SAFETY-VAC

A framework for the post-authorisation SAFETY monitoring and evaluation of VACcines in Europe

Specific Contract No 02 implementing framework contract No EMA/2020/46/TDA/18

Study Report for Objective 3:
Critical review of existing case definitions for immunocompromised populations, and a consolidated approach to identify and characterise such populations in real-world data sources.

Deliverable 4 – September 2024

Main Authors: Carlos E. Durán, Judit Riera-Arnau, Nicoletta Luxi, Fabio Riefolo, Martín Solorzano, Irene Pazos, Elena Ballarín, Lise Skovgaard Svingel, Lorenzo Chiusaroli, Elisa Martín-Merino, Elisa Barbieri, María Lopez-Lasanta, Sima Mohammadi, Denis Rotta, Miriam Sturkenboom

Document version: 1.0
First Release date: 16 September 2024
Second Release date: 21 October 2024
Third Release date: 25 November 2024
Final Release:

Disclaimer & acknowledgement
The research leading to these results was conducted as part of the activities of the EU PE&PV (Pharmacoepidemiology and Pharmacovigilance) Research Network (led by Utrecht University) in collaboration with the Vaccine Monitoring Collaboration for Europe network (VAC4EU).
The scientific work for this project was coordinated by the University Medical Center Utrecht.
The project has received support from the European Medicines Agency under the Framework service contract nr EMA/2020/46/TDA/L5.06.
This report expresses the opinion of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.
This study has been registered in the HMA-EMA Catalogue of Real-World Data under the EU PAS number: EUPAS1000000288.

Study Information

Title	SAFETY-VAC: A framework for the post-authorization SAFETY monitoring and evaluation of VACcines in Europe
Version identifier	v1.0
Date of last version of the final study report	26 November 2024
EUPAS number	EUPAS1000000288
Active substance	Not Applicable
Medicinal product	Not Applicable
Product reference	Not Applicable
Procedure number	Not Applicable
Marketing authorization holder(s)	Not Applicable
Joint PASS	Not Applicable
Research questions and objectives	Providing an immunocompromised phenotype definition that can be applied in multi-database pharmacoepidemiologic studies.
Country (-is) of study	Not Applicable
Authors	Carlos E. Durán, Judit Riera-Arnau, Nicoletta Luxi, Fabio Riefolo, Martín Solorzano, Irene Pazos, Elena Ballarín, Lise Skovgaard Svingel, Lorenzo Chiusaroli, Elisa Martín-Merino, Elisa Barbieri, María Lopez-Lasanta, Sima Mohammadi, Denis Rotta, Miriam Sturkenboom

Table of Contents

1. TITLE

SAFETY-VAC: Critical review of existing case definitions for immunocompromised populations, and a consolidated approach to identify and characterise such populations in real-world data sources.

2. ABSTRACT

2.1. Title

SAFETY-VAC: Critical review of existing case definitions for immunocompromised populations, and a consolidated approach to identify and characterise such populations in real-world data sources.

2.2. Keywords

Immunocompromised host, immunosuppression, diagnostic codes, algorithms, databases, real-world data.

2.3. Rationale and Background

Immunocompromised individuals are characterized by a dysfunctional immune system due to conditions such as Human Immunodeficiency Virus (HIV) / acquired immunodeficiency syndrome (AIDS), organ transplants, use of immunosuppressants, or primary immunodeficiencies. Whereas real-world data (RWD) studies rely on clinical definitions, these must be translated into machine-readable algorithms to identify immunocompromised populations in electronic healthcare records (EHR) data sources. The main challenges arise from the variability in defining such populations in epidemiological studies and the sometimes temporary immunocompromised status, such as in individuals with secondary immunodeficiencies.

2.4. Research questions and objectives

The following research question was formulated: What are the operational definitions used to identify immunocompromised populations when conducting epidemiological studies in population-based EHR databases?

The goal was to produce a phenotype suitable to be applied in multi-database pharmacoepidemiological studies.

2.5. Methods

We performed a scoping literature review in the MEDLINE database using terms related to observational studies, immunosuppression and immunocompromised status, and coding systems. Studies using EHRs and administrative/claims databases and including definitions of immunocompromised populations were included. Data was extracted using RedCap®.

Identified clinical concepts (clinical conditions, medicinal products, tests, and algorithms) were retrieved, organised and assigned into broad clinical categories in consultation with clinical experts.

2.6. Results

From 137 citations initially identified in MEDLINE, a total of 56 studies were finally selected for data charting. Studies used more than one element to identify immunocompromised populations: 91% (n=51) of the studies used a set of diagnostic codes, 78.6% (n=44) used clinical specifications in free text, 42.8% (n=24) used drug exposure codes, and 12.5% (n=7) used algorithms combining diagnoses, drugs, laboratory and imaging tests, number of healthcare visits, and other elements.
The retrieved medical conditions and medicinal products used to define immunocompromised status were classified into 7 diagnostic categories. Category 1 lists genetic and hereditary conditions. Category 2 contains infectious diseases and related conditions (e.g., opportunistic infections). Category 3 includes haematological malignancies, solid organ malignancies and hospitalization for chemotherapy. Category 4 lists diagnoses of solid organ and stem cell transplantations. Category 5 groups clinical conditions that did not fit any other category, (e.g. severe kidney and liver disease, severe malnutrition, severe burns, preterm birth, cryoglobulinemia, and haematological neutropenia). Category 6 lists immunosuppressants and medicines to treat conditions leading to immunosuppressive state. Finally, category 7 lists 19 autoimmune conditions that may lead to immunosuppression only if associated with prescription or dispensing of an immunosuppressant.

2.7. Phenotype to identify immunocompromised populations in EHR databases.

The ultimate goal of this work was to produce a machine-readable phenotype to identify immunocompromised populations across all healthcare settings.

Two clinicians-epidemiologists from the core study team transformed and jointed the retrieved clinical conditions and medicinal products using Boolean terms and logical sequences. The algorithm contained: medical conditions, medicinal products, diagnostic tests, dosage information, and duration of exposure. Finally, the proposed phenotype was reviewed by four clinical specialists, and potential redundancies or missing points were addressed. Our harmonised phenotype algorithm aims to be applicable in a wide variety of databases. In this sense, it is specific but granular. It can be used by pooling all algorithm blocks together, or by selecting individual algorithms, depending on the research question and the population of interest.

2.8. Conclusions

For improving the identification of immunocompromised individuals in epidemiological research based on secondary use of healthcare data, we have developed a phenotype to identify immunocompromised populations when conducting epidemiological and pharmacoepidemiological studies in large EHR databases. Its modular design utilising flexible algorithmic blocks enhances adaptability across diverse database types, and research questions. Future research should focus on its formal validation and applicability across databases with sources of different origins.

3. INVESTIGATORS

Judit Riera-Arnau*, Nicoletta Luxi**, Martín Solorzano***, Sima Mohammadi, Miriam Sturkenboom, Carlos E. Durán.	Data Science and Biostatistics Department, University Medical Center Utrecht, the Netherlands. *Department of Clinical Pharmacology, Vall Hebron Institut de Recerca (VHIR), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. **Department of Medicine, University of Verona, Italy ***IDIAP Jodi Gol, Barcelona, Spain
Fabio Riefolo, Irene Pazos	Teamit Institute, Partnerships, Barcelona Health Hub, 08025 Barcelona, Spain
Elena Ballarín, María Lopez-Lasanta	Department of Clinical Pharmacology, Vall Hebron Institut de Recerca (VHIR), Universitat Autònoma de Barcelona (UAB), Barcelona, Spain.
Denis Rotta	Department of Medicine, University of Verona, Italy
Lise Skovgaard Svingel	Department of Clinical Epidemiology and Department of Clinical Medicine, Aarhus University and Aarhus University Hospital, Aarhus 8200, Denmark
Lorenzo Chiusaroli, Elisa Barbieri	Department for Women's and Children' Health, University of Padua, Italy.
Elisa Martín-Merino	Spanish Agency of Medicines and Medical Devices-AEMPS, Madrid, Spain
SAFETY-VAC consortium	

4. MILESTONES

Start of the project	15 Feb 2024
D1 Project planning virtual meeting	28 Feb 2024
D4 Study report for Objective 3	16 September 2024
D4 Study report for Objective 3 acceptance	04 December 2024

5. RATIONALE AND BACKGROUND

The SAFETY-VAC project

The COVID-19 pandemic emphasised the public health need for comprehensive and rapid post-authorisation vaccine safety surveillance. While safety concerns are expected to arise with novel vaccines, continuous monitoring and evaluation throughout the entire lifecycle remains necessary for all authorised vaccines (1,2). To this aim, networks of real-world data (RWD) sources that are fit-for-purpose and readily accessible are essential. In May 2022, the European Medicines Agency (EMA) and the European Centre for Disease Prevention and Control (ECDC) established the Vaccine Monitoring Platform (VMP) (3). The VMP aims to generate evidence on the safety and effectiveness of vaccines in Europe. VAC4EU and the EU PE&PV research network (EU PE&PV) can provide a framework for the post-authorisation safety and effectiveness evaluation of vaccines. Under the umbrella of these organisations, the SAFETY-VAC project has been settled to deliver an assessment of a network of databases for vaccine safety research purposes (objective 1) (4,5), to estimate incidence rates of flares of 10 selected autoimmune diseases (objective 2) (6), and to provide a phenotype definition framework to identify immunocompromised populations in large electronic healthcare records (EHR) databases (objective 3). This report aims to report on objective 3.

The epidemiology and the identification of immunocompromised populations in EHR databases

Immunocompromised individuals represent a vulnerable population that requires particular considerations when studying the effectiveness and safety of medicinal products, including biologics and vaccines (7). Recently, the prevalence of immunocompromised populations has been reported to vary between 1.4% and 5.4% in population-based studies performed in European healthcare data sources (5,8). There is an association between immunosuppression and increased hospitalization rates which is well-documented across different studies and conditions (9–12). Furthermore, the mortality of ICU ventilator-associated pneumonia has been reported to be higher in immunocompromised than in non-immunocompromised patients (64% vs. 34%), mainly related to multidrug-resistant pathogens (13). Besides, ICU-acquired bloodstream and opportunistic bacteria, viruses and fungi infections are frequent in immunocompromised patients (14).

The terms *immunocompromised* or *immunosuppressed* refer to the host´s inability to combat infections from common pathogens and opportunistic microorganisms that otherwise are considered innocuous for immunocompetent hosts. In this report, both terms are used interchangeably. Immunocompromised individuals include a heterogeneous group of the population with a large range of types and degrees of immunodeficiencies affecting humoral and/or cellular immunity (15). A host´s immunosuppressed status may come from primary immunodeficiency diseases (PIDs) including genetic or hereditary conditions intrinsic to the immune system. Examples include congenital conditions such as severe combined immunodeficiency (SCID), caused by various mutations which can impact several immune cell lineages, and common variable immune deficiency (CVID). Secondary immunodeficiencies include a wide range of medical conditions such as human immunodeficiency virus (HIV) infection or acquired immunodeficiency syndrome (AIDS), organ transplantation, or

haematological malignancies. Exposure to *immunosuppressants* or *immunosuppressive drugs* like corticosteroids or chemotherapy can also contribute to a compromised immune system (7). Moreover, the consumption of immunosuppressive drugs can add complexity to this definition, resulting in different levels of severity of the immunocompromised status, varying from moderate to severe (16,17).

Whereas RWD studies rely on clinical definitions, these concepts must be translated into machine-readable algorithms to identify immunocompromised populations in EHR data sources. The main challenges arise from the high phenotypic variability of the conditions leading to immunosuppression or a compromised immune system, and the temporary status of these conditions in secondary immunodeficiencies. For example, a patient with leukaemia may no longer be immunocompromised once their disease is in remission and leukocyte counts recover, or, a patient receiving biologic immunosuppressive therapy may no longer be immunocompromised once the therapy is completed and the time for immune recovery has elapsed.

Special attention is also needed regarding the effect of pregnancy and ageing on the immune system. During pregnancy, the maternal immune system adapts and changes depending on the stage of pregnancy. These adaptations make the immune system work differently than during non-pregnancy periods. Thus, pregnant people are not considered immunocompromised *per se* but should always be identified as a special population group since the body is not operating the same way as a non-pregnant person (18,19). Regarding ageing, with the beginning of the sixth decade of life, the human immune system undergoes various ageing-related changes, which continuously progress to a state of immunosenescence. Numerous confounding variables hamper a defined contribution of age-related immunologic impairment infections' pathogenesis. However, immune ageing is a specific cofactor of special interest associated with declining protective immunity, increasing incidence of inflammatory diseases, susceptibility to cancers and infections, organ failures and death (20,21). Thus, ≥ 60-year-old individuals should not be included as immunocompromised population by default, but they are an age category of special interest that should be analysed.

6. RESEARCH QUESTION and OBJECTIVES

The aim of this work is to provide a SAFETY-VAC consolidated phenotype to properly identify immunocompromised individuals in EHR databases. Therefore, the first step is to count on a comprehensive overview of the existing operational definitions. The following research question was formulated for this scoping review: What operational definitions are used to identify immunocompromised populations when conducting epidemiological studies in EHR databases?

7. RESEARCH METHODS

7.1. Study design and search strategy

A scoping literature search was conducted to identify studies in the MEDLINE database. The study period ran from inception to 15th August 2024 (the date the literature research started) and there was no language restriction Our search string included key terms related to all the

following: i) observational studies using healthcare databases, ii) immunocompromised host, immunosuppression and immunosuppressants, and iii) coding systems *(***Supplementary Table 1***)*. We also identified potentially eligible articles by reviewing the reference list of the initially selected full-text articles. We followed the Preferred Reporting Items for Scoping Reviews (PRISMA-ScR) guidelines (22). Titles and abstracts were screened independently by five reviewers. Disagreements were solved as a consensus was reached during a dedicated session. The final set of full-text studies was split into five subsets and the relevant information was extracted independently by six reviewers (JRA, CDS, NL, MS, IP, FR). Discrepancies were cross-reviewed and resolved through discussion and mutual agreement among reviewers in two reconciliation sessions.

7.2. Selection criteria

We included titles and abstracts of non-interventional studies in EHR and administrative databases focusing on immunosuppressants, immunosuppressive conditions, and immunocompromised populations.

Then, the full-text reading led to the final manuscript selection based on the availability of definitions, and diagnostic or drug codes used to retrieve immunocompromised individuals. We excluded interventional studies or studies with primary data collection, case reports and studies that did not use human health data. Publications such as reviews, letters, protocols, editorials, conference abstracts or non-peer-reviewed articles were also excluded.

7.3. Data charting

The extraction of information from full-text articles was conducted using a standardized data abstraction tool created in RedCap® (v 14.1.0), a customizable informatics systems-based web software. Main items include: i) identification of the study (first author, year, Digital Object Identifier [DOI] and title), ii) study design (cross-sectional, cohort, case-only, case-control, other, or not reported. If more than one design was used, the one used to answer the main question was selected), iii) main outcome of the study (safety, effectiveness, drug utilization, algorithm validation, predictive model, other, or not reported), and iv) study size, number of included data sources, name and type of the included data sources [administrative/claims, electronic health records, registry, not reported]. We considered a data source to be a combination of more than one type if it was reported to be constituted by linkable data banks from different provenances, like pharmacy reimbursement records linked to medical records or a disease registry), v) clinical definitions as plain text or diagnosis codes using a pre-specified dictionary, vi) therapeutic classes or individual drugs used to define immunocompromised populations (drugs were mapped to the 2024 version of the Anatomical, Therapeutic and Chemical [ATC]/Defined Daily Dose [DDD] classification from World Health Organisation [WHO]) (22), and vii) algorithms (combinations of criteria) used to operationalise the identification of immunocompromised populations. A formal quality assessment of the selected studies was not performed since it was out of the scope of this scoping review and therefore did not impact studies' final selection; the focus of this work was not on the results of the studies themselves but on the methods applied to identify immunocompromised population.

7.4. Data analysis

Once a set of medical conditions (diagnostic codes) and drugs (ATC codes) were extracted from the selected studies, two members of the core study team (CED and FR) proposed a set of categories to group the identified conditions and drugs. Four clinical specialists (LCH, LSS, MLL, DR) from Padova University Hospital, Italy, Aarhus University Hospital, Denmark, Vall d'Hebron University Hospital, Spain, and the Rheumatology Unit of the University of Verona, Italy, provided input about the proposed categories, the assignment of the medical conditions and the medicinal products to the categories, and recommended the inclusions/exclusions of new conditions or drugs not identified in the literature search. Standard descriptive analyses were performed, including counts and percentages. All data analyses were conducted using R (v 4.4.1).

8. RESULTS

8.1. General characteristics of the included studies

A total of 137 citations were identified from the search in MEDLINE. From the total, 2 studies were excluded because of not being observational, 6 because of being reviews, letters or book chapters, and 6 because they did not target humans (and used, for example, cell cultures) (see ***Figure 1***). These 56 studies were finally selected for data charting (see ***Figure 1*** and ***Table 1***). A summary of the main characteristics of the selected studies is presented in ***Table 2.***

39% of the studies (n=22) were descriptive epidemiologic studies, 25.0% (n=14) explored safety outcomes, 17.8% (n=10) applied an association design, 10.7% (n=6) were drug utilisation studies, and 3.5% (n=2) were algorithm validation studies.

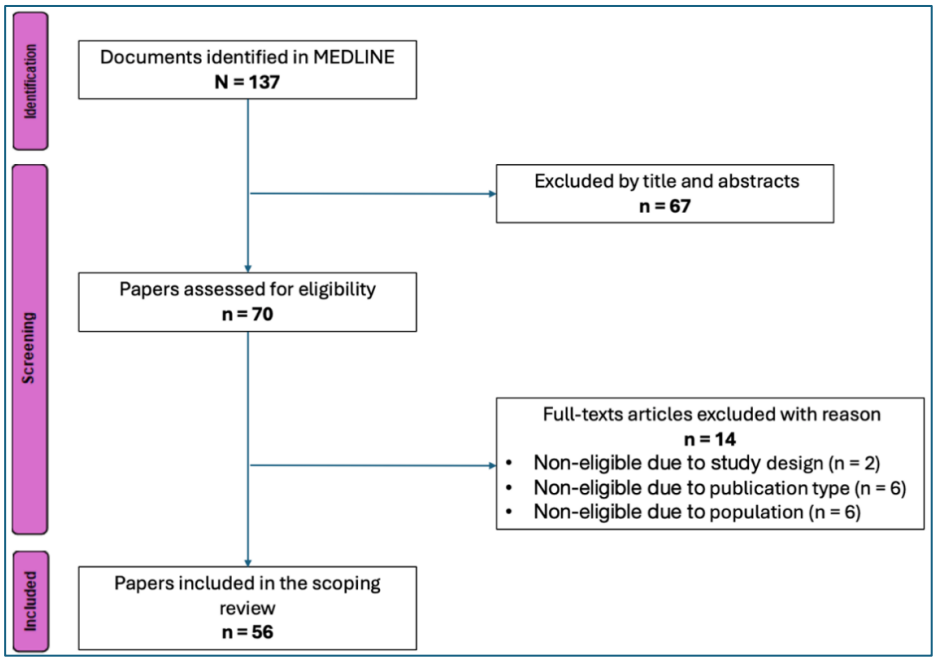

Figure 1. *Flow-chart diagram of the identified studies.*

8.2. Identification of immunocompromised populations in the literature

In 48.2% (n=27) of the selected studies, the main population of interest included immunocompromised individuals, while 51.8% (n=29) performed secondary analyses in these populations. Studies applied more than one strategy to identify immunocompromised populations in large databases: 91% (n=51) of the studies used a set of clinical conditions, 78.6% (n=44) used clinical specifications as plain text, 42.8% (n=24) used drug codes, and 12.5% (n=7) used algorithms (see ***Table 2)***. Amongst these, 14 articles (25%) used ICD-9 system as medical dictionary, followed by 13 articles (23.2%) using ICD-9-CM, 9 (16.0%) using ICD-10, and 6 (10.7%) using ICD-10-CM. 4 studies used other medical dictionaries, and 3 articles stated the use of diagnostic codes but did not report any specific vocabulary. Of the studies that used clinical conditions to identify the immunocompromised populations, 16 used previously validated codes or algorithms.

Table 3 describes 7 studies (12.5%) that used an algorithm to detect immunocompromised populations, combining diagnostic codes with requirements for prescriptions, analytical and/or imaging tests, and the number of healthcare visits. Chin-Fang et.al. (23) studied the burden of invasive fungal infections (IFI) in patients with systemic lupus erythematosus (SLE). IFIs were retrieved by the ICD-9-CM diagnoses of candidiasis, cryptococcosis, histoplasmosis, blastomycosis, aspergillosis, coccidioidomycosis, paracoccidioidomycosis, sporotrichosis, zygomycosis, other and unspecified mycoses, opportunistic mycoses, and pneumonia related to other systemic mycoses. The IFI diagnoses were validated by the evidence of in-hospital prescriptions of antifungal agents. Goldberg et.al. (24) identified end-stage liver disease (ESLD) by combining ICD-9-CM codes of cirrhosis, portal hypertension complications and hepatocellular carcinoma plus another demographic and laboratory criteria. A main limitation of this algorithm was the availability of laboratory data only on a subset (≈20%) of patients, which may be the case on several EHR databases due to the difficulty to extract laboratory data. Joly et al. (25) studied the predisposing conditions and the mortality rate of progressive multifocal leukoencephalopathy (PML). Authors formally validated an algorithm to assess the reliability of PML ICD-10 primary or secondary diagnosis codes, the presence of any predisposing immunosuppressive conditions timely related to the PML diagnosis, and a brain MRI performed 6 months before the PML diagnosis. The study was performed in a large healthcare database (SNDS) covering 98% of the total French population. The algorithm was validated against medical chart review from an University hospital. The PPV of the proposed algorithm was 90%, twice the PPV of the diagnostic code alone. An important limitation of the algorithm was the lack of exhaustive information on drug exposure, including dosing and duration of exposure. Among the selected studies, this was the only study that formally validated an identification algorithm. Lee et.al (26) studied the risk of developing severe bacterial infections in IBD patients taking tumor necrosis factor-alfa (TNFα) inhibitors. To create the study population, authors applied a validated algorithm to detect IBD patients based on ICD-9 dictionary plus the evidence of TNFαs prescription records. The strength of this study lies in the use of a previously validated algorithm for administrative claims data. As in the other algorithm-based studies described here, an important limitation is the unavailability of drug doses and duration of treatment. Yi-Jung et.al. (27) developed an algorithm to identify patients with rheumatoid arthritis by combining the ICD-9-CM code 714.0 and the use of DMARDs for more than 30 days in patients older than 20 years of age. An important limitation of this algorithm was the lack of corticosteroids/DMARDs doses. Liu et.al (28) used a validated algorithm to identify people with HIV in EHR and claims data. It combines diagnostic codes, laboratory tests results, prescription of antiretroviral therapy (ART) and the number of healthcare visits since the diagnosis date. The algorithm reported 98.9% sensitivity and 97.6%

specificity, which is a strength of this study. Finally, Burchell et.al (29) studied the causes of mortality among people with HIV in Ontario, Canada. They identified the interest population through the application of a validated algorithm combining HIV ICD-9 042-044 codes plus a minimum requirement of 3 physician claims for those codes over a period of 3 years. The algorithm reported a sensitivity of 96.2% and a specificity of 99.6%. The application of this algorithm is limited by the absence of laboratory test results, e.g. CD4 count and/or viral load, and the unavailability of information related to the antiretroviral therapy. All these algorithms were adapted into the phenotype presented in ***Box 1***.

Table 1. *Selection of articles included in the study for data charting*

author	year	Main study topic	Study countries	Num. of databases	Subjects included in the study Total	Item to define immunocompromised population		Indication whether the codes have been previously validated or the validation was performed in the current study.
						Item	Vocabularies	
Kulaylat	2017	Safety	US	1	2476	Diagnostic	ICD-9-CM	Not specified
						Medicinal product	Vocabulary not reported	-
James	2017	Drug utilization	US	1	393	Medicinal product	ATC	-
Farraj	2022	Drug utilization	US	1	283970	Diagnostic	ICD-10	Not specified
Liu	2024	Epidemiology, descriptive	US	1	42271	Algorithm	-	Validated
Singh	2022	Safety	US	1	5566	Diagnostic	ICD-9, ICD-10	Not specified
						Medicinal product	ATC	-
Perrone	2020	Drug utilization	Italy	4	41290	Diagnostic	ICD-9-CM, Exemption codes	Not specified
						Medicinal product	ATC	-
Cavanaugh	2015	Epidemiology, descriptive	US	1	1016686	Diagnostic	ICD-9-CM	Not specified
Ahlquist	2023	Safety	US	1	13611	Diagnostic	ICD-9-CM, ICD-10-CM	Not specified
Goldberg	2016	Epidemiology, descriptive and association	US	2	84530	Algorithm		Validated
Cotter	2017	Safety	US	3	937	Diagnostic	Vocabulary not reported	Validated
						Medicinal product		-
						Clinical definition		-
Abdelhay	2023	Epidemiology, descriptive	United Kingdom	1	16136	Diagnostic	ICD-9-CM, ICD-10-CM	Not specified
Liao	2018	Epidemiology, descriptive	Taiwan	1	328	Diagnostic	ICD-9-CM	Not specified
						Medicinal product	ATC	-
Cho	2020	Epidemiology, descriptive	Taiwan	1	12780	Diagnostic	ICD-9-CM	Not specified
						Medicinal product	ATC	-
Kavcic	2013	Epidemiology, descriptive	US	1	2875	Diagnostic	ICD-9-CM	Not validated
						Medicinal product	ATC	-
Bala	2016	Epidemiology, descriptive	US	1	91555	Diagnostic	ICD-9	Not specified
Chiou	2023	Epidemiology, descriptive	US	1	1191	Clinical definition	-	-
Tran	2022	Safety	US	1	224912	Diagnostic	ICD-9-CM, ICD-10-CM	Not specified
						Clinical definition	-	-
Kolbrink	2022	Epidemiology, descriptive	Germany	1	Not reported	Diagnostic	ICD-10	Not specified
						Clinical definition	-	-
Tanenbaum	2018	Epidemiology, descriptive	US	1	514572	Diagnostic	ICD-9-CM	Not specified

author	year	Main study topic	Study countries	Num. of databases	Subjects included in the study Total	Item to define immunocompromised population		Indication whether the codes have been previously validated or the validation was performed in the current study.
						Item	Vocabularies	
						Clinical definition	-	-
Santos	2015	Coding system validation	US	1	393	Diagnostic	ICD-9-CM	Validated
						Clinical definition	-	-
						Population specifications: subjects with confirmed kidney transplantation and readmission to the hospital		-
Joly	2022	Algorithm validation and Epidemiology, descriptive	France	1	584	Diagnostic	ICD-10	Validated
						Population specifications: presence of a predisposing immunosuppressive condition ICD-10 code, either in public or private hospital discharge databases	-	-
						Algorithm		Validated
George	2020	Drug Utilization	US	1	8315	Diagnostic	ICD-9, ICD-10, CSS code	Validated
						Medicinal product	ATC	-
Gregory	2019	Safety	US	1	16005	Diagnostic	ICD-9-CM, ICD-9-CMPCS, CPT-4	Not specified
						Medicinal product	ATC	-
Davy-Mendez	2021	Safety	Canada	1	6997	Diagnostic	ICD-9, ICD-9-CM	Not specified
						Medicinal product	ATC	-
						Clinical definition	-	-
Katrak	2016	Epidemiology, descriptive	US	1	377021	Diagnostic	ICD-9, LOINC	Validated
						Clinical definition	-	-
Grau	2018	Safety	US	1	4717536	Diagnostic	ICD-9-CM	Not specified
						Population specifications: people admitted due to femoral fractures	-	-
King	2012	Epidemiology, descriptive	US	1	119077	Diagnostic	ICD-9, ICD-9-CM	Not specified
						Medicinal product	Vocabulary not reported	-
Moffett	2014	Drug Utilization	US	1	466	Diagnostic	ICD-9	Validated

author	year	Main study topic	Study countries	Num. of databases	Subjects included in the study Total	Item to define immunocompromised population		Indication whether the codes have been previously validated or the validation was performed in the current study.
						Item	Vocabularies	
						Medicinal product	Vocabulary not reported	-
Asfari	2020	Epidemiology, association	US	1	30712524	Diagnostic	ICD-9	Not specified
Chakraborty	2020	Drug Utilization	US	1	27216	Medicinal product	Dictionary not reported	-
						Clinical definition	-	
Santos	2016	Epidemiology, association	US	1	7912	Diagnostic	ICD-9-CM	Validated
						Other	-	-
Chen	2018	Epidemiology, association	Taiwan	1	71650	Diagnostic	ICD-9-CM	Not specified
Chládek	2013	Safety	Czech Republic	1	49	Diagnostic	Vocabulary not reported	Not specified
						Medicinal product	ATC	-
Alqahtani	2018	Epidemiology, descriptive	US	1	1147760	Diagnostic	ICD-9-CM	Not specified
						Clinical definition	-	-
Tsao	2019	Safety	Canada	4	6218	Diagnostic	ICD-9, ICD-10	Not specified
						Medicinal product	ATC	-
						Clinical definition	-	-
Ng	2022	Epidemiology, association	Taiwan	1	319	Diagnostic	ICD-9-CM	Not specified
						Medicinal product	ATC	-
Massicotte-Azarniouch	2020	Diagnostic codes validation	Canada	1	1258	Diagnostic	ICD-10	Validated
Langley	2010	Effectiveness	Canada	2	879	Diagnostic	ICD-9-CM	Validated
						Clinical definition	-	-
Zilberberg	2014	Epidemiology, descriptive	US	1	10839	Diagnostic	ICD-9-CM	Not specified
						Medicinal product	ATC	-
						Clinical definition	-	-
Triant	2007	Epidemiology, association	US	1	2093029	Diagnostic	ICD-9-CM	Not specified
Chow	2014	Epidemiology, association	US	1	39519	Diagnostic	ICD-9-CM	Validated
Mangia	2011	Epidemiology, descriptive	Brazil	1	55370457	Diagnostic	ICD-9, ICD-10	Validated
						Medicinal product	Vocabulary not reported	-
						Clinical definition	-	-
Cammarota	2018	Epidemiology, descriptive	Italy	1	1026	Diagnostic	ICD-9-CM	Not specified
						Clinical definition	-	-
Chang	2020	Epidemiology, association	Taiwan	1	5810	Diagnostic	ICD-9-CM	Validated
						Medicinal product	ATC	-
						Clinical definition	-	-
						Algorithm		-
Rider	2019	Safety	US	1	185892	Diagnostic	ICD-9, ICD-10	Not specified

author	year	Main study topic	Study countries	Num. of databases	Subjects included in the study Total	Item to define immunocompromised population		Indication whether the codes have been previously validated or the validation was performed in the current study.
						Item	Vocabularies	
						Clinical definition	-	-
Edigin	2020	Epidemiology, descriptive	US	1	112	Diagnostic	ICD-9	Validated
Burchell	2019	Epidemiology, descriptive	Canada	6	23043	Algorithm		Validated
Kroner	2019	Epidemiology, descriptive	US	1	433805	Diagnostic	ICD-9-CM	Validated
						Clinical definition	-	-
Wright	2022	Effectiveness	US	1	9667	Diagnostic	ICD-10-CM	Not specified
Lenert	2020	Epidemiology, descriptive	US	1	636	Diagnostic	ICD-9-CM, ICD-10-CM	Not validated
Chin-Fang	2021	Epidemiology, association	Taiwan	1	269951	Diagnostic	ICD-9-CM	Validated
						Medicinal product	ATC	-
						Algorithm	-	-
Orieux	2024	Epidemiology, descriptive	France	1	222			
Tseng	2019	Safety	Taiwan	1	19603	Diagnostic	ICD-9-CM	Not specified
						Medicinal product	ATC	-
Lee	2018	Safety	US	1	10838	Diagnostic	-	
						Medicinal product	-	
						Algorithm	-	Validated
Moein	2023	Safety	US	1	96	Diagnostic	Not reported	Not specified
						Medicinal product		
Dregan	2015	Epidemiology, association	United Kingdom	1	466976	Diagnostic	Vocabulary not reported	Not specified
						Medicinal product		

US: United States

Table 2. *Summary of main characteristics of the selected studies.*

Key characteristics	**Total (n = 56)**
Number of articles by year of publication, n (%)	
Before 2015	9 (16.1)
2015-2020	32 (57.1)
After 2020	15 (26.8)
Main study topic, n (%)	
Safety	14 (25.0)
Effectiveness	2 (3.5)
Epidemiology, descriptive	22 (39.2)
Epidemiology, association	10 (17.8)
Drug utilization	6 (10.7)
Algorithm validation	2 (3.5)
Number of databases involved, median [range]	1 [1-6]
Type of database*, n (%)	
Administrative or Claims	29 (51.8)
EHR	13 (23.2)
Registries [a]	11 (19.6)
More than one type [a]	6 (10.7)
Not reported or unclear	2 (3.6)
Concepts used to define immunocompromised status*, n (%)	
Clinical conditions	51 (91.1)
Clinical specifications (free text)	44 (78.6)
Medicines´ use	24 (42.8)
Algorithms	7 (12.5)

[a]Registries included population registries (such as research patient registries, intensive care units registries or perinatal registries), disease registries (such as transplant or HIV registries) or registries with specific aims (like the Rochester Epidemiology Project).

*For these items the sum of percentages is over 100% as more than one category may apply per article.

Table 3. *Algorithms to identify immunocompromised population in the literature.*

First Author and year of publication	Manuscript title	Algorithm
Chin-Fang S. (2021) (23)	Epidemiology and risk of invasive fungal infections in systemic lupus erythematosus: a nationwide population-based cohort study.	Invasive Fungal Infections diagnoses were further validated by a record of in-hospital prescriptions of systemic antifungal agents available in Taiwan, namely fluconazole, itraconazole, posaconazole, voriconazole, amphotericin B, liposomal amphotericin B, caspofungin, micafungin, anidulafungin, and flucytosine.
Goldberg D. (2016) (24)	Patients With Hepatocellular Carcinoma Have Highest Rates of Wait-listing for Liver Transplantation Among Patients With End-Stage Liver Disease.	Significant Liver Disease (SLD) was defined using algorithms based on ICD-9-CM codes that have been validated to have positive predictive values of >85%. All patients first required a diagnosis of cirrhosis: ≥1 inpatient or ≥2 outpatient ICD-9-CM codes for cirrhosis (571.2, 571.5). Decompensated cirrhosis was then defined by having ≥1 inpatient or ≥2 outpatient ICD-9-CM codes for a complication of portal hypertension (ascites, bleeding esophageal varices, and/or spontaneous bacterial peritonitis) occurring after the diagnosis of cirrhosis. HCC required a diagnosis of cirrhosis and ≥1 inpatient or ≥2 outpatient ICD-9-CM code for HCC (ICD-9-CM code, 155.0). In the subset of patients with cirrhosis without HCC or a complication of portal hypertension, laboratory criteria was used to define hepatocellular dysfunction (calculated MELD score ≥15 and/or a total serum bilirubin ≥3 mg/dL; only available for n = 3499, 20.8% of the HealthCore cohort; available laboratory data were based on capitation to a specific laboratory and not any particular demographic). The age cutoff for inclusion was 18-75 years at ESLD diagnosis. Patients were excluded if they had an extrahepatic malignancy, excluding nonmelanoma skin cancer, diagnosed within 365 days before the ESLD index date.
Joly M. (2023) (25)	Progressive multifocal leukoencephalopathy: epidemiology and spectrum of predisposing conditions.	To assess the reliability of Progressive multifocal leukoencephalopathy (PML) diagnosis code (A81.2 'Multifocal leukoencephalopathy', under A81 'Atypical viral infections of the CNS' in the ICD-10), this algorithm considered that a patient had incident PML if he/she met the following criteria: (i) presence of the A81.2 code as primary diagnosis (PD) code and/or optional related diagnosis (RD) in the public and private hospital discharge database (PMSI); (ii) presence of a predisposing immunosuppressive condition ICD-10 code [including HIV infection, haematological malignancy, chronic inflammatory disease, solid neoplasm, solid organ transplantation (SOT) and primary immune deficiency (PID)] in the PMSI from 2 years before to 1 year following PML diagnosis; (iii) presence of a brain MRI within 6 months before PML diagnosis and in order to select patients with incident PML; and (iv) absence of the A81.2 code in the PMSI before validation study start date.

First Author and year of publication	Manuscript title	Algorithm
Lee WJ. (2018) (26)	Risk of Serious Bacterial Infection Associated With Tumor Necrosis Factor–Alpha Inhibitors in Children and Young Adults With Inflammatory Bowel Disease.	We identified patients age <30 years diagnosed with IBD (ICD-9 code 555.xx or 556.xx) between July 1, 2009, and June 30, 2013 (study period). Eligible subjects had to have ≥2 claims with an IBD diagnosis within 1 year or 1 claim with an IBD diagnosis by a pediatrician or gastroenterologist (using a validated algorithm with a PPV of 93% to 96%) (30). From this group, individuals with at least 1 prescription claim for a TNFI or immunomodulator were identified.
Yi-Jung C. (2020) (27)	Impact of Rheumatoid Arthritis on Alopecia: A Nationwide Population-Based Cohort Study in Taiwan.	Newly diagnosed RA using the ICD-9-CM code 714.0 from 2000-2012 AND use of DMARDs ≥30 days AND age ≥20 years old.
Liu Y. (2024) (28)	Comorbidity burden and health care utilization by substance use disorder patterns among people with HIV in Florida.	The algorithm screens patient's medical records and identifies people with HIV if they had at least one HIV diagnostic code plus at least one of the following: (1) had at least one positive HIV laboratory test, including HIV RNA and antigen/antibody test, (2) had been prescribed ART, or (3) had three or more visits with corresponding HIV diagnostic codes.
Burchell A. (2019) (29)	Cause-specific mortality among HIV-infected people in Ontario, 1995–2014: a population-based retrospective cohort study.	The algorithm required 3 physician claims coded for HIV infection [ICD-9] codes 042–044) over a 3-year period.

Table 4 shows the retrieved medical conditions and medicinal products organised in seven broad concepts categories. Category 1 includes 20 genetic and hereditary conditions. Category 2 contains seven main infectious diseases and related conditions such as opportunistic fungal infections. Category 3 comprises four main diagnostic concepts: haematological malignancies, solid organ malignancies and hospitalization for chemotherapy. Category 4 includes four main diagnostic concepts related to solid organ and stem cell transplantation. Category 5 groups several medical conditions that were not classified into the other categories, it includes conditions such as severe kidney and liver disease, severe malnutrition, severe burns, preterm birth, cryoglobulinemia, asplenia, and haematological neutropenia. Category 6 lists medicinal products split into two subcategories: immunosuppressants (n=14 medicinal products or therapeutic group) and medicines to treat immunosuppressive conditions (n=4 medicinal products or therapeutic groups). Additionally, category 7 lists 19 autoimmune conditions that may produce immunosuppression only under the evidence of prescription or dispensing of an immunosuppressant. The medical conditions, the medicinal products and the categories presented in ***Table 4*** were curated upon discussion and agreement among the study team members and the clinical specialists.

Table 4. *Clinical conditions and medicinal products retrieved in the literature search and curated by the core study team and clinical experts.*

Proposed category	Clinical or drug concept
1. Genetic and hereditary conditions	Immunodeficiency with predominantly antibody defects
	Combined immunodeficiencies
	Common variable immunodeficiency
	Hereditary hypogammaglobulinemia
	Selective immunoglobulin (Ig) M deficiency
	Hyper-IgM syndromes
	Antibody deficiency with near-normal immunoglobulins or with hyperimmunoglobulinemia
	Severe combined immunodeficiencies
	Adenosine deaminase (ADA) deficiency
	Nezelof's syndrome
	Purine nucleoside phosphorylase (PNP) deficiency
	Major Histocompatibility Complex (MHC) class I & II deficiency
	Activated phosphoinositide 3-kinase delta syndrome
	Wiskott-Aldrich syndrome
	Immunodeficiency with short-limbed stature
	Immunodeficiency following hereditary defective response to Epstein-Barr virus
	Hyper-IgE syndromes
	Lymphocyte function antigen-1 (LFA-1) defect
	Defects in the complement system
	Congenital asplenia
2. Infectious diseases and related conditions	Human Immunodeficiency Virus (HIV) infection
	Acquired Immune Deficiency Syndrome (AIDS)
	Immune Reconstitution Inflammatory Syndrome
	Hepatitis C virus (HCV) infection
	Cytomegalovirus (CMV) infection
	Opportunistic mycoses, including infections of the lung and invasive fungal infections (IFI)
	Progressive Multifocal Leukoencephalopathy (PML)
3. Haematological and solid organ malignancies, and related interventions	Haematologic malignancies, including myelodysplastic syndromes
	Solid organ malignancies
	Hospitalisation for chemotherapy
4. Solid organ and haematopoietic stem cell transplantation	Kidney, liver, heart and lung transplant
	Kidney, liver, heart and lung transplant rejection
	Acute and chronic graft-versus-host disease
	Haematopoietic stem cell transplantation
5. Other conditions non-classified elsewhere	End-stage kidney disease, including dialysis dependency
	Significant Liver Disease (SLD)
	Severe malnutrition
	Severe burns
	Extremely and very preterm birth
	Cryoglobulinemia
	Iatrogenic or functional asplenia
	Haematological neutropenia
6. Immunosuppressants and medicines to treat conditions leading to an immunosuppressive state	Corticosteroids for systemic use (H02)
	Cyclophosphamide (L01AA01)
	Rituximab (L01FA01)
	Azathioprine (L04AX01)

Proposed category	Clinical or drug concept
	Methotrexate (L04AX03)
	Selective immunosuppressants (L04AA)
	Tumour necrosis factor alpha (TNF-α) inhibitors (L04AB)
	Interleukin inhibitors (L04AC)
	Calcineurin inhibitors (L04AD)
	Janus-associated kinase inhibitors (L04AF)
	Monoclonal antibodies (L04AG)
	Mammalian target of rapamycin (mTOR) kinase inhibitors (L04AH)
	Complement inhibitors (L04AJ)
	Dihydroorotate dehydrogenase (DHODH) inhibitors (L04AK)
	Direct acting antivirals (J05A)*
	Sulfasalazine
	Mesalazine
	Cobicistat
7. Conditional clinical concepts to be included upon evidence of concomitant use of immunosuppressants	Acquired pure red cell aplasia
	Dermatomyositis
	Pemphigus vulgaris
	Pemphigoid
	Cystic fibrosis
	DiGeorge syndrome
	Autoimmune lymphoproliferative syndrome [ALPS]
	IgG4-related disease
	Hemophagocytic syndromes
	Crohn's disease
	Ulcerative colitis
	Systemic lupus erythematosus
	Systemic sclerosis
	Rheumatoid arthritis
	Psoriatic and enterohepatic arthropathies
	Juvenile arthritis
	Ankylosing spondylitis
	Adult-onset Still's disease
	Systemic arthritis

Italics: included upon clinical specialist recommendation.
* Except: nucleosides and nucleotides excl. reverse transcriptase inhibitors (J05AB)

9. DISCUSSION

9.1. Key Results

This scoping review aimed to provide a comprehensive overview of the existing operational definitions and produce a phenotype identification algorithm for immunocompromised populations in large EHR databases. A proper identification algorithm is fundamental to reducing potential misclassification in vaccine-related pharmacoepidemiological research where, for instance, live attenuated vaccines containing weakened forms of the pathogen may pose a risk of causing the disease in immunocompromised hosts. In contrast, this risk is minimal in those with healthy immune systems (31,32).

We have classified 56 clinical conditions and medicinal products into 7 main clinical categories: genetic and hereditary conditions, infectious diseases and related conditions, haematological malignancies, solid organ malignancies and related conditions, solid organ and haematopoietic stem cell transplantation, other conditions non-classified in other categories, immunosuppressants and medicines to treat conditions leading to an immunosuppressive state, and medical conditions to be included only upon the evidence of concomitant use of an immunosuppressant. The first category *genetic and hereditary conditions* included an important number of PIDs. The increased understanding of human immunology and genetics linked to improved laboratory techniques, mainly genomic tools, has produced an important expansion of new PID phenotypes in recent years (33). Although we are capturing an important spectrum of PIDs, we may have missed very rare or very new PIDs (34). Finally, due to the rarity of several of these diseases, the newest therapeutic alternatives, such as gene therapy or enzyme replacement therapy to treat severe combined immunodeficiencies, have not been identified in our literature search (35). Category 2 groups several infectious diseases which either lead to an immunosuppressed status *per se* (e.g., AIDS), or are an indirect indicator of an individual being immunosuppressed (e.g., opportunistic infections). Categories 3 and 4 include a broad spectrum of clinical conditions condensed as high-level phenotypes. We listed *solid organ malignancies* instead of listing all cancer-related codes identified in the literature search. Category 5 includes 8 different conditions that cannot be classified in other categories, it includes end-stage kidney disease, significant liver disease, and some external causes such as severe burns (36). Category 6 lists several individual medications and medication-classes defined as immunosuppressants. These medications can be used as stand-alone markers to identify immunocompromised populations. Finally, category 7 includes clinical conditions that may be taken into consideration only if linked to an immunosuppressant prescription or dispensing registry.

9.2. Limitations of the scoping review

First, the described search strategy for the scoping review may have some limitations. As the search query (i.e., to capture operational definitions to capture immunocompromised populations) is complex and may not be found explicitly in titles or abstracts (and more frequently in the methods and supplementary files instead), some of the terms in the search string are broad.

This meant an increased risk of retrieving a substantial volume of irrelevant records and the need for a burdening manual filtering, and the risk of missing relevant studies. For example,

alternative keywords, regional variations in terminology, or novel concepts not yet indexed in databases might not be captured. Second, the search strategy heavily relies on predefined codes, terminologies, and phrases, which might not be sensitive elsewhere. Third, we have only searched the PubMed database, so studies from other sources might have been missed. However, this scoping review in which breadth comes at the expense of depth, and some relevant references may have been missed, provides a useful and operational perspective on how the identification of immunocompromised populations is tackled in epi and pharmacoepidemiologic research. So, it is likely that more clinical concepts and algorithms not yet identified, investigated in large-scale observational epidemiologic studies, or captured by the search strategy will need to be added over time. In this sense, future research should focus on refining and testing these algorithms, particularly through the incorporation of up-to-date, dynamically collected data from relevant systems into the development and evaluation of the proposed algorithm. This could involve streams of information from EHRs, registries, or surveillance systems that capture ongoing clinical events, medication use, and health outcomes in near real-time.

10. PHENOTYPE TO IDENTIFY IMMUNOCOMPROMISED POPULATIONS IN EHR DATABASES.

To fulfil the objective of this work to produce a phenotype of immunocompromised populations for studies using EHR databases, the clinical conditions, the medicinal products and the algorithms included in ***Tables 3*** and ***Table 4*** served as building blocks of the phenotype proposed in ***Box 1***.

Briefly, once the set of medical conditions (diagnostic codes), medicinal products (ATC codes) and algorithms were extracted and curated by clinical specialists, two clinicians-epidemiologists from the core study team (JRA and EB) developed the identification algorithm. They transformed and joined the proposed concepts using Boolean terms (i.e., AND, OR, NOT). In addition to the medical conditions and the medicinal products, information on diagnostic tests, dosage information, and duration of exposure were incorporated (37–46). Moreover, in the phenotype section dedicated to extreme and very preterm newborns, additional conditions to identify bacterial, viral, fungal and opportunistic infections were added by the core team to improve the accuracy of the event identification (46, 49, 50). Finally, three core team members (CED, FR, IP) and four clinical specialists (LCH, LSS, MLL, DR) reviewed the algorithm.

Box 1. *Phenotype proposal for identifying immunocompromised populations across electronic healthcare databases.*

For children >12-year-old and adults AND the following:

Exposures [1]	*Prescription or dispensing records of:* (Systemic corticosteroids use (ATC H02) **AND** (≥ 20mg for ≥ 2 weeks, up to 3 weeks after stopping **OR** more than 6 months up to 3 weeks after stopping, independent of the dosage)) (37–39), **OR** (Antineoplastic agents (ATC L01) from the start day, up to 6 months after stopping the drug, independent of the dosage), **OR** (Selective immunosuppressants (ATC L04AA) from 1 month after the start day of the drug, up to 1 month after stopping the drug, independent of the dosage), **OR** (TNF-α (ATC L04AB) from 2 weeks after the start day of the drug, up to 2 months after stopping the drug, independent of the dosage) (40,41), **OR** (Interleukin inhibitors (ATC L04AC) from 1 month after the start day of the of drug, up to 3 months after stopping the drug, independent of the dosage), **OR** (Calcineurin inhibitors (ATC L04AD) from 1 week after the start day of the drug, up to 1 week after stopping the drug, independent of the dosage), **OR** (S1P receptor modulators (ATC L04AE) from 2 weeks after the start day of the drug, up to 2 months after stopping the drug, independent of the dosage), **OR** (JAK inhibitors (ATC L04AF) from 2 weeks after the start day of the drug, up to 1 week after stopping the drug, independent of the dosage). **OR** (Monoclonal antibodies (ATC L04AG) from 2 months after the start day of the drug, up to 3 months after stopping the drug, independent of the dosage), **OR** (mTOR kinase inhibitors (e.g., sirolimus, everolimus) (ATC L04AH) from 2 weeks of after the start day of the drug, up to 2 weeks after stopping the drug, independent of the dosage), **OR** (Complement inhibitors (e.g., eculizumab) (ATC L04AJ) from 3 days after the start day of the drug, up to 3 months after stopping the drug, independent of the dosage), **OR** DHODH inhibitors (e.g., leflunomide) (ATC L04AK), from 1 month after the start day of the drug, up to 5 months after stopping the drug, independent of the dosage), **OR** (Other immunosuppressants (methotrexate, azathioprine, thalidomide, lenalidomide, pirifenidone, pomalidomide, dardvastrocel, dimethyl fumarate and diroximel fumarate) (ATC L04AX), from 1 month after the start day of the drug, up to 1 week after stopping the treatment, independent of the dosage), **OR** (Sulfasalazine (ATC A07EC01) OR mesalazine (ATC A07EC02)) **AND** any of the diagnoses listed in category 7, **Table 4**. **OR**

[1] https://www.ema.europa.eu/en/human-regulatory-overview/marketing-authorisation/product-information-requirements

	(Direct acting antivirals (J05A)* AND any of the diagnoses listed in category 2, **Table 4**.
Diagnoses (from diagnoses included in Table 4)	(Diagnostic code for genetic and hereditary conditions (category 1, Table 4) ≤5 year prior to the study entry date), **OR** ((Diagnostic code for genetic and hereditary conditions **OR** infectious diseases and related conditions **OR** haematological and solid organ malignancies **OR** solid organ and haematopoietic stem cell transplantation **OR** other conditions not classified elsewhere) **AND** (neutropenia, lymphopenia, leukopenia **OR** any opportunistic infection, 1 month apart from the diagnostic code)). **OR** (Diagnostic code for infectious diseases and related conditions (category 2, Table 4) ≤1 year prior to the study entry date. *If HIV disease: ((Individuals with at least 1 HIV diagnostic code, **OR** 1 positive HIV test, **OR** HIV antiretroviral therapy prescription of 3-drug regimens), **AND** (at least 1 year of follow-up, **OR** ≥3 HIV-related visits)) **NOT** (2 consecutive viral load measurements of <400 copies/mL at least 30 days apart within 12 months),* **OR** (Any diagnostic code of sepsis as cause of admission to the hospital NOT present in the previous 3 months) **AND** (exposure to any drug listed in category 6, Table 4) **OR** (Diagnostic code for haematological and solid organ malignancies, and related interventions: *(if non-metastatic solid malignancies: ≤5 year prior to the study entry date **AND** exposure to any drug listed under the sub-header "immunosuppressive drugs" in category 6 in Table 4),* ***OR*** *(if metastatic organ malignancies or haematological malignancies: ≤5 year prior to the study entry date),* ***OR*** *(metastatic organ malignancies or haematological malignancies receiving any of the drugs listed in **Box 1** ±1 month apart from the diagnostic date),* ***OR*** *((any code of malignant tumour as cause of hospital admission or visit to the hospital **NOT** present in the previous year), **AND** (tumour marker, histologically malignant tumour, and other related tests within 1 month before or after the date of the visit), **OR** (biopsy diagnosis within 1 month before or after the date of the initial visit), **OR** (photographing/imaging within 1 month before or after the date of the initial visit), **OR** (surgery within 3 months after the date of the initial visit), **OR** (prescription or dispensing of antineoplastic drugs (ATC L01) within 3 months after the date of the initial visit), **OR** (radiotherapy within 6 months after the date of the initial visit))*, (42) ***OR*** ***(****≥1 procedure code for bone marrow aspirate, or organ biopsy **AND** prescription or dispensing of antineoplastic drugs (ATC L01) within 3 months after the date of the procedure)* (43). ***OR*** *(Diagnostic code for solid organ and haematopoietic stem cell transplantation ≤5 year prior to the study entry date),* ***OR***

	Diagnostic code for other conditions not classified elsewhere: (End-stage kidney disease, including dialysis dependency, **OR** haematological neutropenia, **OR** cryoglobulinemia ≤5 year prior to the study entry date), **OR** (SLD: (diagnosis of cirrhosis with ≥1 inpatient **OR** ≥2 outpatient codes, **OR** decompensated cirrhosis identified by complications like ascites, bleeding oesophageal varices, or spontaneous bacterial peritonitis, **OR** HCC identified by cirrhosis diagnosis) **AND** (≥1 inpatient **OR** ≥2 outpatient codes for HCC, **OR** hepatocellular dysfunction identified by a MELD score ≥15, **OR** total serum bilirubin ≥3 mg/dL), **OR** (Diseases related to external conditions including severe malnutrition OR severe burns, ≤1 year prior to the study entry date) **OR** (PML diagnosis **AND** an MRI exam performed within 6 months prior the PML diagnosis) (44) **OR** IFI diagnostic code **OR** record of in-hospital prescriptions or dispensing of systemic antifungal agents: fluconazole, itraconazole, posaconazole, voriconazole, amphotericin B, liposomal amphotericin B, caspofungin, micafungin, anidulafungin, and flucytosine) (45)
Diagnoses conditioned to exposures	(Acquired pure red cell aplasia, **OR** dermatomyositis, **OR** pemphigus vulgaris, **OR** pemphigoid, **OR** cystic fibrosis, **OR** DiGeorge syndrome, **OR** ALPS, **OR** IgG4-related disease, **OR** hemophagocytic syndromes, **OR** Crohn's disease, **OR** ulcerative colitis, **OR** systemic lupus erythematosus, **OR** systemic sclerosis, **OR** rheumatoid arthritis, **OR** psoriatic and enterohepatic arthropathies, **OR** juvenile arthritis, **OR** ankylosing spondylitis, **OR** adult-onset Still's disease, **OR** systemic arthritis, **OR** Listeria infection) **AND** (Prescription or dispensing records of an immunosuppressant ±1 month apart from diagnosis).

For <12 years-old individuals, everything stated above for >12-year-old individuals, AND the following:

Exposures	When assessing immunosuppression exposures, lower doses than the ones proposed for adults might be considered as sensitivity analysis.
Diagnoses	Diagnostic code for extremely and very preterm birth ≤1 year prior to the study entry date **OR** Diagnostic code for any of the following infections: Bacterial Infections: ((neonate) **AND** ((diagnostic codes of sepsis **OR** pneumonia) **AND** (diagnostic codes of *Streptococcus* Group B infection **OR** *Streptococcus* Group B positive test, **OR** diagnosis of *Escherichia coli* infection **OR** *Escherichia coli* positive test)) **OR** ((diagnostic codes of meningitis **OR** sepsis) **AND** (*Listeria* infection diagnosis **OR** *Listeria monocytogenes* positive test))) (46), **OR** Viral Infections: (neonate) **AND** ((hepatitis **OR** pneumonia **OR** neurologic damage) **AND** (CVM infection diagnosis **OR** CVM positive test)) **OR** (HSV encephalitis **OR** HSV disseminated infection), **OR** Fungal Infections: (neonate) **AND** ((Candida spp., **OR** Aspergillus spp, **OR** Cryptoccocus spp positive test) **OR** (Candida spp., **OR** Cryptoccocus spp, **OR** Aspergillus spp diagnostic code) **OR** (NICU admission)), **OR**

	Opportunistic Infections: (neonate) **AND** (pneumonia diagnostic code AND (*Pneumocystis jirovecii* positve test **OR** *Pneumocystis jirovecii* infection diagnostic code).

ALPS: autoimmune lymphoproliferative syndrome, ATC: Anatomical Therapeutic Chemical classification, CVM: cytomegalovirus, DHODH: Dihydroorotate dehydrogenase, HCC: hepatocellular carcinoma, HIV: Human immunodeficiency virus, HSV: herpes simplex virus, IFI: In-hospital Fungal Infection, JAK: Janus-associated kinase, MELD: Model for End-Stage Liver Disease score, MRI: Magnetic resonance imaging, mTOR: Mammalian target of rapamycin, NICU: Neonatal intensive care unit, PML: Progressive multifocal leukoencephalopathy, S1P: Sphingosine-1-phosphate, SLD: Significant Liver Disease, TNFα: Tumour necrosis factor alpha inhibitors

10.1. Strengths and limitations of the proposed phenotype

To the best of our knowledge and based on the findings of the scoping review, this is the first attempt to combine such a set of complex concepts (genetic diseases, congenital or acquired immunodeficiencies, acute and chronic diseases, tests, and drug use including time windows) to capture the complexity of identifying immunocompromised populations in population-based EHR data sources. Our proposed phenotype has the potential to allow for a broad mapping of related diagnostic codes, tests, and medicinal products. In this sense, it is specific but granular. Depending on the research question and populations of interest, it can be used by pooling all algorithm blocks together or selecting individual algorithms, which makes the algorithm inclusive and adaptable to a diversity of data provenances, i.e. primary care or hospital settings. Furthermore, the modular multi-criteria structure of the algorithm allows users to customize it, tailoring the algorithm to address specific research questions. Additionally, the algorithm's granularity supports specificity in identifying immunocompromised populations by capturing nuances such as corticosteroid dosages, exposure duration, and tests results.

The dynamic nature of acquired immunocompromised status challenges its exact duration identification. For instance, it may be temporary in several conditions, resolved once treatment concludes or when the disease enters remission. The proposed algorithm considers this variability by stating the time period(s) used to define the immunocompromised status. Although dose information was not considered in the algorithm, the dosage information of systemic corticosteroids was included due to the impact of dosages on producing immunosuppression. Corticosteroids can have immunosuppressive effects with dose and duration of use being crucial factors. Doses of prednisone or equivalent ≥ 20 mg/day for ≥ 2 weeks or > 40 mg/day for > 1 week are considered immunosuppressive. These doses can significantly reduce B- and T-lymphocyte subpopulations and affect immune function, potentially lasting for weeks or months after discontinuation. As the drug's dosage information is not always available in data sources, the algorithm combines the possibility to capture information from corticosteroids' dosage and time windows around the prescription or dispensing record, or only based on the exposure time; for instance, by only using the exposure to the ATC H02 starting-codes plus the exposure period, and ignoring the "≥ 20mg for ≥ 2 weeks, up to 3 weeks after stopping".

The proposed algorithm also provides insights into paediatric populations, mainly newborns. As children and newborns might be subject to specific health conditions since the immune system reaches adult maturity after 12 years of age (47,48), further specifications have been created for such populations. Immunocompromised newborns are susceptible to a wide range of pathogens, including common and opportunistic organisms (49,50). While evidence remains scarce, we have proposed an approach considering variations that this population may need,

such as specific infections, neonatal outcomes (such as preterm birth), and the need for potentially adjusting dose and time-window thresholds proposed for adults.

Our phenotype has some limitations that need to be acknowledged. Firstly, immunocompromising status can be transient, resolving as treatment concludes, or diseases enter remission. Even though we have tried to account for this variability including time-sensitive criteria, such as specifying time windows for drug exposure or diagnostic testing, capturing these transitions accurately may require more detailed data than what is often available in specific data sources, and precise duration might be misclassified. Secondly, the algorithm relies on variables not consistently recorded across databases, e.g. ICU records, adding variability to the resulting output. Thirdly, while we accounted for a broad and modular phenotype, it should be appropriately validated in different data source types to ensure its applicability. Researchers can adapt the algorithm to specific database peculiarities or validate its performance using local gold standards. Fourthly, the algorithm's stringent criteria may result in conservative identification, excluding individuals with atypical, mild or less well-documented cases of immunosuppressive conditions. On the other hand, reliance on broad coding definitions might inadvertently include immunocompetent individuals, mainly when supporting clinical features are sparse or when the end of the immunosuppressed period was not adequately captured. Also, while efforts have been made to include the most specific diagnostic codes to minimise misclassification, the potential for misclassification remains inherent in observational data. Some diagnoses are more likely to be recorded incorrectly or listed as provisional (for example, arthropathies), which could affect the algorithm's accuracy.

10.2. Recommendations

The proposed phenotype cannot be seen as a one-size-fits-all tool. As an example, in some southern European countries, oral presentations of corticosteroids are widely used as acute nebulize alternatives in paediatric populations to relieve lower respiratory tract infection symptoms. Hence, the use of systemic corticosteroids is not systemic nor chronic despite the dispensing registry potentially showing it as such. We advise researchers to carefully revise the concepts when applying them to specific data sources. We strongly recommend sensitivity analyses by modifying periods of exposure to drugs or the time gaps for diagnoses' definitions. Finally, although the algorithm has been reviewed by clinical specialists and the conditions, drugs and tests extracted from existent publications, its real-world performance remains to be fully validated across diverse healthcare systems and database types. Finally, mapping across diagnostic and medication coding systems should be incorporated so the algorithm can semantically be seamlessly translated for different databases.

For paediatric populations, although separate criteria are provided for children under 12, the algorithm may require further refinement to fully capture the unique developmental and age-related differences related to the maturity of the immune system, particularly for neonates and infants. The need for inclusion of further specifications for other subpopulations, such as pregnant or ≥ 60-year-old individuals, should be accounted for in the future. As documented elsewhere (18–21), our suggestion is to give specific attention to these two populations, which cannot be considered immunocompromised *per se* but should be studied as special subpopulations in studies referring to an immunosuppressed system.

11. OTHER INFORMATION

List of Supplementary Information/Appendices:

- **Supplementary Table 1**. Search string
- **Supplementary Table 2**. Frequency of citation of diagnoses and drugs or therapeutic groups.

12. CONCLUSION

A proper phenotype to identify immunocompromised populations is crucial in epidemiologic and pharmacoepidemiologic research. Supported by the literature search findings, we have developed a phenotype to identify immunocompromised individuals when conducting research using EHR data sources. Various clinical conditions, medicines and tests were joined using Boolean logic terms. Considering the diversity of EHR databases, the phenotype was built up by putting algorithm blocks together, allowing flexibility in implementation. A primary challenge when attempting to identify a host's immunocompromised status is the dynamic nature of secondary immunodeficiencies since several conditions may be transient, for instance, once the disease enters remission or the immunosuppressive treatment stops. The phenotype presented here deals with this challenge by adding the time period(s) most likely to define the immunocompromised status. Another critical challenge in correctly identifying immunocompromised individuals is the dosage information of the immunosuppressants. We have included a recommendation to deal with this challenge by adding dose and duration recommendations for systemic corticosteroids, a therapeutic group widely used among this population. For the remaining drugs, we have added the immunosuppressive duration of exposure to these drugs but not the dosage information due to the complexity of giving accurate recommendations. The latter becomes a limitation of the phenotype. Another limitation is that several algorithm components are based on variables seldomly recorded in databases, e.g. ICU records, possibly adding variability to the results. Finally, future research should formally test the phenotype across databases with sources of different origins.

13. REFERENCES

1. Salmon DA, Lambert PH, Nohynek HM, Gee J, Parashar UD, Tate JE, et al. Novel vaccine safety issues and areas that would benefit from further research. BMJ Glob Health. 2021 May;6(Suppl 2):e003814.

2. Ricotta EE, Rid A, Cohen IG, Evans NG. Observational studies must be reformed before the next pandemic. Nat Med 2023 298. 2023 Jun;29(8):1903–5.

3. Vaccine Monitoring Platform | European Medicines Agency (EMA) [Internet]. 2024 [cited 2024 Dec 12]. Available from: https://www.ema.europa.eu/en/about-us/what-we-do/crisis-preparedness-management/vaccine-monitoring-platform

4. SAFETY-VAC: Network of Data Sources for Vaccine Safety Evaluation | HMA-EMA Catalogues of real-world data sources and studies [Internet]. [cited 2024 Nov 25]. Available from: https://catalogues.ema.europa.eu/node/4094/administrative-details

5. Durán CE, Gini R, Davide M, Riefolo F, Riera J, Hoxhaj V, et al. SAFETY-VAC study: A framework for the post-authorization SAFETY monitoring and evaluation of the VACcines in Europe. [Internet]. Zenodo; 2024 Aug [cited 2024 Oct 21]. Available from: https://zenodo.org/doi/10.5281/zenodo.13384860

6. SAFETY-VAC: Background incidence estimation of flares of pre-existing chronic diseases using pan-European electronic healthcare data sources. (SAFETY VAC) | HMA-EMA Catalogues of real-world data sources and studies [Internet]. [cited 2024 Nov 25]. Available from: https://catalogues.ema.europa.eu/node/4134/administrative-details

7. Bertini CD, Khawaja F, Sheshadri A. Coronavirus Disease-2019 in the Immunocompromised Host. Clin Chest Med. 2023 Jun;44(2):395–406.

8. Evans RA, Dube S, Lu Y, Yates M, Arnetorp S, Barnes E, et al. Corrigendum to ‘Impact of COVID-19 on immunocompromised populations during the Omicron era: insights from the observational population-based INFORM study’ [The Lancet Regional Health – Europe 35 (2023) 100747]. Lancet Reg Health - Eur. 2024 Sep;44:101008.

9. Patel M, Chen J, Kim S, Garg S, Flannery B, Haddadin Z, et al. Analysis of MarketScan Data for Immunosuppressive Conditions and Hospitalizations for Acute Respiratory Illness, United States. Emerg Infect Dis. 2020 Aug;26(8):1720–30.

10. Davy-Mendez T, Napravnik S, Eron JJ, Cole SR, Van Duin D, Wohl DA, et al. Current and Past Immunodeficiency Are Associated With Higher Hospitalization Rates Among Persons on Virologically Suppressive Antiretroviral Therapy for up to 11 Years. J Infect Dis. 2021 Aug 16;224(4):657–66.

11. Bender DA, Heilbroner SP, Wang TJC, Shu CA, Hyde B, Spina C, et al. Increased rates of immunosuppressive treatment and hospitalization after checkpoint inhibitor therapy in cancer patients with autoimmune disease. J Immunother Cancer. 2020 Dec;8(2):e001627.

12. Bender DA, Spina C, Heilbroner SP, Xanthopoulos E, Wang TJC, Reuscher P, et al. Rates of immunosuppressive treatment and hospitalization after checkpoint inhibitor

therapy in melanoma and lung cancer patients with autoimmune diseases. J Clin Oncol. 2019 May 20;37(15_suppl):e14140–e14140.

13. Moreau AS, Martin-Loeches I, Povoa P, Salluh J, Rodriguez A, Thille AW, et al. Impact of immunosuppression on incidence, aetiology and outcome of ventilator-associated lower respiratory tract infections. Eur Respir J [Internet]. 2018 Feb 8 [cited 2024 Nov 12]; Available from: https://publications.ersnet.org/content/erj/early/2018/01/18/13993003.01656-2017

14. Kreitmann L, Helms J, Martin-Loeches I, Salluh J, Poulakou G, Pène F, et al. ICU-acquired infections in immunocompromised patients. Intensive Care Med. 2024 Mar 1;50(3):332–49.

15. Addendum to the Guideline on clinical development of vaccines to address clinical trials in immunocompromised individuals.

16. CDC. Centers for Disease Control and Prevention. 2020 [cited 2024 Nov 25]. COVID-19 and Your Health. Available from: https://archive.cdc.gov/www_cdc_gov/coronavirus/2019-ncov/need-extra-precautions/people-who-are-immunocompromised.html

17. Antinori A, Bausch-Jurken M. The Burden of COVID-19 in the Immunocompromised Patient: Implications for Vaccination and Needs for the Future. J Infect Dis. 2023 Aug 1;228(Supplement_1):S4–12.

18. Mor G, Cardenas I. REVIEW ARTICLE: The Immune System in Pregnancy: A Unique Complexity. Am J Reprod Immunol. 2010;63(6):425–33.

19. Vale AJM, Fernandes ACL, Guzen FP, Pinheiro FI, de Azevedo EP, Cobucci RN. Susceptibility to COVID-19 in Pregnancy, Labor, and Postpartum Period: Immune System, Vertical Transmission, and Breastfeeding. Front Glob Womens Health [Internet]. 2021 Feb 17 [cited 2024 Nov 25];2. Available from: https://www.frontiersin.org/journals/global-womens-health/articles/10.3389/fgwh.2021.602572/full

20. Weyand CM, Goronzy JJ. Aging of the Immune System. Mechanisms and Therapeutic Targets. Ann Am Thorac Soc. 2016 Dec;13(Supplement_5):S422–8.

21. Saltzman RL, Peterson PK. Immunodeficiency of the Elderly. Rev Infect Dis. 1987 Nov 1;9(6):1127–39.

22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467–73.

23. Su CF, Lai CC, Li TH, Chang YF, Lin YT, Chen WS, et al. Epidemiology and risk of invasive fungal infections in systemic lupus erythematosus: a nationwide population-based cohort study. Ther Adv Musculoskelet Dis. 2021 Jan 1;13:1759720X211058502.

24. Goldberg D, French B, Newcomb C, Liu Q, Sahota G, Wallace AE, et al. Patients With Hepatocellular Carcinoma Have Highest Rates of Wait-listing for Liver Transplantation

Among Patients With End-Stage Liver Disease. Clin Gastroenterol Hepatol. 2016 Nov 1;14(11):1638-1646.e2.

25. Joly M, Conte C, Cazanave C, Le Moing V, Tattevin P, Delobel P, et al. Progressive multifocal leukoencephalopathy: epidemiology and spectrum of predisposing conditions. Brain. 2023 Jan 4;146(1):349–58.

26. Lee WJ, Lee TA, Calip GS, Suda KJ, Briars L, Schumock GT. Risk of Serious Bacterial Infection Associated With Tumor Necrosis Factor–Alpha Inhibitors in Children and Young Adults With Inflammatory Bowel Disease. Inflamm Bowel Dis. 2018 Mar 19;24(4):883–91.

27. Chang YJ, Lee YH, Leong PY, Wang YH, Wei JCC. Impact of Rheumatoid Arthritis on Alopecia: A Nationwide Population-Based Cohort Study in Taiwan. Front Med [Internet]. 2020 Apr 28 [cited 2024 Nov 25];7. Available from: https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2020.00150/full

28. Liu Y, Manavalan P, Siddiqi K, Cook RL, Prosperi M. Comorbidity Burden and Health Care Utilization by Substance use Disorder Patterns among People with HIV in Florida. AIDS Behav. 2024 Jul 1;28(7):2286–95.

29. Burchell AN, Raboud J, Donelle J, Loutfy MR, Rourke SB, Rogers T, et al. Cause-specific mortality among HIV-infected people in Ontario, 1995–2014: a population-based retrospective cohort study. CMAJ Open. 2019 Jan 1;7(1):E1–7.

30. Liu L, Allison JE, Herrinton LJ. Validity of computerized diagnoses, procedures, and drugs for inflammatory bowel disease in a northern California managed care organization. Pharmacoepidemiol Drug Saf. 2009;18(11):1086–93.

31. Alnaimat F, Sweis JJG, Jansz J, Modi Z, Prasad S, AbuHelal A, et al. Vaccination in the Era of Immunosuppression. Vaccines. 2023 Sep 1;11(9):1446.

32. Pollard AJ, Bijker EM. A guide to vaccinology: from basic principles to new developments. Nat Rev Immunol. 2021 Feb;21(2):83–100.

33. Milner JD, Holland SM. The cup runneth over: lessons from the ever-expanding pool of primary immunodeficiency diseases. Nat Rev Immunol. 2013 Sep;13(9):635–48.

34. Parvaneh N, Casanova JL, Notarangelo LD, Conley ME. Primary immunodeficiencies: A rapidly evolving story. J Allergy Clin Immunol. 2013 Feb;131(2):314–23.

35. Wadbudhe AM, Meshram RJ, Tidke SC. Severe Combined Immunodeficiency (SCID) and Its New Treatment Modalities. Cureus [Internet]. 2023 Oct 26 [cited 2024 Oct 21]; Available from: https://www.cureus.com/articles/184143-severe-combined-immunodeficiency-scid-and-its-new-treatment-modalities

36. Moins-Teisserenc H, Cordeiro DJ, Audigier V, Ressaire Q, Benyamina M, Lambert J, et al. Severe Altered Immune Status After Burn Injury Is Associated With Bacterial Infection and Septic Shock. Front Immunol. 2021 Mar 2;12:586195.

37. Fan PT, Yu DT, Clements PJ, Fowlston S, Eisman J, Bluestone R. Effect of corticosteroids on the human immune response: comparison of one and three daily 1 gm intravenous pulses of methylprednisolone. J Lab Clin Med. 1978 Apr;91(4):625–34.

38. Zwar NA. Travel and immunosuppressant medication. Aust J Gen Pract. 2020 Mar 1;49(3):88–92.

39. Rubin LG, Levin MJ, Ljungman P, Davies EG, Avery R, Tomblyn M, et al. Executive Summary: 2013 IDSA Clinical Practice Guideline for Vaccination of the Immunocompromised Host. Clin Infect Dis. 2014 Feb 1;58(3):309–18.

40. De La Torre I, Valor L, Nieto JC, Montoro M, Carreño L. Minimum Effective Dosages of Anti-TNF in Rheumatoid Arthritis: A Cross-sectional Study. Reumatol Clínica. 2014 Mar;10(2):101–4.

41. Curtis JR, Patkar N, Xie A, Martin C, Allison JJ, Saag M, et al. Risk of serious bacterial infections among rheumatoid arthritis patients exposed to tumor necrosis factor α antagonists. Arthritis Rheum. 2007 Apr;56(4):1125–33.

42. Nishikawa A, Yoshinaga E, Nakamura M, Suzuki M, Kido K, Tsujimoto N, et al. Validation Study of Algorithms to Identify Malignant Tumors and Serious Infections in a Japanese Administrative Healthcare Database. Ann Clin Epidemiol. 2022;4(1):20–31.

43. Brandenburg NA, Phillips S, Wells KE, Woodcroft KJ, Amend KL, Enger C, et al. Validating an algorithm for multiple myeloma based on administrative data using a SEER tumor registry and medical record review. Pharmacoepidemiol Drug Saf. 2019 Feb;28(2):256–63.

44. Joly M, Conte C, Cazanave C, Le Moing V, Tattevin P, Delobel P, et al. Progressive multifocal leukoencephalopathy: epidemiology and spectrum of predisposing conditions. Brain. 2023 Jan 5;146(1):349–58.

45. Su CF, Lai CC, Li TH, Chang YF, Lin YT, Chen WS, et al. Epidemiology and risk of invasive fungal infections in systemic lupus erythematosus: a nationwide population-based cohort study. Ther Adv Musculoskelet Dis. 2021 Jan;13:1759720X211058502.

46. Maródi L. Neonatal Innate Immunity to Infectious Agents. Infect Immun. 2006 Apr;74(4):1999–2006.

47. Moraes-Pinto MI de, Suano-Souza F, Aranda CS. Immune system: development and acquisition of immunological competence. J Pediatr (Rio J). 2021 Mar 1;97:S59–66.

48. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. Proc R Soc B Biol Sci. 2015 Dec 22;282(1821):20143085.

49. Sampah MES, Hackam DJ. Dysregulated Mucosal Immunity and Associated Pathogeneses in Preterm Neonates. Front Immunol. 2020 May 15;11:899.

50. Ng PC, Fok TF. Infections in the neonate. Curr Opin Infect Dis. 1996 Jun;9(3):181–6.

14. SUPPLEMENTARY INFORMATION (Appendices)

Supplementary Table 1. Search strings.

#1	diagnostic codes[All Fields] OR "ATC"[All Fields] OR "coding system"[All Fields] OR "standard code"[All Fields] OR "medical vocabulary"[All Fields] OR "medical dictionary"[All Fields] OR "International Classification of Diseases"[All Fields] OR (("anatomic"[All Fields] OR "anatomical"[All Fields] OR "anatomically"[All Fields]) AND ("therapeutical"[All Fields] OR "therapeutically"[All Fields] OR "therapeuticals"[All Fields] OR "therapeutics"[MeSH Terms] OR "therapeutics"[All Fields] OR "therapeutic"[All Fields]) AND ("chemical"[All Fields] OR "chemical s"[All Fields] OR "chemically"[All Fields] OR "chemicals"[All Fields]) AND "ATC"[All Fields] AND ("clinical coding"[MeSH Terms] OR ("clinical"[All Fields] AND "coding"[All Fields]) OR "clinical coding"[All Fields] OR "code"[All Fields]) AND ("system"[All Fields] OR "system s"[All Fields] OR "systems"[All Fields])) OR "atc code*"[All Fields] OR "anatomical therapeutic chemical code"[All Fields] OR "medicin* code"[All Fields] OR "drug code"[All Fields] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "systematized nomenclature of medicine"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "vocabulary, controlled"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "International Classification of Diseases"[MeSH Terms] OR "unified medical language system"[MeSH Terms] OR "unified medical language system"[MeSH Terms] OR "logical observation identifiers names and codes"[MeSH Terms] OR "standardized nursing terminology"[MeSH Terms] OR "standardized nursing	73,540

	terminology"[MeSH Terms] OR "standardized nursing terminology"[MeSH Terms] OR "standardized nursing terminology"[MeSH Terms] OR "diagnostic and statistical manual of mental disorders"[MeSH Terms] OR "diagnostic and statistical manual of mental disorders"[MeSH Terms] OR "diagnostic and statistical manual of mental disorders"[MeSH Terms] OR "diagnostic and statistical manual of mental disorders"[MeSH Terms] OR "current procedural terminology"[MeSH Terms] OR "current procedural terminology"[MeSH Terms] OR "current procedural terminology"[MeSH Terms] OR "current procedural terminology"[MeSH Terms] OR "current procedural terminology"[MeSH Terms] OR "current procedural terminology"[MeSH Terms]	
#2	"Immunocompromised Host"[Mesh] OR "Immunosup*"[Title/Abstract] OR "Immunosuppressive Agents"[Mesh] OR "L04*"[Title/Abstract] OR "Immunosuppression Therapy"[Mesh] OR "Immunocompr*"[Title/Abstract] OR "Immunosup*"[All fields] OR "corticoster*"[Title/Abstract] OR "antineoplas*"[Title/Abstract] OR "chemother"[Title/Abstract] OR "antiretrovir*"[Title/Abstract] OR "AIDS"[Title/Abstract] OR "HIV"[Title/Abstract] OR "transplant*"[Title/Abstract] OR "trasplant*"[Title/Abstract] OR "immunodef*"[Title/Abstract]	1,472,526
#3	"epidemiolog*"[MeSH Terms] OR "Pharmacoepidemiology"[MeSH Terms] OR "retrospective"[Title/Abstract] OR "cohort*"[Title/Abstract] OR "longitudinal studies"[Title/Abstract] OR "cross sectional"[Title/Abstract] OR "cross-sectional"[Title/Abstract] OR "pharmacoepidemiologic"[Title/Abstract] OR "pharmacoepidemiological"[Title/Abstract] OR "case-control"[Title/Abstract] OR "case control"[Title/Abstract] OR "case-crossover"[Title/Abstract] OR "case crossover"[Title/Abstract] OR "case time-control"[Title/Abstract] OR "case-time-control"[Title/Abstract] OR "case-time control"[Title/Abstract] OR "self-controlled case series"[Title/Abstract] OR "self controlled case series"[Title/Abstract] OR "self-controlled risk interval"[Title/Abstract] OR "SCRI"[Title/Abstract] OR "SCCS"[Title/Abstract] OR "times series"[Title/Abstract] OR "new user active comparator design"[Title/Abstract] OR "new-user active comparator design"[Title/Abstract]	2,210,208
#4	"database*"[Title/Abstract] OR "Databases"[Title/Abstract] OR "data bases"[Title/Abstract] OR "computerized data"[Title/Abstract] OR "administrative claims"[Title/Abstract] OR "administrative data"[Title/Abstract] OR "claims data*"[Title/Abstract] OR "healthcare records"[Title/Abstract] OR "health records"[Title/Abstract] OR "electronic health records"[Title/Abstract] OR "data bases"[Title/Abstract] OR "databases pharmaceutical"[Title/Abstract] OR "electronic healthcare database"[Title/Abstract] OR "healthcare databases"[Title/Abstract] OR "electronic health records"[Title/Abstract] OR "drug utilisation database"[Title/Abstract] OR "drug use database"[Title/Abstract] OR "Databases, Factual"[Mesh] OR "multiple databases"[Title/Abstract] OR "databases"[Title/Abstract] OR "multi-database"[Title/Abstract] OR "Multi-Database"[Title/Abstract] OR "multidatabase"[Title/Abstract] OR "multi-database"[Title/Abstract] OR "multi-source"[Title/Abstract] OR "multi cent*"[Title/Abstract] OR "multinational"[Title/Abstract] OR "multi-cohort"[Title/Abstract] OR "multi-site"[Title/Abstract] OR "multiple sites"[Title/Abstract] OR "distributed data*"[Title/Abstract] OR "distributed network"[Title/Abstract] OR "distributed data network"[Title/Abstract] OR "database network"[Title/Abstract] OR "data network"[Title/Abstract] OR "research network"[Title/Abstract] OR "safety network"[Title/Abstract] OR "MDBS"[Title/Abstract] OR "MDPES"[Title/Abstract] OR "cohorts"[Title/Abstract]	1,070,552

#5	"clinical trials"[Title/Abstract] OR "pre-clinical"[Title/Abstract] OR "in vitro"[Title/Abstract] OR "preclinical"[Title/Abstract] OR "Phase I"[Title/Abstract] OR "Phase II"[Title/Abstract] OR "Phase III"[Title/Abstract] OR "Phase 1"[Title/Abstract] OR "Phase 2"[Title/Abstract] OR "Phase 3"[Title/Abstract] OR "in-vitro"[Title/Abstract] OR "in silico"[Title/Abstract] OR "double-blind"[Title/Abstract] OR "placebo-controlled"[Title/Abstract] OR "single centre"[Title] OR "single center"[Title] OR "Single-Centre"[Title] OR "Single-Center"[Title] OR "pilot trial"[Title/Abstract] OR "randomized controlled trial"[Title/Abstract] OR "randomized controlled trials"[Title/Abstract] OR "randomised controlled trial"[Title/Abstract] OR "randomised controlled trials"[Title/Abstract] OR "clinical trial"[Title/Abstract] OR "controlled clinical trial"[Title/Abstract] OR "controlled trial"[Title/Abstract] OR "randomized clinical trial"[Title/Abstract] OR "randomised clinical trial"[Title/Abstract] OR "animal"[Title/Abstract] OR "RCT"[Title/Abstract] OR "experimental"[Title/Abstract] OR "cell"[Title/Abstract] OR "celular"[Title/Abstract] OR "clinical practice guideline"[Title/Abstract] OR "case series"[Title/Abstract] OR "systematic review"[Title/Abstract] OR "systemic review"[Title/Abstract] OR "literature review"[Title/Abstract] OR "narrative review"[Title/Abstract] OR "scoping review"[Title/Abstract] OR "documentary search"[Title/Abstract] OR "Editorial"[Publication Type] OR "Review"[Publication Type] OR "Practice Guideline"[Publication Type] OR "Published Erratum"[Publication Type] OR "surveys and questionnaires"[MeSH Terms] OR "data collection"[MeSH Terms] OR "protocol"[Title] OR "guideline"[Title] OR "case report"[Title/Abstract] OR "case reports"[Title/Abstract] OR "Case Reports"[Publication Type] OR "Validation Study"[Publication Type] OR "Consensus Development Conference"[Publication Type] OR "Clinical Conference"[Publication Type] OR "Congress" [Publication Type] OR "Congress"[Publication Type] OR "Abstracts"[Publication Type] OR "Book Review"[Publication Type] OR "Guideline"[Publication Type] OR "Meeting Abstract"[Publication Type] OR "News"[Publication Type] OR "Monograph"[Publication Type] OR "Letter"[Publication Type] OR "Review"[Publication Type] OR "randomized controlled trial"[Publication Type] OR "case reports"[Publication Type] OR "editorial"[Publication Type] OR "review"[Publication Type] OR "commentaries"[Publication Type] OR "systematic review"[Publication Type] OR "book review"[Publication Type] OR "primary collection"[All fields] OR "*omic*"[Title/Abstract] OR "*econom*"[Title/Abstract] OR "*pharmacoeconom*"[Title/Abstract] OR "Economics, Pharmaceutical"[Mesh] OR "Expert Opinion"[Mesh] OR "Expert Opinions"[Mesh] OR "Health Care Economics and Organizations"[Mesh]	17,539,666
	(#1 AND #2 AND #3 AND #4) NOT #5	137

Supplementary Table 2. *Frequency of citation of diagnoses and drugs or therapeutic groups*

Num.	Clinical and drug entities	Num. of studies citing the entity (%)	References
1	HIV infection or AIDS	10 (17.8%)	10.1016/j.jse.2016.02.033, 10.1016/j.jvs.2021.09.034, 10.1093/ofid/ofw173, 10.1212/WNL.0000000000000958, 10.9778/cmajo.20180159, 10.1097/BOT.0000000000001286, 10.1210/jc.2006-2190, 10.2147/CEOR.S162625, 10.1016/j.lanepe.2022.100400, 10.1007/s10461-024-04325-y
2	Kidney, liver, heart and lung transplant status or rejection	8 (14.3%)	10.1016/j.arth.2023.05.028, 10.1016/j.lanepe.2022.100400, 10.1177/2054358120977390, 10.1016/j.arth.2015.09.003, 10.1016/j.jpeds.2012.11.038, /10.1016/j.mayocpiqo.2019.03.006, 10.1111/pace.13498, 10.1016/j.transproceed.2015.04.087
3	Methotrexate (L04AX03)	7 (12.5%)	10.1016/j.jdin.2020.05.002, 10.1093/ecco-jcc/jjy148, 10.1136/rmdopen-2022-002343, 10.3389/fmed.2020.00150, 10.1007/s40744-020-00218-3, 10.3233/jad-150171, 10.1093/rheumatology/kez622,
4	Corticosteroids for systemic use (H02)	6 (10.7%)	10.1007/s40744-020-00218-3, 10.3233/jad-150171, 10.3389/fmed.2020.00150, 10.1093/rheumatology/kez622, /10.1186/s13054-023-04774-2, 10.1002/jgh3.12841
5	Selective immunosuppressants (L04AA)	6 (10.7%)	10.1093/ecco-jcc/jjy148, 10.1136/rmdopen-2022-002343, 10.6002/ect.2023.0137, 10.1007/s10620-021-07073-4, 10.1007/s40744-020-00218-3, 10.1016/j.jcma.2018.04.003
6	TNF-α inhibitors (L04AB)	6 (10.7%)	10.1093/ibd/izx080, 10.1136/bmjopen-2018-023714, 10.1136/rmdopen-2022-002343, 10.1007/s10620-021-07073-4, 10.1007/s40744-020-00218-3, 10.1093/ecco-jcc/jjy148,
7	Calcineurin inhibitors (L04AD)	6 (10.7%)	10.1007/s40744-020-00218-3, 10.1016/j.jcma.2018.04.003, 10.1136/rmdopen-2022-002343, 10.3233/jad-150171, 10.6002/ect.2023.0137 /10.1093/rheumatology/kez622
8	End-stage kidney disease, including dialysis dependency	5 (8.9%)	10.1016/j.arth.2015.09.003, 10.1016/j.arth.2023.05.028, /10.1186/s13054-023-04774-2, 10.1016/j.lanepe.2022.100400, 10.1111/pace.13498
9	Azathioprine (L04AX01)	4 (7.1%)	10.1093/ecco-jcc/jjy148, 10.1093/ibd/izx080, 10.3233/jad-150171, /10.1093/rheumatology/kez622
10	Rituximab (L01FA01)	3 (5.4%)	10.1016/j.jpeds.2012.11.038, 10.1136/rmdopen-2022-002343, 10.1007/s40744-020-00218-3
11	Solid organ malignancies	3 (5.4%)	10.1016/j.lanepe.2022.100400, 10.1016/j.jpeds.2012.11.038, /10.1186/s13054-023-04774-2
12	Haematologic malignancies, including myelodysplastic syndromes	3 (5.4%)	10.1016/j.lanepe.2022.100400, 10.1016/j.jpeds.2012.11.038, /10.1186/s13054-023-04774-2
13	Hepatitis C virus (HCV) infection	3 (5.4%)	10.1093/ofid/ofw173, 10.1097/BOT.0000000000001286, 10.2147/CEOR.S162625
14	Combined immunodeficiencies	2 (3.6%)	10.1016/S2213-2600(22)00042-X, [#]
15	Cytomegalovirus (CMV) infection	2 (3.6%)	10.1016/j.transproceed.2015.04.087, 10.1016/j.jvs.2021.09.034
16	Opportunistic mycoses, including infections of the lung and invasive fungal infections (IFI)	2 (3.6%)	10.1177%2F1759720X211058502, 10.1016/j.jvs.2021.09.034
17	Interleukin inhibitors (L04AC)	2 (3.6%)	10.1136/rmdopen-2022-002343, 10.1007/s40744-020-00218-3
18	Immunodeficiency with predominantly antibody defects	1 (1.8%)	10.1016/S2213-2600(22)00042-X
19	Hereditary hypogammaglobulinemia	1 (1.8%)	10.1016/S2213-2600(22)00042-X
20	Common variable immunodeficiency	1 (1.8%)	10.1016/S2213-2600(22)00042-X

21	Selective immunoglobulin (Ig) M deficiency	1 (1.8%)	10.1016/S2213-2600(22)00042-X
22	Hyper-IgM syndromes	1 (1.8%)	10.1016/S2213-2600(22)00042-X
23	Antibody deficiency with near-normal immunoglobulins or with hyperimmunoglobulinemia	1 (1.8%)	10.1016/S2213-2600(22)00042-X
24	Severe combined immunodeficiencies	1 (1.8%)	10.1016/S2213-2600(22)00042-X
25	ADA deficiency	1 (1.8%)	10.1016/S2213-2600(22)00042-X
26	Nezelofs syndrome	1 (1.8%)	10.1016/S2213-2600(22)00042-X
27	PNP deficiency	1 (1.8%)	10.1016/S2213-2600(22)00042-X
28	MHC class I & II deficiency	1 (1.8%)	10.1016/S2213-2600(22)00042-X
29	Activated phosphoinositide 3-kinase delta syndrome	1 (1.8%)	10.1016/S2213-2600(22)00042-X
30	Wiskott-Aldrich syndrome	1 (1.8%)	10.1016/S2213-2600(22)00042-X
31	Immunodeficiency with short-limbed stature	1 (1.8%)	10.1016/S2213-2600(22)00042-X
32	Immunodeficiency following hereditary defective response to Epstein-Barr virus	1 (1.8%)	10.1016/S2213-2600(22)00042-X
33	Hyper-IgE syndromes	1 (1.8%)	10.1016/S2213-2600(22)00042-X
34	LFA-1 defect	1 (1.8%)	10.1016/S2213-2600(22)00042-X
35	Defects in the complement system	1 (1.8%)	10.1016/S2213-2600(22)00042-X
36	Sickle cell disease	1 (1.8%)	10.1093/infdis/jiaa786
37	Immune Reconstitution Inflammatory Syndrome	1 (1.8%)	10.1016/S2213-2600(22)00042-X
38	PML	1 (1.8%)	10.1093/brain/awac237
39	Hospitalisation for chemotherapy	1 (1.8%)	10.1093/infdis/jiaa786
40	Haematopoietic stem cell transplantation	1 (1.8%)	10.1016/j.lanepe.2022.100400
41	Acute and chronic graft-versus-host disease	1 (1.8%)	10.1016/S2213-2600(22)00042-X
42	SLD	1 (1.8%)	10.1016/j.cgh.2016.06.019
43	Cryoglobulinemia	1 (1.8%)	10.1016/S2213-2600(22)00042-X
44	Haematological neutropenia	1 (1.8%)	10.1093/infdis/jiaa786
45	Cyclophosphamide (L01AA01)	1 (1.8%)	10.1136/rmdopen-2022-002343
46	Janus-associated kinase inhibitors (L04AF)	1 (1.8%)	10.1007/s40744-020-00218-3
47	Monoclonal antibodies (L04AG)	1 (1.8%)	10.1007/s10620-021-07073-4
48	DHODH inhibitors (L04AK)	1 (1.8%)	10.1093/rheumatology/kez622

49	Inflammatory Bowel Disease	1 (1.8%)	10.1016/j.spinee.2017.11.007

ADA: Adenosine deaminase, AIDS: acquired immunodeficiency syndrome, DHODH: Dihydroorotate dehydrogenase, HIV: Human Immunodeficiency Virus, LFA-1: Lymphocyte function antigen-1, MHC: Major Histocompatibility Complex, PML: Progressive Multifocal Leukoencephalopathy, PNP: Purine nucleoside phosphorylase, SLD: Significant Liver Disease, TNF-α: Tumor necrosis factor alpha

Supplementary Table 2 presents the frequency of citation of different clinical conditions and medicinal products retrieved in the studies. The diagnostic codes most frequently cited were HIV infection- or AIDS-related, accounting for 9 of the 56 selected studies (16%). It was followed by codes related to solid organ transplantation status or rejection in 8 studies (14.3%). Medicinal product codes of methotrexate, corticosteroids, selective immunosuppressants, TNF-alfa inhibitors, and calcineurin inhibitors were most frequently retrieved between 6 (10.7%) and 7 (12.5%) times. Among the 10 most cited entities, 7 corresponded to medicinal products.