cancer				
DK-DHR	5,862,553	1,629	N/A	2.80 (2.60–2.90)
DK-DHR (sensitivity)	5,862,553	1,267	77.8%	2.20 (2.20–2.30)
Soft-tissue sarcoma				
DK-DHR	5,862,553	2,450	N/A	4.2 (4.00–4.30)
DK-DHR (sensitivity)	5,862,553	2,059	84.0%	3.50 (3.40–3.70)

DK-DHR=Danish Data Health Registries, N/A=Not applicable.
[1]DK-DHR (sensitivity) includes cases originating from the Danish National Cancer Registry. Data from the Danish National Cancer Registry was available up to the end of 2022.

Table S12. Estimated 5-year partial point prevalence of selected cancers in 2022, considering all recorded cases vs. cases originating from the Cancer Registry in NAJS.

Data source name[1]	Estimate name			
	Denominator (N)	Outcome (N)	Outcomes captured (%)[1]	Prevalence per 10,000 [95% CI]
Acute lymphoid leukaemia				
NAJS	4,238,792	471	N/A	1.10 (1.00–1.20)
NAJS (sensitivity)	4,238,792	237	50.3%	0.60 (0.50–0.60)
Acute myeloid leukaemia				
NAJS	4,238,792	673	N/A	1.60 (1.50–1.70)
NAJS (sensitivity)	4,238,792	384	57.1%	0.90 (0.80–1.00)
Chronic lymphocytic leukaemia				
NAJS	4,238,792	2,053	N/A	4.80 (4.60–5.10)
NAJS (sensitivity)	4,238,792	760	37.0%	1.80 (1.70–1.90)
Diffuse large B-cell lymphoma				
NAJS	4,238,792	1,266	N/A	3.00 (2.80–3.20)
NAJS (sensitivity)	4,238,792	760	60.0%	1.80 (1.70–1.90)
Follicular lymphoma				
NAJS	4,238,792	1,693	N/A	4.00 (3.80–4.20)
NAJS (sensitivity)	4,238,792	469	27.7%	1.10 (1.00–1.20)
Multiple myeloma				
NAJS	4,238,792	1,893	N/A	4.50 (4.30–4.70)
NAJS (sensitivity)	4,238,792	889	47.0%	2.10 (2.00–2.20)
Pancreatic cancer				
NAJS	4,238,792	1,594	N/A	3.80 (3.60–3.90)
NAJS (sensitivity)	4,238,792	728	45.7%	1.70 (1.60–1.80)
Soft-tissue sarcoma				
NAJS	4,238,792	1,747	N/A	4.10 (3.90–4.30)
NAJS (sensitivity)	4,238,792	478	27.4%	1.10 (1.00–1.20)

NAJS=Croatian National Public Health Information System; N/A=Not applicable
[1]NAJS (sensitivity) includes cases originating from the Cancer Registry at the Croatian Institute of Public Health.

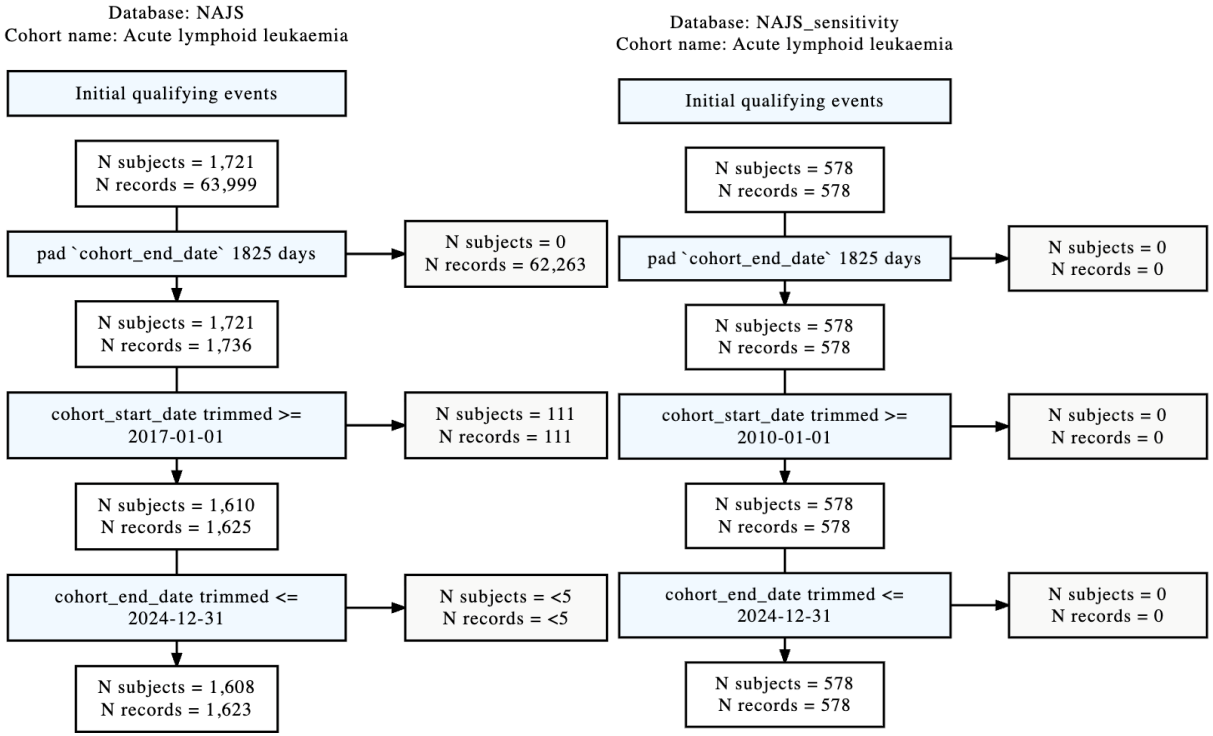

Figure S9. Flowchart depicting individuals with acute lymphoid leukaemia in NAJS, in the main (study period restriction starting from 2017 onward) vs. sensitivity analysis (no study period start restriction; cases restricted to those originating from the Cancer Registry).

NAJS=Croatian National Public Health Information System

Table S13. Estimated 5-year partial point prevalence per 10,000 of rare blood cancers in 2020: Comparison of current estimates with those obtained in the prior DARWIN EU® study.

	Prevalence per 10,000 [95% CI]											
	Reference study (EUPAS50800)[1]					Current study (EUPAS1000000715)						
	IQVIA LPD Belgium	IQVIA DA Germany	IPCI	SIDIAP	CPRD GOLD	NAJS	NAJS (sensitivity)[2]	DK-DHR	DK-DHR (sensitivity)[2]	InGef RDB	NCR	CRN
Acute lymphoid leukaemia	0.44 (0.27–0.71)	0.65 (0.59–0.71)	N/A	0.57 (0.51–0.641)	0.49 (0.42–0.58)	1.70 (1.60–1.90)	0.50 (0.50–0.60)	0.70 (0.60–0.70)	0.50 (0.50–0.60)	1.60 (1.50–1.70)	0.52 (0.48–0.55)	0.61 (0.54–0.68)
Acute myeloid leukaemia	N/A	0.99 (0.92–1.06)	N/A	1.03 (0.95–1.12)	0.72 (0.62–0.82)	1.90 (1.80–2.10)	0.90 (0.80–1.00)	1.00 (0.90–1.10)	0.80 (0.80–0.90)	1.80 (1.70–1.90)	0.77 (0.73–0.81)	0.66 (0.59–0.73)
Chronic lymphocytic leukaemia	2.83 (2.34–3.43)	4.13 (3.98–4.28)	N/A	3.82 (3.66–3.98)	2.98 (2.79–3.18)	6.40 (6.20–6.70)	1.80 (1.60–1.90)	4.80 (4.60–4.90)	3.70 (3.50–3.80)	3.20 (3.10–3.30)	2.90 (2.82–2.98)	3.02 (2.87–3.17)
Diffuse large B-cell lymphoma	N/A	0.47 (0.42–0.52)	N/A	1.73 (1.62–1.84)	0.91 (0.80–1.02)	3.60 (3.40–3.80)	1.70 (1.60–1.80)	3.80 (3.60–4.00)	2.70 (2.50–2.80)	3.10 (3.00–3.20)	2.74 (2.67-2.82)	2.58 (2.45-2.72)
Follicular lymphoma	1.47 (1.13–1.92)	0.90 (0.83–0.97)	N/A	2.83 (2.70–2.97)	1.35 (1.22–1.49)	6.50 (6.30–6.80)	1.10 (1.00–1.30)	3.30 (3.20–3.50)	2.80 (2.70–3.00)	2.30 (2.20–2.40)	1.69 (1.63–1.75)	2.00 (1.88–2.12)
Multiple myeloma	2.15 (1.73–2.68)	4.27 (4.12–4.42)	3.02 (2.72–3.34)	2.89 (2.76–3.04)	2.67 (2.49–2.86)	5.50 (5.30–5.70)	2.10 (2.00–2.30)	3.90 (3.70–4.00)	3.30 (3.20–3.50)	3.60 (3.50–3.70)	2.84 (2.76–2.92)	3.16 (3.01–3.32)

CPRD GOLD=Clinical Practice Research Datalink GOLD; DK-DHR=Danish Data Health Registries; InGef RDB=InGef Research Database; IPCI=Integrated Primary Care Information; IQVIA DA Germany=IQVIA Disease Analyzer Germany; IQVIA LPD Belgium= IQVIA Longitudinal Patient Database Belgium; NAJS=Croatian National Public Health Information System; N/A=Not applicable; SIDIAP=The Information System for Research on Primary Care.

[1] All results are available in EUPAS50800. All data sources contained outpatient electronic health records, except for SIDIAP, which also had hospital linkage. Data sources were from the following countries: Belgium (IQVIA LPD Belgium), Germany (IQVIA DA Germany), the Netherlands (IPCI), Spain (SIDIAP), and the United Kingdom (CPRD GOLD).
[2] NAJS (sensitivity) includes cases originating from the Cancer Registry at the Croatian Institute of Public Health.
[3] DK-DHR (sensitivity) includes cases originating from the Danish National Cancer Registry.

ANNEX V: Glossary

Additional definitions are available in the EMA Glossary of terms https://www.ema.europa.eu/en/about-us/glossaries.

Aggregated Data

Data collected and combined from multiple sources to generate summary information, typically anonymised.

Benefit-Risk Assessment

Evaluation of the positive therapeutic effects of a medicine compared to its risks (e.g., side effects).

Common Data Model (CDM)

A standardized data structure that enables data from multiple sources to be harmonized, making analysis consistent and reproducible. DARWIN EU® utilises the OMOP CDM maintained by the OHDSI community.

Complex Studies (C3)

Studies requiring the development or customisation of specific study designs, protocols, and Statistical Analysis Plans (SAPs), with extensive collection or extraction of data. Examples include etiological studies measuring the strength and determinants of an association between an exposure and the occurrence of a health outcome in a defined population considering sources of bias, potential confounding factors, and effect modifiers.

Coordination Centre (CC)

The central hub responsible for managing and overseeing the activities within DARWIN EU®. It is based at Erasmus University Medical Centre in Rotterdam, the Netherlands.

Data Access

The process of obtaining permission to use specific datasets for regulatory or scientific studies.

Data Quality Framework

A set of standards and procedures to ensure accuracy, completeness, timeliness, and consistency of data used in DARWIN EU®.

Data Source

A database or repository of structured health-related data, such as electronic health records (EHRs), insurance claims, or registries.

DARWIN EU®

The European Medicines Agency's (EMA) federated network of real-world data sources designed to generate evidence to support regulatory decision-making.

EMA (European Medicines Agency)

The regulatory body responsible for the evaluation and supervision of medicinal products in the EU, overseeing DARWIN EU®.

Evidence Generation

The process of analysing real-world data to produce scientific information that can inform healthcare or regulatory decisions.

Federated Network

A data infrastructure where data remain at their original location but can be analysed in a harmonised way across multiple partners using a common model and tools.

GDPR (General Data Protection Regulation)

The EU regulation governing the protection of personal data and privacy, crucial to how DARWIN EU® handles health data.

Health Technology Assessment (HTA)

A systematic evaluation of properties and impacts of health technology, often using DARWIN EU® data to support assessments.

Metadata

Descriptive information about a data source (e.g., its content, quality, and structure), essential for identifying relevant databases in DARWIN EU® studies.

Off-the-Shelf Studies (OTS)

Studies for which a standard protocol per study/analysis type and standardised analytics may be developed and applied or adapted, typically relating to a descriptive research question. This includes studies on disease epidemiology, for example, the estimation of the prevalence or incidence of health outcomes in defined time periods and population groups, or drug utilisation studies at the population or patient level.

OHDSI (Observational Health Data Sciences and Informatics)

An open-science collaborative community that develops tools and standards (including the OMOP CDM) to enable large-scale analytics of observational health data. OHDSI provides the technical and scientific foundation for DARWIN EU®'s analytical ecosystem.

Patient-Level Data

Data related to individuals, de-identified, used for longitudinal or detailed analyses.

OMOP (Observational Medical Outcomes Partnership)

A common data model (CDM) that standardises the structure and content of observational healthcare data, enabling systematic analysis across disparate datasets. DARWIN EU® uses the OMOP CDM to ensure interoperability and consistency in real-world evidence generation.

Real-World Data (RWD)

Data relating to individual health status or healthcare delivery that is collected from routine clinical practice rather than from randomised controlled trials.

Real-World Evidence (RWE)

Clinical evidence derived from the analysis of RWD, used to inform decisions by regulators, payers, or clinicians.

Regulatory Decision-Making

The process by which authorities like EMA assess data to authorise, monitor, or modify the use of medicines in the EU.

Routine Repeated Studies (RR)

Studies that are either Off-the-Shelf or Complex studies repeated on a regular basis, following the same protocol and study code, but with updated data and/or different data partners.

Study Protocol

A detailed plan describing how a specific real-world study will be conducted, including objectives, design, data sources, and analyses.

Very Complex Studies (C4)

Studies which cannot rely only on electronic health care databases, or which would require complex methodological work, for example, due to the occurrence of events that cannot be defined by existing diagnosis codes, including events that do not yet have a diagnosis code, where it may be necessary to combine a diagnosis code with other data such as results of laboratory investigations. These studies might require the collection of data prospectively, or the inclusion of new (not previously onboarded) data sources.