SAFETY-VAC

A framework for the post-authorization safety monitoring and evaluation of the vaccines in Europe.

Study protocol for the occurrence of flares of Graves´ disease, Hashimoto´s thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, systemic lupus erythematosus, and ulcerative colitis using electronic healthcare data sources.

Protocol V3.0
September 16, 2024

Title	SAFETY-VAC **A framework for the post-authorisation safety monitoring and evaluation of vaccines in the EU**
Protocol version identifier	3.0
Date of last version of protocol	September 16, 2024
EU PAS register number	
Active substance	*NA*
Medicinal product	*NA*
Product reference	*NA*
Procedure number	*NA*
Marketing authorisation holder(s)	*NA*
Research question and objectives	The SAFETY-VAC project assesses the feasibility of participating data sources to participate in vaccine safety studies using electronic healthcare databases in European countries. The primary objective is this specific study is i) to estimate the incidence rates and 6- and 12-month risks of flares of Graves´ disease, Hashimoto´s thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, lupus erythematosus and ulcerative colitis using electronic healthcare data sources. The secondary objectives are i) to assess the contribution of different provenances of data to the incidence rates of flares, and ii) to estimate the incidence of flares of selected auto-immune diseases in selected subgroups, including pregnancy.
Country(-ies) of study	United Kingdom Spain Denmark Finland Norway Italy France
Author	Carlos E. Durán (University Medical Center Utrecht, the Netherlands) Miriam Sturkenboom (University Medical Center Utrecht, the Netherlands)

TABLE OF CONTENTS

1 TITLE

SAFETY-VAC: A framework for the post-authorisation safety monitoring and evaluation of vaccines in the EU.

Study Protocol for the occurrence of flares of Graves´ disease, Hashimoto´s thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, systemic lupus erythematosus, and ulcerative colitis using electronic healthcare data sources.

Document version: 3.0

2 MARKETING AUTHORISATION HOLDER

Not applicable (N/A).

3 RESPONSIBLE PARTIES

3.1 Investigators

University Medical Center Utrecht (UMCU), Utrecht, The Netherlands	Dr. Carlos E. Durán, Prof. Dr. Miriam Sturkenboom, Dr. Judit Riera, Nicoletta Luxi Vjola Hoxhaj,
Universiteit Utrecht (UU), Utrecht, The Netherlands CPRD data	Prof. Dr. Olaf Klungel, Dr. Patrick Souverein, Dr. Satu Johanna Siiskonen
VAC4EU	Dr. Sima Mohammadi
Teamit Institute	Dr. Fabio Riefolo, Dr. Irene Pazos
Agenzia Regionale di Sanitá Toscana (ARS) Data Tuscany region	Dr. Rosa Gini, Davide Messina, Dr. Giuseppe Roberto
Società Servizi Telematici -Pedianet	Prof. Dr. Carlo Giaquinto, Dr. Elisa Barbieri, Luca Stona
Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAP JGol)	Dr. Felipe Villalobos, Dr. Martín Solorzano, Carlo Alberto Bissacco
Instituto Aragonés de Ciencias de la Salud (IACS)	Dr. Antonio Gimeno, Dr. Beatriz Poblador, Dr. Mercedes Aza, Dr. Aida Moreno, Alejandro Santos
Department of Clinical Epidemiology, Aarhus University and Aarhus University Hospital	Prof. Vera Ehrenstein, Lise Skovgaard Svingel, Benjamin Randeris Johannesen

Bordeaux PharmacoEpi platform (BPE) & ADERA	Dr. Cécile Droz-Perroteau, Laure Carcaillon-Bentata,
University of Eastern Finland	Prof. Anna-Mija Tolppanen, Prof. Sirpa Hartikainen, Dr Thuan Vo, Dr Anne Paakinaho, Blair Rajamaki
University of Oslo (UiO), Norway Norwegian linked registry data	Prof. Dr. Hedvig Nordeng, Saeed Hayati, Mahmoud Zidan
Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO) – Valencia health system Integrated Database (VID)	Dr. Juan José Carreras Martínez, Dr. Arantxa Urchueguía Fornes, Elisa Correcher Martínez, Dr. Javier Díez-Domingo
Spanish Agency on Medicines and Medical Devices (AEMPS) -BIFAP database	Dr. Mar Martin, Dr. Patricia Garcia-Poza, Dr. Airam de Burgos, Belén Castillo-Cano, Dr. Elisa Martín-Merino

3.2 Other parties

Universiteit Utrecht (UU), Utrecht, The Netherlands	Dr. Helga Gardarsdottir,
University of Padova, Italy	Lorenzo Chiusaroli, Giulia Sturniolo
University of Verona. Italy	Ylenia Ingrasciotta, Massimo Carollo, Denise Rotta, Maurizio Rossini, Chiara Bellitto, Ombretta Viapiana
Vall d' Hebron University Hospital, Spain	Maria López Lasanta Alba Erra, Carlos Zafón, Mar Riveiro-Barciela, Natalia Borruel, Andreu Vilaseca
Drug Safety Research Unit (DSRU), United Kingdom	Kathryn Morton, Sandeep Dhanda, Debabrata Roy, Miranda Davies

4 ABSTRACT

4.1 Title with subtitles including version and date of the protocol and name and affiliation of main author

SAFETY-VAC: A framework for the post-authorisation safety monitoring and evaluation of vaccines in the EU.

Study Protocol for the occurrence of flares of Graves´ disease, Hashimoto´s thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, systemic lupus erythematosus, and ulcerative colitis using electronic healthcare data sources.

Document version: 3.0

Main authors:

- Dr. C.E. Durán, University Medical Center Utrecht, The Netherlands.
- Prof. dr. M.C.J.M. Sturkenboom, University Medical Center Utrecht, The Netherlands.

4.2 Rationale and background

The European Medicines Agency (EMA) together with the European Centre for Disease Prevention and Control (ECDC) established the Vaccine Monitoring Platform (VMP), of which one of the objectives is to generate a real-world evidence (RWE) framework for post-authorisation safety evaluation that can be leveraged in case of a new public health emergency or a safety concern occurring with a novel, or a more characterised, vaccine authorized in the European Union (EU) and the European Economic Area (EEA).
This study is part of this feasibility assessment to assess whether we can detect and provide background incidence rates of flares of auto-immune chronic diseases using European electronic healthcare record databases.

4.3 Research question and objectives

The research question is whether we can estimate the background incidence of flares of selected auto-immune disease in electronic health record data.

Primary objective:

To estimate the background (independent of vaccination) incidence rates of flares and the 6- and 12-month cumulative incidence of flares in patients with Graves' disease, Hashimoto's thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, systemic lupus erythematosus, and ulcerative colitis using European electronic healthcare data sources.

Secondary objectives:

1. To assess the contribution of different provenances of data as well as the contribution of the different components of the flare definition to the incidence rates of flares.

2. To estimate the background incidence rates and the 6 and 12-month cumulative incidence of flares of selected auto-immune diseases stratified by subgroups of age (< 18, 18-59, 60 years and older), sex and selected subpopulations.

4.4 Study design

A multi-database cohort design conducted during the period from January 1st, 2017, till the last data availability, specific for each data source. Study cohorts start with a first diagnosis of the selected auto-immune diseases, during the study period. We will access data from 10 different electronic health record data sources from the EU PE&PV and VAC4EU networks of 7 European countries: BIFAP, VID, EPICHRON and SIDIAP from Spain, PEDIANET from Italy, SNDS from France, Danish national registries (DNR) from Denmark, NHR from Norway, Finnish registries, and CPRD from the United Kingdom.

4.5 Population

The source population comprises all persons in the data sources who can be potentially included in the study.

4.5.1 Inclusion, exclusion criteria and follow-up

Ten disease-specific study cohorts will be created, one for each event**.** Persons will be included in the incident disease-specific study cohorts if they have:

- At least 1 day of observation after 1/1/2017.
- The first recorded diagnosis of the disease of interest after 1/1/2017

Persons will be excluded when:

- No birth year and sex information.
- Less than 365 days look-back at first diagnosis of the disease of interest, for those not born in the study period.

Additional exclusion criteria will be implemented to the cohorts of Hashimoto thyroiditis, multiple sclerosis and erythema nodosum.
For Hashimoto thyroiditis, all levothyroxine use registered within the look-back period will be excluded.
For multiple sclerosis, all interferon beta 1a and 1b, glatiramer and teriflunomide use registered within the look-back period will be excluded.
For erythema nodosum, all panniculitis, including Löfgren syndrome, diagnoses identified within the look-back period will be excluded.

Follow-up will start when study persons have completed 90 days after the disease diagnosis, except for autoimmune hepatitis and erythema nodosum where follow-up will start 30, and 60 days after the diagnosis, respectively. This delay is needed for the first episode to be ended and to be at risk of a flare. Upon each flare, the same eligibility delay for a next flare will be applied.

Follow-up for incidence rates will be interrupted when vaccination with any vaccine of interest occurs, from date of vaccination to 90 days later (to ensure we do not include rates in post-

vaccination risk windows), and finish at death, disenrollment, or last data extraction. If follow-up is interrupted because of a vaccination, follow-up will be resumed at day 91 after vaccination until follow-up finishes. An interruption of follow-up will also happen upon a flare, with the same eligibility period as after the first diagnosis. If a vaccination happens during the eligibility interruption, the maximum duration of interruption (eligibility and post-vaccination) is used. The occurrence of a flare-up of the corresponding event will not be considered a censoring criterion since a new flare-up can occur after a gap of 90 days. For estimation of cumulative incidence, follow-up will be censored upon the earliest of a vaccination, a flare, or end of follow-up.

4.6 Variables

All variables will be identified in the data sources by using diagnostic or procedure codes, medicines, and vaccines.

4.7 Data sources

The study will use data from 10 secondary electronic health record databases that are population-based in 7 countries in Europe (UK, Spain, Denmark, Finland, Norway, Italy and France). The characteristics of each of the participating DAPs are summarized in the following table:

Table 1. Data sources to be included in this study.

Country	**Data Source**	**Data Provider**	**Estimated Population size**
Spain (ES)	BIFAP	BIFAP	17 million
Spain (ES)	SIDIAP	IDIAP JGol	5.8 million
Spain (ES)	VID	FISABIO	5.0 million
Spain (ES)	EPICHRON	IACS	1.3 million
Italy (IT)	PEDIANET	So.Se.Te	50.000
Denmark (DK)	Danish national registries (DNR)	Aarhus University	5.9 million
Norway (NO)	Norwegian national registers	University of Oslo	5.3 million
United Kingdom (UK)	CPRD	Utrecht University	16 million
France (FR)	SNDS	BPE & ADERA	6.7 million (10% sample of the total population)
Finland (FI)	Finnish national registers	University of Eastern Finland	2.9 million (50% random sample of total population)

4.8 Study size

The study will include all eligible subjects in the data sources.

4.9 Data analysis

All analyses will be conducted using R version R-4.03 or higher or STATA:

Descriptive analyses will be shown through histogram plots with distance from the start of follow-up to the new occurrence of potential flare for visual inspection.

Incidence rates will be calculated based on the occurrence of a flare requiring a lag-time of 90 days (in order to be eligible) after its incident diagnosis and each flare, except for autoimmune

hepatitis and erythema nodosum (30 and 60 days, respectively), and allowing re-occurrence after the last flare-up event tailored in time-period to the condition of interest and the flare-specific identification component. Cumulative incidence of a first flare will be calculated to estimate six - and 12-months cumulative risk of flares.

Component analysis will describe the data diversity of each component of the flare identification algorithm in a data source-specific manner (1). Each identification component may differ in rates due to data source characteristics and data meanings. For each data sources, the different *component algorithms* and their data meanings for a same flare identification component will be identified. Flare cases, cumulative incidence of each *component algorithm* and their combinations in one study year will be measured. Incidence rates with the same "component"/"composite" algorithms will be compared across data sources, facilitating results benchmarking.

4.10 Milestones

Milestones for SAFETY-VAC: A framework for the post-authorisation safety monitoring and evaluation of vaccines in the EU.

Table 2. Milestones for SAFETY-VAC.

Contract signature	15 Feb 2024
Start of the project	15 Feb 2024
Study protocol*	16 Sep 2024
Study report	15 Oct 2024
Manuscript on flares of chronic conditions	14 Feb 2025

*Deliverable last submitted to EMA

5 AMENDMENTS AND UPDATES

None

6 MILESTONES

Milestones for SAFETY-VAC: A framework for the post-authorisation safety monitoring and evaluation of vaccines in the EU.

Table 3. Milestones for SAFETY-VAC.

Contract signature	15 Feb 2024
Start of the project	15 Feb 2024
Start of data collection	N/A (data instances start in 01 Jan 2017)
End of data collection	Latest available
Study protocol*	16 Sep 2024

Study report	15 Oct 2024
Manuscript on flares of chronic conditions	14 Feb 2025

*Deliverable last submitted to EMA

7 RATIONALE AND BACKGROUND

The COVID-19 pandemic emphasized the public health need for comprehensive and rapid post-authorisation vaccine safety surveillance. An increasing number of vaccine products are based on novel technologies, for which safety experience is limited to pre-authorisation clinical trials. While new safety concerns are expected to arise with these novel vaccines, continuous monitoring and evaluation throughout the entire lifecycle remains necessary for authorized vaccines (2,3). To this aim, networks of real-world data sources that are fit-for-purpose are essential, offering up-to-date data that can be readily accessed.

In May 2022, the EMA and the ECDC established the VMP with the perspective of generating RWE on the safety and effectiveness of vaccines in the EU and the European EEA(4). The VMP research agenda, endorsed in July 2023 by the Immunisation and Vaccine Monitoring Advisory Board (IVMAB), confirmed the need for RWE capacity, capability and readiness to allow the timely evaluation of vaccine safety concerns. Therefore, the EMA required the service of a contractor to provide a framework for post-authorisation safety evaluation that can be leveraged in case of a new public health emergency or a safety concern occurring with a novel or a more characterised vaccine. This study is part of this feasibility assessment to assess whether we can detect and provide background incidence rates of flares of auto-immune chronic diseases using electronic healthcare record databases.

Among few others, the new COVID-19 vaccines (5,6) and the recombinant zoster vaccine (7) have been recently proposed as potential triggers of flares of immune-mediated inflammatory diseases. In general, a flare-up is a transient exacerbation of symptoms of an existing disease or condition. Traditionally, flare-ups have been investigated through direct questionaries to patients and/or physicians. Measuring flares of chronic immune-mediated disease in electronic health record data sources is challenging for several reasons; for instance, in electronic health records, there is, most of the time, an adequate registration of the data of the diseases, specially dates and diagnoses, however, it could not exist a proper tracking of their evolution (ending, worsening, flares, etc), and the flare may be diagnosed/treated in a different clinical setting (e.g., GP) than where it was originally diagnosed, and multiple data banks need to be consulted. In this study, ten chronic conditions have been selected to investigate the background occurrence (i.e. not linked with vaccines exposure) (8) of flare-ups identified in European electronic health care record databases. Additionally, an analysis of the different components of the proposed algorithm to identify flare-up episodes will be performed to provide an approximate quantification of the rates underestimation per database, allowing a better interpretation of study findings (1).

8 RESEARCH QUESTION AND OBJECTIVES

The overarching goal of the SAFETY-VAC project is to create a framework for the post-authorisation safety monitoring and evaluation of vaccines in Europe that can conduct near real-time studies on new or existing vaccines.

The research question is whether we can estimate the background incidence of flares of selected auto-immune disease in electronic health record data. Background means not in the risk period of vaccination.

Primary objective:

To estimate the background (independent of vaccination) incidence rates of flares and 6- and 12-month cumulative incidence of flares in patients with Graves´ disease (GD), Hashimoto´s thyroiditis (HT), polyarteritis nodosa (PAN), autoimmune hepatitis (AIH), rheumatoid arthritis (RA), psoriatic arthritis (PsA), multiple sclerosis (MS), erythema nodosum (EN), systemic lupus erythematosus (SLE), and ulcerative colitis (UC) using European electronic healthcare data sources

Secondary objective(s):

1. To assess the contribution of different provenances of data as well as the contribution of the different components of the flare definition to the incidence rates of flares.

2. To estimate the background incidence and 6 and 12 months cumulative incidence of flares of selected auto-immune diseases stratified by subgroups of age (< 18, 18-59, 60 years and older), sex and selected subpopulations.

9 RESEARCH METHODS

9.1 Study design

A multi-database cohort design conducted during the period from January 1st, 2017, till the last data availability, specific for each data source. Study cohorts of incident diagnosis of the selected auto-immune diseases will be constructed.

9.2 Study Setting

The study will be conducted using electronic health record data from 10 data sources in 7 European countries (Spain: BIFAP, SIDIAP, VID, and EPICHRON; Denmark: Danish National Registries (DNR); Finland: Finnish National Registers; Norway: Norwegian National Registers; Italy: PEDIANET; France: SNDS, and United Kingdom: CPRD), see section 9.5 for further details on data sources. The source population comprises all persons in the data sources who can be potentially included in the study.

9.3 Source and study population

The source population comprises all persons included in the data sources. The study population will be selected from the source population after application of the inclusion and the exclusion criteria. Ten incident disease-specific study cohorts will be created, one for each auto-immune disease: Graves' disease, Hashimoto´s thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, systemic lupus erythematosus, and ulcerative colitis.

Persons will be included in the disease-specific incident cohorts if they have:

- At least 1 day of observation after 1 January 2017.
- The first recorded diagnosis of the disease of interest after 1 January 2017

Persons will be excluded when:

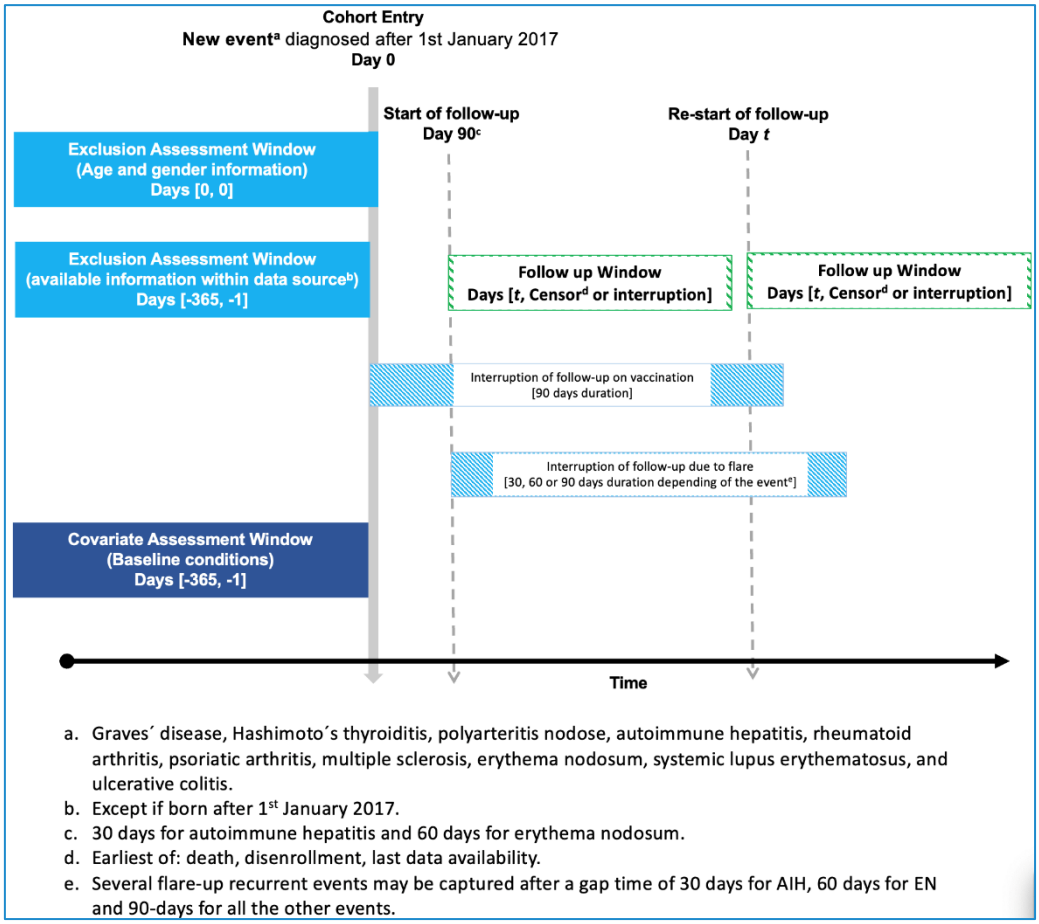
- No birth year and sex information.
- Less than 365 days look-back at first identification of the disease, for those not born in the study period.
- Persons with a diagnosis indicating that the disease is prevalent/chronic at first identification.

Additional exclusion criteria will be implemented to the cohorts of Hashimoto thyroiditis, multiple sclerosis and erythema nodosum.
For Hashimoto thyroiditis, all levothyroxine use registered within the look-back period will be excluded.
For multiple sclerosis, all interferon beta 1a and 1b, glatiramer and teriflunomide use registered within the look-back period will be excluded.
For erythema nodosum, all panniculitis, including Löfgren syndrome, diagnoses identified within the look-back period will be excluded.

Follow-up will start when study persons have completed 90 days after the disease diagnosis, except for autoimmune hepatitis and erythema nodosum where follow-up will start 30, and 60 days after the diagnosis, respectively. This delay is needed for the first episode to be ended and to be at risk of a flare. Upon each flare the same eligibility delay for a next flare will be applied.

Follow-up for incidence rates will be interrupted when vaccination with any vaccine of interest occurs, from date of vaccination to 90 days later (to ensure we do not include rates in post-vaccination risk windows), and finish at death, disenrollment, or last data extraction. If follow-up is interrupted because of a vaccination, follow-up will be resumed at day 91 after vaccination until follow-up finishes. An interruption of follow-up will also happen upon a flare, with the same eligibility period as after the first diagnosis. If a vaccination happens during the eligibility interruption, the maximum duration of interruption (eligibility and post-vaccination) is used. The occurrence of a flare-up of the corresponding event will not be considered a censoring criterion since a new flare-up can occur after a gap of 90 days. For estimation of cumulative incidence, follow-up will be censored upon the earliest of a vaccination, a flare, or end of follow-up.

Figure 1. Representation of study participants eligibility and follow-up during the study period.



9.4 Variables

Outcomes

This study will provide an estimation of the occurrence of flares of 10 selected pre-existing chronic diseases (Graves´ disease, Hashimoto´s thyroiditis, polyarteritis nodosa, autoimmune hepatitis, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, erythema nodosum, systemic lupus erythematosus, and ulcerative colitis) using electronic healthcare record data sources (**objective 1**).

For each event, the clinical definition is shown in Table 4 and a specific Event Definition Form will be developed during the study**.** The clinical definition reported in in Table 4, as well as the Event Definition Forms that will be generated during the study, are based on literature review, a clinical definition of the flare for a given condition, and a phenotype algorithm proposal. The event definitions presented in this protocol and developed during the study have been created by the central study team, together with a representative of the VAC4EU code list task force, and University Hospitals of University of Verona (Italy), University of Padova (Italy), Vall d'Hebron Research Institute (Barcelona, Spain), and the Drug Safety Research Unit (UK). Table 4 presents the flare-up operationalisation algorithm per event. The operationalisation proposal presented in Table 4 is based on the current availability of the ConcePTION CDM to ETL healthcare-related information. We will use the start of a new drug treatment and the

switching from one drug to another one as a proxy of the flare, increases in dose cannot be done reliably, since many DAPs do not have posology data. Table 4 presents the approach to identify disease flares. Further details will be included in the Event Definition Form.

To create the incident disease-specific study cohorts (see section 9.3), code lists created and provided by VAC4EU will be used. Briefly, study variables are named in a standard VAC4EU hierarchical fashion based on the body system. Each code within the code list is tagged as *narrow* or *possible* by two medical reviewers from the VAC4EU code list taskforce based on standard VAC4EU work instructions. Comments are consolidated in the VAC4EU medical code list taskforce. The code lists are subsequently compiled in a CSV file through a standard R code. This is the output to be included in the analysis script.

Table 4. Clinical definition of the events and flare-up operationalization.

Event	Clinical definition	Flare-up definition	Flare-up algorithm
Graves' disease (GD)	GD is a common form of hyperthyroidism with a diffuse hyperplastic goiter. It is an autoimmune disorder that produces thyrotropin-receptor antibodies (TRAb) against the thyroid stimulating hormone (TSH) receptor. These autoantibodies activate the TSH receptor, thereby stimulating the thyroid gland and hypersecretion of thyroid hormones.	A flare-up/relapse of GD can be defined as both, the worsening of hyperthyroidism symptoms and thyroid hormones in patients with GD who are currently undergoing medical treatment, and the re-occurrence of hyperthyroidism symptoms and thyroid hormones in patients with GD after completion of medical treatment.	Graves´ disease diagnosis (through diagnosis codes) **AND** at least 90 days after (9): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes). OR • Prescription or dispensing of antithyroid drugs (ATD) after a period of 90 days of no drug exposure, OR • Radioactive iodine ablation after a period of >90 days after last ATDs prescription or dispensing, OR • Thyroidectomy (through procedure codes) after a period of >90 days since last ATDs prescription or dispensing, OR • Thyroid storm (through code list), OR • Any emergency room visit, or hospitalization related to Graves' disease (primary diagnosis).
Hashimoto's thyroiditis (HT)	HT is an autoimmune disease in which thyroid cells are destroyed via cell- and antibody-mediated immune processes. It is characterized by the presence of high serum thyroid autoantibodies, goitre and hypothyroidism.	A HT flare-up is a period of worsening and intense hypothyroidism symptoms. HT flare-up can seldomly be characterized by a hyperthyroid state.	Diagnosis of HT (through diagnosis codes) **AND** (at least 90 days after (9)): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes) OR • Start of levothyroxine use (evidence of no use of levothyroxine since the disease diagnosis). OR • Any emergency room visit, or hospitalization related to HT (primary diagnosis).
Polyarteritis nodosa (PAN)	PAN is a rare form of necrotizing non-granulomatous inflammation occurring primarily in medium-sized arteries, often with microaneurysms. It is characterized by muscle, joint, and abdominal pain resulting from arterial infarction and scarring in affected organs, neurological and cutaneous symptoms.	PAN relapse is defined as recurrence or new onset of disease attributable to active vasculitis. Relapse is defined as the reappearance of manifestations attributable to active vasculitis occurring after at least one-month symptom-free period.	Diagnosis of PAN (using diagnosis codes) AND (at least 90 days after (11)): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes), OR • Switch from a corticosteroid to another immunosuppressant drug, OR • Any emergency room visit, or hospitalization related to PAN (primary diagnosis).
Autoimmune hepatitis (AIH)	AIH is a chronic inflammatory disease of the liver that is characterized by circulating autoantibodies and elevated serum immunoglobulin G	The exacerbation of autoimmune hepatitis refers to a sudden and marked increase in inflammatory activity in the liver. This exacerbation can manifest with an	Diagnosis code of AIH **AND** (at least 30 days after): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes), OR

Event	Clinical definition	Flare-up definition	Flare-up algorithm
	(IgG) levels. The disease may start as acute hepatitis or asymptomatic increase of liver enzymes, that can lead to the development of liver cirrhosis.	elevation of liver enzymes and IgG in blood tests with or without symptoms.	• Initiation of glucocorticoids, mycophenolate mofetil, tacrolimus, sirolimus, everolimus, antiTNF, rituximab or cyclosporin, OR • Any hospitalization, or emergency room visit related to AIH (primary diagnosis). **NOT** • Diagnosis code of hepatitis infectious in the last 30 days. • Diagnosis code of alcoholic liver disease in the last 30 days. • Presence of dispensing or prescription of any drug with a DILI rank as "Most-DILI concern" (12) in the last 30 days.
Rheumatoid arthritis (RA)	RA is a chronic inflammatory disease characterized by joint swelling, joint tenderness, and destruction of synovial joints, leading to severe disability and premature mortality. The cause of RA is unknown. RA is considered an autoimmune disease.	RA flare-up generally refers to a worsening of symptoms/disease activity (e.g., increased joint pain, swelling, stiffness, and fatigue) following a period of partial or complete remission that would, if persistent or of sufficient duration/intensity, in most cases lead to initiation, change or increase in therapy as cannot be self-managed.	Diagnosis of RA (using diagnosis codes) **AND** (at least 90 days after diagnosis (13–15)): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes), OR • Any initiation of a NSAID, OR • Any initiation of a corticosteroids (16,17) OR • Any initiation of any Disease-modifying antirheumatic drugs (DMARDs) (conventional nonbiologic/ synthetic, targeted synthetic, or biologic), OR • Identification of a code related to RA complications (scleritis, tunnel carpal syndrome, cervical myelopathy, Felty syndrome, rheumatoid vasculitis, pleuritis, pericarditis, rheumatoid nodules) if no similar code has been identified in the previous 90 days, OR • Any hospitalization, or emergency room visit related to RA (primary diagnosis).
Psoriatic arthritis (PsA)	PsA is a type of inflammatory arthritis associated with psoriasis, whose main clinical manifestations are peripheral arthritis, dactylitis, enthesitis and axial involvement. It is characterized by the presence of HLA-B27-associated spondyloarthropathy, and the absence of rheumatoid factor.	PsA flare-up can be broadly considered as a period of worsening in disease activity, with significant effect on patients' quality of life.	Diagnosis of PsA (using diagnosis codes) **AND** (at least 90 days after): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes) • Any initiation of a NSAID, OR • Any initiation of a corticosteroids, OR • Any initiation of any DMARDs (conventional nonbiologic/ synthetic, targeted synthetic, or biologic), OR • Adding any DMARDs to preexisting treatment, OR

Event	Clinical definition	Flare-up definition	Flare-up algorithm
			• Any hospitalization or emergency room visit related to PsA (primary diagnosis).
Multiple sclerosis (MS)* *Excluding primary progressive MS (PPMS)	MS is an autoimmune disorder mainly affecting young adults and characterized by destruction of myelin in the central nervous system (CNS). Pathologic findings include multiple sharply demarcated areas of demyelination throughout the white matter of the CNS. Clinical manifestations include visual loss, extra-ocular movement disorders, paraesthesia, loss of sensation, muscle weakness, dysarthria, spasticity, ataxia, and bladder dysfunction. The usual pattern is one of recurrent attacks followed by partial recovery, but acute fulminating and chronic progressive forms also occur.	A monophasic clinical episode with patient-reported symptoms and objective findings typical of MS, reflecting a focal or multifocal inflammatory demyelinating event in the CNS, developing acutely or subacutely, with a duration of at least 24 h, with or without recovery, and in the absence of fever or infection(18).	Diagnosis code of MS (using diagnosis codes) **AND** (at least 90 days after (10)): • Identification of a general code indicating a relapse of symptoms (only for SNOMED-based codes), OR • Any hospitalization or emergency room visit related to MS (primary diagnosis), OR • Any initiation of interferon beta 1a and 1b, glatiramer, teriflunomide, other immunosuppressants (ATC L04AX), monoclonal antibodies, cladribine, S1P modulators, and mitoxantrone.
Erythema nodosum (EN)	EN is a type of panniculitis, an inflammatory disorder affecting subcutaneous fat. It is a delayed-type hypersensitivity reaction that most often presents as erythematous tender nodules on the anterior shins. Less commonly, they affect the thighs and forearms.	The exacerbation of EN refers to a sudden and intense increase in symptoms associated with this dermatological condition. It manifests through an inflammatory response of the subcutaneous fat which leads to painful red nodules on the skin, commonly on the legs. The nodules are slightly raised and typically 2 to 5 cm in diameter. The nodules develop over several days and may follow a prodrome of fatigue, fever, malaise, arthralgias, or upper respiratory infection symptoms by one to three weeks.	Diagnosis (through diagnosis codes) of EN **AND** (at least 60 days after) (19): • Identification of a general code indicating a relapse of symptoms (only for SNOMED-based codes), OR • Acute erythema nodosum (diagnosis codes), OR • Panniculitis (diagnosis codes), including Löfgren syndrome, OR • Debridement procedure (through procedure codes), OR • Initiation of one of the following drugs after a 60 days gap of no prescription or dispensing of the same drug: NSAIDs, colchicine, dapsone, systemic corticosteroids, tetracyclines, TNF-α inhibitors, thalidomide, OR. • Hospitalization or emergency room visit related to EN (primary diagnosis).
Systemic lupus erythematosus (SLE)	SLE is a chronic, relapsing, inflammatory, and often febrile multisystemic autoimmune disease of connective tissue, primarily	SLE flare-up is a measurable increase in disease activity in one or more organ systems involving new or worse clinical findings/laboratory measurements, that is	Diagnosis of SLE **AND** (at least 90 days after): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes). • Any initiation of corticosteroids or NSAID, OR

Event	Clinical definition	Flare-up definition	Flare-up algorithm
	characterized by involvement of the skin, joints, kidneys, and serosal membranes.	usually accompanied by a change or an increase in the treatment, involving stronger agents or higher dosages.	• Any initiation of any DMARDs (conventional nonbiologic/ synthetic, targeted synthetic, or biologic), OR • Any initiation of a monoclonal antibody (anifrolumab, belimumab), OR • Any initiation of voclosporin, OR • Any initiation of cyclophosphamide, OR • Hospitalization or emergency room visit related to SLE (primary diagnosis)
Ulcerative colitis (UC)	UC is a type of inflammatory bowel disease (IBD) that is characterized by continuous and diffuse inflammation which is limited to the colonic mucosa and extends proximally from the rectum.	UC relapse is defined as flare-up of symptoms in a patient who is in clinical remission. Typically, these patients have rectal bleeding, increase in stool frequency and abnormal mucosa at sigmoidoscopy. The relapse is considered early if it occurs within 3 months after achieving remission with previous therapy. The pattern of relapse could be infrequent (less than or equal to 2 episodes/year), frequent (more than twice per year), or continuous (persistent symptoms of active UC without a period of remission).	Diagnosis of UC (through diagnosis codes) **AND** (at least 90 days after): • Identification of a code indicating a relapse of symptoms (only for SNOMED-based codes), OR • Rectal bleeding (through diagnosis codes), OR • Anemia (through diagnosis codes), OR • Blood transfusion or apheresis (through procedures codes), OR • Any intestinal surgery or procedure (through diagnosis codes) linked to the UC diagnosis (colectomy, proctocolectomy, restorative proctocolectomy with ileal pouch-anal anastomosis), OR • Any drug switching (as from the event definition from), OR • Hospitalization or emergency room visit related to UC (primary diagnosis).

For each flare-up identification algorithm, the component strategy will be applied to describe the impact of the different components of the event identification strategy on the observed rates of flares in each data source. A detailed description of the component analysis strategy is presented in section 9.8 Data analysis.

Covariates

All covariates will be identified during the look-back period (one year) prior to first diagnosis, see Figure 1. Covariates include risk factors, comorbidities, and co-medications. Table 5 presents the comorbidities per event.

Table 5. List of demographic characteristics and comorbidities per event and ConcePTION source table.

Covariate	Events									
	Graves' disease	**Hashimoto's thyroiditis**	**Polyarteritis nodose**	**Autoimmune hepatitis**	**Rheumatoid arthritis**	**Psoriatic arthritis**	**Multiple sclerosis**	**Erythema nodosum**	**Systemic lupus erythematosus**	**Ulcerative colitis**
Age	X	X	X	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X	X	X	X
Number of GP visits	X	X	X	X	X	X	X	X	X	X
Number of hospitalizations	X	X	X	X	X	X	X	X	X	X
Immunocompromised status*	X	X	X	X	X	X	X	X	X	X
Pregnancy	X	X	X	X	X	X	X	X	X	X
Hypertension					X	X			X	
Malignancies	X	X	X	X	X	X	X	X	X	X
HIV			X			X		X		
Cardiocerebrovascular disease					X	X				
Diabetes type 1					X	X				
Inflammatory bowel disease				X		X		X		
Infection	X			X	X					
Alcohol abuse				X	X					
Sepsis										
Chronic renal disease			X							
Herpes simplex				X				X		
Influenza							X			
Hepatitis C	X	X	X	X				X		
Rheumatoid arthritis				X						
SLE				X						
Sjogren's syndrome		X		X						
Myasthenia gravis		X								
Pernicious anemia		X								
Autoimmune hepatitis		X								
Celiac disease		X		X					X	
Hepatitis B			X	X				X		
Psoriasis				X		X				
Gout						X				

Covariate	Events									
	Graves' disease	**Hashimoto's thyroiditis**	**Polyarteritis nodose**	**Autoimmune hepatitis**	**Rheumatoid arthritis**	**Psoriatic arthritis**	**Multiple sclerosis**	**Erythema nodosum**	**Systemic lupus erythematosus**	**Ulcerative colitis**
Atopic dermatitis									X	X
Nonalcoholic fatty liver										X
Obesity					X		X			

* Combination of: inflammatory bowel disease, diabetes type 1, gout, AIDS, Sjogren syndrome, systemic lupus erythematosus, transplant recipient, psoriasis, psoriatic arthropathy, rheumatoid arthritis, spondylarthritis, multiple sclerosis, hematological cancer, multiple immunodeficiencies, immunosuppressants

Co-medication will be used as a proxy for comorbidities as well. The following co-medications will be assessed during the 12 months before time zero by calculating prevalence of use during the look back period for all study cohorts:

Comedication (dispensed/prescribed)

- Immunosuppressive agents (L04, H02)
- Immunostimulants (L03)
- PD-1/PD-L1 (Programmed cell death protein 1/death ligand 1) inhibitors (L01FF)
- Analgesics (N02)
- Systemic corticosteroids (H02)
- Antithrombotic agents (B01A)
- Sex hormones (G03)
- Immunosuppressants (L04, H02)
- Diabetes medications (A10A, A10B)
- Antibiotics (J01)
- Antiviral drugs (J05)
- Antimycotics (J02)
- Non-steroidal anti-inflammatory drugs (M01)
- Drug to treat mental health diseases (N05A, N06A, N06D)
- Lipid lowering drugs (C10)
- Cardiovascular meds (C01, C03, C04, C05, C07, C08, C09, B01AC, C02, C10)
- Oncologic drugs (L01, L02B, L02A)
- Anti-epileptics (N03)
- Diuretics (C03, C02L, C07B, C07C, C07D, C08G, C09BA, C09DA)
- Tumor necrosis factor (TNF) inhibitor (L04AB).

Vaccines
In this study, the exposure to the following vaccines will be collected for interruption of follow-up, alone or in combination:

- Coronavirus
- Influenza
- RSV
- Meningococcal
- HPV
- Rotavirus
- Pneumococcal
- Varicella
- Mpox
- Diphtheria
- Pertussis
- Tetanus
- Poliomyelitis
- HiB
- Hepatitis B
- Measles
- Mumps
- Rubella
- Tuberculosis (BCG)

- Herpes Zoster

DAPs will transform and load their local data on vaccinations into the VACCINE table of the ConcePTION CDM.

9.5 Data sources

The study will use data from secondary electronic health record databases that are population-based. The characteristics of each of the participating DAPs are summarized in Table 6 and further detailed below.

Table 6. Data provider and data sources.

Country	Data Source	Data Provider	Estimated Population size
Spain (ES)	BIFAP*	BIFAP	17 million
Spain (ES)	SIDIAP	IDIAP JGol	5.8 million
Spain (ES)	VID	FISABIO	5.0 million
Spain (ES)	EPICHRON*	IACS	1.3 million
Italy (IT)	PEDIANET	So.Se.Te	50.000
Denmark (DK)	Danish national registries (DNR)	Aarhus University	5.9 million
Norway (NO)	Norwegian national registers (NHR)	University of Oslo	5.3 million
United Kingdom (UK)	CPRD	Utrecht University	16 million
France (FR)	SNDS	BPE & ADERA	6.7 million (10% sample of the total population)
Finland (FI)	Finnish national registers	University of Eastern Finland	2.9 million (50% random sample of total population)

*EPICHRON data source covers data from one of the nine countries that are covered by the BIFAP data source. Thus, potential overlapping of information between the two data sources may happen and will be discussed in the study report. As this is a study based on an overall feasibility assessment, including both data sources will not affect the results validity. Furthermore, a metanalysis of results will not be performed.

ES: BIFAP (Several regions)

BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en el Ámbito público), a computerized database of medical records of primary care is a non-profit research project funded by the Spanish Agency for Medicines and Medical Devices (AEMPS). Information collected by PCPs includes administrative, socio-demographic, lifestyle, and other general data, clinical diagnosis and health problems, results of diagnostic procedures, interventions, vaccines administration (by nurses) and prescriptions/dispensations. Diagnoses are classified according to the International Classification of Primary Care (ICPC)-2, ICD-9 and SNOMED-CT system, and a variable proportion of clinical information is registered in "medical notes" in free text fields in the electronic medical records. Additionally, information on hospital discharge diagnoses coded in ICD-9 and ICD-10 terminology is linked for a subset of periods and regions participating in the database. All information on prescriptions of medicines by the PCP is incorporated and linked by the PCP to a health problem (episode of care), and information on the dispensation of medicines at pharmacies is extracted from the e-prescription system that is widely implemented in Spain. Vaccines administered in primary care setting are systematically recorded and some vaccinations in public settings (such as covid-19 vaccinations or RSV immunisations) may also be linked.

The project started in 2001 and the participating version of the database with information until April 2022 includes clinical information of more than 14,810 primary care physicians and paediatricians' practices (PCPs). Nine participant autonomous regions send their data to BIFAP every year including anonymized clinical and prescription/dispensing and vaccination data from around 20 million (17 active population) patients representing ≥92% of all patients of those regions participating in the database, and ≥32% of the Spanish population. Mean duration of follow-up in the database is 9 years.

ES: SIDIAP (CATALUÑA)

The Information System for Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAP) in Catalonia, Spain, is a primary care database set up by the Institute of Research in Primary Care (Fundació Institut Universitari D'Investigació en Atenció Primària Jordi Gol [IDIAP JGol]) and Catalan Institute of Health (Institut Català de la Salut). The database collects information from 278 primary health care centres and includes more than 5.8 million patients covered by the Catalan Institute of Health (approximately 78% of the Catalan population) and is highly representative of the Catalan population.

SIDIAP data comprise the clinical and referral events registered by primary care health professionals (i.e., GPs, paediatricians, and nurses) and administrative staff in electronic medical records, comprehensive demographic information, community pharmacy invoicing data, specialist referrals, and primary care laboratory test results. Health professionals gather this information using International Classification of Diseases, 10th Revision (ICD-10) codes, ATC codes, and structured forms designed for the collection of variables relevant to primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, and blood urine test results. In relation to vaccines, information on all routine childhood and adult immunisations is included in addition to the antigen and the number of administered doses.

SIDIAP is listed under the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources databases. SIDIAP was characterised in the IMI-ADVANCE project and considered fit for purpose for vaccine coverage, benefits, and risk assessment. An algorithm to identify pregnancies has been previously used within SIDIAP. The algorithm uses diagnosis codes recorded in primary healthcare records during pregnancy and information recorded in the sexual and reproductive healthcare registries, including LMP, gestational week, expected date of delivery, actual date of delivery or termination, and pregnancy outcomes. Approximately 50% to 60% of pregnant women in Catalonia are attended in the sexual and reproductive healthcare centres that contribute data to SIDIAP. Approximately 70% of infant records can be linked to maternal records and used for research. The protocol will be evaluated by the SIDIAP Scientific Committe and by the IDIAPJGol Ethics Committee, the approval can take up to 4 weeks. The timeframe for data availability after the approval by the two local Committees is one month.

ES: VID (VALENCIA)

The Valencia health system integrated database (VID) is a set of multiple, public, population-wide electronic databases for the Valencia Region, the fourth most populated Spanish region, with ≈5 million inhabitants and an annual birth cohort of 48000 new-borns, representing 10.7% of the Spanish population and around 1% of the European population. The VID provides exhaustive longitudinal information including sociodemographic and administrative data (sex, age, nationality, etc.), clinical (diagnoses, procedures, diagnostic tests, imaging, etc.), pharmaceutical (prescription, dispensation) and healthcare utilization data from hospital care, emergency departments, specialized care (including

mental and obstetrics care), primary care and other public health services. It also includes a set of associated population databases and registries of significant care areas such as cancer, rare diseases, vaccines, congenital anomalies, microbiology and others, and public health databases from the population screening programmers. All electronic health systems in the VID use the ICD-9-CM and the ICD-10-CM. All the information in the VID databases can be linked at the individual level through a single personal identification code. The databases were initiated at different moments in time, but all in all the VID provides comprehensive individual-level data fed by all the databases from 2008 to date. Information on PCR test results as well as serological/antibody tests results for the whole population of the Valencia region is available and linkable from the Microbiological Surveillance Network (RedMIVA). The Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO) is Data Access Provider for Valencia Integrated Databases (VID).

ES: EPICHRON (ARAGON)

The EPICHRON database links sociodemographic and clinical anonymised information from 2010 to present for all the users of the public health system in Aragón (approximately 98% of the reference population). This database is built from the BIGAN platform, which integrates a technical infrastructure and a data lake gathering individual patient data from the regional health service information systems, including primary care, specialised care, hospitalisations, emergency department visits, drug prescriptions, image diagnosis, laboratory tests, diagnostics (ICD-10), vaccination, medical history, and demographics from the users of the public health system of Aragon, which comprises about 2 million individuals historic data and an active population of 1.3 million individuals.

IT: PEDIANET

Pedianet is a national population database that contains anonymous patient-level data of more than 500,000 children since 2004, corresponding to around 4% of the annual paediatric population who received healthcare from family paediatricians (FPs) in Italy who were part of the PEDIANET network.

The network links FPs distributed throughout several Italian regions designated by the Italian NHS, including Friuli-Venezia Giulia, Liguria, Lombardia, Piemonte, Veneto, Lazio, Marche, Toscana, Abruzzo, Campania, Sardegna, and Sicilia, and who use the same software (Junior Bit®) (Padova, Italy) in their professional practice.

According to the Italian NHS, each child is assigned to a FP, who is the primary referral for health-related matters. In Italy, there is a tax-funded public healthcare system with universal access, and patients do not incur direct costs related to primary care visits. The Pedianet database captures several types of patient-level information, including the reason for accessing healthcare, health status, demographic data, diagnosis, and clinical symptoms (free text or ICD-9-CM codes), drugs (Anatomical-Therapeutical-Chemical codes), specialist appointments, diagnostic procedures, hospital or emergency room (ER) admissions, growth parameters, and clinical outcome data. Informed consent is required from children's parents to enter the data in the database. The data collected from the child's parents/tutors by paediatricians enters the dedicated cloud already encrypted and anonymised. Pedianet researchers do not know the process to anonymise the data and cannot know the owner of the data in any way.

DK: DANISH NATIONAL REGISTRIES (DNR)

All Danish registries used in this study have a nationwide coverage and an almost 100% capture of contacts covering information on currently 5.9 million inhabitants plus historical information[2]. Unambiguous person-level linkage across all data sources is possible via a unique identifier used in all Danish public records. Linked data from the following registries are

available for the current project: the Danish Civil Registration System (identifier for linkage, age, sex, births, deaths, migrations); the Danish National Prescription Registry (outpatient dispensing in community pharmacies, no data on drugs administered in hospitals); the Danish National Health Service Register (general practitioner contacts including vaccinations other than COVID-19); the Danish National Patient Registry (diagnoses – ICD10 Danish version, and procedures from all hospital encounters); the Danish Vaccination Register (COVID-19 vaccinations only). Data are linked using a unique pseudonymized identifier on the servers of the Danish Health Data Authority (SDS). Individual-level data will be analysed by uploading and running of analytic scripts on the SDS servers and aggregate data that does not allow backtracking to individuals in accordance with the data regulation will be used for reporting. The Danish national registries are listed as a resource in the EU PAS Register.

NO: NORWEGIAN NATIONAL LINKED REGISTERS AT UIO (NHR)

The core data that the University of Oslo (UiO) has access to are the health care administrative data banks of the entire Norwegian population, which amounts to approximately 5.3 million inhabitants. Norway has a universal public health care system, consisting of primary health care services and specialist healthcare services. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. The mandatory national health registries were established to maintain national functions. They are used for health analysis, health statistics, improving the quality of healthcare, research, administration and emergency preparedness. The Norwegian data sources, in this project are the national, mandatory Norwegian Surveillance System for Communicable Diseases (MSIS), which will be linked to five national health registries, i.e. the Medical Birth Registry, the National Patient Register, Norway Control and Payment of Health Reimbursement, the Norwegian Immunisation Registry, and the National Prescription Registry.

Information about all Norwegian National Registries can be found here: www.fhi.no/en/more/access-to-data/about-the-national-health-registries2/.

In this project, University of Oslo is Data Access Provider for Norwegian national registry data. Their current Norwegian health registry data will be used, capitalizing on the existing ETL's and quality checked data instance. In specific, UiO will contribute with ETL'd data on all residents in Norway between 1.1.2017- 31.12.2022, with historical data on these individuals back to 2010. Consequently, we will not be able to provide analysis as a near real-time analysis. Some ICD10 codes are not at the 4-digit level.

UK: CPRD-AURUM

The Clinical Practice Research Database (CPRD) from the UK collates the computerised medical records of GPs in the UK who act as the gatekeepers of health care and maintain patients' life-long electronic health records. Accordingly, GPs are responsible for primary health care and specialist referrals, and they also store information about specialist referrals and hospitalisations. General practitioners act as the first point of contact for any non-emergency health-related issues, which may then be managed within primary care and/or referred to secondary care, as necessary. Secondary care teams also provide information to GPs about their patients, including key diagnoses. The data recorded in the CPRD include demographic information, prescription details, clinical events, preventive care, specialist referrals, hospital admissions, and major outcomes, including death. Most of the data is coded using Read or SNOMED codes. Data validation with original records (specialist letters) is also available. The population in the data bank is generalisable to the UK population based on age, sex, socioeconomic class, and national geographic coverage CPRD Aurum versions is used. There are currently approximately 59 million individuals (acceptable for research purposes) -17

millions of whom are active (ie, still alive and registered with the GP practice)- in over 2,000 primary care practices (https://cprd.com/Data). Data include demographics, all GP/health care professional consultations (e.g., phone calls, letters, e- mails, in surgery, at home), diagnoses and symptoms, laboratory test results, treatments (including all prescriptions), all data referrals to other care providers, hospital discharge summary (date and Read/SNOMED codes), hospital clinic summary, preventive treatment and immunisations, and death (date and cause). For a proportion of the CPRD panel practices (> 80%), the GPs have agreed to permit the CPRD to link at the patient level to HES data. The CPRD is listed under the ENCePP resources database, and access will be provided by University Utrecht).

FR: SNDS

The *Système National des Données de Santé* (SNDS) is the French nationwide healthcare database. It currently covers the overall French population (about 67 million persons) from birth (or immigration) to death (or emigration), even if a subject changes occupation or retires. Using a unique pseudonymized identifier, the SNDS merges all reimbursed outpatient claims from all French health care insurance schemes (SNIIRAM database), hospital-discharge summaries from French public and private hospitals (PMSI database), and the national death registry. In this study, SNDS will contribute with a random sample of 10% of the database population. SNDS data are available since 2006 and contains information on:

- General characteristics: gender, year of birth, area of residence, deprivation index, etc.
- Death: month, year, and cause.
- Long-term disease registration associated with an ICD-10 diagnostic code.
- Outpatient reimbursed healthcare expenditures with dates and codes (ICD-10) (but not the medical indication nor result): visits, medical procedures, nursing acts, physiotherapy, lab tests, dispensed drugs and medical devices, etc. For each expenditure, associated costs, prescriber and caregiver information (specialty, private/public practice) and the corresponding dates are provided.
- Inpatient details: primary, related and associated ICD-10 diagnostic codes resulting from hospital discharge summaries with the date and duration of the hospital stay, the performed medical procedures (but no results), lab tests (but no results) and the related costs. Drugs included in the diagnosis related group cost are not captured. However, expensive drugs (i.e., the one charged in addition to the group cost) are.

Outpatient data (SNIIRAM) are uploaded to the SNDS throughout the year. It is admitted that a lag of around 6 months is required to catch 90% of the dispensing. Inpatient data (PMSI) are uploaded in one time, at the end of the following year. Hence, we consider that complete SNDS data of year Y are available in January of the year Y+2.

SNDS access is regulated: each study involving the human person with or without data extraction from the SNDS needs approval from the *Comité Ethique et Scientifique pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé* (CESREES) in charge of assessing scientific quality of the project, and authorization from the *Commission Nationale de l'Informatique et des Libertés* (CNIL) which is the French data protection authority, and then an agreement with the SNDS data holder (CNAM) for data extraction.

FN: FINNISH NATIONAL REGISTERS

Finnish national data registers account for a total population of 5.4 million inhabitants. In this study, data from 50% (2.9 million inhabitants) random sample of total population will be used. Main linkable data banks are: 1. *Hospital discharge register*: use of in- and outpatient services. Diagnoses (ICD10 Finland) for each admission are made by the attending physician. The

register contains the following information on each hospital visit: dates, reason for hospital stay, specialty of the caring unit, date of operation, up to five operational codes (NOMESCO classification), where the patient was discharged to and assessment of need for assistance in activities of daily life. Since 2009, the data bank contains outpatient visits to specialised healthcare and since 2011 to primary healthcare. Laboratory and physiological measurements are available since 2015. 2. *Kanta electronic prescriptions:* all prescribed medicines purchased by an individual. Medicines used in hospitals are not included, but the register covers prescriptions written by hospital physicians and dispensed in community settings. Data on dispensing date, number of packages, tablets and defined daily dose (DDD) are available. Medicines are classified according to Anatomical Therapeutic Chemical (ATC)–classification system. 3. *Special reimbursement register:* entitlement to special reimbursement due to severe chronic diseases such as Alzheimer's disease, diabetes, psychosis, epilepsy, asthma, c*hronic obstructive pulmonary disease and* several cardiovascular diseases. The diagnoses are based on explicit predefined criteria. 4. *Statistics Finland* is the statistical authority of Finland, producing the majority of official statistics and conducting the population census, which has solely been based on the register data since 1990. These censuses include indicators of socioeconomic position (e.g. education, occupational status and taxable income). The causes of death register are compiled from death certificate data containing underlying, direct, intervening, and contributing causes. Death certificates are issued by physicians and if an autopsy is required, by a medicolegal officer.

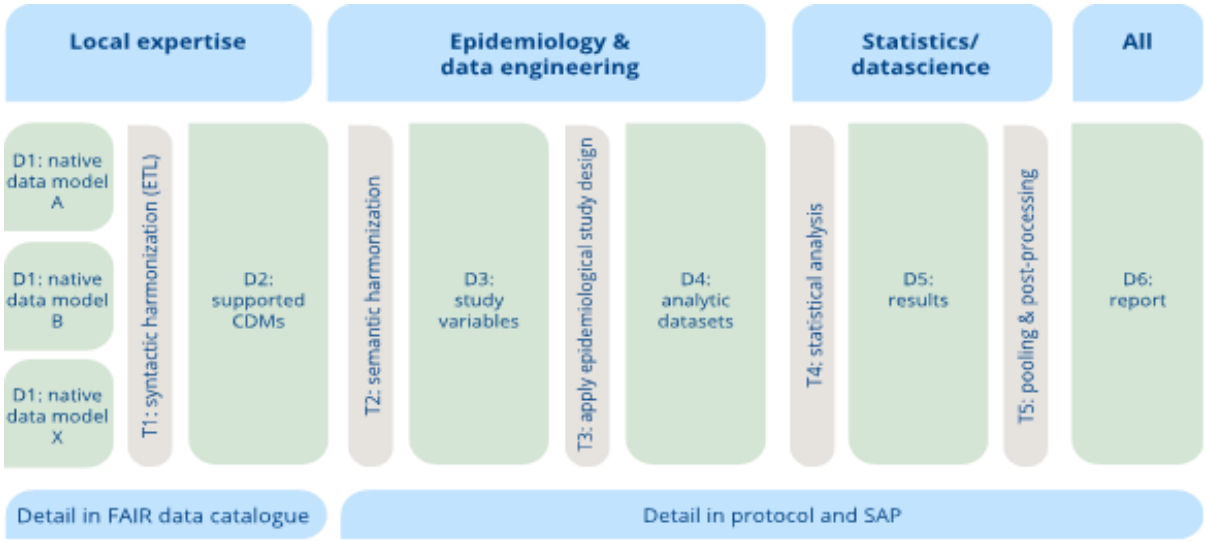
9.6 Study size

The study will include all eligible subjects in the data sources.

9.7 Data management

The study will be conducted in a distributed manner using the UMCU and VAC4EU tools, procedures, and pipeline. This pipeline can be viewed from a programming perspective (see Figure 2) or tool perspective (Figure 3). Figure 2 specifies the data sets (D) and transformation processes (T), programming follows this pipeline, with involvement of different types of experts.

Figure 2. Data Management from the data transformation perspective.



D1: Original data can be in any native format

The RWD-RWE pipeline used by VAC4EU, and EU PE&PV starts with data banks that are controlled by the DAP, these can be in any format. This stays local. The ETL design is shared in a searchable FAIR VAC4EU catalogue. The VAC4EU FAIR Molgenis data catalogue is a meta-data management tool designed to contain searchable meta-data describing organisations that can provide access to specific data sources.

T1: Syntactic harmonisation (ETL)

T1: Syntactic harmonisation is conducted through an extraction, transformation, and loading (ETL) process of native data into the requested CDM. To harmonise the structure of the data sets stored and maintained by each data partner, a shared syntactic foundation is used. The ETL process has various structured steps as described by Thurin et al (20):

- DAPs are asked to share the data dictionaries of their data banks (selected tables and variable names/structure)
- Metadata (descriptive data about the data sources and databanks) & data dictionaries, are uploaded in FAIR data catalogue (Molgenis).

D2: Common data model

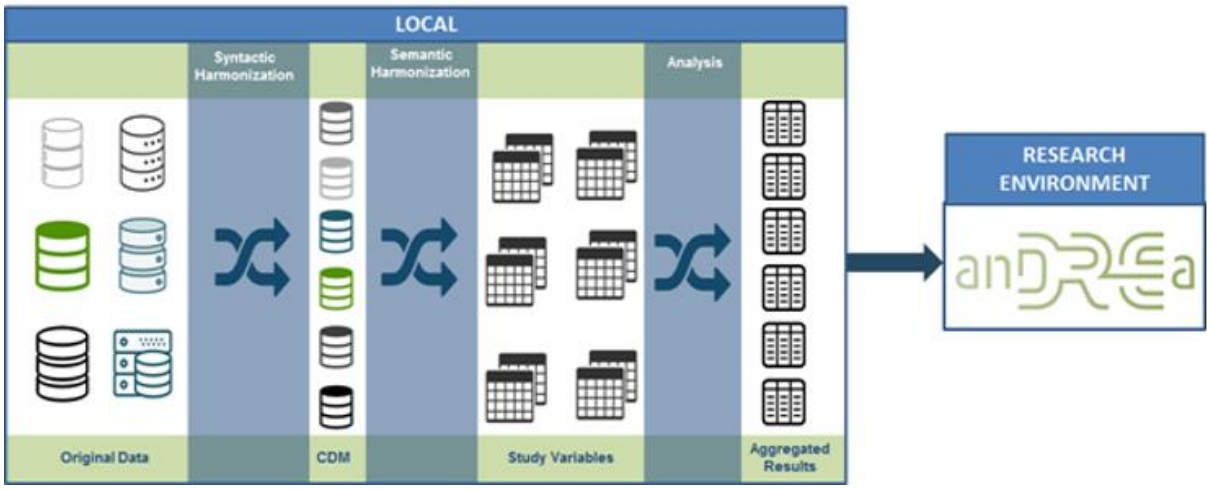
For this project the CDM (D2) is the ConcePTION common data model (CDM). The CDM version that is used is v2.2, which is available as an open-source CDM. In this CDM, data are represented in a common structure, but the values of the data remain in their original language (e.g. codes will have either ICD9/10/ICPC/SNOMED values).

T2: Semantic harmonisation

During the T2 step, many data transformations occur related to the completion of missing features in the data. Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g., medications), one or more ***phenotype algorithms*** are constructed (typically one sensitive, or broad, algorithm and one specific, or narrow, algorithm) to operationalise the identification and measurement of each event. In this step we conduct time anchoring (observation periods, look back periods), clean the data such as the dose of vaccines,

sort on record level, aggregate across multiple records, and combine concepts for implantation of algorithms, and rule-based creation of study variables.
In this phase of creation of study variables, semantic mapping is conducted. This semantic mapping across different vocabularies is conducted as part of the R-study script using different functionalities. To reconcile differences between different terminologies and native data availability, machine readable code lists are used that comprise the terminologies that are used in the network (e.g. ICD-9, ICD10, SNOMED, ICPC and DAP specific adaptations). This is combined with the BRIDGE metadata file that defines risk windows, look back periods, and algorithms for each study variable (21).

Figure 3. Data management from a systems and location perspective



D3: Study variables
D3 datasets are interim data sets with information on study variables for each study participant, the unit may be a person, a medicine, or episode of time. The design of these datasets is described in codebooks. Examples of D3 datasets are the outputs of the ConcePTION pregnancy algorithm, functions that define smoking. Multiple functions/packages exist within the VAC4EU, for different study variables.

T3: application of epidemiological design
In the T3 step epidemiological designs are applied such as sampling, matching (on specific variables and/or propensity scores), and selection based on inclusion and exclusion criteria using the study variables in the D3 datasets. The designs will be implemented for the various study objectives using R-scripts, and these may use the existing functions (R-cran) or functions that have been developed in the VAC4EU community.

D4: Analytical data set
D4 is an analytical dataset, and multiple D4 data sets may be produced based on the objectives of the study. The format is described initially in a code book for communication between programmers and statisticians.

T4: Statistical analysis
This step in the data transformation pipeline will produce statistical estimates such as descriptives (counts, percentages), distributions (mean, percentiles), rates (prevalence,

incidence), regression coefficients, or other relevant estimates. This will be conducted using R.

D5: Results
D5 is the set of estimands, tables or aggregate data that is transferred from the DAPs to the Digital Research Environment (DRE) The aggregated results produced by these scripts at the DAPs site will be uploaded to the UMCU Digital Research Environment (DRE) for post-processing, pooling, and visualization (see Figure 3). The DRE is made available through UMCU. The DRE is a cloud-based, globally available research environment where data are stored and organized securely and where researchers can collaborate. The DRE is made available through UMCU. The DRE applies double authentication where researchers can collaborate using data that is stored and organised securely. UMC Utrecht is responsible for data processing and data security.
All researchers who need access to the DRE will be granted access to study-specific secure workspaces by UMCU. Access to the workspaces will be possible only after double authentication using an identification code and password together with the user's mobile phone for authentication.
The uploading of files will be possible for all researchers with access to the workspace within the DRE. Downloading of files will be possible only after requesting and receiving permission from a workspace member with an "owner" role, who will be a UMCU team member.

T5: Post-processing/pooling
In this step, the result from different DAPs is pooled and converted into tables and figures for reporting.

Scripting and deployment
The analytical R scripts that produce the T2-T4 steps are produced on VAC4EU GitHub for version control. Links to the latest script will be distributed to DAPs for local deployment. Any issues can be notified on the private GitHub, and the data engineers who are responsible for the R code will work with the local DAP to resolve issues if they occur. After the final report is accepted the script will be made publicly available through GitHub and get a digital object identifier through Zenodo.

9.8 Data analysis

All analyses will be conducted using R version R-4.03 or higher (Foundation for Statistical Computing, Vienna Austria) or STATA.

Descriptive analyses
Histogram plots with distance from the start of follow-up to the new occurrences of potential flares will be created for visual inspection. Crude counts of flares during the lag time of 90 days (from day 0 to start of follow-up), except for 30 days for AIH and 60 days for EN, will be reported.

Incidence rates and cumulative risk
Incidence rates of flares per 100,000 PY will be calculated using the number of flares in the numerator (requiring a lag time of at least the 90 days (with the exception of 30 days for AIH and 60 days for EN), between repeated events to avoid misclassification due to repeat visits for a same event). Upon the occurrence of the first flare event, follow-up time will not be censored,

thus recurrence of a new flare after a new gap of 30, 60 or 90 days (depending on the disease) will be allowed. Kaplan Meier analysis will be used to estimate the cumulative incidence (incidence proportion) of a first flare 6- and 12-months after start of follow-up. To estimate the cumulative incidence, follow-up will be censored upon the earliest of a vaccination, a flare, or end of follow-up.
Age-specific estimates of incidence and risk of flares will be calculated in the following categories: < 18, 18-59, 60 years and older, by sex, as well as in selected subgroups.

Component analysis

For each auto-immune condition of interest, the component strategy will be applied to describe the contribution and data diversity (22) of each component of the flare identification algorithm in a data source-specific manner.
As an example, the flare-up of multiple sclerosis (MS) (see section 9.4) will be identified in each participating data source using at least one of the following components: i) "a general code indicating a relapse of symptoms", or ii) "any hospitalization or emergency room visit due to MS as primary diagnosis", or iii) "any switch or addition of specific drugs for MS". Using each component of the MS flare identification algorithm may result in a different flare rate depending on the data source, possibly due to differences in the underlying data source-specific characteristics (e.g., healthcare system organization and policies) and/or the available data meanings (e.g., different record prompts and/or healthcare settings of data collection) (20). For instance, the MS flare component iii) "any switch or addition of specific drugs for MS" can be equally applied to data sources that captures this information through different data meanings, e.g., hypothetically:

- A "Data source A" might use records of drug prescriptions from primary care practices, thus missing prescription from specialistic care,
- A "Data source B" might use dispensing records of drug reimbursed by the National Healthcare Service for outpatient use, missing inpatient use and/or non- reimbursed drugs,
- A "Data source C" might use records of drug prescriptions from both primary and specialist care, missing inpatient prescriptions that possibly lead to a slight overestimation of flare rate due to drugs prescribed but never dispensed.

Such data diversity will be considered and described by identifying the different "component algorithms" that apply to a same component of the flare identification algorithm for each participating data source (e.g., the different component algorithms that can identify iii) "any switch or addition of specific drugs for MS" are used to describe the data diversity within the participating data sources). Each "component algorithm" has a direct dependence on the specific data meanings (e.g., prescription from primary care) as well as on the specific code list used to catch the event.[1] In particular, the code list for identifying the event of interest will be selected according to the expected sensitivity and Positive Predictive Value (PPV) to the event identification. Codes will be classified as "narrow" if they certainly imply the condition of interest, or "possible" if that code may be observed not only in the presence of the condition of interest but also within others. Identification algorithms only based on narrow codes are expected to have a high PPV and lower sensitivity. If algorithms are based on both narrow and possible codes (named "broad algorithms"), they have higher sensitivity and lower PPV.

[1] *Future publication, abstract accepted as an oral presentation to 2024 ISPE Annual Meeting in Berlin: "Assessing the impact of data diversity on background incidence rates of adverse events of special interest: the component strategy from the COVID-19 Vaccine Monitoring project" Roberto G. et al.*

We will measure the number of flare cases identified and the cumulative incidence of each "component algorithm" in one study year. The combination of two or more "component algorithms" (situation named as "composite algorithms") will also be assessed. Incidence rates of an event of interest that are obtained from the application of the same component or composite algorithm will be compared across different data sources, thus facilitating results benchmarking. The extent of the contribution of different component and composite algorithms to the observed flare rate within the same data source, as well as the overlaps between different component or composite algorithms, will be also quantified. Altogether, this approach, which we referred to as *the component algorithm strategy*, will provide evidence for quali-quantitative appraisal of possible causes of incidence rates' inaccuracy, allowing us to get insights into data diversity and interpretation of findings.

9.9 Quality control

Rigorous quality-control (QC) procedures will be used. Data transformation into the ConcePTION CDM will be conducted by each subcontracted research partner in its associated database, using the processes described in the following sections (see below), each of these steps is fully transparent and will be signed of/reviewed by local and central teams.
Standard operating procedures or internal process guidance at each research centre will be used to guide the conduct of the study. These procedures include rules for secure and confidential data storage, backup, and recovery; methods to maintain and archive project documents; QC procedures for programming; standards for writing analysis plans; and requirements for scientific review by senior staff.

Quality of protocol, SAP, reports and manuscript
All these products are and will be reviewed by the entire consortium with many domain experts, using version control on SharePoint.

Quality of data checks
Data partners will be asked to provide INSIGHT data quality checks outputs for level 1 (completeness) and 2 (consistency) (or any EMA quality framework checks when available); level 3 is checking for study variables and assess whether data are fit for purpose.
Generic open-source data quality check scripts are available from the that are publicly available on the UMCU-RWE GitHub.

Level 1 - Data completeness
The purpose of the level 1 check is to verify the completeness of the ETL process and the data in the variables. Examples of tests are:

- Presence of variables in each of the CDM tables in D2
- Checks for misspelling and letter case in variable names in each of the CDM tables
- Verification of vocabularies
- Check date formats
- Check conventions of values
- Missing data analysis
- Frequency tables for categorical variables

Level 2 - Data logic/consistency

Real data is not random but follow certain logical constraints that reflect rules governing real-world situations. Examples of indicators generated by level 2 checks are:

- Event dates before date of birth
- Event dates after date of death
- Event dates out of observation periods
- Subjects having an observation but not present in the PERSONS table
- Observations associated with a visit id and occurred before/after the visit start/end date
- Subjects younger than 12 years old reported as parents
- Age at the observation period older than 115 years old

Level 3

Level 3 checks review patterns of study variables over time, age within and between data sources. We will use 7 modules, which may be used depending on the study variables.

- Source and study population.
- Medicines
- Vaccines
- Diagnoses
- Pregnancy
- Populations of interest.
- Health-seeking behaviour and lifestyle factors.

General approach to quality of R-coding

Data Management will follow standard operating procedures. R programs will be made available by UMCU (INSIGHT data quality checks) and ARS. ARS will create clear documentation (graphical and in Excel spreadsheet) of the data management steps, beginning with describing the required variables from the CDM and each of the subsequent transformation steps and intermittent data tables.

Coding conventions (process quality)

VAC4EU GitHub (and the underlying GitHub version control system) will be used to collaborate with multiple parties. At its core, GitHub tracks all changes and shows which, when, who and why changes were made. In the chain of events, any previous state can be recovered easily. Regarding proposed changes or potential bugs, GitHub provides a platform to discuss details. Using GitHub Actions, standard workflows will be defined and executed after a submitted change. An example is executing unit tests to ensure that scripts are correct.
The main coordinator of the GitHub is VAC4EU who creates a repository for the study and provides the 'main' functions to be used in each study:

- A readme file is initialized with relevant information about the scripts.
- For each study, a 'branch' is created in which scripts are tailored to the respective study.
- After each version update, the coordinator requests from all teams to incorporate the changes using 'merge'. One responsible from each team is appointed and allowed access to the repository. In case the main scripts contain an error, the 'issues' functionality is used to report the bug. If possible, a bug fix can be proposed by creating a 'pull request'. The `issues` platform also provides a means to ask for further clarification regarding new versions.

We will use one set of standard conventions for all parties to facilitate collaboration and minimize bugs in scripts. Coding conventions are categorized into three parts:

- Notation (e.g., name scripts, functions, and objects).
- Syntax (e.g., spacing, braces, indentation)
- Documentation (e.g., writing comments, dividing code into sections)

Script names will be informative, where words are separated with an underscore or a hyphen. For scripts that are executed sequentially, the names are prefixed with numbers that indicate the order. For naming functions and objects, we suggest adopting the "snake style", where words are separated with an underscore. For syntax rules, we will implement the tidy verse style guide found at https://style.tidyverse.org/. To facilitate implementing these rules, we will use the 'formatR' R package. This package automatically restyles R code to adhere to these rules. For documentation, comments will be provided that explain each part of the code. Each script file will start with a title, author, date, and version number. Comments are placed to describe functions and objects.
Scripts will be private during development and will be made public though the VAC4EU Github and Zenodo upon acceptance of the final report.

Standard/bespoke analyses script creation, testing and release
Study scripts connect and package functions using a structured design and follow the statistical analysis plan. Study scripts will be created in 4 steps:

1. Defining a map of the script, which includes specification of the folder structure, data model, graphical representation of the steps, use of functions, allocation of responsibilities and timelines, plus review schedules.
2. Programming of the code by a programmer plus statistician. Test with code profiler to monitor bottlenecks in the code.
3. Test script on one real data partner before making it available for all the DAPS.
4. Take the script into deployment.

For each update, these steps are repeated.

Quality of study conduct
The work in this protocol will be conducted according to the guidelines for Good Pharmacoepidemiology Practice (GPP) (International Society for Pharmacoepidemiology 2008) and according to the ENCePP code of conduct (European Medicines Agency 2018). All partners and principal investigators have experience in conducting pharmacovigilance/pharmacoepidemiological research and research is done by researchers trained in pharmacoepidemiology or pharmacovigilance. Utrecht University and University Medical Center Utrecht (data management), and Teamit (project management) are working according to a quality management system based on ISO 9001 principles and are certified.
All partners are ENCePP centers and the majority are members of the VAC4EU network, an ENCePP-listed network. Collectively these partners have registered more than 100 PASS studies on the EU PAS register, several with ENCePP seal.
The quality management system is system and process-oriented and based on continuous improvement. The system is based upon standard operating procedures implemented throughout the divisions with regular internal audits as well as external audits that lead to certification. The quality management system is based on national and international external quality requirements where available and pertinent, including the guidelines for Good

Pharmacoepidemiological Practices, RECORD-PE, ENCePP Guide on Methodological Standards in Pharmacoepidemiology, Good Clinical Practice, and Good Clinical Data management. Practice as well as national and international guidelines and legislation concerning data-handling and privacy issues.
All deliverables will be reviewed by project partners, we have multiple expert organizations in public health and epidemiology who will contribute to methods and review.

9.10 Limitations of the research methods

First, this study uses available data sources, which capture different databanks. Not all data sources capture the same type of information which may impact on the ability to identify chronic diseases and their flares. Therefore, we conducted a pre-feasibility assessment and included data sources which mostly cover both general practices as well as outpatient and inpatient diagnosis, medicines, and vaccine data.
Second, although changes in the doses of the recommended drugs to treat the selected chronic diseases might be useful information to identify potential flare-up episodes, posology information is not available in most data sources, which may impact the identification of flare ups leading to some misclassification. Finally, our operational phenotype algorithms are defined based on agreed clinical definitions. Although these phenotype algorithms have been developed by a team of epidemiologist and clinical experts, they have not been formally validated.

10 PROTECTION OF HUMAN SUBJECTS

This is a non-interventional study using secondary data collection and does not pose any risks for individuals. Each data source research partner will apply for an independent ethics committee review according to local regulations. Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants.

Patient information
This study involves data that exists in an anonymized structured format and contains no patient personal information.
All parties will comply with all applicable laws, including laws regarding the implementation of organizational and technical measures to ensure the protection of patient personal data. Such measures will include omitting patient names or other directly identifiable data in any reports, publications, or other disclosures, except where required by applicable laws.
Patient personal data will be stored at DAPs in encrypted electronic form and will be password protected to ensure that only authorized study staff have access.
DAPs will implement appropriate technical and organizational measures to ensure that personal data can be recovered in the event of a disaster. In the event of a potential personal data breach, DAPs shall be responsible for determining whether a personal data breach has in fact occurred and, if so, providing breach notifications as required by law.

Patient consent
As this study does not involve data subject to privacy laws according to applicable legal requirements, obtaining informed consent from individuals is not required.

11 MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

For studies in which the research team uses only data from automated healthcare databases, according to the International Society for Pharmacoepidemiology Guidelines for GPP.
"Aggregate analysis of database studies can identify an unexpected increase in risk associated with a particular exposure. Such studies may be reportable as study reports, but typically do not require reporting of individual cases. Moreover, access to automated databases does not confer a special obligation to assess and/or report any individual events contained in the databases. Formal studies conducted using these databases should adhere to these guidelines."
For non-interventional study designs that are based on secondary use of data, such as studies based on medical chart reviews or electronic health records, systematic reviews, or meta-analyses, reporting of adverse events/adverse drug reactions is not required. Reports of adverse events/adverse drug reactions should only be summarized in the study report, where applicable.
According to the EMA Guideline on GVP, Module VI – Management and Reporting of Adverse Reactions to Medicinal Products,
"All adverse events/reactions collected as part of [non-interventional post-authorization studies with a design based on secondary use of data], the submission of suspected adverse reactions in the form of [individual case safety reports] is not required. All adverse events/reactions collected for the study should be recorded and summarized in the interim safety analysis and in the final study report."
Module VIII – Post-Authorization Safety Studies echoes this approach. Legislation in the EU further states that for certain study designs such as retrospective cohort studies, particularly those involving electronic health records, it may not be feasible to make a causality assessment at the individual case level.

12 PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

As per EMA GVP Module VIII, the study and its protocol will be registered in the EU PAS Register prior to the start of data collection.
Results of analyses and interpretation will be delivered in report form.
Study results will be published following guidelines, including those for authorship, established by the ICMJE. When reporting the results of this study, the appropriate Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist and the RECORD-PE extension will be followed. Independent publication rights will be granted to the research team in line with Section VIII.B.5., Publication of study results, of the EMA *Guideline on Good Pharmacovigilance Practices (GVP) Module VIII: Post-Authorisation Safety Studies*.
Upon study completion and finalisation of the study report, the results of this study will be submitted for publication, preferably in a relevant peer-reviewed journal and posted in the EU PAS Register.
Analytical programs will be posted in a public Github & Zenodo repository after acceptance of the report, reports will also be publicly posted on Zenodo (VAC4EU, EU PE&PV) and cross linked to the EU PAS register.

13 REFERENCES

1. Gini R, Dodd CN, Bollaerts K, Bartolini C, Roberto G, Huerta-Alvarez C, et al. Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project. Vaccine [Internet]. 2020 Dec 22 [cited 2024 May 10];38 Suppl 2:B56–64. Available from: https://pubmed.ncbi.nlm.nih.gov/31677950/

2. 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 [Internet]. 2021 May 1 [cited 2024 May 8];6(Suppl 2):e003814. Available from: https://gh.bmj.com/content/6/Suppl_2/e003814

3. Ricotta EE, Rid A, Cohen IG, Evans NG. Observational studies must be reformed before the next pandemic. Nature Medicine 2023 29:8 [Internet]. 2023 Jun 7 [cited 2024 May 8];29(8):1903–5. Available from: https://www.nature.com/articles/s41591-023-02375-8

4. Vaccine Monitoring Platform | European Medicines Agency [Internet]. [cited 2024 May 8]. Available from: https://www.ema.europa.eu/en/about-us/what-we-do/crisis-preparedness-management/vaccine-monitoring-platform

5. Xie Y, Liu Y, Liu Y. The Flare of Rheumatic Disease After SARS-CoV-2 Vaccination: A Review. Front Immunol [Internet]. 2022 Jul 4 [cited 2024 Jun 21];13:919979. Available from: www.frontiersin.org

6. Paik JJ, Sparks JA, Kim AHJ. Immunogenicity, breakthrough infection, and underlying disease flare after SARS-CoV-2 vaccination among individuals with systemic autoimmune rheumatic diseases. Curr Opin Pharmacol. 2022 Aug 1;65:102243.

7. Lenfant T, Jin Y, Kirchner E, Hajj-Ali RA, Calabrese LH, Calabrese C. Safety of recombinant zoster vaccine: a retrospective study of 622 rheumatology patients. Rheumatology [Internet]. 2021 Nov 3 [cited 2024 Jun 21];60(11):5149–57. Available from: https://dx.doi.org/10.1093/rheumatology/keab139

8. Black SB, Law B, Chen RT, Dekker CL, Sturkenboom M, Huang WT, et al. The critical role of background rates of possible adverse events in the assessment of COVID-19 vaccine safety. Vaccine. 2021 May 6;39(19):2712–8.

9. Graves' hyperthyroidism in nonpregnant adults: Overview of treatment - UpToDate [Internet]. [cited 2024 Jun 21]. Available from: https://www.uptodate.com/contents/graves-hyperthyroidism-in-nonpregnant-adults-overview-of-treatment

10. Treatment of acute exacerbations of multiple sclerosis in adults - UpToDate [Internet]. [cited 2024 Jun 20]. Available from: https://www.uptodate.com/contents/treatment-of-acute-exacerbations-of-multiple-sclerosis-in-adults

11. Wolff L, Horisberger A, Moi L, Karampetsou MP, Comte D. Polyarteritis Nodosa: Old Disease, New Etiologies. International Journal of Molecular Sciences 2023, Vol 24, Page 16668 [Internet]. 2023 Nov 23 [cited 2024 Jun 20];24(23):16668. Available from: https://www.mdpi.com/1422-0067/24/23/16668/htm

12. Chen M, Suzuki A, Thakkar S, Yu K, Hu C, Tong W. DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. Drug Discov Today. 2016 Apr 1;21(4):648–53.

13. Di Matteo A, Bathon JM, Emery P. Rheumatoid arthritis. The Lancet. 2023 Nov 25;402(10416):2019–33.

14. Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis [Internet]. 2023 Jan 1 [cited 2024 Jun 20];82(1):3–18. Available from: https://ard.bmj.com/content/82/1/3

15. Bingham CO, Pohl C, Woodworth TG, Hewlett SE, May JE, Rahman MU, et al. Developing a Standardized Definition for Disease "Flare" in Rheumatoid Arthritis (OMERACT 9 Special Interest Group). J Rheumatol [Internet]. 2009 Oct 1 [cited 2024 Jun 20];36(10):2335–41. Available from: https://www.jrheum.org/content/36/10/2335

16. Konzett V, Kerschbaumer A, Smolen JS, Kristianslund EK, Provan SA, Kvien TK, et al. Definition of rheumatoid arthritis flare based on SDAI and CDAI. Ann Rheum Dis [Internet]. 2024 Feb 1 [cited 2024 Jun 20];83(2):169–76. Available from: https://ard.bmj.com/content/83/2/169

17. Hua C, Buttgereit F, Combe B. Glucocorticoids in rheumatoid arthritis: current status and future studies. RMD Open [Internet]. 2020 Jan 1 [cited 2024 Jun 20];6(1):e000536. Available from: https://rmdopen.bmj.com/content/6/1/e000536

18. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol [Internet]. 2018 Feb 1 [cited 2024 May 15];17(2):162–73. Available from: http://www.thelancet.com/article/S1474442217304702/fulltext

19. Pérez-Garza DM, Chavez-Alvarez S, Ocampo-Candiani J, Gomez-Flores M. Erythema Nodosum: A Practical Approach and Diagnostic Algorithm. Am J Clin Dermatol [Internet]. 2021 May 1 [cited 2024 Jun 21];22(3):367–78. Available from: https://link.springer.com/article/10.1007/s40257-021-00592-w

20. Thurin NH, Pajouheshnia R, Roberto G, Dodd C, Hyeraci G, Bartolini C, et al. From Inception to ConcePTION: Genesis of a Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding. Clin Pharmacol Ther [Internet]. 2022 Jan 1 [cited 2024 May 8];111(1):321–31. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cpt.2476

21. Royo AC, Elbers R, Weibel D, Hoxhaj V, Kurkcuoglu Z, Sturkenboom MC, et al. Real-World Evidence BRIDGE: a tool to connect protocol with code programming. medRxiv [Internet]. 2024 May 8 [cited 2024 May 10];2024.05.08.24306833. Available from: https://www.medrxiv.org/content/10.1101/2024.05.08.24306833v1

22. Gini R, Pajouheshnia R, Gardarsdottir H, Bennett D, Li L, Gulea C, et al. Describing diversity of real world data sources in pharmacoepidemiologic studies: The DIVERSE scoping review. Pharmacoepidemiol Drug Saf [Internet]. 2024 May 1 [cited 2024 Jun 20];33(5):e5787. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/pds.5787