Study Report

P4-C1-019

DARWIN EU® - Characteristics of individuals with acute graft vs. host disease and acute graft vs. host disease with intestinal involvement

29/04/2026

Version 4.0

Authors: Katia Verhamme, Marzyeh Amini, Guido van Leeuwen

Public

CONTENTS

Study title	DARWIN EU® - Characteristics of individuals with acute graft vs. host disease and acute graft vs. host disease with intestinal involvement
Study report version	V4.0
Date	29/04/2026
EUPAS number	EUPAS1000000878
Active substance	• Systemic steroids • Second line pharmacological treatment of aGvHD: • Ruxolitinib • Basiliximab • Brentuximab • Etanercept • Infliximab • IL-2 • Mycophenolate • Rituximab • Tocilizumab • Vedolizumab • Calcineurin inhibitors • Anti-thymocyte globulin (ATG) • Inolimomab • Alemtuzumab • Pentostatin • Methotrexate • mROR inhobotpr (sirolimus/rapamycin)
Medicinal product	Ruxolitinib
Research question and objectives	Research question What are the characteristics of individuals with acute graft vs. host disease (aGvHD) and aGvHD with intestinal involvement? Objectives 1. Describe the baseline characteristics (age, sex, transplant type, indication of transplant, treatment (systemic corticosteroids, ruxolitinib, and other second/third line treatment for aGvHD), and mortality) of patients with aGvHD. 2. Describe the baseline characteristics (age, sex, transplant type, indication of transplant, treatment (systemic corticosteroids, ruxolitinib, and other second/third line treatment for aGvHD), and mortality) of patients with aGvHD with intestinal involvement.
Countries of study	France, Germany, Spain
Authors	Katia Verhamme, k.verhamme@darwin-eu.org Marzyeh Amini, m.amini@darwin-eu.org

LIST OF ABBREVIATIONS

Acronyms/terms	Description
aGvHD	Acute Graft versus Host Disease
AML	Acute Myeloblastic Leukaemia
allo-HSCT	Allogeneic hematopoietic stem cell transplantation
ALL	Acute Lymphoblastic Leukaemia
APHM	Assistance publique Hôpitaux de Marseille
ATC	Anatomical Therapeutic Chemical
ATG	Anti-thymocyte globulin
CC	Coordination centre
CDM	Common Data Model
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DRE	Digital Research Environment
ED	Emergency Department
EHR	Electronic Health Records
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
EUPAS	EU Post-Autorisation Studies Register
GDPR	General Data Protection Regulation
GI	Gastrointestinal
GP	General Practitioner
GvHD	Graft versus Host Disease
HSCT	Hematopoietic stem cell transplantation
ICD	International Classification of Diseases
ID	Index Date
IL	Interleukin
InGef RDB	InGef Research Database
IP	Inpatient
IRB	Institutional Review Board
IQR	Interquartile Range
MDS	Myelodysplastic Syndrome
MSCs	mesenchymal stem cells
mTOR	Mechanistic Target of Rapamycin
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient

Acronyms/terms	Description
OPS	Operation and Procedure Classification System
RxNorm	Medical prescription normalised
SNOMED	Systemised Nomenclature of Medicine
VID	Valencia Health System Integrated Dataset
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Characteristics of individuals with acute graft vs. host disease and acute graft vs. host disease with intestinal involvement

2. DESCRIPTION OF THE STUDY TEAM

Study team role	Names	Organisation
Principal Investigator	Katia Verhamme Marzyeh Amini Guido van Leeuwen	Erasmus MC
Data Scientist	Ioanna Nika Ger Inberg Maarten van Kessel Cesar Barboza Ross Williams	Erasmus MC
Study Manager	Natasha Yefimenko	Erasmus MC
Data source	**Names**	**Data Partner Organisation***
APHM	Laurent Boyer Vanessa Pauly Dorian Grousset	Assistance publique Hôpitaux de Marseille
InGef RDB	Raeleesha Norris Annika Vivirito Alexander Harms	InGef - Institute for Applied Health Research Berlin GmbH
VID	Celia Robles Cabaniñas Fran Llopis Cardona Gabriel Sanfélix Gimeno	Valencia Health System Integrated Dataset

*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® - Characteristics of individuals with acute graft vs. host disease and acute graft vs. host disease with intestinal involvement

Rationale and background

Acute graft-versus-host disease (aGvHD) is a serious and potentially life-threatening complication that can occur after allogeneic hematopoietic stem cell transplantation (allo-HSCT).

Intestinal involvement in aGvHD is particularly severe and is associated with poor prognosis. This form of aGvHD is often resistant to first-line corticosteroid therapy, and failure of second-line treatments like ruxolitinib further complicates management and has a poor prognosis.

Given the clinical severity and poor prognosis often associated with aGvHD with intestinal involvement, there is a need to generate real-world data on the characteristics, treatment use, and outcomes of these patients.

Research question and objectives

What are the characteristics of individuals with acute graft vs. host disease (aGvHD) and acute graft vs. host disease (aGvHD) with intestinal involvement?

1. Describe the baseline characteristics (age, sex, transplant type, indication of transplant, treatment (corticosteroids, ruxolitinib, and other second/third line treatment for aGvHD), and mortality) of patients with aGvHD.
2. Describe the baseline characteristics (age, sex, transplant type, indication of transplant, treatment (corticosteroids, ruxolitinib, and other second/third line treatment for aGvHD), and mortality) of patients with aGvHD with intestinal involvement.

Methods

A retrospective cohort study was conducted to characterise individuals with aGvHD (with or without intestinal involvement).

Index date was the date of diagnosis of aGvHD (with or without intestinal involvement), and individuals were followed up for up to 365 days following diagnosis of aGvHD, loss of follow-up, mortality, or end of observation period, whichever came first.

The following covariates were assessed: demographic characteristics (sex, age), drug prescriptions (systemic corticosteroids, ruxolitinib, and other second/third line treatment of aGvHD), transplant type, indication of use, and mortality.

Data were used from three sources: a hospital data source (Assistance publique Hôpitaux de Marseille (APHM), France), a claims data source (InGef Research Database (InGef RDB), Germany), and a linked primary/secondary care data sources (Valencia Health System Integrated Dataset (VID), Spain).

Covariates of interest were summarized using counts and proportions. Patient demographics characterisation was conducted at index date. Registration of the type of transplant and indication for transplant was assessed in the window of 100 days prior to the index date. Use of corticosteroids, ruxolitinib, and other second/third line treatment was assessed in 3 windows namely 30 days following index date, from 31 days up to 90 days following the index date, and in the window from 91 to 365 days following index date. Furthermore, mortality was assessed within these respective windows.

Results

Across three real-world data sources, 955 individuals with incident aGvHD were identified (APHM: 74; InGef RDB: 669; VID: 212), of whom 595 (62.3%) had aGvHD with intestinal involvement. APHM represented a predominantly paediatric population (median age 5 years; 95.9% <18 years), whereas InGef RDB and VID mainly included adults (median age 52–57 years), with males comprising 56.6%–66.2% of patients across cohorts. Stem cell transplantation via peripheral blood was the most common stem cell source in adult datasets (≈72%–79%), while bone marrow and cord blood transplants were infrequent. The type of allogeneic transplantation was not recorded in the paediatric data source.

Indication for transplant was dominated by haematological malignancies, with marked variation across data sources and age groups. In the InGef RDB, acute myeloblastic leukaemia was recorded in 47.4% of individuals with aGvHD (50.3% with intestinal involvement) followed by acute lymphoblastic leukaemia (15.3% and 12.1%), myelodysplastic syndrome (14.7% and 15.3%), and lymphoma (11.5% and 12.1%), respectively. Similar patterns were observed in VID, with acute myeloblastic leukaemia predominating (29.3%; 34.5%). In APHM, acute lymphoblastic leukaemia was most frequent (36.5%).

Line of therapy was not assessed; instead, treatment proportions were described in predefined time windows after aGvHD diagnosis. Systemic corticosteroids were the most commonly recorded treatment in the included data sources, although not all patients would be expected to receive them as disease severity was not assessed. In the overall aGvHD cohort, corticosteroid use within 0–30 days after diagnosis was 59.5% in InGef RDB and 28.3% in VID, with continued use over longer follow-up (48.6% and 41.4% during days 91–365, respectively). Among individuals with aGvHD with intestinal involvement, early steroid use remained high, (InGef RDB: 60.8%; VID: 32.2%). In APHM, in the overall aGvHD cohort, systemic corticosteroid use ranged from 48.7% to 52.8% in the first 90 days, declining to 31.8% thereafter. Ruxolitinib use increased over time after diagnosis (InGef RDB: 15.1% in days 0–30 to 29.1% in days 91–365; VID: 7.1% to 12.2%) and was more common in patients with intestinal involvement (InGef RDB: 33.9%; VID: 17.7% in days 91–365). Calcineurin inhibitors were commonly prescribed early after diagnosis, (InGef RDB: 72.8% and APHM: 66.2% in days 0–30), with declining use over time (48.8% and 34.9% in days 91–365, respectively); similar early use was observed in intestinal involvement (64.8% in InGef RDB, days 0–30). Mycophenolate was also commonly used early after diagnosis (42.2% in InGef RDB and 13.7% in VID during days 0–30).

Mortality increased over time and was consistently higher among patients with aGvHD and intestinal involvement than among the overall aGvHD population, with early mortality within 30 days ranging from 6.0%–7.8% versus 3.8%–6.0%, and the highest mortality observed during days 91–365 (up to 19.1% in VID). The proportion of patients who died in the paediatric cohort (APHM) could not be assessed, with fewer than five deaths reported.

Discussion

In this multi-database real-world study, heterogeneity in patient characteristics and pre-transplant conditions was observed across paediatric and adult populations. The recorded treatment patterns were generally in line with expected use of corticosteroids and later use of ruxolitinib but should be interpreted cautiously because disease severity was not assessed. Intestinal involvement was common and was associated with more intensive treatment use and persistently higher mortality over time, particularly beyond 90 days. These findings highlight the significant burden of aGvHD with intestinal involvement and underscore the ongoing need for improved therapeutic strategies and outcomes in these high-risk patients.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study deliverable	Timelines (planned)	Timelines (actual)
Final Study Protocol	December 2025	16 January 2026
Creation of Analytical code	December 2025	December 2025
Execution of Analytical Code on the data	January 2026	15 January 2026
Draft Study Report	1 February 2026	1 February 2026
Final Study Report	May 2026	To be confirmed by EMA

6. RATIONALE AND BACKGROUND

Acute graft-versus-host disease (aGvHD) is a serious and potentially life-threatening complication that can occur after allogeneic hematopoietic stem cell transplantation (allo-HSCT).(1)

Intestinal involvement in aGvHD is particularly severe and is associated with poor prognosis. This form of aGvHD is often resistant to first-line corticosteroid therapy, and failure of second-line treatments like ruxolitinib further complicates management and has a reserved prognosis.(2-4)

Given the high mortality associated with advanced gastrointestinal (GI) aGvHD, especially after failure of standard treatments, there is a need to generate data on use of treatment for aGvHD with intestinal involvement in real life.

7. RESEARCH QUESTION AND OBJECTIVES

Research questions

What are the characteristics of individuals with aGvHD and aGvHD with intestinal involvement?

Research objectives

The specific objectives of this study were:

1. Describe the baseline characteristics (age, sex, transplant type, indication of transplant, treatment (corticosteroids, ruxolitinib, and other second/third line treatment for aGvHD), and mortality) of patients with aGvHD.
2. Describe the baseline characteristics (age, sex, transplant type, indication of transplant, treatment (corticosteroids, ruxolitinib, and other second/third line treatment for aGvHD), and mortality) of patients with aGvHD with intestinal involvement.

8. RESEARCH METHODS

8.1. Study design

A cohort study was conducted using routinely collected health data from 3 data sources from 3 countries across Europe. The study was comprised of:

- A characterisation study to address objectives 1 and 2.

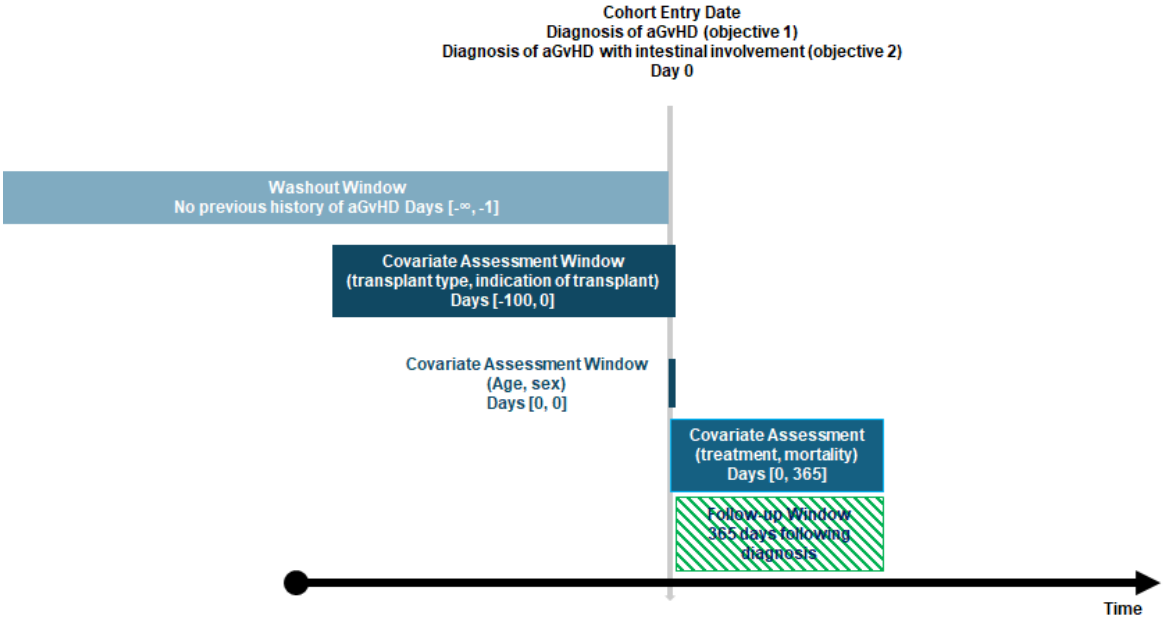

Figure 1. Graphical depiction of the study design for individuals with aGvHD and individuals with aGvHD with intestinal involvement.

8.2. Follow-up

To assess the proportion of individuals receiving treatment or who die within 30 days, within 31–90 days, and within 91–365 days of the index date (i.e., following diagnosis of aGvHD (objective 1) or diagnosis of aGvHD with intestinal involvement (objective 2)):

Follow-up started on the date of first diagnosis of aGvHD (objective 1) with intestinal involvement (objective 2). Follow-up ended on the earliest of loss to follow-up, death, end of observation period (the latest available data), or on 365 days following diagnosis of aGvHD (objective 1) with intestinal involvement (objective 2).

8.3. Study population with inclusion and exclusion criteria

Inclusion criteria:

For objective 1, the study population consisted of all individuals diagnosed with aGvHD or aGvHD with intestinal involvement.

Exclusion criteria:

- Individuals with a history of aGvHD prior to the start of the study period

The phenotype of aGvHD with intestinal involvement was either comprised of individuals with the presence of a disease code for ‘Graft versus host disease of intestine – concept ID 37167528’ or of individuals with a disease code of ‘Acute graft vs host disease – concept ID 439416’ in combination with intestinal symptoms

during the same visit occurrence or within -7/+7 days of the index date. More details are provided in **Annex IV**.

8.4. Study setting and data sources

This study was conducted using routinely collected data from 3 data sources in the DARWIN EU® network of data partners from 3 European countries. All data were mapped a priori to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM).

Table 1. Data sources.

Country	Name of Data source	Health Care setting	Type of Data	Number of active individuals	Calendar period covered by each data source	Contributing to
France	APHM	Inpatient and Outpatient Hospital Care	EHR	249,900	February 2014 – Dec 2024	Objective 1 and 2
Germany	InGef RDB	Outpatient General Practitioner Care Inpatient and Outpatient Hospital Care	Claims	7,432,600	January 2016 – September 2025	Objective 1 and 2
Spain	VID	Outpatient General Practitioner Care Inpatient and Outpatient Hospital Care	EHR	5,575,800	January 2009 – December 2024	Objective 1 and 2

EHR = electronic health records, APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.

Data sources

1. France: Assistance publique Hôpitaux de Marseille (APHM)
2. Germany: InGef Research Database (InGef RDB)
3. Spain: Valencia Health System Integrated Dataset (VID)

Data sources selection

These data sources fulfil the criteria required in terms of data quality, completeness, timeliness, and representativeness while covering different regions of Europe (**Annex II**).

8.5. Study period

The study period was from 1st January 2020 until 31st December 2025 or the most recent data available for each contributing data source.

8.6. Variables

8.6.1. Exposure

None.

8.6.2. Outcome

Mortality was assessed in three time windows following index date, both in the aGvHD and aGvHD with intestinal involvement cohort (irrespective of treatment).
These windows were:

- In the window from index date (day 0) to 30 days following index date
- In the window from 31 days up to 90 days following index date
- In the window from 91 to 365 days following index date

For InGef, mortality was assessed in a subpopulation of patients with aGvHD, namely in patients diagnosed with aGvHD during the study period 2020–2023 as information on mortality might be incomplete for patients diagnosed after 2023.

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

aGvHD and aGvHD with intestinal involvement:

The diagnosis was based on disease codes of aGvHD with or without intestinal involvement. In case source data was not granular enough (VID and APHM), aGvHD with intestinal involvement was based on presence of aGvHD in combination with intestinal symptoms in the same visit occurrence or within -7/+7 days of the index date. For patients with only a disease code for GvHD, an additional condition of the code occurring within 100 days since the transplant was added to the definition of aGvHD to differentiate from chronic GvHD.

The covariates to characterise aGvHD (overall and with intestinal involvement) were as follows:

- Sex
  - Female/male
- Age: Median age at index date and age category (<18 years, 18–65 years, and >65 years).
- Concomitant medications:

  The proportion of individuals with aGvHD and aGvHD with intestinal involvement on treatment with systemic steroids, ruxolitinib, or other second/third line therapies was assessed in three time windows following diagnosis, namely:
  - In the window from index date (day 0) to 30 days following index date
  - In the window from 31 days up to 90 days following index date
  - In the window from 91 to 365 days following index date

  The following medications were investigated:
  - Systemic steroids (ATC code H02)
  - Second/third line treatment:
    - ruxolitinib (ATC code L01EJ01)
    - Ruxolitinib plus corticosteroids (defined as prescription of systemic steroids within -30 days/+30 days of a ruxolitinib prescription)

    - Additional systemic immunosuppressants
        - Basiliximab (ATC code L04AC02)
        - Brentuximab (ATC code L01FX05)
        - Etanercept (ATC code L04AB01)
        - Infliximab (ATC code L04AB02)
        - IL-2 (aldesleukin – ATC code L03AC01 and oprelvekin L03AC02)
        - Mycophenolate (ATC code L04AA06)
        - Rituximab (ATC code L01FA01)
        - Tocilizumab (ATC code L04AC07)
        - Vedolizumab (ATC code L04AG05)
        - Calcineurin inhibitors (i.e., tacrolimus (ATC code L04AD02) or cyclosporin (ATC code L04AD01))
        - Anti-thymocyte Immunoglobulin (ATG) (ATC code L04AA04)
        - Inolimomab (ATC code L04AC)
        - Alemtuzumab (ATC code L04AG06)
        - Pentostatin (ATC code L01XX08)
        - Mechanistic Target of Rapamycin (mTOR) inhibitor (sirolimus (ATC code L04AH01)/rapamycin (ATC code L01EG04)
        - Methotrexate (ATC code L01BA01 or ATC code L04AX03)
- Non-medical treatment of aGvHD with intestinal involvement:
    - Extracorporeal photopheresis
    - Mesenchymal stem cells (MSCs)
- Comorbidities
    - Type of Haemopoietic stem cell transplant
        - Marrow - Allogeneic bone marrow transplantation (concept ID 4242257)
        - Peripheral blood - Allogeneic peripheral blood stem cell transplant (concept ID 4143404)
        - Cord - Cord blood-derived stem-cell transplantation, allogeneic (concept ID 2721124)
    - Indication for transplant: (2, 5-7)
        - Acute myeloblastic leukaemia
        - Acute lymphoblastic leukaemia
        - Myelodysplastic syndrome (MDS)
        - Aplastic anaemia
        - Haemoglobinopathies
        - Lymphoma

The transplant type and indication for transplant were assessed in the window -100 days to the date of the index date.

The concept sets used for the identification of covariates are described in **Annex IV**. These codes were refined during the study execution following the DARWIN EU® phenotyping standard processes, which involve the review of code lists by clinical experts and the review of phenotypes after their execution in the participating data sources.

8.7. Study size

No sample size was calculated, as this is a descriptive disease epidemiology study which did not test a specific hypothesis. In addition, the study was based on secondary use of data (i.e., data already collected for other purposes than research) to characterise individuals with aGvHD (with or without intestinal involvement). Based on a preliminary feasibility assessment, the expected number of person counts for ruxolitinib (all ruxolitinib use irrespective of indication of use) in the data sources included in this study ranged from 253 (APHM) to 3,670 (InGef RDB) and from 247 (APHM) to 1,400 (InGef RDB) for aGvHD (750 individuals with aGvHD with intestinal involvement). These numbers were based on the overall number of conditions, drugs or observation registries in each data source with no filter by study period or inclusion and exclusion criteria.

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 III**), 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

None.

8.9.2. Main statistical methods

Descriptive statistics were used to summarise demographic and clinical characteristics at different time windows as described in **Section 8.6.3**.

Categorical variables (e.g., sex, indication of use, medication use, and mortality) were described using counts and percentages.

Continuous variables (e.g., age at diagnosis) were described using means, standard deviations, medians, and interquartile ranges.

8.9.3. Missing values

None.

8.9.4. Sensitivity analysis

None.

8.10. Deviations from the protocol

None.

9. RESULTS

The full set of results for this study is available through an interactive web-application ShinyApp at EUPAS1000000878

9.1. Participants

Table S1 presents the study attrition process across the three data sources, detailing the stepwise selection from all individuals with GvHD to incident aGvHD cases and the subset of aGvHD with intestinal involvement during the study period. Following application of the eligibility criteria, a total of 954 individuals with incident aGvHD were included in the analysis (APHM: 74; InGef RDB: 669; VID: 212), of whom 595 (APHM: 5; InGef RDB: 503; VID: 87) were identified as aGvHD with intestinal involvement.

Based on **Table 2**, individuals with aGvHD and those with aGvHD with intestinal involvement showed marked differences across data sources in terms of age distribution, while sex distribution was relatively consistent. As APHM is a predominantly paediatric population, the aGvHD cohort was predominantly comprise of children, with a median age of 5 years (IQR 2–11) and 95.9% of individuals were younger than 18 years. A similar pattern was observed for aGvHD with intestinal involvement (median age 4 years). In contrast, in the InGef RDB and VID data sources, both cohorts largely consisted of adults, with median ages ranging from 52 to 57 years. Across all data sources, males constituted a slightly higher proportion than females in both cohorts, with male representation ranging from approximately 56.6% to 66.2%.

Table 2. Distribution of baseline characteristics among individuals with aGvHD and individuals with aGvHD with intestinal involvement) by data source.

Characteristics	aGvHD			aGvHD with intestinal involvement		
	APHM	InGef RDB	VID	APHM	InGef RDB	VID
Overall, N	74	669	212	5	503	87
Median age (IQR) years	5 (1 – 11)	56 (40 – 63)	52 (31 – 63)	4 (1 – 15)	57 (44 – 64)	54 (34 – 64)
Mean age (SD) years	7.2 (8.2)	49.5 (19.3)	45.4 (20.8)	7.2 (7.7)	51.7 (17.6)	48.1 (19.0)
Age category, N (%)						
<18 years	71 (95.9%)	73 (10.9%)	35 (16.5%)	5 (100.0%)	37 (7.4%)	10 (11.5%)
18–65 years	<5	479 (71.6%)	149 (70.3%)	0	368 (73.1%)	65 (74.7%)
>65 years	0	117 (17.5%)	28 (13.2%)	0	98 (19.5%)	12 (13.8%)
Sex, N (%)						
Male	49 (66.2%)	403 (60.2%)	120 (56.6%)	<5	305 (60.6%)	51 (58.6%)
Female	25 (33.8%)	266 (39.8%)	92 (43.4%)	<5	198 (39.4%)	36 (41.4%)

aGvHD = acute graft versus host disease; IQR=Interquartile Range, SD=Standard Deviation, APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.

9.2. Descriptive data

None.

9.3. Main results

Indications for transplant and stem cell source

As shown in **Table 3**, the most frequently recorded conditions in the 100 days prior to aGvHD varied across data sources but were largely dominated by haematological malignancies.

In the InGef RDB, acute myeloblastic leukaemia was the most commonly recorded condition prior to aGvHD (47.4% among individuals with aGvHD and 50.3% among those with intestinal involvement), followed by acute lymphoblastic leukaemia (15.3% and 12.1%), myelodysplastic syndrome (14.7% and 15.3%), lymphoma (11.5% and 12.1%), and lastly, aplastic anaemia (5.8% and 6.4%). Similarly, in the VID data source, acute myeloblastic leukaemia was most frequent (29.3% in aGvHD and 34.5% with intestinal involvement), followed by acute lymphoblastic leukaemia (18.4% and 13.8%), myelodysplastic syndrome (17.9% and 25.3%), lymphoma (16.5% and 19.5%), and aplastic anaemia (5.7% and 8.1%). In APHM, which represents a paediatric population, acute lymphoblastic leukaemia was the most frequently recorded condition among individuals with aGvHD (36.5%), followed by acute myeloblastic leukaemia (14.9%) and aplastic anaemia (9.5%).(**Table 3**).

Based on the results presented in **Table 3**, peripheral blood was the most frequently recorded stem cell source among adults with aGvHD, including those with aGvHD with intestinal involvement, in both InGef RDB and VID. Peripheral blood accounted for 78.2% and 77.4% of transplants in the overall aGvHD population, and 72.4% and 79.3% among those with intestinal involvement, respectively. Bone marrow transplants were less common (2.8%–9.6%), and cord blood transplants were rare. The source type of stem cell transplantation was unknown in APHM.

Table 3. Distribution of type of transplant and conditions recorded before transplant among individuals with aGvHD and individuals with aGvHD with intestinal involvement (number and %) by data source (assessed within 100 days of to the index date (i.e., date of diagnosis of aGvHD).

Characteristics	aGvHD			aGvHD with intestinal involvement		
	APHM N=74	InGef RDB N=669	VID N=212	APHM N=5	InGef RDB N=503	VID N=87
Stem cell source, N (%)*						
Bone Marrow	0	64 (9.6%)	6 (2.8%)	0	31 (6.2%)	<5
Peripheral blood	0	523 (78.2%)	164 (77.4%)	0	364 (72.4%)	69 (79.3%)
Umbilical Cord	0	0	<5	0	0	<5
Unknown	74 (100%)	82 (12.2)	40 (18.9)	0	108 (21.4%)	14 (16.1%)
Conditions recorded before transplant, N (%)						
Acute myeloblastic leukaemia	11 (14.9%)	317 (47.4%)	62 (29.3%)	<5	253 (50.3%)	30 (34.5%)
Acute lymphoblastic leukaemia	27 (36.5%)	102 (15.3%)	39 (18.4%)	<5	61 (12.1%)	12 (13.8%)
Myelodysplastic syndrome	<5	98 (14.7%)	38 (17.9%)	<5	77 (15.3%)	21 (24.1%)
Aplastic anaemia**	7 (9.5%)	39 (5.8%)	12 (5.6%)	<5	32 (6.4%)	7 (8.1%)
Haemoglobinopathies	<5	<5	0	0	0	0
Lymphoma	<5	77 (11.5%)	35 (16.5%)	0	61 (12.1%)	17 (19.5%)
Unknown/Other	20 (27.3%)#	33 (4.9%)#	26 (12.3%)	<5	19 (3.8%)	0

APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.
* Stem cell source and indication was assessed in the window of -100 days to the index date. Note: The 100% 'unknown' in APHM is due to the use of a broad code, and the specific transplant type is not reported.
** This excludes aplastic anaemia caused by treatment and coded as such.
# Cells reported as <5 were imputed using the midpoint integer value of 3 to estimate counts for the "unknown/other" category; resulting proportions were calculated based on this midpoint estimate.

Treatments

Lines of treatment were not evaluated, instead all treatments recorded after diagnosis were reported in specific timeframes.

Based on **Table 4**, treatment patterns varied across data sources, cohorts, and follow-up periods, with systemic corticosteroids being the most frequently used therapy in both aGvHD and aGvHD with intestinal involvement. In InGef RDB, systemic steroids were recorded in 59.5% of individuals with aGvHD within the first 30 days following index date, 40.1% during days 31–90, and 48.6% during days 91–365 among those still under follow-up in each window. Corresponding proportions observed in the VID data source were 28.3%, 48.2%, and 41.4%, respectively. Among individuals with aGvHD with intestinal involvement, steroid use remained high, particularly in InGef RDB (60.8% in days 0–30) and VID (32.2% in days 0–30). In APHM, systemic corticosteroid use was 48.7% during days 0–30, 52.8% during days 31–90, and 31.8% during days 91–365 in individuals with aGvHD.

Ruxolitinib use was also recorded in each follow-up window. Among individuals with aGvHD, ruxolitinib use was recorded in 15.1% during days 0–30, 17.7% during days 31–90, and 29.1% during days 91–365 in InGef RDB; corresponding proportions in VID were 7.1%, 10.1%, and 12.2%. Among individuals with aGvHD with intestinal involvement, ruxolitinib was recorded in 19.9%, 21.7%, and 33.9% of individuals in InGef RDB, and in 8.1%, 19.5%, and 17.7% in VID, across the same time windows. These results reflect recorded use within

each time window among patients who remained under follow-up and should not be interpreted as treatment trajectories in the same fixed group of patients. Combination therapy with ruxolitinib and systemic steroids followed comparable trends but with slightly lower proportions (**Table 4**).

Calcineurin inhibitors were frequently used among individuals with aGvHD, particularly early after diagnosis, but these records may reflect ongoing prophylactic/background immunosuppression as well as treatment used during aGvHD management. In InGef RDB, calcineurin inhibitors were recorded in 72.8% of individuals during days 0–30, 63.8% during days 31–90, and 48.8% during days 91–365; in APHM, the corresponding proportions were 66.2%, 63.9%, and 34.9%. Among individuals with aGvHD with intestinal involvement, calcineurin inhibitors were recorded in 64.8%, 62.8%, and 49.9% of individuals in InGef RDB across the three follow-up windows. Mycophenolate was also frequently used, especially in InGef RDB, where it was prescribed to 42.2% of individuals with aGvHD and to 36.8% of those with intestinal involvement within 0–30 days. Other therapies, including biologics and non-pharmacological treatments such as extracorporeal photopheresis, were infrequently recorded across data sources (**Table 4**).

Table 4. Distribution of treatment received among individuals with aGvHD and aGvHD with intestinal involvement (number and %), irrespective of treatment start (all lines of therapy).

	aGvHD									aGvHD with intestinal involvement								
	APHM			InGef RDB			VID			APHM			InGef RDB			VID		
	N=74	N=71	N=65	N=669	N=594	N=543	N=212	N=199	N=181	N=Supp	N=Supp	N=Supp	N=503	N=425	N=387	N=87	N=77	N=68
Treatment received N (%)	Days 0–30 following ID	Days 31–90 following ID	Days 91–365 following ID	Days 0–30 following ID	Days 31–90 following ID	Days 91–365 following ID	Days 0–30 following ID	Days 31–90 following ID	Days 91–365 following ID	Days 0–30 following ID	Days 31–90 following ID	Days 91–365 following ID	Days 0–30 following ID	Days 31–90 following ID	Days 91–365 following ID	Days 0–30 following ID	Days 31–90 following ID	Days 91–365 following ID
Systemic steroids N (%)	36 (48.7)	38 (52.8)	21 (31.8)	398 (59.5)	238 (40.1)	264 (48.6)	60 (28.3)	96 (48.2)	75 (41.4)	<5	5 (100.0)	5 (100.0)	306 (60.8)	181 (42.6)	189 (48.8)	28 (32.2)	33 (42.9)	18 (26.5)
Ruxolitinib N (%)	<5	12 (16.7)	11 (16.7)	101 (15.1)	105 (17.7)	158 (29.1)	15 (7.1)	20 (10.1)	22 (12.2)	<5	<5	0	100 (19.9)	92 (21.7)	131 (33.9)	7 (8.1)	15 (19.5)	12 (17.7)
Ruxolitinib in combination with systemic steroids* N (%)	<5	12 (16.7)	9 (13.6)	89 (13.3)	87 (14.7)	118 (21.7)	8 (3.8)	19 (9.6)	19 (10.5)	<5	<5	<5	86 (17.1)	79 (18.6)	97 (25.1)	6 (6.9)	12 (15.6)	10 (14.7)
Etanercept N (%)	0	0	0	5 (0.8)	<5	5 (0.9)	0	0	0	0	0	0	5 (1.0)	<5	5 (1.3)	0	0	0
Mycophenolate N (%)	18 (24.3)	11 (15.3)	0 (0.0)	282 (42.2)	51 (8.6%	30 (5.5)	29 (13.7)	33 (16.6)	11 (6.1)	<5	<5	0	185 (36.8)	36 (8.5)	23 (5.9)	7 (8.1)	11 (14.3)	6 (8.8)
Rituximab N (%)	0	0	0	22 (3.3)	20 (3.4)	18 (3.3)	0	0	0	0	0	0	16 (3.2)	16 (3.8)	13 (3.4)	0	0	0
Vedolizumab N (%)	0	0	<5	0	<5	5 (0.9)	0	0	0	0	0	<5	<5	<5	<5	0	0	0
Calcineurin inhibitor N (%)	49 (66.2)	46 (63.9)	23 (34.9)	487 (72.8)	379 (63.8)	265 (48.8)	11 (5.2)	16 (8.0)	18 (9.9)	<5	<5	<5	326 (64.8)	267 (62.8)	193 (49.9)	<5	<5	8 (11.8)

	aGvHD									aGvHD with intestinal involvement								
ATG N (%)	14 (18.9)	0 (0.0)	<5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mTOR inhibitor (sirolimus/rapamycin) N (%)	0	<5	0	13 (1.9)	8 (1.4)	10 (1.8)	27 (12.7)	56 (28.1)	46 (25.4)	0	0	0	9 (1.8)	7 (1.7)	7 (1.8)	8 (9.2)	18 (23.4)	17 (25.0)
Methotrexate N (%)	0	0	0	<5	<5	7 (1.3)	0	0	0	0	0	0	<5	<5	5 (1.3)	0	0	0
Extracorporeal photopheresis N (%)	0	0	0	37 (5.5)	48 (8.1)	58 (10.7)	0	0	0	0	0	0	33 (6.6)	38 (8.9)	46 (11.9)	0	0	0

APHM = Assistance publique Hôpitaux de Marseille, InGef RDB = InGef Research Database, VID = Valencia Health System Integrated Dataset; mTOR = Mechanistic Target of Rapamycin; aGvHD = acute graft versus host disease; MSCs = mesenchymal stem cells; ATG = Anti-thymocyte globulin.
* Defined as prescription of systemic steroids -30 days/+30 days of a prescription of ruxolitinib, ID= Index Date (date of diagnosis of aGvHD).
** Defined as prescription of aldesleukin with ATC code L03AC01 and oprelvekin with ATC code L03AC.
Note: No counts or counts <5 were observed for Inolimomab, alemtuzumab, pentostatin, basiliximab, brentuximab, infliximab, IL-2, MSCs, and tocilizumab.
Note: Denominators vary across follow-up windows because only patients still under follow-up contributed to each window. Therefore, the same patients did not necessarily contribute to all windows.

Mortality

Table 5 provides the proportion of individuals who died within specified windows following diagnosis of aGvHD and aGvHD with intestinal involvement. Mortality increased over time in both cohorts and in all data sources, with consistently higher proportions observed among individuals with aGvHD with intestinal involvement compared with those with aGvHD overall. Within the first 0–30 days after the index date, mortality among individuals with aGvHD was 6.0% (30 individuals) in InGef RDB and 3.8% (8 individuals) in VID. In the corresponding intestinal involvement cohort, early mortality was higher, reaching 7.8% (29 individuals) in InGef RDB and 6.9% (6 individuals) in VID, with no deaths reported in APHM in any of the cohorts.

During the 31–90 days window, mortality ranged from 2.5% to 5.5% in the aGvHD cohort and from 3.2% to 9.1% in the intestinal involvement cohort across InGef RDB and VID. In the 91–365 day window, mortality further increased, reaching 5.8% in InGef RDB and 14.4% in VID for aGvHD, and 6.2% and 19.1%, respectively, for aGvHD with intestinal involvement. Mortality in APHM could not be assessed as numbers were low.

Table 5. Mortality among individuals with aGvHD and individuals with aGvHD with intestinal involvement (number and %) by data source.

	aGvHD			aGvHD with intestinal involvement		
	APHM N=74	InGef RDB N=500*	VID N=212	APHM N=5	InGef RDB N=372*	VID N=87
Day 0 to Day 30 following ID (N (%))	0	30 (6.0%)	8 (3.8%)	0	29 (7.8%)	6 (6.9%)
Day 31 to Day 90 following ID (N (%))	<5	11 (2.5%)	11 (5.5%)	0	10 (3.2%)	7 (9.1%)
Day 91 to Day 365 following ID (N (%))	<5	24 (5.8%)	26 (14.4%)	<5	18 (6.2%)	13 (19.1%)

aGvHD = acute graft versus host disease; APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset, ID= index date (date of diagnosis of aGvHD/aGvHD with intestinal involvement). Numbers in cells are mutually exclusive.
Note: The denominators were not constant over the different windows in case the individuals were no longer in follow-up (e.g., exclusion of patients who died in the previous window from the denominator of interest).
* For InGef, mortality was assessed in a subpopulation of patients with aGvHD, namely in those patients diagnosed with aGvHD during the study period 2020 – 2023 as information on mortality might be incomplete for those patients diagnosed after 2023.

10. DISCUSSION

10.1. Key results

Across the three data sources, a total of 955 individuals with incident aGvHD were included in the analysis (APHM: 74; InGef RDB: 669; VID: 212), of whom 595 individuals (APHM: 5; InGef RDB: 503; VID: 87) were identified as having aGvHD with intestinal involvement. Demographic differences were observed across data sources: APHM represented a predominantly paediatric population, with a median age of 5 years (IQR 2–11) and 95.9% of patients were younger than 18 years, whereas InGef RDB and VID mainly included adults, with median ages ranging from 52 to 57 years and a substantial proportion of patients aged over 65 years. Across all cohorts, males constituted a higher proportion than females, accounting for approximately 56.6% to 67.1% of patients.

Conditions recorded before aGvHD (potential indications) were mainly haematological malignancies. In the InGef RDB, acute myeloblastic leukaemia was most frequent (47.4% overall; 50.3% with intestinal involvement), followed by acute lymphoblastic leukaemia (15.3%; 12.1%), myelodysplastic syndrome (14.7%; 15.3%), and lymphoma (11.5%; 12.1%), while aplastic anaemia was less common (5.8%; 6.4%). In the VID data source, acute myeloblastic leukaemia was the most common condition (29.3% in aGvHD and 34.5% with intestinal involvement), followed by acute lymphoblastic leukaemia (18.4% and 13.8%) and myelodysplastic syndrome (17.9% and 25.3%). Aplastic anaemia was observed in 5.7% of patients with aGvHD and 8.1% of those with intestinal involvement. In the paediatric data source APHM, acute lymphoblastic leukaemia was the most frequently recorded indication (36.5%), followed by acute myeloblastic leukaemia (14.9%) and aplastic anaemia (9.5%). Peripheral blood was the main stem cell source in adult datasets (72.4%–78.2%), whereas the type of stem cell source was not captured in APHM.

Treatment patterns varied across data sources, cohorts, and follow-up periods. Systemic corticosteroids were the most frequently used treatment across all cohorts, particularly within 0–30 days after diagnosis (59.5% in InGef RDB and 28.3% in VID), with continued use over longer follow-up (48.6% and 41.4% during days 91–365, respectively). Among individuals with aGvHD with intestinal involvement, steroid use remained high, particularly in InGef RDB (60.8% in days 0–30) and VID (32.2% in days 0–30). In APHM, use ranged from 48.7% to 52.8% in the first 90 days, declining to 31.8% thereafter. Ruxolitinib use increased over time after diagnosis, suggesting later-line use, rising in InGef RDB from 15.1% (days 0–30) to 29.1% (days 91–365), and in VID from 7.1% to 12.2%. Ruxolitinib was more common in patients with intestinal involvement, reaching 33.9% in InGef RDB and 17.7% in VID during days 91–365. Calcineurin inhibitors were commonly prescribed early after diagnosis (72.8% in InGef RDB during days 0–30), with decreasing use over time (48.8% during days 91–365). Among those with intestinal involvement, early use was also common (64.8% in days 0–30 in InGef RDB), with lower proportions over longer follow-up. While mycophenolate was also frequently used early after diagnosis (42.2% in InGef RDB and 13.7% in VID during days 0–30).

Mortality increased over time across data sources and was consistently higher among individuals with aGvHD with intestinal involvement than among those with aGvHD overall. Early mortality within 30 days was higher in the intestinal involvement cohort (7.8–6.9%) compared with aGvHD overall (3.8–6.0%), and this difference persisted at longer follow-up, with the highest mortality observed during 91–365 days, particularly in VID (19.1% vs. 14.4%). Mortality in the paediatric APHM data source was low, with fewer than five deaths reported in either cohort.

10.2. Strengths and limitations of the research methods

The study was informed by routinely collected health care data, so data quality issues must be considered. The study population of this study consists of individuals with aGvHD and individuals with aGvHD with intestinal involvement. Outcome misclassification may occur due to coding limitations. First, at time of study start, not all source data were mapped to aGvHD (with intestinal involvement). For those where data

was already (partially) mapped, there was the potential for misclassification between aGvHD and chronic GvHD in case source data did not capture sufficient granularity to distinguish between acute and chronic graft vs host disease. To minimize this misclassification, we applied an additional criteria classifying any GvHD diagnosis as aGvHD diagnosis if the diagnosis occurs within 100 days of transplantation. More importantly, not all data sources have source codes for aGvHD with intestinal involvement (this is only the case for InGef RDB). For those data sources where this level of granularity is missing within the source code (VID and APHM), the phenotype of aGvHD with intestinal involvement was based on disease codes for aGvHD in combination with intestinal symptoms during the same visit occurrence.

As part of characterisation, we described the type of transplant procedure (peripheral blood, bone marrow, peripheral blood) but this information might be incomplete within the data sources. Furthermore, we may misclassify the indication of allo-HSCT in case the diseases of interest (acute myeloblastic leukaemia, acute lymphoblastic leukaemia, myelodysplastic syndrome, aplastic anaemia, haemoglobinopathies, lymphoma, and other malignant conditions) are not reported. Regarding aplastic anaemia, some misclassification is possible in case the source code lacked granularity to differentiate between aplastic anaemia as an indication for the transplant vs. aplastic anaemia as a result of the treatment of the malignant condition.

The proportion of treated individuals with aGvHD may be underestimated in InGef RDB data source, because the Drug_Exposure table primarily captures medications dispensed at pharmacies, such as prescriptions provided to patients in outpatient settings and subsequently billed to health insurance. Some inpatient medication use is also recorded, including both oral and parenteral drugs (e.g. Operation and Procedure Classification System (OPS) code 6-009.4 for oral ruxolitinib); however, the recording of inpatient treatments depends on their relevance for hospital billing and whether a specific OPS code exists to document them. Misclassification of ruxolitinib in InGef RDB is low as it is a new and expensive drug. In contrast, systemic steroids are relatively well captured in the data, as they are frequently prescribed in outpatient settings and dispensed through pharmacies, resulting in pharmacy dispensing records in the data source.

In the InGef RDB, the date of death is recorded as the last day of the respective quarter in which the death occurred rather than the exact date. For this study, this means that deaths may be assigned to a later time window than the one in which they actually occurred, and some deaths occurring near the end of the 365-day follow-up period may not be captured within the 1-year analysis window. Therefore, mortality estimates from InGef RDB, particularly those by follow-up window and near 1 year after index, should be interpreted with caution. In addition, completeness death data for more recent calendar years (2024–2025) may be reduced due to delays in source data availability. To reduce this issue, mortality analyses in InGef RDB were restricted to patients with index dates between 2020 and 2023, for whom complete 1-year follow-up was expected. Importantly, this misclassification is not differential; therefore a comparative measure will still have some value.

In APHM hospital data source, vital status is retrieved from healthcare coverage details and captures deaths occurring within the hospital setting; therefore, deaths occurring outside the hospital may not be fully captured, potentially leading to underestimation of mortality. We do not expect this effect to be large, as these patients are not likely to be discharged soon after diagnosis.

Additionally, the results of this study only reflect the populations from the included data sources. Electronic health records have certain inherent limitations because they were collected for clinical purposes rather than primarily for research use. Consequently, using 3 types of data sources from France, Germany, and Spain limits generalisability to those countries.

10.3. Interpretation

This multi-database real-world analysis shows aGvHD treatment patterns largely consistent with current clinical practice. Systemic corticosteroids were the predominant first-line therapy, with increasing use of

ruxolitinib over time. Across all data sources, intestinal involvement was associated with higher mortality, highlighting this subgroup as a major driver of poor outcomes and an area of persistent unmet need.

The observed age pattern likely reflects the underlying populations captured by each data source rather than a common age distribution across all patients with aGvHD. APHM is a reference centre for paediatric oncology explaining the younger population (median ~4–5 years; ~all <18), while InGef RDB and VID are predominantly adult (median ~52–57 years; most 18–65, then >65). Sex distribution is comparatively stable across sources (male predominance ~57–67%). This is consistent with the transplant and GvHD literature, where outcomes are influenced more by disease characteristics and transplant-related factors than by sex, and risk assessment focuses mainly on disease severity and organ involvement, such as GI involvement, rather than demographics alone.(8)

Proper line of therapy was not assessed; instead, treatment proportions were described across predefined time windows following aGvHD diagnosis. The treatment patterns observed across data sources were in line with current clinical practice for aGvHD. Systemic corticosteroids were the most frequently used first-line therapy, reflecting guideline recommendations that steroids remain the standard initial treatment for clinically significant aGvHD.(9) Use of ruxolitinib increased, particularly in later follow-up periods, which is in line with its established role in steroid-refractory aGvHD following evidence from pivotal clinical trials demonstrating improved response rates compared with best available therapy.(4, 10, 11)

Calcineurin inhibitors and mycophenolate were commonly recorded, especially shortly after diagnosis, consistent with their routine use as background immunosuppression and adjunctive therapy in aGvHD management.(12) Other treatments, such as extracorporeal photopheresis or biologic agents, were less frequently observed, likely reflecting their use in selected refractory cases and variable availability and coding in real-world data.(12)

The observed higher mortality and more intensive treatment patterns among individuals with aGvHD with intestinal involvement are consistent with the recognized clinical severity of GI aGvHD, which is associated with poorer prognosis, higher steroid refractoriness, and increased risk of complications.(13) The increasing mortality over time, particularly beyond 90 days, is likely because of the cumulative burden of treatment-related toxicity, infections, and chronic organ damage following allogeneic transplantation.(14, 15)

The observed distribution of transplant conditions aligns with established clinical guidelines and the known role of haematopoietic stem cell transplantation in managing high-risk haematological malignancies. Transplant indications were predominantly haematological malignancies across the adult data sources, particularly acute myeloblastic leukaemia and myelodysplastic syndrome. This is expected, as allogeneic haematopoietic stem cell transplantation remains a potentially curative option for high-risk or relapsed acute leukaemia and advanced myelodysplastic syndromes due to its graft-versus-leukaemia effect.(16, 17) In the paediatric cohort, acute lymphoblastic leukaemia was most frequent, which is in line with current practice where haematopoietic stem cell transplantation is recommended for children with high-risk features or relapsed disease to improve long-term survival.(18)

Hematopoietic stem cell transplant via peripheral blood is the dominant recorded graft source in InGef RDB and VID (roughly three-quarters of transplants), with blood marrow less common and cord blood rare. This reflects current allogeneic hematopoietic stem cell transplantation practice, where peripheral blood stem cell transplantation is commonly used because of practical advantages, such as faster hematopoietic recovery.(19) However, studies comparing peripheral blood and bone marrow transplantation show important trade-offs, with peripheral blood often associated with different GvHD profiles, including a higher risk of chronic GvHD. Effects on relapse and survival vary depending on the clinical context.(20)

Overall, these findings shows that real-world data can be used to describe patients with aGvHD and aGvHD with intestinal involvement across different healthcare data sources. However, the findings should be interpreted with caution because of important limitations, including heterogeneity across data sources,

differences in age and care setting, possible misclassification of aGvHD and intestinal involvement, incomplete capture of some treatments, and lack of information on disease severity and line of therapy. Therefore, the results should be seen as descriptive and source-specific rather than as a basis for direct clinical decision-making. Within these limitations, the consistently higher mortality observed among patients with intestinal involvement suggests a higher burden in this subgroup.

10.4. Generalisability

The results reflect only the populations captured in the included data sources and should be interpreted as source-specific descriptions. The study populations were highly heterogeneous, particularly with respect to age and care setting: APHM mainly captured a paediatric hospital-based population, whereas InGef RDB and VID mainly captured adult populations. In addition, differences in healthcare setting, coding practices, and data capture across data sources limit direct comparability and reduce the generalisability of the findings beyond these specific populations and healthcare systems.

11. CONCLUSION

This study provides a descriptive overview of patients with aGvHD and aGvHD with intestinal involvement in three European data sources. The results show variation in patient profiles, recorded treatments, and mortality across settings, with higher recorded mortality generally observed among patients with intestinal involvement. These findings add to the limited real-world evidence in this area but should be interpreted in light of the descriptive study design and the limitations of the underlying data.

12. REFERENCES

1. Jacobsohn DA, Vogelsang GB. Acute graft versus host disease. Orphanet Journal of Rare Diseases. 2007;2(1):35.
2. Hou CY, Xu LL, Chen H, Liu N, Jiang M, Wang GQ, et al. Intestinal aGVHD and infection after hematopoietic stem cell transplantation. Med Sci Monit. 2013;19:802-6.
3. Johnson BH, Taylor A, Kim G, Drahos J, Yang J, Akbari M, et al. Clinical Outcomes and Healthcare Resource Utilization for Gastrointestinal Acute Graft-versus-Host Disease after Allogeneic Transplantation for Hematologic Malignancy: A Retrospective US Administrative Claims Database Analysis. Biology of Blood and Marrow Transplantation. 2019;25(4):834-41.
4. Zeiser R, von Bubnoff N, Butler J, Mohty M, Niederwieser D, Or R, et al. Ruxolitinib for Glucocorticoid-Refractory Acute Graft-versus-Host Disease. N Engl J Med. 2020;382(19):1800-10.
5. Giralt S, Bishop MR. Principles and overview of allogeneic hematopoietic stem cell transplantation. Cancer Treat Res. 2009;144:1-21.
6. Kroger N, Bacigalupo A, Barbui T, Ditschkowski M, Gagelmann N, Griesshammer M, et al. Indication and management of allogeneic haematopoietic stem-cell transplantation in myelofibrosis: updated recommendations by the EBMT/ELN International Working Group. Lancet Haematol. 2024;11(1):e62-e74.
7. Snowden JA, Sanchez-Ortega I, Corbacioglu S, Basak GW, Chabannon C, de la Camara R, et al. Indications for haematopoietic cell transplantation for haematological diseases, solid tumours and immune disorders: current practice in Europe, 2022. Bone Marrow Transplant. 2022;57(8):1217-39.
8. MacMillan ML, DeFor TE, Holtan SG, Rashidi A, Blazar BR, Weisdorf DJ. Validation of Minnesota acute graft-versus-host disease Risk Score. Haematologica. 2020;105(2):519-24.
9. Malard F, Huang XJ, Sim JPY. Treatment and unmet needs in steroid-refractory acute graft-versus-host disease. Leukemia. 2020;34(5):1229-40.
10. Abedin S, Rashid N, Schroeder M, Romee R, Nauffal M, Alhaj Moustafa M, et al. Ruxolitinib resistance or intolerance in steroid-refractory acute graft-versus-host disease - a real-world outcomes analysis. Br J Haematol. 2021;195(3):429-32.
11. Jagasia M, Perales MA, Schroeder MA, Ali H, Shah NN, Chen YB, et al. Ruxolitinib for the treatment of steroid-refractory acute GVHD (REACH1): a multicenter, open-label phase 2 trial. Blood. 2020;135(20):1739-49.
12. Ruutu T, Gratwohl A, de Witte T, Afanasyev B, Apperley J, Bacigalupo A, et al. Prophylaxis and treatment of GVHD: EBMT-ELN working group recommendations for a standardized practice. Bone Marrow Transplant. 2014;49(2):168-73.
13. Farhadfar N, Leather HL, Wang S, Burton N, IrizarryGatell V, Itzler R, et al. Severity of Acute Graft-versus-Host Disease and Associated Healthcare Resource Utilization, Cost, and Outcomes. Transplant Cell Ther. 2021;27(12):1007 e1- e8.
14. Styczynski J, Tridello G, Koster L, Iacobelli S, van Biezen A, van der Werf S, et al. Death after hematopoietic stem cell transplantation: changes over calendar year time, infections and associated factors. Bone Marrow Transplant. 2020;55(1):126-36.
15. Bhatia S, Dai C, Landier W, Hageman L, Wu J, Schlichting E, et al. Trends in Late Mortality and Life Expectancy After Allogeneic Blood or Marrow Transplantation Over 4 Decades: A Blood or Marrow Transplant Survivor Study Report. JAMA Oncol. 2021;7(11):1626-34.

16. Oran B, Popat U, Andersson B, Champlin R. Allogeneic Hematopoietic Stem Cell Transplantation for Myelodysplastic Syndromes. Clinical Lymphoma Myeloma and Leukemia. 2013;13:S282-S8.

17. Snowden JA, Sánchez-Ortega I, Corbacioglu S, Basak GW, Chabannon C, de la Camara R, et al. Indications for haematopoietic cell transplantation for haematological diseases, solid tumours and immune disorders: current practice in Europe, 2022. Bone Marrow Transplantation. 2022;57(8):1217-39.

18. Fagioli F, Quarello P, Zecca M, Lanino E, Rognoni C, Balduzzi A, et al. Hematopoietic stem cell transplantation for children with high-risk acute lymphoblastic leukemia in first complete remission: a report from the AIEOP registry. Haematologica. 2013;98(8):1273-81.

19. Ringden O, Remberger M, Runde V, Bornhauser M, Blau IW, Basara N, et al. Faster engraftment of neutrophils and platelets with peripheral blood stem cells from unrelated donors: a comparison with marrow transplantation. Bone Marrow Transplant. 2000;25 Suppl 2:S6-8.

20. Kiene S, Albrecht M, Theurich S, Scheid C, Skoetz N, Holtick U. Bone marrow versus peripheral blood allogeneic haematopoietic stem cell transplantation for haematological malignancies in adults. Cochrane Database Syst Rev. 2024;11(11):CD010189.

13. ANNEXES

ANNEX I. Results

Table S1. Study attrition of individuals included in each cohort during the study period within each data source.

Cohort definition	APHM		InGef RDB		VID	
	Current_n	Excluded	Current_n	Excluded	Current_n	Excluded
All GvHD in the data source (all ages)	220	0	1,326	0	938	0
Incident GvHD within study period (Jan 2020–Dec 2025)	89	131	753	573	356	582
Incident aGvHD within study period (Jan 2020–Dec 2025)	73	94	669	493	212	382
Incident aGvHD with intestinal involvement within study period (Jan 2020–Dec 2025)	5	11	503	337	87	140

aGvHD = acute graft versus host disease; APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.

ANNEX II. Data sources description

Assistance publique Hôpitaux de Marseille (APHM)

#	Section	Description
1	Data source identification and country	APHM (Assistance publique Hôpitaux de Marseille) France
2	Data partner information section	Assistance Publique – Hôpitaux de Marseille Public Health
3	Coverage and timespan	Data collection since 2014 Extent: Regional. The data covers all inpatients and outpatients treated at the Assistance Publique – Hôpitaux de Marseille (APHM), which includes five public university hospitals, all located in Marseille, France. The APHM serves not only the local population of Marseille and the surrounding Bouches-du-Rhône department but also attracts patients from the broader Provence-Alpes-Côte d'Azur (PACA) region and Corsica, representing a combined population of over 5 million inhabitants. Additionally, the hospital's specialized services attract patients from across France and abroad, including foreign nationals, all of whom are integrated into the OMOP CDM.
4	Healthcare setting / type of data	Secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care. The data source used in this study includes all hospital stays across various care settings—acute care, psychiatric care, rehabilitation care, and home hospitalization. The EHR system covers diagnoses, procedures, drug prescription and administration, medical and paramedical notes, such as hospitalization reports, radiology, EEG, endoscopy, and consultation summaries, and laboratory data.
5	Data collection process	Insurance/administrative claims, and Outpatient electronic health records, and Inpatient hospital electronic health records, and Inpatient hospital billing systems, and Registries, and Biobank. Data is entered by clinicians into the hospitals EHR system, consisting of several pieces of software. Diagnoses and procedures are managed via the CORA software, drug prescription and administration data, through the PHARMA software. Additionally, reports, radiology and consultation summaries are recorded using the AXIGATE software. These systems are integrated in the IATROS database for secondary use.
6	General representativeness	The database population are limited to patients visiting a specialised hospital.
7	Data content /source coding	Diagnoses are coded using ICD-10 and procedures are recorded using CCAM, in line with the French DRG system. Drug prescription and administration use UCD drug codes, ATC classifications, quantities, and dosages. Additionally, medical, and paramedical notes, such as hospitalization reports, radiology, EEG, endoscopy, and consultation summaries are recorded using the AXIGATE software. Laboratory data, covering both prescriptions and test results, is also included.
8	Data Harmonization	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. Each patient is assigned a unique patient number, which remains consistent throughout their care. In the rare event of duplicate records, a dedicated team is responsible for merging these records and ensuring quality control. Additionally, in France, patients have a unique national identifier (INS), which will eventually allow seamless linkage between hospital data and the SNDS (national health data system).
9	Quality control (data source specific)	Each software used for data collection undergoes quality verification before allowing validation and integration into the databases. These processes are managed by the hospital's IT department. Rigorous quality control is performed at multiple stages by various stakeholders, including the IT department, the medical information department, and internal controls. Quality assurance in the source systems is managed through a series of checks. These include

#	Section	Description
		validation loops when studies are conducted, ensuring that data is research-ready and meets required standards. Additionally, these controls help in identifying and resolving any data inconsistencies or errors before the data is made available for research purposes.
10	Linkage	Patient-Reported Experience Measures (PREMs) and Patient-Reported Outcome Measures (PROMs) can be linked. However, this linkage is not exhaustive across all domains. While the data allows for connections between medicine usage and some health outcomes, further development is required to achieve comprehensive linkage across all patient records and conditions. In particular, using non-structured data, such as clinical notes, could be improved through Natural Language Processing (NLP) for specific conditions. The implementation of additional linkages will need to be done on a case-by-case basis. In addition, it is possible to link socioeconomic information, e.g., for indicators like the FDEP (FDep or French Deprivation Index), a measure of neighbourhood deprivation.
11	Vital status	Vital status is retrieved from healthcare coverage details and covers date of death for all patients if they died inside the hospital.
12	Limitations	No database-specific limitations documented. General limitations for the data type applicable.
13	Main references	Fond G, Pauly V, Orleans V, Antonini F, Fabre C, Sanz M, Klay S, Jimeno MT, Leone M, Lancon C, Auquier P, Boyer L "Increased in-hospital mortality from COVID-19 in patients with schizophrenia." L'Encephale (2021): 32933762
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111141 Website: http://ap-hm.fr/

InGef Research Database (InGef RDB)

#	Section	Description
1	Database Identification and country	InGef Research Database (InGef RDB) Germany (DE)
2	Data partner information section	InGef – Institute for Applied Health Research Berlin GmbH InGef is a research service provider within the German statutory health insurance system.
3	Coverage and timespan	The Research Database contains approximately 10.5 million insured persons from 50 of the 94 statutory health insurances in Germany (as of Oct 2025). All individuals in the research database are included in the OMOP CDM. Due to data protection laws for personal data in Germany, only the last 10 years of data are available in the RDB and thus in the OMOP CDM.
4	Healthcare setting / type of data	In general, the data in the RDB reflects most of the health services paid for by statutory health insurances. In the OMOP CDM, the health services are reduced to primary and secondary care. This means that treatments by General Practitioners (GPs) and specialists (e.g., pediatricians), prescription medicines dispensed by pharmacies, hospital inpatient and outpatient care, as well as information about all insured individuals are included. The following data elements are presented in the OMOP CDM: demographic information, diagnoses, procedures, dispensing drugs and advanced therapy medicinal products, vaccinations, pregnancy data (via diagnoses and procedures) and contraception.
5	Data collection process	In Germany, data exchange between medical service providers and statutory health insurances is organized via central data collection centers. In addition to receiving and forwarding data, these centers also store data and provide access to third parties under strict regulations in data warehouses. The RDB contains selected, condensed and anonymized information from one of these data warehouses. Data sovereignty remains with the individual health insurance companies.

#	Section	Description
6	General representativeness	The RDB covers about 11% of the German population and is comparable to the German population in terms of the distribution of age and sex. Most health insurances that contribute to the RDB have nationwide coverage, meaning that the database covers all regions of Germany. Since almost all services covered by statutory health insurances are specified in national legislation, healthcare provision all over Germany is well represented in the RDB. Additionally, in Germany it is very common to stay with the same health insurance throughout life, which results in a good longitudinal coverage over the entire period of 10 years.
7	Data content /source coding	The coding in the research database complies with national classification and coding rules in Germany. Diagnoses are coded according to ICD-10-GM. Inpatient and outpatient surgeries or procedures are recorded as OPS codes (German classification of Operations and Procedures). The dispensing of drugs in pharmacies is recorded using the PZN (pharmaceutical registration number). For drugs that miss a PZN-to-RxNorm mapping, the ATC code is used instead. In some cases, dispensed drugs can be coded using OPS codes (e.g., in hospitals) or EBM codes (fee schedule for outpatient treatments).
8	Data Harmonization	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.
9	Quality control (database specific)	The data transmitted by healthcare providers complies with the standardized requirements and formats of the Association of Statutory Health Insurances (GKV-SV). Before being imported into the research database, the data elements are checked for data format, completeness and plausibility. After each update of the research database, various counts are compared with the previous update to verify completeness. Due to the anonymity of the database, direct validation of the data (e.g., using medical records as the gold standard) is not possible.
10	Linkage	Due to the anonymization of the source data, linkage is not possible.
11	Vital status	The date of death is recorded as the last day of the quarter in which the death occurred (i.e., 30/31st of Mar/Jun/Sept/Dec) as reported to the health insurance (no linkage to death registry). The cause of death is not available.
12	Limitations	Ambulatory diagnoses and procedures are summarised in the source on a quarterly basis. Both are mapped to the observation table with the date set to the last day of the respective quarter (i.e., 30/31st of Mar/Jun/Sept/Dec) and the concept “History of event within 3 months” (observation_concept_id 1340222), with the actual diagnosis or procedure concept_id recorded in the field “value_as_concept_id”. There is no vocabulary for the German pharmaceutical product codes (PZN). A direct source-to-standard-mapping has been done manually by InGef but is incomplete. The drug exposure duration is unknown. Following OMOP conventions, the end date is always set to dispensing date + 29. Outpatient and inpatient procedures are recorded as OPS codes (German Procedure Classification), for which the vocabulary is incomplete.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): PMID 26530279 | doi: 10.1002/pds.3895 Ludwig M, Enders D, Basedow F, Walker J, Jacob J: Sampling strategy, characteristics and representativeness of the InGef research database. Public Health 2022. doi: 10.1016/j.puhe.2022.02.013
14	Link to HMA-EMA catalogue and database webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

Valencia Health System Integrated Dataset (VID)

#	Section	Description
1	Data source identification and country	VID (Valencia Health System Integrated Dataset) Communitas Valenciana, Spain
2	Data partner information section	FISABIO Health Services Research & Pharmacoepidemiology Unit
3	Coverage and timespan	Data collection since 2009 Extent: Regional. The VID covers the general population of the Valencia region, comprising 10.7% of the Spanish population. The total population is estimated to be around 5,300,000.
4	Healthcare setting / type of data	Primary care – General Practitioner, and primary care specialists (e.g., paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and other (specify). Both primary and secondary care settings are covered, where visits, diagnoses, medications, measurements, and procedures are recorded. The population information system collects sociodemographic, health coverage, and mortality data. The Electronic prescription and dispensing system capture all information related to medication (active ingredient, strength, duration, indication, etc.). Emergency department and hospital admissions are registered, providing information on dates, diagnoses, and procedures. Measurements are captured additionally from the vaccine information system and the Microbiological surveillance network. Mortality is also captured.
5	Data collection process	Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. Data extraction is performed by clinical IT personnel. Data is released by the health authorities on a project basis and can only be used for such purposes.
6	General representativeness	The population captured by the VID should represent the Valencia region well, as the VID contains data of the general population covered by the universal public health care system. About 97% of the population in this region is covered by public care.
7	Data content /source coding	Prescribed and dispensed medications are coded with the ATC system. The indications of each prescription, as well as procedures are coded using ICD9CM and ICD10ES.
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. Patients have a unique id between practices.
9	Quality control (data source specific)	The data is reviewed carefully with the IT personnel who perform the extraction of data and then by a senior researcher with expertise in RWD management in the HSRP unit. Several quality check scripts are run against the received data. Finally, a senior researcher with RWD and clinical expertise assesses the completeness, consistency, and quality of the data extraction. If any inconsistency or error is detected, the dataset is requested and extracted again.
10	Linkage	VID also contains hospital discharge records, emergency care discharge records, birth registry, congenital anomaly registry, perinatal mortality registry, cancer registry, pharmacy prescription and dispensing records, vaccine records, and microbiology records. . Mother- and father-child linkage is also available. Most databases are updated daily, but certain registries, such as the congenital anomaly registry and perinatal mortality registries, are updated yearly.
11	Vital status	Mortality dates and causes of death are available in the mortality registry and the perinatal mortality registry.
12	Limitations	A subgroup of women (born before 1953) is not mapped into OMOP CDM yet. Note that another DARWIN Data Partner, BIFAP, also covers the Valencia region and patient information will overlap with VID. Biological sex is not captured, only has the legal gender. The last year's information for gender is used, and can change upon data refresh.

#	Section	Description
13	Main references	García-Sempere A, Orrico-Sánchez A,Muñoz-Quiles C,Hurtado I,Peiró S,Sanfélix-Gimeno G,Diez-Domingo J "Data Resource Profile: The Valencia Health System Integrated Database (VID)." International journal of epidemiology (2020): 31977043
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111174 Website: https://www.san.gva.es/ca/web/salut-publica

ANNEX III. Fitness for use assessment

Data source justification for inclusion and key characteristics:

France - Assistance publique Hôpitaux de Marseille (APHM)

The data source used in this study includes all hospital stays across various care settings (acute care, psychiatric care, rehabilitation care, and home hospitalisation) capturing approximately 300,000 stays annually. It is a public university hospital network that has a transplant unit, meaning it is the right setting for this study.

APHM was chosen for this study based on size of the population diagnosed with aGvHD: 247 patients with aGvHD, of which a proportion also have aGvHD with intestinal involvement. The number of individuals on ruxolitinib was 253 (any indication)).

APHM informed that Institutional Review Board (IRB) approval is fast (monthly meetings at least) and that it was possible to execute the code to have results by mid-January 2026. Therefore, APHM was selected for this study.

Germany - InGef Research Database (InGef RDB)

The setting of care was appropriate for this study, including hospital care.

Data availability and follow-up were sufficient, with records were available from 01/01/2015, and the most recent data extraction on 18/04/2025, fully aligned with the study period.

The size of the population with aGvHD: size of the population diagnosed with aGvHD 1,400 patients with aGvHD, of which 750 had aGvHD with intestinal involvement. Number of individuals on ruxolitinib was 3,670, of which approximately 300 with diagnosis GvHD with intestinal involvement.

As InGef RDB had blanket approval and were able to execute the code and have results by mid-January 2026, it was selected for this study.

Spain - Valencia Health System Integrated Dataset (VID)

The Valencia Health System Integrated Dataset (VID) is a set of multiple, public, population-wide electronic databases for the Valencia Region, the fourth most populated Spanish region, with about 5 million inhabitants and an annual birth cohort of 48,000 new-borns, representing 10.7% of the Spanish population and around 1% of the European population.

VID contained the right setting of care for this study, including hospital and ICU stay data.

VID was chosen for this study based on the size of the predicted study population diagnosed with aGvHD: 754 patients with aGvHD, of which 268 also have aGvHD with intestinal involvement. The number of individuals on ruxolitinib was 711 (any indication).

As VID informed that IRB approval is possible within December 2025 and were able to execute the code and have results by mid-January 2026, it was selected for this study.

Table S2. Fitness-for-use assessment of data sources.

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
Study population	aGvHD and aGvHD with intestinal involvement	Disease codes of aGvHD with or without intestinal involvement. In case only disease codes for GvHD have been reported, and in case this disease code is entered within maximum 100 days of a transplant procedure, it will also be considered as aGvHD In case source data are not granular enough, aGvHD with intestinal involvement will be based on presence of aGvHD + intestinal symptoms in the same visit occurrence	High	100% of included individuals will have a diseases code for aGvHD. The number of individuals as based on the feasibility assessment for aGvHD was 247 for APHM, 1400 for InGef, and 754 in VID. A subgroup of these consists of aGvHD with intestinal involvement	N/A	N/A
Treatment/exposure	Systemic steroids, ruxolitinib, and other 2nd and 3rd line treatment for aGvHD	Prescriptions of Systemic steroids, ruxolitinib, and other 2nd line and 3rd line treatment for aGvHD (RxNorm code)	High	N/A	N/A	N/A
Comparator group (not relevant)	N/A	N/A	N/A	N/A	N/A	N/A
Covariates	Demographic and clinical characteristics	Concept IDs for sex, age, and various clinical characteristics for the indication of the transplant (see "**Section 8.6.4 - Covariates, including confounders, effect modifiers, and other variables**" and mock Table 2)	Medium	N/A	N/A	N/A
Confounders	N/A	N/A	N/A	N/A	N/A	N/A
Follow-up time (if relevant)	Follow-up will be censored upon 365 days following index date (first date of diagnosis of aGvHD)	Will be derived from the observation period for each individual	High	N/A	N/A	N/A

Design elements	Operational definition	Data elements for valid capture	Criticality of the quality of the element, including justification where relevant	Suggested extensiveness assessment	Suggested assessment of other quality dimensions	Suggested substantiation by documentation
Death information	Patient with information on death within OMOP CDM. Whether patient dies will be assessed within specific windows (0–30 days, 31–90 days, and 91–365 days following index date)	Information on death will be derived from the "death table" from the OMOP CDM	High – important to have information on death	N/A	N/A	N/A

aGvHD = acute graft versus host disease; OMOP CDM = Observational Medical Outcomes Partnership Common Data Model; N/A: Not Applicable.
EMA Data Quality Framework for EU medicines regulation: application to Real-World Data for more information (https://www.ema.europa.eu/system/files/documents/other/data-quality-framework-eu-medicines-regulation-application-real-world-data_en.pdf).

ANNEX IV. 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 standardized 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.

The output files are stored in the DARWIN EU® Digital Research Environment (DRE). These output files do not contain any data that allow identification of subjects included in the study. The DRE implements further security measures to ensure a high level of stored data protection to comply with the local implementation of the General Data Protection Regulation (GDPR) (EU) 679/20161 in the various member states.

QUALITY CONTROL

Data source quality control

When defining drug cohorts, non-systemic products will be excluded from the list of included codes summarised on the ingredient level.

When defining cohorts for indications, a systematic search of possible codes for inclusion will be 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) and *DrugExposureDiagnostics* (https://cran.r-project.org/web/packages/DrugExposureDiagnostics/index.html) R packages will be run, if needed, to assess the use of different codes across the data sources contributing to the study and identify any codes potentially omitted in error. The *DrugExposureDiagnostics* package evaluates ingredient-specific attributes and patterns in drug exposure records.

The study code will be based on DARWIN EU® R packages: *IncidencePrevalence* to estimate Incidence and Prevalence, *DrugUtilisation* to characterise the drug use, and *CohortCharacteristics* to characterise the cohort by indication. These packages will include 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.

ANNEX V. List of stand-alone documents

Table S3. List of phenotypes, concept names, and concept IDs for aGvHD.

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
aGvHD	aGvHD	439416	-	SNOMED
aGvHD with intestinal involvement	Graft versus host disease of intestine (disorder)	37167528	-	SNOMED
Intestinal symptoms	Gastrointestinal hemorrhage; Nausea; Ulcerative colitis; Diarrhea; Abdominal pain; Vomiting; Malabsorption syndrome; Disorder of intestine; Inflammatory disorder of digestive tract; Moderate protein-calorie malnutrition (weight for age 60-74 percent of standard); Loss of appetite	192671, 31967, 81893, 196523, 200219, 441408, 4138253, 201618, 4043371, 4098458, 442165	-	SNOMED
All stem cell transplant	Administration @ Circulatory @ Transfusion @ Central Artery @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Artery @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Artery @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, T-cell Depleted Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, T-cell Depleted Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Percutaneous @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, Cord Blood;	4120445,2841327,2899295,2899293,2885573,2899294,2899295,2859130,42639209,2845880,2845877,2853604,1523592,42639687,2845881,2828433,1523822,2841329,2820541,2805699,42639500,2899158,2841321,2859124,1523795,42639742,2893411,2867444,1523338,42897996,42897994,42898013,42898014,2788705,2788961,42897995,42898015		SNOMED, ICD10PCS

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, T-cell Depleted Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, T-cell Depleted Hematopoietic; Allogeneic bone marrow transplant from matched unrelated donor; Allogeneic bone marrow transplant from unmatched unrelated donor; Allogeneic bone marrow transplantation; Allogeneic bone marrow transplantation with purging; Allogeneic bone marrow transplantation without purging; Allogeneic cord blood transplant to bone marrow; Allogeneic hematopoietic stem cell transpant without purging; Allogeneic hematopoietic stem cell transplant with purging; Allogeneic imperfect T-cell depleted allogeneic bone marrow transplant; Allogeneic lymphocyte infusions; Allogeneic peripheral blood stem cell transplant; Allogeneic related bone marrow transplant; Allogeneic T-cell depleted allogeneic bone marrow transplant; Allogeneic unrelated bone marrow transplant; Allograft of bone marrow from haploidentical donor; Allograft of bone marrow from sibling donor; Allograft of bone marrow from unmatched unrelated donor; Cord blood harvesting for transplantation, allogeneic; Cord blood stem cell transplant; Cord blood transplant to bone marrow; Cord blood-derived stem-cell transplantation, allogeneic; Hematopoietic progenitor cell (HPC); allogeneic transplantation per donor; Hematopoietic progenitor cell (HPC); HPC boost; Hemopoietic stem cell transplant; High-dose chemotherapy with stem cell transplant; Other specified graft of bone marrow; Other specified graft of cord blood stem cells to bone marrow; Peripheral blood stem cell transplantation; stem cell transplant for crohns disease; Syngeneic bone marrow transplant; Syngeneic peripheral blood stem cell transplantation; Transfusion of Allogeneic Related Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Related Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Related Cord Blood Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Related Cord Blood Stem Cells into Central Vein, Percutaneous Approach; Transfusion of			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Allogeneic Related Cord Blood Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Related Cord Blood Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Related Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Related Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Related Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Related Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Related T-cell Depleted Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Related T-cell Depleted Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Related T-cell Depleted Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Related T-cell Depleted Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Cord Blood Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Unrelated Cord Blood Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Cord Blood Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Unrelated Cord Blood Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Unrelated Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Unrelated Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated T-cell Depleted Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Unrelated T-cell Depleted Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated T-cell Depleted Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Unrelated T-cell Depleted Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Cord Blood Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Unspecified Cord Blood Stem Cells into			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Cord Blood Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Unspecified Cord Blood Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Unspecified Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Unspecified Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified T-cell Depleted Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Allogeneic Unspecified T-cell Depleted Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified T-cell Depleted Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Allogeneic Unspecified T-cell Depleted Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Embryonic Stem Cells into Central Vein, Open Approach; Transfusion of Embryonic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Embryonic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Embryonic Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Nonautologous Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Nonautologous Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Central Artery, Open Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Central Artery, Percutaneous Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Central Vein, Open Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Peripheral Artery, Open Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Peripheral Artery, Percutaneous Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Peripheral Vein, Open Approach; Transfusion of Nonautologous Cord Blood Stem Cells into Peripheral Vein, Percutaneous Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into Central Artery, Open Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into Central Vein, Open Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into Central Vein, Percutaneous Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into Peripheral Artery, Open Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into			

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Peripheral Artery, Percutaneous Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into Peripheral Vein, Open Approach; Transfusion of Nonautologous Hematopoietic Stem Cells into Peripheral Vein, Percutaneous Approach; Transplantation of allogeneic hematopoietic stem cell; Transplantation of bone marrow; Unrelated cord blood stem cell transplant			
Haemopoietic stem cell transplant – allogenic bone marrow transplant	Transplantation of bone marrow; Transfusion of Allogeneic Unspecified Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Related Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Related Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Nonautologous Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Nonautologous Bone Marrow into Central Vein, Percutaneous Approach; Transfusion of Allogeneic Unrelated Bone Marrow into Peripheral Vein, Percutaneous Approach; Transfusion of Allogeneic Unspecified Bone Marrow into Central Vein, Percutaneous Approach	4028623, 42897996, 42897994, 42898013, 42898014, 2788705, 2788961, 42897995, 42898015	-	SNOMED, ICD10PCS
Haemopoietic stem cell transplant – peripheral blood	Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Artery @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Percutaneous @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Artery @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, T-cell Depleted Hematopoietic; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, T-cell Depleted Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, Hematopoietic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, T-cell Depleted Hematopoietic; Administration @	2841327, 2899295, 2899293, 2885573, 2899295, 42639209, 2853604, 1523592, 42639687, 1523822, 2820541, 42639500, 2859124, 1523795, 42639742, 1523338, 4081380	-	SNOMED, ICD10PCS

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Hematopoietic Stem/Progenitor Cells, Genetically Modified; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, T-cell Depleted Hematopoietic; Peripheral blood stem cell transplantation			
Haemopoietic stem cell transplant – cord	Administration @ Circulatory @ Transfusion @ Central Artery @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Artery @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Vein @ Open @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Central Vein @ Percutaneous @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Artery @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Open @ Stem Cells, Embryonic; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, Cord Blood; Administration @ Circulatory @ Transfusion @ Peripheral Vein @ Percutaneous @ Stem Cells, Embryonic; Cord blood transplant to bone marrow; Cord blood stem cell transplant	2899294, 2859130, 2845880, 2845877, 2845881, 2828433, 2841329, 2805699, 2899158, 2841321, 2893411, 2867444, 4139690, 2002365	-	SNOMED, ICD10PCS
Acute myeloblastic leukaemia	Acute myeloid leukemia, minimal differentiation of lymph node, NOS; Acute myeloid leukemia without maturation of lymph node, NOS; Acute myeloid leukemia with abnormal marrow eosinophils of lymph nodes of inguinal region or leg; Acute myeloid leukemia, t(8;21)(q22;q22) of lymph nodes of head, face and neck; Acute myeloid leukemia without maturation of lymph nodes of head, face and neck; Acute myeloid leukemia with maturation of lymph nodes of inguinal region or leg; Acute myeloid leukemia with myelodysplasia-related changes of lymph node, NOS; Acute myeloid leukemia without maturation of lymph nodes of multiple regions; Acute myeloid leukemia with abnormal marrow eosinophils of lymph nodes of head, face and neck; Acute myeloid leukemia, t(8;21)(q22;q22) of intrathoracic lymph nodes; Acute myeloid leukemia without maturation of intrathoracic lymph nodes; Acute myeloid leukemia without maturation of intra-abdominal lymph nodes; Acute myeloid leukemia with abnormal marrow eosinophils of intra-abdominal lymph nodes; Acute myeloid leukemia, t(8;21)(q22;q22) of lymph nodes of multiple regions; Acute myeloid leukemia, minimal differentiation of spleen; Acute myeloid leukemia,	36546092, 36524587, 36549181, 36564542, 36523523, 36519450, 36548112, 36562529, 36521596, 36521545, 36553288, 36555336, 36535883, 36522560, 36533846, 36535976, 36525753, 36550331, 36528799, 36532889, 36556432, 36567759, 36520645, 36525886,	-	SNOMED, ICDO3

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	minimal differentiation of intra-abdominal lymph nodes; Acute myeloid leukemia, 11q23 abnormalities of lymph nodes of inguinal region or leg; Acute myeloid leukemia with maturation of spleen; Acute myeloid leukemia with myelodysplasia-related changes of intrathoracic lymph nodes; Acute myeloid leukemia with abnormal marrow eosinophils of lymph nodes of axilla or arm; Acute myeloid leukemia with myelodysplasia-related changes of pelvic lymph nodes; Acute myeloid leukemia, minimal differentiation of lymph nodes of head, face and neck; Acute myeloid leukemia with myelodysplasia-related changes of intra-abdominal lymph nodes; Acute myeloid leukemia without maturation of pelvic lymph nodes; Acute myeloid leukemia with maturation of lymph node, NOS; Acute myeloid leukemia, 11q23 abnormalities of intra-abdominal lymph nodes; Acute myeloid leukemia, t(8;21)(q22;q22) of intra-abdominal lymph nodes; Acute myeloid leukemia, 11q23 abnormalities of lymph nodes of axilla or arm; Acute myeloid leukemia, t(8;21)(q22;q22) of lymph node, NOS; Acute myeloid leukemia, 11q23 abnormalities of lymph nodes of multiple regions; Acute myeloid leukemia with myelodysplasia-related changes of lymph nodes of multiple regions; Acute myeloid leukemia with maturation of lymph nodes of head, face and neck; Acute myeloid leukemia with maturation of lymph nodes of multiple regions; Acute myeloid leukemia, t(8;21)(q22;q22) of lymph nodes of inguinal region or leg; Acute myeloid leukemia with abnormal marrow eosinophils of pelvic lymph nodes; Acute myeloid leukemia, 11q23 abnormalities of intrathoracic lymph nodes; Acute myeloid leukemia, 11q23 abnormalities of lymph node, NOS; Acute myeloid leukemia, 11q23 abnormalities of pelvic lymph nodes; Acute myeloid leukemia with maturation of pelvic lymph nodes; Acute myeloid leukemia with abnormal marrow eosinophils of spleen; Acute myeloid leukemia with maturation of lymph nodes of axilla or arm; Acute myeloid leukemia, 11q23 abnormalities of spleen; Acute myeloid leukemia with myelodysplasia-related changes of lymph nodes of head, face and neck; Acute myeloid leukemia with myelodysplasia-related changes of lymph nodes of inguinal region or leg; Acute myeloid leukemia with abnormal marrow eosinophils of lymph nodes of multiple regions; Acute myeloid leukemia, t(8;21)(q22;q22) of spleen; Acute myeloid leukemia with abnormal marrow eosinophils of lymph node, NOS; Acute myeloid leukemia, minimal differentiation of intrathoracic lymph nodes; Acute myeloid leukemia, minimal differentiation of lymph nodes of axilla or arm; Acute myeloid leukemia, t(8;21)(q22;q22) of pelvic lymph nodes; Acute myeloid leukemia, minimal differentiation of lymph nodes of multiple regions; Acute myeloid leukemia without maturation of lymph nodes of axilla or arm; Acute myeloid leukemia with abnormal marrow eosinophils	36553526, 36522761, 36522759, 36536188, 36530037, 36562801, 36555611, 36535120, 36552622, 36532145, 36545420, 36551581, 36519835, 36557716, 36551575, 36527103, 36563915, 36533252, 36550752, 36523177, 36560043, 36551864, 36532401, 36557966, 36535450, 36554900, 36564205, 36523246, 36550889, 36521194, 36532533, 36532530, 36529483, 36538708, 36521303, 36558245, 36566508, 36529641, 36522484, 140352, 37151872		

Phenotype	Concept name	Concept ID (including descendants)	Exclude concept ID	Vocabulary
	of intrathoracic lymph nodes; Acute myeloid leukemia with myelodysplasia-related changes of spleen; Acute myeloid leukemia, 11q23 abnormalities of lymph nodes of head, face and neck; Acute myeloid leukemia, minimal differentiation of lymph nodes of inguinal region or leg; Acute myeloid leukemia with maturation of intrathoracic lymph nodes; Acute myeloid leukemia without maturation of spleen; Acute myeloid leukemia, t(8;21)(q22;q22) of lymph nodes of axilla or arm; Acute myeloid leukemia with maturation of intra-abdominal lymph nodes; Acute myeloid leukemia, minimal differentiation of pelvic lymph nodes; Acute myeloid leukemia without maturation of lymph nodes of inguinal region or leg; Acute myeloid leukemia with myelodysplasia-related changes of lymph nodes of axilla or arm; Acute myeloid leukemia, disease; Acute myeloid leukemia			
Acute lymphoblastic leukaemia	Common acute lymphoblastic leukemia; Philadelphia chromosome-positive acute lymphoblastic leukemia; B-cell acute lymphoblastic leukemia; Philadelphia chromosome-positive acute lymphoblastic leukemia; Precursor B-cell acute lymphoblastic leukemia; T-lymphoblastic leukemia/lymphoblastic lymphoma; Acute lymphoid leukemia	4079280, 4138008, 4173963, 4143821, 4082461, 4221907, 134305	-	SNOMED
Myelodysplastic syndrome (MDS)	Myelodysplastic syndrome; allogenic bone-marrow-derived mesenchymal stem cells	40571982, 1465072	-	SNOMED
Aplastic anaemia	Aplastic anaemia	137829	4146088, 4101583, 4184200, 4184758, 4098027	SNOMED
Haemoglobinopathies	Haemoglobinopathies	432868	-	SNOMED
Lymphoma	Malignant lymphoma	432571	-	SNOMED

Table S4. List of drugs of interest.

Substance Name	Concept name	Class	ATC code	Concept ID	Include descendants
Systemic steroids	Systemic steroids	ATC 2nd	H02	21602722	Yes
ruxolitinib	ruxolitinib	Ingredient	L01EJ01	40244464	Yes
Basiliximab	Basiliximab	Ingredient	L04AC02	19038440	Yes
Brentuximab	Brentuximab	Ingredient	L01FX05	40241969	Yes
Etanercept	Etanercept	Ingredient	L04AB01	1151789	Yes
Infliximab	Infliximab	Ingredient	L04AB02	937368	Yes
aldesleukin	aldesleukin	Ingredient	L03AC01	1309770	Yes
oprelvekin	oprelvekin	Ingredient	L03AC02	1318030	Yes
Mycophenolate	Mycophenolate	Ingredient	L04AA06	19068900	Yes

Substance Name	Concept name	Class	ATC code	Concept ID	Include descendants
Rituximab	Rituximab	Ingredient	L01FA01	1314273	Yes
Tocilizumab	Tocilizumab	Ingredient	L04AC07	40480263	Yes
Vedolizumab	Vedolizumab	Ingredient	L04AG05	45774639	Yes
Calcineurin inhibitor	Calcineurin inhibitor	Ingredient	L04AD02, L04AD01	950637, 19010482	Yes
Anti-thymocyte Immunoglobulin (ATG) (ATC code L04AA04)	Anti-thymocyte Immunoglobulin	Ingredient	L04AA04	19136207, 19136041	Yes
Inolimomab (ATC code L04AC)	Inolimomab	Ingredient	L04AC	36857739	Yes
Alemtuzumab (ATC code L04AG06)	Alemtuzumab	Ingredient	L04AG06	1312706	Yes
Pentostatin (ATC code L01XX08)	Pentostatin	Ingredient	L01XX08	19031224	Yes
mTOR inhibitor (sirolimus (ATC code L04AH01)/rapamycin (ATC code L01EG04)	mTOR inhibitor	Ingredient	L04AH01, L01EG04	19034726	Yes
Methotrexate (ATC code L01BA01 or ATC code L04AX03)	Methotrexate	Ingredient	L01BA01, L04AX03	1305058	Yes

Table S5. Non-medical treatment.

Non-medical treatment	Concept name	Class	Concept ID	Include descendants
Extracorporeal photopheresis	Extracorporeal photopheresis	Procedure	4309050	Yes
mesenchymal stem cells (MSCs)	MSCs	Body Structure	4170237	Yes

ANNEX VI. 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.