Study Protocol

P4-C1-019

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

25/02/2026

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 with intestinal involvement
Protocol version	V4.0
Date	25/02/2026
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 inhibitor intensification • Anti-thymocyte globulin (ATG) • Inolimomab • Alemtuzumab • Pentostatin • Methotrexate • mTOR inhibitor (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 acute graft vs host disease (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 acute GvHD 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 acute GvHD 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
AA	Aplastic Anaemia
aGvHD	Acute Graft versus Host Disease
AML	Acute Myeloblastic Leukaemia
allo-HSCT	Allogeneic hematopoietic stem cell transplantation
ALL	Acute Lympholastic Leukaemia
APHM	Assistance publique Hôpitaux de Marseille
allo-SCT	Allogeneic stem cell transplantation
ATC	Anatomical Therapeutic Chemical
ATG	Anti-thymocyte globulin
CC	Coordinating centre
CDM	Common Data Model
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DOI	Declaration of Interests
DQD	Data Quality Dashboard
DRE	Digital Research Environment
DTZ	Data Transfer Zone
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-Authorisation Studies Register
GDPR	General Data Protection Regulation
GI	Gastrointestinal
GvHD	Graft versus Host Disease
GP	General Practitioner
HSCT	Hematopoietic stem cell transplantation
ICD	International Classification of Diseases
InGef RDB	InGef Research Database
IL	Interleukin
IP	Inpatient
IRB	Institutional Review Board
MDS	Myelodysplastic Syndrome
MSCs	mesenchymal stem cells
mTOR	Mechanistic Target of Rapamycin
OHDSI	Observational Health Data Sciences and Informatics

Acronyms/terms	Description
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
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	Josephine Jacob Raeleesha Norris Annika Vivirito Alexander Harms	InGef Research Database
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 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 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.

Research question and objectives

Research question

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

Objectives

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

Study design

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

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

Population

The study population will include individuals with a first occurrence of aGvHD in the study period.

Variables

Relevant covariates:

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 sources

1. France: Assistance publique Hôpitaux de Marseille (APHM)
2. Germany: InGef Research Database (InGef RDB)

3. Spain: Valencia Health System Integrated Dataset (VID)

Study size

No sample size has been calculated, as this is an exploratory study which will not test a specific hypothesis. Based on a preliminary feasibility assessment, the expected number of persons counts for ruxolitinib in the data sources included in this study range from 253 (APHM) to 3670 (InGef RDB) and for aGvHD from 247 (APHM) to 1,400 (InGef RDB) (750 individuals with aGvHD with intestinal involvement).

Statistical analysis

Covariates of interest will be reported as counts and proportions.

Patient characterisation will be conducted at index date, including individual demographics, transplant type, and treatment for aGvHD (corticosteroids, ruxolitinib, and other second/third line treatment). Registration of the type of transplant will be assessed in the window of 100 days prior to the index date. Use of corticosteroids, ruxolitinib, and other second/third line treatment will be 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. Also, mortality will be assessed within these respective windows.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study milestones and deliverables	Planned dates*
Final Study Protocol	December 2025
Creation of Analytical code	December 2025
Execution of Analytical Code on the data	January 2026
Draft Study Report	1 February 2026
Final Study Report	May 2026

*Planning dates are conditional upon completion of activities including IRB approvals, code performance assessments, and document reviews, and may therefore be adjusted.

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 and 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 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 acute graft vs. host disease (aGvHD) and acute graft vs host disease (aGvHD) with intestinal involvement?

Research objectives

The specific objectives of this study are:

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 will be conducted using routinely collected health data from 3 data sources from 3 countries across Europe. The study will comprise of:

- A characterisation study will be conducted to address objectives 1 and 2.

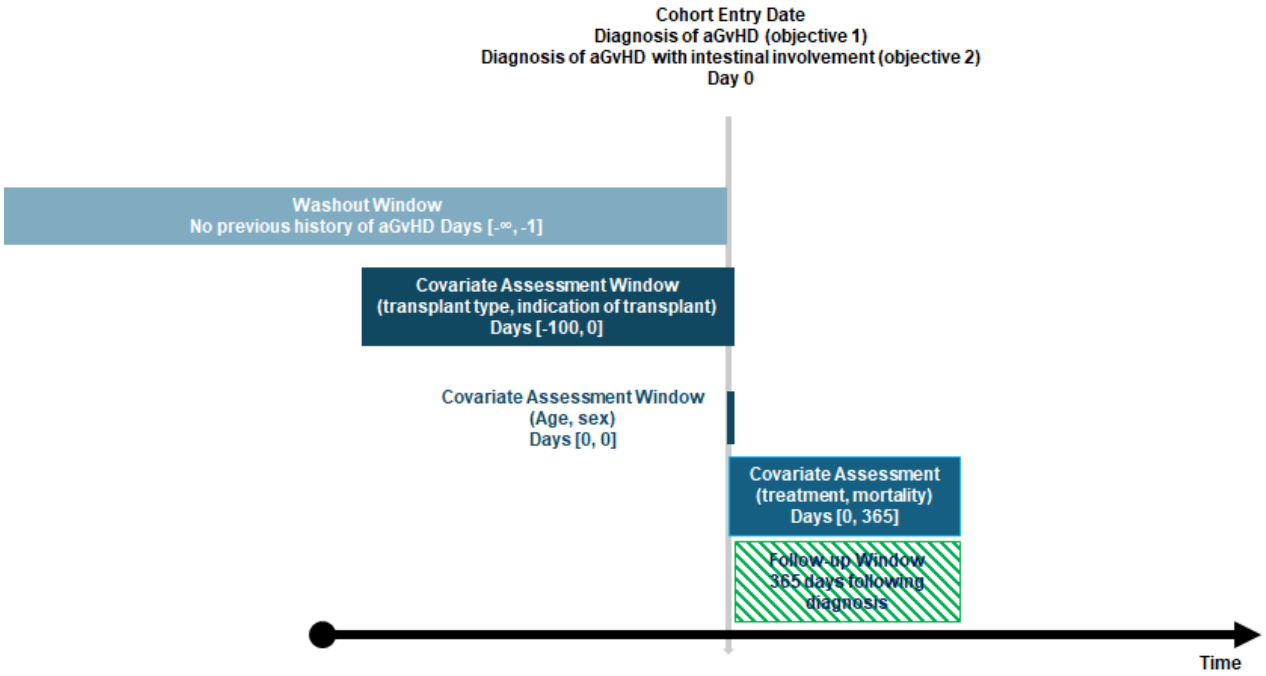

Figure 1. Graphical depiction of the study design for individuals with aGvHD and individuals with aGvHD with intestinal involvement.

8.2. Follow-up

To assess the proportion of individuals receiving treatment or that 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 will start on the date of first diagnosis of aGvHD (objective 1) with intestinal involvement (objective 2). Follow-up will end on the earliest of loss to follow-up, death, end of observation period (the latest available data), or on day 365 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 will consist 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 will either consist of individuals with presence of a disease code for 'Graft versus host disease of intestine – concept ID 37167528' or will consist 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 will be conducted using routinely collected data from 3 data sources in the DARWIN EU® network of data partners from 3 European countries. All data were *a priori* mapped to the 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 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 is 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 will be assessed in three time windows following index date, both in the aGvHD and aGvHD with intestinal involvement cohort (irrespective of treatment)
These windows are:

- 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

8.6.3. Intercurrent events (only for causal studies)

None.

8.6.4. Covariates, including confounders, effect modifiers, and other variables

aGvHD and aGvHD with intestinal involvement.

The diagnosis will be based on disease codes of aGvHD with or without intestinal involvement. In case source data are not granular enough (VID and APHM), aGvHD with intestinal involvement will be 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 those patients with only a disease code for GvHD, an additional condition of the code occurring within 100 days since the transplant will be added to the definition of aGvHD to differentiate from chronic GvHD.

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

- Sex
    - Female/male
- Age: Median age at index date and age category (<18 years, 18–65 years, and >65 years)
- To assess the proportion of individuals receiving treatment or that die within 30 days, 31–90 days, and 91–365 days of the index date (i.e., following diagnosis of aGvHD (objective 1) or diagnosis of aGvHD with intestinal involvement (objective 2)
- 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 will be 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 exposures will be 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 inhibitor intensification (i.e., tacrolimus (ATC code L04AA05) or cyclosporin (ATC code L04AD01))
        - Anti-thymocyte Immunoglobulin (ATG) (ATC code L04AA04)
        - Inolimomab (ATC code L04AC)
        - Alemtuzumab (ATC code L04AG06)
        - Pentostatin (ATC code L01XX08)
        - 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
        - Other/Unknown

The transplant type and indication for transplant will be assessed in the window -100 days to the date of the index date. Since aplastic anaemia also occurs after conditioning before transplant, and cannot be differentiated from the actual indication, this indication will be considered only if none of the other indications are present.

The preliminary concept sets used for the identification of covariates are described in **Annex IV**. These codes will be 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 has been calculated, as this is a descriptive disease epidemiology study which will not test a specific hypothesis. In addition, the study will be 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 persons counts for ruxolitinib in the data sources included in this study range from 253 (APHM) to 3,670 (InGef RDB) and for aGvHD from 247 (APHM) to 1,400 (InGef RDB) (750 individuals with aGvHD with intestinal involvement).These numbers are based on the overall number of conditions or observations registries in each data source with no filter by study period or inclusion and exclusion criteria.

8.8. Analysis

8.8.1. Federated network analyses

All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing individuals' data.

Before sharing the study package, test runs of the analytics will be performed on a subset of the data sources, and quality control checks will be performed. After all the tests are passed (see **Annex III. Operational and reporting considerations**), the final package will be released in a version-controlled study repository for execution against all the participating data sources.

8.8.2. Data privacy protection

The data partners will locally execute the analytics against the OMOP CDM in R Studio and review and approve the default aggregated results. They will then be made available to the Principal Investigators and study team in secure online repository of DTZ (Data Transfer Zone). All results will be locked and timestamped for reproducibility and transparency. The study results of all data sources will be checked, after which they are made available to the team, and the Study Dissemination Phase can start. All analyses will be conducted separately for each data source, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level data. Cell counts <5 will be suppressed when reporting results to comply with the data source's privacy protection regulations.

8.8.3. Statistical model specification and assumptions of the analytical approach considered

Objective 1 and objective 2

Descriptive statistics will be 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) will be described using counts and percentages.

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

8.8.4. Output

Output will include a PDF report including an executive summary, and the following table(s) and figure(s).

- Table 1. Distribution of baseline characteristics among individuals with aGvHD (overall) and individuals with aGvHD with intestinal involvement (number and %, median and IQR, mean and SD) by data source.
- Table 2. Distribution of transplant type and indication for transplant within 100 days prior to diagnosis and including the index date among individuals with aGvHD (overall) and individuals with aGvHD with intestinal involvement (number and %) by data source.
- Table 3. Distribution of treatment in the following windows from index date (day 0 to 30 days following the index date, day 31 to 90 days following index date, and 91 to 365 days following index date)

- Table 4. Proportion of individuals dying within the following windows (day 0 to 30 days, day 31 to 90 days, and 91 to 365 days) following diagnosis of aGvHD among individuals with aGvHD and individuals with aGvHD with intestinal involvement (number and %) by data source.
- Table S1. Study attrition of individuals included in each cohort during the study period within each data source.

An interactive dashboard will be generated by incorporating all the results (tables and figures) included in the PDF report mentioned above.

Table 1. Distribution of baseline characteristics among individuals with aGvHD and individuals with aGvHD with intestinal involvement (number and %, median and IQR, mean and SD) by data source.

Characteristics	aGvHD			aGvHD with intestinal involvement		
	APHM	InGef RDB	VID	APHM	InGef RDB	VID
Overall, N						
Median age (IQR)						
Mean age (SD)						
Age category, N (%): <18 years 18–65 years >65 years						
Sex, N (%)						
Male Female						

IQR=Interquartile Range, SD=Standard Deviation, APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.

Table 2. Distribution of type of transplant and indication of transplant among individuals with aGvHD and individuals with aGvHD with intestinal involvement (number and %) by data source (assessed within 100 days up to the index date).

Characteristics	aGvHD			aGvHD with intestinal involvement		
	APHM	InGef RDB	VID	APHM	InGef RDB	VID
Stem cell source, N (%) *						
Hematopoietic stem cell transplant (overall)						
Marrow						
Peripheral blood						
Cord						
Other/Unknown						
Indication for transplant, N (%)						
Acute myeloblastic leukaemia						
Acute lymphoblastic leukaemia						
Myelodysplastic syndrome (MDS)						
Aplastic anaemia						

	aGvHD			aGvHD with intestinal involvement		
Haemoglobinopathies						
Lymphoma						
Other/Unknown						

APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.
*Stem cell source and indication will be assessed in the window of -100 days to the index date.

Table 3. Distribution of treatment received among individuals with aGvHD and aGvHD with intestinal involvement (number and %).

Characteristics, N (%)	aGvHD									aGvHD with intestinal involvement								
	APHM			InGef RDB			VID			APHM			InGef RDB			VID		
	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																		
Ruxolitinib																		
Ruxolitinib in combination with systemic steroids*																		
Basiliximab																		
Brentuximab																		
Etanercept																		
Infliximab																		
IL-2**																		
Mycophenolate																		
Rituximab																		
Tocilizumab																		
Vedolizumab																		
Calcineurin inhibitor intensification																		

	aGvHD									aGvHD with intestinal involvement								
Anti-thymocyte globulin (ATG)																		
Inolimomab																		
Alemtuzumab																		
Pentostatin																		
mTOR inhibitor (sirolimus/rapamycin)																		
Methotrexate																		
Extracorporeal photopheresis																		
mesenchymal stem cells (MSCs)																		

APHM= Assistance publique Hôpitaux de Marseille, InGef RDB= InGef Research Database, VID= Valencia Health System Integrated Dataset.
* 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 L03AC02.

Table 4. Survival among individuals with aGvHD and individuals with aGvHD with intestinal involvement (number and %) by data source.

Characteristics	aGvHD			aGvHD with intestinal involvement		
	APHM	InGef RDB	VID	APHM	InGef RDB	VID
Mortality within Day 0 to Day 30 following ID (N (%))						
Mortality within Day 31 to Day 90 following ID (N (%))						
Mortality within Day 91 to Day 365 following ID (N (%))						

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).

ANNEX RESULTS

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

Cohort definitions	aGvHD					
	APHM		InGef RDB		VID	
	Current_n	Excluded	Current_n	Excluded	Current_n	Excluded
All GvHD in the database (all ages)						
Incident GvHD within study period (Jan 2020–Dec 2025)						
Incident aGvHD within study period (Jan 2020–Dec 2025)						
Incident aGvHD with intestinal involvement within study period (Jan 2020–Dec 2025)						

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

8.9. Evidence synthesis

Results from analyses described in **Section 8.8.3** will be presented separately for each data source. No meta-analysis of results will be conducted.

9. STRENGTHS AND LIMITATIONS

The study will be informed by routinely collected health care data and 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 is a potential for misclassification between aGvHD and chronic GvHD in case source data do 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 do have source codes for aGvHD with intestinal involvement (this is only the case for InGef RDB). For those DPs where this level of granularity is missing within the source code (VID and APHM), the phenotype of aGvHD with intestinal involvement will be based on disease codes for aGvHD in combination with intestinal symptoms during the same visit occurrence.

As part of characterisation, we will describe the type of transplant procedure (peripheral blood, marrow, blood) but this information might be missing within the data sources. Furthermore, we might 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)

Additionally, the results of this study will 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.

In the InGef RDB claims data source, the date of death is recorded as the last day of the quarter in which the death occurred. This misclassification can lead to a time lag of up to three months between the date of actual occurrence and the recorded date in the data source. This will impact results regarding survival time and proportion of individuals dying within specific time windows.

10. 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.

11. ANNEXES

ANNEX I. Description of data sources

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,

#	Section	Description
		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 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	Data source identification and country	InGef RDB (InGef Research Database) Germany
2	Data partner information section	Institut für angewandte Gesundheitsforschung Berlin GmbH
3	Coverage and timespan	Data collection since 2014 Extent: Nation-wide. The data source contains information from the statutory health insurances (SHI), which insure a total of about 89% (~73 million individuals) of the German population. Since the InGef RDC currently includes about ten million individuals, it covers about 13% of the total population insured in one of the German SHIs. The data in the database depicts all health care use which has been reimbursed by the SHI.
4	Healthcare setting / type of data	Primary care – General Practitioner, and community pharmacists, and primary care specialists (e.g. paediatricians), and secondary care – specialists (ambulatory or hospital outpatient care), and hospital inpatient care, and claims data. The following data elements are 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	Insurance/administrative claims. The data in the database depicts all health care use which has been reimbursed by the SHI (statutory health insurances).
6	General representativeness	The 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.

#	Section	Description
		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. No. In the German statutory health system, a person can only be enrolled in one health insurance at a time. However, if a person changes from one contributing insurer to another, a new ID number will be generated.
9	Quality control (data source specific)	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. Approx. 10.5 Million insurees are included in the database, 7.8 Million of these actively insured in 2024. This corresponds to 7% of the total German population. Data are longitudinally linked over a period of currently ten years.
13	Main references	Andersohn F, Walker J "Characteristics and external validity of the German Health Risk Institute (HRI) Database." Pharmacoepidemiology and drug safety (2016): 26530279
14	Link to HMA-EMA catalogue and data source webpage	HMA-EMA Catalogue entry: https://catalogues.ema.europa.eu/data-source/1111207 Website: https://www.ingef.de/en/

Valencia Health System Integrated Dataset (VID)

#	Section	Description
1	Data source identification and country	VID (Valencia Health System Integrated Dataset) Comunitat 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). Outpatient electronic health records, and Inpatient hospital electronic health records, and Registries, and Other. Both primary and secondary care settings are covered, where visits, diagnoses, medications, measurements, and procedures are recorded. The population information system collects sociodemographics, 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.
5	Data collection process	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 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. 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 II. 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 IRB approval is fast (monthly meetings at least) and that it is 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 is appropriate for this study, including hospital care.

Data availability and follow-up are sufficient, with records 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 have 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 has blanket approval and are 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 contains 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 are able to execute the code and have results by mid-January 2026, it was selected for this study.

Table S1. 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, in case this disease code is entered within maximum 100 days of a transplant procedure, this 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

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
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
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

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 III. Operational and reporting considerations

DATA MANAGEMENT

Data management

All data sources have previously mapped their data to the OMOP common data model. This enables the use of standardised analytics and using DARWIN EU® tools across the network, since the structure of the data and the terminology system is 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 will be written in R and will use standardized analytics wherever possible. Each data partner will execute the study code against their data source containing patient-level data and then return the results (csv files), which will only contain aggregated data. The results from each of the contributing data sites will then be combined in tables and figures for the study report.

Data storage and protection

For this study, participants from various EU member states will process personal data from individuals that is collected in 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 are 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 are adhered to. In agreement with these regulations, rather than combining person level data and performing only a central analysis, local analyses will be 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 RRE 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: *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.

PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

A PDF report including an executive summary, and the specified tables and/or figures will be submitted to EMA by the DARWIN EU® CC upon completion of the study.

An interactive dashboard incorporating all the results (tables and figures) will be provided alongside the PDF report. The full set of underlying aggregated data used in the dashboard will also be made available, if requested.

ANNEX IV. List of stand-alone documents

Table S2. Preliminary 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	Malabsorption, diarrhea, GI hemorrhage, pancolitis	40317605; 196523; 46273183, 26727, 40479839	-	SNOMED
Haemopoietic stem cell transplant – allogenic bone marrow transplant	Allogeneic bone marrow transplantation	4242257	-	SNOMED
Haemopoietic stem cell transplant – peripheral blood	Allogeneic peripheral blood stem cell transplant	4143404	-	SNOMED
Haemopoietic stem cell transplant – cord	Cord blood-derived stem-cell transplantation, allogeneic	2721124	-	SNOMED
Acute myeloblastic leukaemia	Acute myeloblastic leukaemia	40367590	-	SNOMED
Acute lymphoblastic leukaemia	Acute lymphoblastic leukaemia	40641318	-	SNOMED
Myelodysplastic syndrome (MDS)	Myelodysplastic syndrome (MDS)	40571982	-	SNOMED
Aplastic anaemia	Aplastic anaemia	137829	-	SNOMED
Haemoglobinopathies	Haemoglobinopathies	3533303	-	SNOMED
Lymphoma	Malignant lymphoma	37156205	-	SNOMED
Other/Unknown	N/A Will be assessed via code	No conceptIDs or presence of conceptIDs different than the ones of the concepsets as described above		SNOMED

N/A= Not Applicable.

This is a preliminary set of conditions and related concept IDs. Proper phenotyping will be conducted, and these tables will be updated in the report.

Table S3. Preliminary 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
Rituximab	Rituximab	Ingredient	L01FA01	1314273	Yes
Tocilizumab	Tocilizumab	Ingredient	L04AC07	40480263	Yes
Vedolizumab	Vedolizumab	Ingredient	L04AG05	45774639	Yes
Calcineurin inhibitor intensification (i.e. increasing dose of tacrolimus (ATC code L04AA05) or cyclosporin (ATC code L04AD01))	Calcineurin inhibitor intensification	Ingredient	L04AA05, 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 S4. 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 V. ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

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

EU PAS Register® number: EUPAS1000000878
Study reference number (if applicable): P4-C1-019

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	8.5
1.1.2 End of data collection[2]	☒	☐	☐	8.5
1.1.3 Progress report(s)	☐	☐	☒	
1.1.4 Interim report(s)	☐	☐	☒	
1.1.5 Registration in the EU PAS Register®	☐	☒	☐	
1.1.6 Final report of study results.	☐	☒	☐	

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	6
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	6
2.1.2 The objective(s) of the study?	☒	☐	☐	7
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	7
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	8.1

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	8.4
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☐	☐	☒	
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☐	☐	☒	
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	8.4
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	8.5
4.2.2 Age and sex	☐	☐	☒	
4.2.3 Country of origin	☒	☐	☐	8.4
4.2.4 Disease/indication	☒	☐	☐	8.6
4.2.5 Duration of follow-up	☒	☐	☐	8.2
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	8.3

Comments:

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☐	☐	☒	
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☐	☒	
5.3 Is exposure categorised according to time windows?	☐	☐	☒	
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☐	☐	☒	

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☐	☐	☒	
5.6 Is (are) (an) appropriate comparator(s) identified?	☐	☐	☒	

Comments:

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☐	☐	☒	
6.2 Does the protocol describe how the outcomes are defined and measured?	☐	☐	☒	
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☐	☒	
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☐	☐	☒	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	☒	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☐	☐	☒	

Comments:

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☐	☐	☒	

Comments:

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☐	☐	☒	
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☐	☐	☒	
9.1.3 Covariates and other characteristics?	☒	☐	☐	8.4
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☐	☐	☒	
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☐	☐	☒	
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	8.4
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☐	☒	
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☐	☐	☒	
9.3.3 Covariates and other characteristics?	☒	☐	☐	8.4, Annex I
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	Annex I

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	8.8
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	8.7
10.3 Are descriptive analyses included?	☒	☐	☐	8.8.3
10.4 Are stratified analyses included?	☐	☐	☒	
10.5 Does the plan describe methods for analytic control of confounding?	☐	☐	☒	
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☐	☐	☒	

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.7 Does the plan describe methods for handling missing data?	☐	☐	☒	
10.8 Are relevant sensitivity analyses described?	☐	☐	☒	

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	Annex IV
11.2 Are methods of quality assurance described?	☒	☐	☐	Annex IV
11.3 Is there a system in place for independent review of study results?	☐	☐	☒	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☐	☐	☒	
12.1.2 Information bias?	☐	☐	☒	
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☐	☒	
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	8.7, Annex II

Comments:

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	8.4
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	Annex IV

Comments:

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	☒	☐	☐	4

Comments:

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	☒	☐	☐	Annex IV
15.2 Are plans described for disseminating study results externally, including publication?	☐	☐	☒	

Comments:

Name of the main author of the protocol: Katia Verhamme

Date: 16th January 2026

Signature: 

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 data source 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 data sources 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 data sources, 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.