Study Protocol

P3-C1-014

DARWIN EU® – Azathioprine - user characteristics

15/10/2024

Version 4.0

This document represents the views of the DARWIN EU® Coordination Centre only and cannot be interpreted as reflecting those of the European Medicines Agency or the European Medicines Regulatory Network.

Table of Contents

Study Title	DARWIN EU® - Azathioprine - user characteristics
Protocol version	V4.0
Date	15/10/2024
EUPAS number	EUPAS1000000322
Active substance	Azathioprine Therapeutic Drug class: L04AX
Medicinal product	Azathioprine
Research question and objectives	The aim of this study it to characterise individuals treated with Azathioprine. The specific study objectives are: 1. Characterisation of patients newly treated with Azathioprine by sex and age at first prescription, overall and per indication 2. Characterisation of patients newly treated with Azathioprine by indication: a. Organ transplantation b. Severe rheumatoid arthritis or chronic polyarthritis c. Inflammatory bowel disease d. Autoimmune hepatitis e. Systemic lupus erythematosus f. Dermatomyositis g. Polyarteritis nodosa h. Pemphigus vulgaris and bullous pemphigoid i. Behcet's disease j. Refractory autoimmune haemolytic anaemia k. Refractory idiopathic thrombocytopenic purpura l. Polymyositis m. Pyoderma gangrenosum n. Multiple sclerosis o. Myasthenia gravis p. None of the above/missing 3. Large-scale characterisation overall and per indication of drugs and conditions within one year prior to the index date (-365 to -1 day to the index date) and on the index date will be assessed. 4. Characterisation of patients newly treated with Azathioprine in terms of treatment duration, overall and for each indication
Country(ies) of study	This study will include 5 databases, representing 4 countries in Europe: • Germany: IQVIA DA Germany • United Kingdom: CPRD GOLD • Netherlands: IPCI • Spain: IMASIS and SIDIAP
Author	Guido van Leeuwen (g.vanleeuwen@darwin-eu.org) Katia Verhamme (k.verhamme@darwin-eu.org)

LIST OF ABBREVIATIONS

Acronyms/term	Description
CDM	Common Data Model
CC	Coordinating centre
CPRD	Clinical Practice Research Datalink
DA	Disease Analyzer
DARWIN EU®	Data Analysis and Real-World Interrogation Network
DRE	Digital Research Environment
DQD	Data Quality Dashboard
DUS	Drug Utilisation Study
EHR	Electronic Health Records
ED	Emergency department
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EU	European Union
GDPR	General Data Protection Regulation
ICD	International Classification of Diseases
ID	Index date
IMASIS	Institut Municipal Assistència Sanitària Information System
IP	Inpatient
IPCI	Integrated Primary Care Information
LPD	Longitudinal Patient Database
MA	Marketing Authorisation
MS	Multiple Sclerosis
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OP	Outpatient
PRAC	Pharmacovigilance Risk Assessment Committee
PSMAR	Consorci Mar Parc de Salut Barcelona
SD	Standard deviation
SIDIAP	The Information System for Research in Primary Care
SLE	Systemic Lupus Erythematosus
SNOMED	Systematized Nomenclature of Medicine
WHO	World Health Organisation

1. TITLE

DARWIN EU® - Azathioprine – user characteristics

2. RESPONSIBLE PARTIES – STUDY TEAM

Study team role	Names	Organisation
Study Project Manager/ Principal Investigator	Guido van Leeuwen Katia Verhamme	Erasmus MC
Data Scientist	Ross Williams Maarten van Kessel Cesar Barbosa Ger Inberg Adam Black	Erasmus MC
Epidemiologist/Clinical Domain Expert	Guido van Leeuwen Katia Verhamme	Erasmus MC
Data Partner*	**Names**	**Organisation**
CRPD GOLD	Antonella Delmestri	University of Oxford
IPCI	Katia Verhamme	Erasmus MC
IQVIA DA Germany	James Brash Isabella Kaczmarczyk Dina Vojinovic	IQVIA
IMASIS	Miguel-Angel Mayer Maria Angeles Leis Machin Juan Manuel Ramirez Anguita	Consorci Mar Parc de Salut Barcelona (PSMar)
SIDIAP	Talita Duarte-Salles Anna Palomar Agustina Giuliodori Picco	IDIAPJGol

*Data partners' role is only to execute code at their data source, review and approve their results. They do not have an investigator role. Data analysts/programmers do not have an investigator role and thus declaration of interests (DOI) for them is not needed.

3. ABSTRACT

Title

DARWIN EU® – Azathioprine – user characteristics

Rationale and background

Azathioprine is a purine analogue and prodrug of mercaptopurine that is used as an immunosuppressive medication alone or in combination with other immunosuppressive therapy to prevent rejection following organ transplantation and to treat certain autoimmune diseases, where it is considered a steroid-sparing agent. The Pharmacovigilance Risk Assessment Committee (PRAC) recently discussed a signal procedure regarding the association between treatment with azathioprine and non-cirrhotic portal hypertension/porto-sinusoidal vascular disease (PSVD).

Through this study we aim to characterize patients newly treated with azathioprine, to contextualise the signal assessment.

Research question and objectives

This is a study in incident users of azathioprine aiming to characterise new users of azathioprine with respect to indications for treatment, age at treatment initiation, and sex, and to summarise the treatment durations with azathioprine for all indications combined, and for each individual treatment indication.

Study specific objectives are as following:

1. Characterisation of patients newly treated with Azathioprine by sex and age at first prescription, overall and per indication

2. Identify potential indications for azathioprine, and the percentage of azathioprine-treated patients for each pre-defined approved indication:

a) Organ transplantation
b) Severe rheumatoid arthritis or chronic polyarthritis
c) Inflammatory bowel disease
d) Autoimmune hepatitis
e) Systemic lupus erythematosus
f) Dermatomyositis
g) Polyarteritis nodosa
h) Pemphigus vulgaris and bullous pemphigoid
i) Behçet's disease
j) Refractory autoimmune haemolytic anaemia
k) Refractory idiopathic thrombocytopenic purpura
l) Polymyositis
m) Pyoderma gangrenosum
n) Multiple sclerosis
o) Myasthenia gravis.
p) None of the above/missing

3. Large-scale characterisation overall and per indication of drugs and conditions within one year prior to the index date (-365 to -1 day to the index date) and on the index date will be assessed.

4. Estimate and summarise duration of treatment with azathioprine, overall, and stratified per indication.

Methods

Study design

Patient-level drug utilisation study

Study period

01/01/2000 – 31/12/2023

Population

For this study we have one cohort, namely:

- Population of individuals newly treated with Azathioprine

Data source

1. Clinical Practice Research Datalink (CPRD) GOLD, United Kingdom
2. Integrated Primary Care Information (IPCI), Netherlands
3. IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany
4. Institut Municipal Assistència Sanitària Information System (IMASIS), Spain
5. The Information System for Research in Primary Care (SIDIAP), Spain

Sample size

Based on a preliminary feasibility assessment, the expected person counts for Azathioprine were 900 for IMASIS, 6,000 for IPCI, 23,700 for IQVIA DA Germany, 24,500 for SIDIAP and 52,100 for CPRD GOLD.

Exposure of interest

Azathioprine use.

Outcomes of interest

- Characteristics (sex and age)
- Indication
    - Organ transplantation
    - Severe rheumatoid arthritis or chronic polyarthritis
    - Inflammatory bowel diseases
    - Autoimmune hepatitis
    - Systemic lupus erythematosus
    - Dermatomyositis
    - Polyarteritis nodosa
    - Pemphigus vulgaris and bullous pemphigoid
    - Behcet's disease
    - Refractory autoimmune haemolytic anaemia
    - Refractory idiopathic thrombocytopenic purpura
    - Polymyositis
    - Pyoderma gangrenosum

  - Multiple sclerosis
  - Myasthenia gravis
  - None of the above/missing

- Large-scale characterisation: overall and per indication of drugs and conditions within one year prior to the index date (-365 to -1 day to the index date) and on the index date
- Treatment duration, overall and by indication

Statistical analysis

Characterisation of individuals newly initiating treatment with Azathioprine will be done using the *CohortCharacteristics* and *CohortDiagnostics* R packages. For the second and third objective, we will use the *DrugUtilisation* package to characterise Azathioprine use including counts (%) for each indication, and treatment duration.

4. AMENDMENTS AND UPDATES

None.

5. MILESTONES

Study deliverables	Timelines
Draft Study Protocol	23/08/2024
Final Study Protocol	25/09/2024
Creation of Analytical code	August/September 2024
Execution of Analytical Code on the data	September/October 2024
Draft Study Report	15/10/2024
Final Study Report	15/11/2024

6. RATIONALE AND BACKGROUND

Azathioprine is a purine analogue and prodrug of mercaptopurine that is used as an immunosuppressive medication alone or in combination with other immunosuppressive therapy to prevent rejection following organ transplantation and to treat certain autoimmune diseases, where it is considered a steroid-sparing agent. Examples of autoimmune diseases that could be treated with azathioprine are rheumatoid arthritis, inflammatory bowel disease, systemic lupus erythematosus (SLE), polymyositis, dermatomyositis, multiple sclerosis (MS) and myasthenia gravis(1-7).

Azathioprine is associated with minor usually transient and asymptomatic elevations of aminotransferase levels and with rare instances of acute cholestatic liver injury and with long-term use, portal hypertension may occur(8, 9).

A signal procedure regarding a potential association between azathioprine and non-cirrhotic portal hypertension/portosinusoidal vascular disease (PSVD) - which is a rare disorder characterised by signs of portal hypertension in the absence of an identifiable aetiology, such as cirrhosis - is needed. A liver biopsy is mandatory for the diagnosis of PSVD.(10) Specific histologic signs include obliterative portal venopathy, nodular regenerative hyperplasia and incomplete septal fibrosis (10).

This study is intended to support the signal evaluation by providing information about the use of azathioprine, including the most frequent indications, the age- and sex distribution at initiation of treatment, and the treatment duration for all indications combined, and for each indication separately.

7. RESEARCH QUESTION AND OBJECTIVES

The aim of this study is to characterise individuals treated with Azathioprine.

The specific study objectives are:

1. Characterise azathioprine initiators by sex and age at first prescription, overall and per indication

2. Identify potential indications for azathioprine, and the percentage of azathioprine-treated patients for each pre-defined approved indication:

a) Organ transplantation
b) Severe rheumatoid arthritis or chronic polyarthritis
c) Inflammatory bowel disease
d) Autoimmune hepatitis
e) Systemic lupus erythematosus
f) Dermatomyositis
g) Polyarteritis nodosa
h) Pemphigus vulgaris and bullous pemphigoid
i) Behçet's disease
j) Refractory autoimmune haemolytic anaemia
k) Refractory idiopathic thrombocytopenic purpura
l) Polymyositis
m) Pyoderma gangrenosum
n) Multiple sclerosis
o) Myasthenia gravis
p) None of the above/missing

3. Large-scale characterisation overall and per indication of drugs and conditions within one year prior to the index date (-365 to -1 day to the index date) and on the index date will be assessed.

4. Estimate and summarise duration of treatment with azathioprine, overall, and stratified per indication

Description of the proposed objectives are described in (**Table 1**).

Table 1. Primary and secondary research questions and objective.

A. Primary research question and objective.

Objective 1,2,3 and 4:	To characterise individuals newly treated with Azathioprine by age and sex. In addition, new users of Azathioprine will be characterised in terms of indication and treatment of use. Characterisation will be done in terms of demographics, indication, treatment duration and prior drug use and comorbidities
Hypothesis:	N/A
Population *(mention key inclusion-exclusion criteria):*	All patients present in the databases with at least 365 days of prior history and newly treated with Azathioprine
Exposure:	Exposure to Azathioprine

Comparator:	N/A
Outcome:	N/A
Time *(when follow up begins and ends):*	Study period will be from 1st January 2000 to 31st December 2023. Within this study period, we will identify individuals newly treated with azathioprine. To generate information on duration of use, follow-up will be from the first date of azathioprine during the study period until 1) censuring or loss to follow-up, 2) end of data availability, or 3) the end date of exposure (whichever occurs first).
Setting:	Primary care data (CPRD GOLD, IPCI, IQVIA DA Germany and SIDIAP) and outpatient secondary care (IQVIA DA Germany)
Main measure of effect:	Counts

8. RESEARCH METHODS

8.1 Study type and study design

The proposed design for this study is a patient-level DUS study as "off-the-shelf", as described in the DARWIN EU® Complete Catalogue of Standard Data Analyses. We will also perform a patient-level characterisation.

Table 2. Description of potential study types and related study designs.

Study type	Study design	Study classification
Patient Level DUS	New drug/s user cohort	Off the shelf

8.2 Study setting and data sources

This study will use routinely collected health data from 5 databases in the DARWIN EU® network of data partners from 4 European countries. All databases were previously mapped to the OMOP CDM.

Data sources

1. Clinical Practice Research Datalink (CPRD) GOLD, United Kingdom
2. Integrated Primary Care Information (IPCI), Netherlands
3. IQVIA Disease Analyzer Germany (IQVIA DA Germany), Germany
4. Institut Municipal Assistència Sanitària Information System (IMASIS), Spain
5. The Information System for Research in Primary Care (SIDIAP), Spain

These databases fulfil the criteria required for patient-level drug utilisation, and characterisation allowing for large-scale characterisation while covering different regions of Europe. The selection of databases was based on the availability of data on the selected drugs of interest, the conditions of interest, and other

variables needed to perform the described analyses. As information needs to be provided in a timeline manner, data partners not requiring IRB approval or DPs with fast IRB review time (for instance because of basket protocol approval) have been selected. Detailed information on the selected data sources is described in **Table 3**.

When it comes to assessing the reliability of data sources, the data partners are asked to describe their internal data quality process on the source data as part of the DARWIN EU® onboarding procedure. To further ensure data quality, we utilised the Achilles tool, which systematically characterises the data and generates data characteristics such as age distribution, condition prevalence per year, data density, measurement value distribution can be compared against expectations for the data. Additionally, the data quality dashboard (DQD) provides more objective checks on plausibility consistently across the data sources. In terms of relevance, more general-purpose diagnostic tools, CohortDiagnostics and DrugExposureDiagnostics, were developed. The CohortDiagnostics package provides additional insights into cohort characteristics, record counts and index event misclassification. The *DrugExposureDiagnostics* package assesses ingredient specific diagnostics for drug exposure records. Furthermore, data is maintained up to date by extracting the release dates for each dataset in the network and monitoring when data are out-of-date with the expected refresh cycle (typically quarterly or half-yearly). In addition, it is important to have clear understanding of the time covered by each released database, as this can vary across different domains. To facilitate this, the C*DMOnboarding* (and *Achilles*) packages contain a 'data density' plot. This plot displays the number of records per OMOP domain monthly. This allows us to get insights when data collection started, when new sources of data were added and when until when data was included.

Table 3. Description of the selected data sources.

Country	Name of Database	Justification for Inclusion	Health Care setting (e.g. primary care, specialist care, hospital care)	Type of Data (EHR, claims, registries)	Number of active subjects	Feasibility count of exposure (Azathioprine)*	Data lock for the last update
United Kingdom	CPRD GOLD	Adequate number of patients exposed to Azathioprine, availability of medical history for cohort characterisation. Able to perform study and produce results before mid-October deadline.	Primary care	EHR	17m	199,700	01/01/2024
Netherlands	IPCI	Adequate number of patients exposed to Azathioprine, availability of medical history for cohort characterisation. Able to perform study and produce results before mid-October deadline.	Primary Care	EHR	1.39m	15,700	30/04/2024
Spain	SIDIAP	Adequate number of patients exposed to Azathioprine, availability of medical history for cohort characterisation.	Primary care	EHR	5.8m	29,200	30/06/2023
Germany	IQVIA DA Germany	Adequate number of patients exposed to Azathioprine, availability of medical history for	Primary care and outpatient	EHR	43m	80,400	30/09/2023

Country	Name of Database	Justification for Inclusion	Health Care setting (e.g. primary care, specialist care, hospital care)	Type of Data (EHR, claims, registries)	Number of active subjects	Feasibility count of exposure (Azathioprine)*	Data lock for the last update
		cohort characterisation. Able to perform study and produce results before mid-October deadline.	secondary care				
Spain	IMASIS	Adequate number of patients exposed to Azathioprine, availability of medical history for cohort characterisation. Able to perform study and produce results before mid-October deadline.	Inpatient hospital care and secondary outpatient care	EHR	1.1m	2,100	10/02/2024

Clinical Practice Research Datalink GOLD, United Kingdom (University of Oxford)

The Clinical Practice Research Datalink (CPRD) GOLD is a database of anonymised electronic health records (EHR) from General Practitioner (GP) clinics in the UK that use the Vision® software system for their management.(11) The source population encompasses 98% of the UK, registered with GPs responsible for non-emergency care and referrals. Participating GPs provide CPRD EHR for all registered patients who did not specifically request to opt out of data sharing. Covering 4.6% of the current UK population, GOLD includes 4.9% of contributing GP practices, providing comprehensive information within its defined source population. GOLD contains data from all four UK constituent countries and the current regional distribution of its GP practices is 5.7% in England, 55.6% in Scotland, 28.4% in Wales, and 10.2% in Northern Ireland (May 2022).

GOLD data include patient's demographic, biological measurements, clinical symptoms and diagnoses, referrals to specialist/hospital and their outcome, laboratory tests/results, and prescribed medications. GOLD has been assessed and found broadly representative of the UK general population in terms of age, gender, and ethnicity.(11) GOLD has been widely used internationally for observational research to produce nearly 3,000 peer-reviewed publications, making GOLD the most influential UK clinical database so far(12-14).

The Integrated Primary Care Information (IPCI), the Netherlands

The Integrated Primary Care Information (IPCI) database is a longitudinal observational database containing routinely collected data extracted from computer-based patient records of a selected group of general practitioners (GPs) across the Netherlands.(15) IPCI was started in 1992 by the department of Medical Informatics of the Erasmus University Medical Center in Rotterdam. The current database includes patient records from 2006 on, when the size of the database started to increase significantly. The demographic composition of the IPCI population mirrors that of the general Dutch population in terms of age and sex. Although the geographical spread is limited, GP practices are located in urban and non-urban areas.

Patient-level data includes demographic information, patient's complaints and symptoms, diagnoses, laboratory test results, lifestyle factors and correspondence with secondary care, such as referral and discharge letters. For complaints, symptoms and diagnoses, Dutch GPs use International Classification of Primary Care (ICPC-1) coding, an international standard developed and updated by the World Organization of Family Doctors' (WONCA) International Classification Committee.

IPCI data quality has been previously documented and IPCI has proved valuable for epidemiological studies.(16-20) In terms of quality control, extensive quality control steps are performed prior to each data release. These include comparison of patient characteristics between practices and checks to identify abnormal temporal data patterns in practices. Additional checks include over 200 indicators related to population characteristics (e.g. reliability of birth and mortality rates) and medical data (e.g. availability of durations of prescriptions, completeness of laboratory results, availability of hospital letters and prescriptions, proportion of patients with blood pressure measurement, etc.(15) Based on this information, two quality scores have been created. Practices with low scores have been excluded.

Information System for Research in Primary Care (SIDIAP), Spain

The Information System for Research in Primary Care (SIDIAP) is a dynamic database of pseudo-anonymized electronic health records of the primary care patient population in Catalonia, Spain.(21) (REF 16 doxy) It contains data of approximately 80% of the Catalan population registered in over 280 primary care practices throughout Catalonia since 2005.

The database contains data recorded in primary care centres on a daily basis. Additionally, it integrates data from external sources including biomarkers data from laboratories and records of drug prescription and dispensation. The dataset covers demographics, all-cause mortality, disease diagnoses classified under the International Classification of Diseases 10th revision (ICD-10), prescription and dispensation records of drugs, results of laboratory tests, socio-economic indicators, vaccination records, lifestyle information, parent–child linkage and various clinical parameters. Additional data from other data sources such as hospital discharges, mental health centres or specific disease registries can be obtained through diverse linkages. The demographic composition within SIDIAP closely mirrors that of the broader Catalan population, encompassing a representative spectrum of geographic distribution, age, and sex proportions. The database is updated every 6 months.

SIDIAP data quality has been previously documented and SIDIAP has proved valuable for epidemiological studies.(22-30) In terms of data integrity and reliability, SIDIAP has been subject to rigorous evaluation. Quality checks have been implemented including central identification of duplicate patient ID and visual inspection for temporal patterns in the registry of a certain variable. Furthermore, the data undergoes assessment for availability (longitudinally and reliability), plausibility (range checks and unusual values) and consistency using visualization tools. Specifically, for biochemistry data, consistency for measurements taken in different laboratories is assessed, and unit conversion is undertaken when needed.

IQVIA Disease Analyser (DA) Germany, Germany

IQVIA Disease Analyzer (DA) Germany is a database of de-identified electronic medical records from specialized and general primary practices (GP) in Germany since 1992. This dataset encompasses approximately 3% of all outpatient practices within Germany, ensuring a substantial representation of the national healthcare landscape.(31, 32) The sampling methods used for practice selection, taking into account physician's demographics, specialty focus, community size category and federal state location, was instrumental in constructing a database that accurately mirrors the diverse spectrum of healthcare providers in the country.(31, 32) Consequently, data within IQVIA DA Germany database has been demonstrated to be representative of general and specialised practices throughout Germany.

The database contains demographics records, basic medical data, disease diagnosis according to International Classification of Diseases, 10th revision (ICD-10), and prescription records.(32) While the database partly records information on deaths and procedures, it currently does not support linkage with external data sources and therefore information on mortality is incomplete. Routine updates are conducted at regular intervals. The quality of data is assessed based on several criteria including completeness of information and correctness (e.g. linkage between diagnosis and prescriptions). IQVIA DA Germany is suitable for pharmacoepidemiologic and pharmacoeconomic studies as previously demonstrated.(25, 32, 33)

The Institut Municipal Assistència Sanitària Information System (IMASIS), Spain

The Institut Municipal Assistència Sanitària Information System (IMASIS) is the Electronic Health Record (EHR) system of Parc de Salut Mar Barcelona (PSMar) which is a complete healthcare services organisation. Currently, this information system includes and shares the clinical information of two general hospitals (Hospital del Mar and Hospital de l'Esperança), one mental health care centre (Centre Dr. Emili Mira) and one social-healthcare centre (Centre Fòrum) including emergency room settings, that are offering specific and different services in the Barcelona city area (Spain). At present, IMASIS includes clinical information from around 1 million patients with at least one diagnosis and who have used the services of this healthcare system since 1990 and from different settings such as admissions, outpatients, emergency room and major ambulatory surgery. The diagnoses are coded using The International Classification of Diseases ICD-9-CM and ICD-10-CM. The average follow-up period per patient in years is 6.37 (SD±6.82). IMASIS-2 is

the anonymised relational database of IMASIS which is used for mapping to OMOP including additional sources of information such as the Tumours Registry.

8.3 Study period

The study period was from 01/01/2000 to 31/12/2023 (when the study end date was not reached in a data source, the study end for that source was the last date of available data) (see **Table 3** for more details).

8.4 Follow-up

Study participants will be followed up from the date of first exposure to Azathioprine until date of:

1) censuring or loss to follow-up, 2) end of data availability, or 3) the end date of exposure (whichever occurs first).

Table 4. Operational definition of time 0 (index date) and other primary time anchors.

Study population name(s)	Time Anchor Description (e.g. time 0)	Number of entries	Type of entry	Washout window	Care Setting[1]	Code Type[2]	Diagnosis position	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Individuals initiating treatment with Azathioprine (objective 1,2, 3 and 4)	Date of first treatment with Azathioprine during study period	Single entry	Incident	Any time prior to study entry date	OP	RxNorm	N/A	Exposure	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

8.5 Study population with inclusion and exclusion criteria

The study population will include all individuals observed in one of the participating data sources during the study period (1st January 2000 to 31st December 2023) and having at least 1 year of database history.

Patients will be excluded if the start date in the database falls within 365 days prior to the index date, as sufficient data availability is required.

The operational definitions of the inclusion and exclusion criteria are presented by means of **Table 5** and **Table 6** , respectively.

Table 5. Operational definitions of inclusion criteria.

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position[2]	Applied to study populations:
Prior database history	Study participants will be required to have 365 days of prior history observed before the index date	Prior	[-365,-1]	OP	N/A	N/A	All individuals within selected databases
Observation period in the database during the period	All individuals present in the 2000-2023 (or the latest date available)	after	N/A	OP	N/A	N/A	All individuals within selected databases
User of Azathioprine (exposure)	Prescription record of azathioprine	after	Study period	OP	RxNorm	N/A	All individuals within selected databases

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

Table 6. Operational definitions of exclusion criteria.

Criterion	Details	Order of application	Assessment window	Care Settings[1]	Code Type	Diagnosis position[2]	Applied to study populations:	Measurement characteristics/ validation	Source for algorithm
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.6 Variables

Preliminary concept/code lists used for the identification of exposure/s and/or outcomes are included as Supplementary Documents in **Appendix I**. These 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 databases. The concepts sets already contain more broad concepts because the some of the indications have quite low granularity.

8.6.1 Exposure/s

Azathioprine exposure consists of a prescription record of azathioprine (for systemic use thus oral or parenteral use). The operational definition of exposure is described by means of **Table 7**.

Table 7. Operational definitions of exposure.

Exposure group name(s)	Details	Washout window	Assessment Window	Care Setting[1]	Code Type	Diagnosis position[2]	Applied to study populations	Incident with respect to…	Measurement characteristics/ validation	Source of algorithm
Azathioprine	Preliminary code lists provided in Appendix I	[-inf,-1], for incident use only	Anytime post start of the study date	OP	RxNorm	N/A	All eligible individuals in the database	Azathioprine use	N/A	N/A
Azathioprine + indication of use	Preliminary code lists provided in Appendix I	n/a = not applicable	Indication of use in -90/+30 days of index date	OP	SNOMED codes to assess indication of use	Yes (indication of use)	All incident users of azathioprine	Azathioprine use	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.6.2 Outcome/s

The study outcome(s) are described conceptually, and the context or rationale for the choices are provided in this section. The operational definition of the outcomes is presented in the **Table 8**.

Table 8. Operational definitions of outcome.

Outcome name	Details	Primary outcome ?	Type of outcome	Washout window	Care Settings[1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics/ validation	Source of algorithm
N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable
[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.6.3 Other covariates, including confounders, effect modifiers and other variables

The study covariates are described conceptually, and the context or rationale for the choices are provided in this section. The operational definition of the covariates is described in the **Table 9**.

Covariates for the patient-level characterisation

To characterise the new initiators of Azathioprine, the covariates will include:

- age at the index date (first prescription of Azathioprine) presented as a mean/median age as well as proportion of patient stratified within age groups (<18, 18-44, 45-64, 65-79, and >=80), overall and per indication
- sex (male/ female), overall and per indication.
- Conditions considered indications for Azathioprine use, will be assessed within –90/+30 days of the index date. These indications will be identified based on the presence of disease codes.
  The preliminary code list for each condition of interest is available in **Appendix I**.
  Indications:
    - Organ transplantation
    - Severe rheumatoid arthritis or chronic polyarthritis
    - Inflammatory bowel diseases
    - Autoimmune hepatitis
    - Systemic lupus erythematosus
    - Dermatomyositis
    - Polyarteritis nodosa
    - Pemphigus vulgaris and bullous pemphigoid
    - Behcet's disease
    - Refractory autoimmune haemolytic anaemia
    - Refractory idiopathic thrombocytopenic purpura
    - Polymyositis

- Pyoderma gangrenosum
- Multiple sclerosis
- Myasthenia gravis
- None of the above/missing

- Large-scale characterisation: overall and per indication of drugs (by RxNorm code at ingredient level) and conditions (by standard SNOMED code) within one year prior to the index date (-365 to -1 day to the index date) and at the index date will be assessed. While all conditions and medications above 1% will be available within the Shiny App, only the 10 most frequent conditions and drugs will be described within the report. Drugs will be reported by ingredient and not by strength or formulation
- Treatment duration overall and per indication, will be assessed from the index date until date of: 1) censuring or loss to follow-up, 2) end of data availability, or 3) the end date of exposure (whichever occurs first). Furthermore, to provide more context, also the proportion (%) of patients with a duration that fall within a prespecified treatment duration group will be calculated and provided, both for the overall treatment duration as per indication. The prespecified treatment duration groups will be: <12 months, 12-24 months, 2-3 years, 3-4 years, 4-5 years, 5-10 years and ≥10 years.

Table 9. Operational definitions of covariates.

Characteristic	Details	Type of variable	Assessment window	Care Settings [1]	Code Type	Diagnosis Position[2]	Applied to study populations	Measurement characteristics / validation	Source for algorithm
Demographics	Age and sex at index date	Numeric, continuous, binary	All history	OP	N/A	N/A	Incident azathioprine users	N/A	N/A
Indications	Prespecified indications of interest -90/+30 days of the index date	Count, binary	[-90, +30]	OP	SNOMED	N/A	Incident azathioprine users	N/A	N/A
Medication use, comorbidities	All history prior to the index date (large-scale characterisation)	Count, binary	[-365, -1] and [0,0]	OP	SNOMED, RxNorm	N/A	Incident Azathioprine users	N/A	N/A
Treatment duration	Duration of treatment with Azathioprine for the: - new users of azathioprine use - New users of azathioprine per indication	Numeric, continuous	[0, end of exposure]	OP	N/A	N/A	Incident azathioprine users	N/A	N/A

[1] IP = inpatient, OP = outpatient, ED = emergency department, OT = other, n/a = not applicable

[2] Specify whether a diagnosis code is required to be in the primary position (main reason for encounter)

8.7 Study size

No sample size has been calculated as this is a descriptive Drug Utilisation Study where we are interested in the characteristics of all incident cases of new users of Azathioprine in each database. Based on a preliminary feasibility assessment, the expected person counts for Azathioprine were 900 for IMASIS, 6,000 for IPCI, 23,700 for IQVIA DA Germany, 24,500 for SIDIAP and 52,100 for CPRD GOLD.

The study period is from 2000-2023 but it should be noted that not all DPs do have observations period starting from 2000 on. For IPCI, observation period starts from 2008 and for SIDIAP from 2005.

8.8 Analysis

All analyses will be conducted separately for each database, and will be carried out in a federated manner, allowing analyses to be run locally without sharing patient-level 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 the final package will be released in a version-controlled study repository for execution against all the participating data sources.

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 (Data Transfer Zone). All results will be locked and timestamped for reproducibility and transparency.

Cell counts <5 will be suppressed to comply with the database's privacy protection regulations.

The type of analysis in this study is described in **Table 10**.

Table 10. Description of study types and type of analysis.

Study type	Study classification	Type of analysis
Patient Level DUS	Off-the-shelf	- Characterisation of patient-level features - Frequency and % of indication/s - Large-scale characterisation of drug use and comorbidities - Estimation of minimum, p25, median, p75, and maximum treatment duration

8.8.1 Statistical model specification and assumptions of the analytical approach considered

R-packages

We will use the R package "DrugUtilization" package to calculate the duration of interest and the *PatientProfiles* and *CohortDiagnostics* R packages to describe the characteristics of individuals initiating treatment with Azathioprine.

Drug exposure calculation

Drug eras are defined as follows: Exposure starts at date of the first prescription, e.g., the index date the person entered the cohort. For each prescription, the estimated duration of use is retrieved from the drug exposure table in the CDM, using the start and end date of the exposure. Subsequent prescriptions are combined into continuous exposed episodes (drug eras) using the following specifications:

Two drug eras are merged into one continuous drug era if the distance in days between end of the first era and start of the second era is ≤ 30 days. The time between the two joined eras is considered as exposed as shown in **Figure 1**, first three rows.

Gap era joint mode	Schematics	Dose in between	Cumulative dose	Cumulative time
"first"		d_1	$d_1 \cdot (x_1 + x_{12}) + d_2 \cdot x_2$	$x_1 + x_{12} + x_2$
"second"		d_2	$d_1 \cdot x_1 + d_2 \cdot (x_2 + x_{12})$	$x_1 + x_{12} + x_2$
"zero"		0	$d_1 \cdot x_1 + d_2 \cdot x_2$	$x_1 + x_{12} + x_2$
"join"	first exposure time = x_1, dose = d_1 gap time = x_{12} second exposure time = x_2, dose = d_2	NA	$d_1 \cdot x_1 + d_2 \cdot x_2$	$x_1 + x_2$

Figure 1. Gap-era joint mode.

For each prescription, the estimated duration of use is retrieved from the drug exposure table in the CDM, using the start and end date of the exposure, and calculated as the duration of the first treatment era of the azathioprine during the study period. Treatment duration is summarized providing the minimum, p25, median, p75, and maximum treatment duration.

8.8.2 Methods to derive parameters of interest

Calendar time

Calendar time will be based on the calendar year of the index prescription.

Age

Age at index date will be calculated using January 1st of the year of birth as a proxy for the actual birthday. Age will be presented as a mean/median age as well as proportion of patient stratified within age groups (<18, 18-44, 45-64, 65-79, and >=80).

Sex

Results will be presented stratified by sex (men/women).

8.8.3 Patient-level characteristics on/before index date

Characterisation of patients newly treated with Azathioprine: We will use the R package "PatientsProfiles" for the patient-level characterisation of demographics and predefined clinical characteristics, as well as large-scale characterisation. The co-variates to be presented in a summary baseline characteristics table have been described in section 8.6.3.

Characterisation of treatment with azathioprine in a cohort of new users: The "DrugUtilisation" package will be used to characterise the use of Azathioprine.

The number and % of patients receiving Azathioprine for a prespecified list of indications including duration of treatment.

8.8.4 Output

New drug user, patient-level characterisation:

- Table 1. Study attrition of individuals included in each cohort per database.
- Table 2. Number and proportion of participants treated with azathioprine. Distribution (number and %) overall and stratified by sex, age and age-groups Number of participants per pre-specified strata will be included where necessary/applicable.
- Table 3. Distribution of prespecified indications in participants with azathioprine. Number of participants per pre-specified strata will be included where necessary/applicable.
- Table 4. Large-scale characterisation: top 10 drugs (ingredient level) in the year prior to index date ([-365, -1]).
- Table 5. Large-scale characterisation: top 10 conditions in the year prior to index date ([-365, -1]).
- Table 6. Large-scale characterisation: top 10 drugs (ingredient level) on index date ([0,0]).
- Table 7. Large-scale characterisation: top 10 conditions on index date ([0,0]).
- Table 8: Duration of treatment for Azathioprine overall and per prespecified indication. Proportion (%) of patients with a duration that fall within the prespecified treatment duration group (<12 months, 12-24 months, 2-3 years, 3-4 years, 4-5 years, 5-10 years and ≥10 years.), both overall and per indication.

8.9 Evidence synthesis

Results from analyses described in section 8.8 will be presented separately for each database. No meta-analysis will be conducted.

9. DATA MANAGEMENT

9.1 Data management

All databases are mapped to the OMOP common data model. This enables the use of standardised analytics and tools across the network since the structure of the data and the terminology system is harmonised. The OMOP CDM is 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. Each data partner will execute the study code against their database containing patient-level data and will then return the results set 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.

9.2 Data storage and protection

For this study, participants from various EU member states will process personal data from individuals which is collected in national/regional electronic health record databases. 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 databases 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 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 in order 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.

10. QUALITY CONTROL

General database quality control

A number of open-source quality control mechanisms for the OMOP CDM have been developed (see Chapter 15 of The Book of OHDSI http://book.ohdsi.org/DataQuality.html). In particular, data partners have run the OHDSI Data Quality Dashboard tool (https://github.com/OHDSI/DataQualityDashboard). This tool provides numerous checks relating to the conformance, completeness and plausibility of the mapped data. Conformance focuses on checks that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions, completeness in the sense of data quality is solely focused on quantifying missingness, or the absence of data, while plausibility seeks to determine the believability or truthfulness of data values. Each of these categories has one or more subcategories and are evaluated in two contexts: validation and verification. Validation relates to how well data align with external benchmarks with expectations derived from known true standards, while verification relates to how well data conform to local knowledge, metadata descriptions, and system assumptions.

Study specific quality control

When defining specific drugs, conditions, and co-morbidities, a systematic search of possible codes for inclusion was previously identified using CodelistGenerator R package (https://github.com/darwin-eu/CodelistGenerator). This software allows the user to define a search strategy and, using this will then query the vocabulary tables of the OMOP CDM to find potentially relevant codes. The codes returned will then be reviewed by two clinical epidemiologists to consider their relevance. In addition, the CohortDiagnostics R package (https://github.com/OHDSI/CohortDiagnostics) will be run to assess the use of different codes across the databases contributing to the study and identify any codes potentially omitted in error. This will allow for a consideration of the validity of the study cohort of patients with the selected conditions, drugs, and co-morbidities in each of the databases and inform decisions around whether multiple definitions are required.

11. LIMITATIONS OF THE RESEARCH METHODS

Regarding azathioprine use, the recording of the prescription of the drug (e.g. concept id at clinical drug level or at ingredient level), and the prescriptions themselves may vary across databases. In addition, a recording of a prescription or dispensation does not mean that the patient actually took the drug. In addition, assumptions around the duration of drug use will be unavoidable and might lead to overestimating treatment duration of treatment (considering a 30-day drug gap era).

For this study, the indication of use is important and predefined indications of use have been identified. Not all of these indications of interest might be present within the database (i.e. in DP with less granular source coding such as IPCI). For that reason, it might be interesting to use more broad definitions of the indication of use (e.g. look for “inflammatory disease of liver” and not necessarily “autoimmune hepatitis”. These broad searches might impose a potential of misclassification however, as we search in a relatively small window, 3 months before and 1 month after, around the index date (first prescription of azathioprine during study period) this risk of misclassification is probably minimal. Also, the large-scale characterisation (looking for top 10 disease codes) might provide insight into the indication of use on top of the indications which have been predefined.

The study period is from 2000 until 2023 however not all DPs have that long observation time. For IPCI, observation period starts from 2008 and for SIDIAP from 2005.

Small cell counts may affect the analysis for some subgroup strata. If the numbers are too low, counts will not be disclosed for governance reasons.

12. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Adverse events/adverse reactions will not be collected or analysed as part of this evaluation. The nature of this non-interventional evaluation, through the use of secondary data, does not fulfil the criteria for reporting adverse events, according to module VI, VI.C.1.2.1.2 of the Good Pharmacovigilance Practices (https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf).

Only in case of prospective data collection, there is a need to describe the procedures for the collection, management and reporting of individual cases of adverse events/adverse reactions.

13. GOVERNANCE BOARD ASPECTS

All data sources require approval from their respective IRB boards, except IQVIA DA Germany, which will not require any further specific approvals to undertake this study.

14. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

14.1 Study Report

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.

15. OTHER ASPECTS

None.

16. REFERENCES

1. Abu-Shakra M, Shoenfeld Y. Azathioprine therapy for patients with systemic lupus erythematosus. Lupus. 2001;10(3):152-3.
2. Bunch TW. Prednisone and azathioprine for polymyositis: long-term followup. Arthritis Rheum. 1981;24(1):45-8.
3. Casetta I, Iuliano G, Filippini G. Azathioprine for multiple sclerosis. Cochrane Database of Systematic Reviews. 2007(4).
4. Gordon PA, Winer JB, Hoogendijk JE, Choy EH. Immunosuppressant and immunomodulatory treatment for dermatomyositis and polymyositis. Cochrane Database of Systematic Reviews. 2012(8).
5. Palace J, Newsom-Davis J, Lecky B, Group MGS. A randomized double-blind trial of prednisolone alone or with azathioprine in myasthenia gravis. Neurology. 1998;50(6):1778-83.
6. de Boer NKH, Peyrin-Biroulet L, Jharap B, Sanderson JD, Meijer B, Atreya I, et al. Thiopurines in Inflammatory Bowel Disease: New Findings and Perspectives. J Crohns Colitis. 2018;12(5):610-20.
7. Woodland J, De Saintonge DC, Evans S, Sharman V, Currey H. Azathioprine in rheumatoid arthritis: double-blind study of full versus half doses versus placebo. Annals of the Rheumatic Diseases. 1981;40(4):355-9.
8. Björnsson ES, Gu J, Kleiner DE, Chalasani N, Hayashi PH, Hoofnagle JH. Azathioprine and 6-Mercaptopurine-induced Liver Injury: Clinical Features and Outcomes. J Clin Gastroenterol. 2017;51(1):63-9.
9. Koller T, Galambosova M, Filakovska S, Kubincova M, Hlavaty T, Toth J, et al. Drug-induced liver injury in inflammatory bowel disease: 1-year prospective observational study. World J Gastroenterol. 2017;23(22):4102-11.
10. De Gottardi A, Rautou PE, Schouten J, Rubbia-Brandt L, Leebeek F, Trebicka J, et al. Porto-sinusoidal vascular disease: proposal and description of a novel entity. Lancet Gastroenterol Hepatol. 2019;4(5):399-411.
11. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.
12. Carey IM, Nirmalananthan N, Harris T, DeWilde S, Chaudhry UAR, Limb E, et al. Prevalence of co-morbidity and history of recent infection in patients with neuromuscular disease: A cross-sectional analysis of United Kingdom primary care data. PLoS One. 2023;18(3):e0282513.
13. Fahmi A, Wong D, Walker L, Buchan I, Pirmohamed M, Sharma A, et al. Combinations of medicines in patients with polypharmacy aged 65-100 in primary care: Large variability in risks of adverse drug related and emergency hospital admissions. PLoS One. 2023;18(2):e0281466.
14. Wigglesworth S, Neligan A, Dickson JM, Pullen A, Yelland E, Anjuman T, et al. The incidence and prevalence of epilepsy in the United Kingdom 2013-2018: A retrospective cohort study of UK primary care data. Seizure. 2023;105:37-42.
15. de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, et al. Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands. Int J Epidemiol. 2022;51(6):e314-e23.
16. Ali MS, Berencsi K, Marinier K, Deltour N, Perez-Guthann S, Pedersen L, et al. Comparative cardiovascular safety of strontium ranelate and bisphosphonates: a multi-database study in 5 EU countries by the EU-ADR Alliance. Osteoporos Int. 2020;31(12):2425-38.

17. Baan EJ, van den Akker ELT, Engelkes M, de Rijke YB, de Jongste JC, Sturkenboom M, et al. Hair cortisol and inhaled corticosteroid use in asthmatic children. Pediatr Pulmonol. 2020;55(2):316-21.
18. Berencsi K, Sami A, Ali MS, Marinier K, Deltour N, Perez-Gutthann S, et al. Impact of risk minimisation measures on the use of strontium ranelate in Europe: a multi-national cohort study in 5 EU countries by the EU-ADR Alliance. Osteoporos Int. 2020;31(4):721-55.
19. Engelkes M, Baan EJ, de Ridder MAJ, Svensson E, Prieto-Alhambra D, Lapi F, et al. Incidence, risk factors and re-exacerbation rate of severe asthma exacerbations in a multinational, multidatabase pediatric cohort study. Pediatr Allergy Immunol. 2020;31(5):496-505.
20. James G, Collin E, Lawrance M, Mueller A, Podhorna J, Zaremba-Pechmann L, et al. Treatment pathway analysis of newly diagnosed dementia patients in four electronic health record databases in Europe. Soc Psychiatry Psychiatr Epidemiol. 2021;56(3):409-16.
21. Recalde M, Rodríguez C, Burn E, Far M, García D, Carrere-Molina J, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). Int J Epidemiol. 2022;51(6):e324-e36.
22. Braeye T, Emborg HD, Llorente-García A, Huerta C, Martín-Merino E, Duarte-Salles T, et al. Age-specific vaccination coverage estimates for influenza, human papillomavirus and measles containing vaccines from seven population-based healthcare databases from four EU countries - The ADVANCE project. Vaccine. 2020;38(16):3243-54.
23. Burn E, Tebé C, Fernandez-Bertolin S, Aragon M, Recalde M, Roel E, et al. The natural history of symptomatic COVID-19 during the first wave in Catalonia. Nat Commun. 2021;12(1):777.
24. Lane JC, Butler KL, Poveda-Marina JL, Martinez-Laguna D, Reyes C, de Bont J, et al. Preschool Obesity Is Associated With an Increased Risk of Childhood Fracture: A Longitudinal Cohort Study of 466,997 Children and Up to 11 Years of Follow-up in Catalonia, Spain. J Bone Miner Res. 2020;35(6):1022-30.
25. Ly NF, Flach C, Lysen TS, Markov E, van Ballegooijen H, Rijnbeek P, et al. Impact of European Union Label Changes for Fluoroquinolone-Containing Medicinal Products for Systemic and Inhalation Use: Post-Referral Prescribing Trends. Drug Saf. 2023;46(4):405-16.
26. Monteagudo M, Nuñez A, Solntseva I, Dhalwani N, Booth A, Barrecheguren M, et al. Treatment Pathways Before and After Triple Therapy in COPD: A Population-based Study in Primary Care in Spain. Arch Bronconeumol (Engl Ed). 2021;57(3):205-13.
27. Ortega Y, Aragonès E, Piñol JL, Basora J, Araujo A, Cabré JJ. Impact of depression and/or anxiety on the presentation of cardiovascular events in a cohort with metabolic syndrome. StreX project: Five years of follow-up. Prim Care Diabetes. 2018;12(2):163-71.
28. Ramos R, Comas-Cufí M, Martí-Lluch R, Balló E, Ponjoan A, Alves-Cabratosa L, et al. Statins for primary prevention of cardiovascular events and mortality in old and very old adults with and without type 2 diabetes: retrospective cohort study. Bmj. 2018;362:k3359.
29. Recalde M, Davila-Batista V, Díaz Y, Leitzmann M, Romieu I, Freisling H, et al. Body mass index and waist circumference in relation to the risk of 26 types of cancer: a prospective cohort study of 3.5 million adults in Spain. BMC Med. 2021;19(1):10.
30. Troncoso-Mariño A, López-Jiménez T, Roso-Llorach A, Villén N, Amado-Guirado E, Guisado-Clavero M, et al. Medication-related problems in older people in Catalonia: A real-world data study. Pharmacoepidemiol Drug Saf. 2021;30(2):220-8.
31. Rathmann W, Bongaerts B, Carius HJ, Kruppert S, Kostev K. Basic characteristics and representativeness of the German Disease Analyzer database . Int J Clin Pharmacol Ther. 2018;56(10):459-66.
32. Zappacosta S, Cascarano A, Konrad M, Tanislav C, Kostev K. Association between COVID-19 and subsequent vascular events in primary care patients in Germany. Public Health. 2022;213:107-13.
33. Tanislav C, Rosenbauer J, Zingel R, Kostev K. No increased incidence of venous thrombosis or pulmonary embolism after SARS-CoV-2 vaccination in Germany. Public Health. 2022;207:14-8.

17. ANNEXES

Appendix I: List of Stand-Alone documents (e.g. lists with concept definitions (conditions & drugs), validation procedures, questionnaires etc.).

Appendix II: ENCePP checklist for study protocols.

Appendix III: Additional Information.

APPENDIX I:

Table S1. Code list for drug of interest .

Concept set name	IDs (including descendants)
Azathioprine	19014878

Table S2. Code list for indication definitions.

Concept set name	Included concepts	IDs (including descendants)
Organ transplantation	Transplantation Transplant follow-up Disorder related to transplantation	4300185 4081759 4179244
Severe rheumatoid arthritis or chronic polyarthritis	Rheumatoid arthritis Chronic polyarthritis	80809 3168431
Inflammatory bowel diseases	Inflammatory bowel disease Ulcerative colitis Crohn's disease	4074815 81893 201606
Autoimmune hepatitis	Autoimmune hepatitis	200762
Systemic lupus erythematosus	Systemic lupus erythematosus Cutaneous lupus erythematosus	257628 4324123
Dermatomyositis	Dermatomyositis	80182
Polyarteritis nodosa	Polyarteritis nodosa	320749
Pemphigus vulgaris and bullous pemphigoid	Pemphigus Bullous pemphigoid	135338 4298692
Behcet's disease	Behcet's syndrome	436642
Refractory autoimmune haemolytic anaemia	Hemolytic anemia	435503
Refractory idiopathic thrombocytopenic purpura	Immune thrombocytopenia Thrombocytopenic purpura	4103532 4119134
Polymyositis	Polymyositis	80800
Pyoderma gangrenosum	Pyoderma gangrenosum	133283
Multiple sclerosis	Multiple sclerosis	374919
Myasthenia gravis	Myasthenia gravis	76685

APPENDIX II: ENCePP checklist for study protocols

ENCePP Checklist for Study Protocols (Revision 4)

Study title: DARWIN EU® - Azathioprine – user characteristics

EU PAS Register® number: EUPAS1000000322
Study reference number: P3-C1-014

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1. Does the protocol specify timelines for				
1.1.1 Start of data collection				
1.1.2 End of data collection	X			5. Milestones,
1.1.3 Progress report(s)				8.2 Data Sources
1.1.4 Interim report(s)				
1.1.5 Registration in the EU PAS Register®				
1.1.6 Final report of study results.	X			

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:				
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)	X			7. Research question and objectives
2.1.2 The objective(s) of the study?	X			
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	X			8. Research methods
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)	X			8.1 Study type and Study Design
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	X			8.2 Study Setting and Data Sources

3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	X			8.8 Analysis
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))			X	
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)			X	

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	X			8.5 Study Population
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	X			8.3 Study Period
4.2.2 Age and sex	X			8.6.3. Other covariates
4.2.3 Country of origin				8.2 Study Setting and Data Sources
4.2.4 Disease/indication	X			8.6.1. Exposures
4.2.5 Duration of follow-up	X			8.4 Follow-up
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	X			8.5 Study Population with inclusion and exclusion criteria

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)	X			8.6.1. Exposures And 8.8 Analysis
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)			X	
5.3 Is exposure categorised according to time windows?			X	
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	X			8.6.3. Other covariates And 8.8 Analysis
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?			X	
5.6 Is (are) (an) appropriate comparator(s) identified?			X	

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?	X			8.6.2. Outcomes
6.2 Does the protocol describe how the outcomes are defined and measured?	X			8.6.2. Outcomes
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)		X		
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)		X		

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)			X	
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)			X	
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)			X	

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)			X	

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)	X			8.2 Study Setting and Data Sources
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	X			8.2 Study Setting and Data Sources
9.1.3 Covariates and other characteristics?	X			8.6.3. Other covariates
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)	X			8.2 Study Setting and Data Sources

9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	X			8.2 Study Setting and Data Sources
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	X			8.2 Study Setting and Data Sources And 8. 6.3. Other covariates
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	X			
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	X			8.6.2. Outcomes
9.3.3 Covariates and other characteristics?	X			8.6.3. Other covariates
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)			X	

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	X			8.8 Analysis
10.2 Is study size and/or statistical precision estimated?			X	
10.3 Are descriptive analyses included?	X			8.8.2 Descriptive statistics
10.4 Are stratified analyses included?	X			8.8 Analysis
10.5 Does the plan describe methods for analytic control of confounding?			X	
10.6 Does the plan describe methods for analytic control of outcome misclassification?			X	
10.7 Does the plan describe methods for handling missing data?	X			8.6 variables and 8.8 analysis
10.8 Are relevant sensitivity analyses described?			X	

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)	X			9. Data management
11.2 Are methods of quality assurance described?	X			10. Quality Control
11.3 Is there a system in place for independent review of study results?			X	

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).	X			11. Limitations of the research 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)	X			Table 8.2. Description of the selected Data Sources.

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?	X			13. Governance board aspects
13.2 Has any outcome of an ethical review procedure been addressed?			X	
13.3 Have data protection requirements been described?	X			9.2 Data storage and protection

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?	X			4. Amendments and updates

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)?	X			14. Plans for disseminating and communicating study results
15.2 Are plans described for disseminating study results externally, including publication?	X			14. Plans for disseminating and communicating study results

Comments:

Name of the main author of the protocol: Guido Van Leeuwen

Date: 25/09/2024

Signature: