Covid-Vaccine-Monitor

Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care datasources

Protocol

V2.1

29 July 2021

Title	Rapid Safety Assessment of SARS-CoV-2 vaccines in 5 EU Member States using 10 electronic health care datasources
Protocol version identifier	2.1
Date of last version of protocol	July 29, 2021
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	This study will create readiness for and assess the association between COVID-19 vaccines and adverse events of special interest using electronic healthcare databases in European countries
Country(-ies) of study	Netherlands, United Kingdom, Norway, Italy and Spain
Author	Miriam Sturkenboom (University Medical Center Utrecht)
Contributors	Estel Plana, Alejandro Arana, Bradley Layton, Alison Kawai, Rachael Ross RTI-HS Barcelona, Spain Ursula Kirchmayer, Valeria Belleudi, Department of Epidemiology, Lazio Regional Health Service, ASL Roma1 Felipe Villalobos, Boni Bolibar, IDIAPJGol, Spain, Ylenia Ingrasciotta, Gianluca Trifirò, Valentina Lentile (INSPIRE srl) Karin Swart, Jetty Overbeek, Ron Herings, PHARMO Institute, Netherlands Ian Douglas, Anna Schultze, London School of Hygiene & Tropical Medicine Olaf Klungel, Utrecht University, The Netherlands Ainara Mira-Iglesias (FISABIO), Juan José Carreras-Martínez (FISABIO), Angela Lupattelli (University of Oslo)

1. Table of contents

2. Abbreviations

ACCESS	vACCine covid-19 monitoring readinESS
ADVANCE	Accelerated Development of VAccine beNefit-risk Collaboration in Europe
AESI	Adverse Event of Special Interest
ARDS	Acute respiratory distress requiring ventilation
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CDM	Common Data Model
CI	Confidence interval
DAP	Data Access Provider
DRE	Digital Research Environment
EMA	European Medicines Agency
EMR	Electronic Medical Records
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance.
ETL	Extract, Transform, and Load
EU PAS	The European Union electronic Register of Post-Authorisation Studies
GDPR	General Data Protection Regulation
GP	General Practitioner
GPP	Good Participatory Practice
HIV	Human Immunodeficiency Virus
ICD	International Classification of Diseases
ICMJE	International Committee of Medical Journal Editors
ICU	Intensive Care Unit
MIS-C	Multisystem Inflammatory Syndrome in children
mRNA	messenger Ribonucleic acid
NHS	National Health Service
QC	Quality Control
RNA	Ribonucleic acid
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SPEAC	Safety Platform for Emergency vACcines
VAC4EU	Vaccine monitoring Collaboration for Europe

3. Responsible parties

University Medical Center Utrecht, The Netherlands
Prof. dr. Miriam Sturkenboom Dr. D Weibel Dr. C Duran Drs. V Hoxhaj Dr. Daniel Oberski Prof. dr. Eijkemans Dr. R van de Bor Drs. R Elbers Drs. E Alsina

Key collaborators and roles in CVM study

Name	Organization
Prof. dr. Miriam Sturkenboom, Dr. D Weibel,Drs. V Hoxhaj, Dr. Daniel Oberski, Prof. dr. Eijkemans, Dr. R van de Bor Drs. R Elbers Drs. E Alsina	University Medical Center Utrecht, Utrecht, The Netherlands (UMCU)
Prof. dr. Olaf Klungel Dr. Helga Gardarsdottir, Dr. Romin Pajouheshnia, Dr. Patrick Souverein, Svetlana Belitser Dr. Satu Johanna Siiskonen	Universiteit Utrecht (UU), Utrecht, The Netherlands CPRD data
Dr. Daniel Weibel Dr. Patrick Mahy Dr. Angela Bastidas	VAC4EU
Dr. Rosa Gini, Claudia Bartolini	Agenzia Regionale di Sanitá Toscana (ARS) Data Tuscay region
Dr. Ursula Kirchmayer, Valeria Belleudi	Department of Epidemiology, Lazio Regional Health Service, ASL Roma1 (DEP Lazio)
Dr. A. Arana, Dr. S. Perez-Gutthann, Dr. Alison Kawai, Ms Lia Gutierrez, Ms. Estel Plana, Rachel Weinrib, Dr. X. Garcia de Albeniz, Dr. B. Layton	RTI Health Solutions (RTI-HS), Spain & US
Prof. Dr. RMC Herings, Jetty Overbeek, Karin Swart	PHARMO Institute - PHARMO Database Network
Prof. Dr. Gianluca Trifirò, Ylenia Ingrasciotta, Valentina Ientile	INSPIRE srl, Messina Caserta database
Prof. Hedvig Nordeng Dr. Angela Lupattelli	University of Oslo (UiO), Norway Norwegian data
Dr. Mar Martin, Dr. Patricia Garcia-Poza Dr. Consuelo Huerta, Dr. Dolores Montero, Maria Martinez, Airam de Burgos	Spanish Agency on Medicines and Medical Devices (AEMPS) -BIFAP database

Prof. Ian Douglas Dr. Anna Schultze	London School of Hygiene and Tropical Medicine (LSHTM): method
Eva Molero, Dr. Fabio Riefolo	Team-it Institute SL (TEAMIT), coordination
Felipe Villalobos, Boni Bolibar	IDIAP Jordi Gol, SIDIAP data
Ainara Mira-Iglesias, Juan José Carreras-Martínez, Carlos Vergara-Hernández, Valle Morales-Cuenca, Javier Díez-Domingo	Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO) – Valencia health system Integrated Database (VID)

Sponsor:

This protocol has been developed as a deliverable of the framework contract No EMA/2018/28/PE (SC01, Lot 3) with the European Medicines Agency.

4. Abstract

Title:

Rapid Safety Assessment of SARS-CoV-2 vaccines in EU Member States using electronic health care datasources

Main author:

Prof. dr. M.C.J.M. Sturkenboom, University Medical Center Utrecht, The Netherlands.

Rationale and background:

The global rapid spread of COVID-19 caused by the SARS-CoV2 triggered the need for developing vaccines to control this pandemic. This study will create readiness and allows for rapid assessment of the association of adverse events of special interest (AESI) following COVID-19 vaccination.

Research question and objectives:

Readiness

The readiness phase will include the following objectives:

- To provide an overview of the methods for identification of COVID-19 vaccine exposure in the data sources
- To monitor the number of individuals exposed to any COVID-19 vaccine and to compare this to COVID-19 vaccine exposure (benchmark: https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#uptake-tab
- To quantitatively evaluate different algorithms to identify adverse events by provenance in electronic health care data
- To conduct time-to-onset analyses for the AESI with respect to time since vaccination
- To assess the association between and the vaccines of interest and negative control events using the SCRI to estimate systematic bias (unmeasured confounding)
- To test the impact of different comparators in the cohort design, using the negative control outcomes
- To generate information for testing of methodological questions around misclassification of events/exposure

Rapid assessment studies

Primary objective

The primary objective for this rapid assessment study is to assess the potential association between the occurrence of specific AESIs and vaccination with COVID-19 vaccines within disease-specific risk periods in individuals exposed to the COVID-19 vaccines compared to other COVID-19 vaccine exposed individuals or compared to a control window within the same individual.

Secondary objectives

The secondary objectives for the rapid assessments studies are:

- To assess the potential association between the occurrence of specific AESIs and vaccination with COVID-19 vaccines in the following subgroups
    - immunocompromised persons
    - persons with the presence of co-morbidities elevating the risk of serious COVID-19
    - persons with a history of diagnosed COVID-19 disease
    - pregnant women
    - age groups (<18, 18-64, 65 years and older)
    - patients with a prior history (ever) of that event more than a year before.

Study design:

A retrospective, multi-database, self-controlled risk interval or cohort study, conducted during the study period ranging from December 1, 2020, to the latest availability of data. (For readiness we will start the study period on January 1, 2019). The self-controlled study will compare against non-vaccinated, and the cohort analysis against another COVID-19 vaccine.

As part of the methods development work in the CVM project, we will explore the implementation of the use of an active comparator in the SCRI allowing comparison with the estimates from the cohort analysis,[1] and different comparators for the cohort design. This will first be tested using the negative control events in the readiness phase and results may inform sensitivity analyses for the rapid evaluation studies. For death, the SCRI design will be adapted and start at the date of vaccination with a comparison in different intervals of the postvaccination window. For events with a high fatality rate, a sensitivity analysis will be conducted with those that survive both the control and vaccination risk window.

Population:

All subjects in the source population in the participating data sources who were in follow-up for at least 365 days on December 1, 2020 (January 1, 2019 for readiness study) or were born into the cohort during the study period, and for whom vaccination data would be able to be obtained/linked.

Variables:

Variables of interest will be:

- Person-time: birth and death dates as well as periods of observation.
- Events: dates of medical and/or procedure and/or prescription/dispensing codes to identify AESI, COVID-19, and at-risk medical conditions.
- Vaccines: vaccine brands

Table 1: The list of AESI (as per EMA June 8 communication) is listed below, this list may be updated if new issues occur.

[1] Hallas J, Whitaker H, Delaney JA, Cadarette SM, Pratt N, Maclure M. The Use of active Comparators in self-controlled Designs. Am J Epidemiol. 2021 Apr 16:kwab110. doi: 10.1093/aje/kwab110. Epub ahead of print. PMID: 33861309.

Event	ACCESS	SCRI	cohort	Naïve period to estimate new onset	Primary Risk period*
Multisystem inflammatory syndrome	✓	✓	✓	365 days	28 days
Acute respiratory distress syndrome	✓	✓	✓	365 days	28 days
Acute cardiovascular injury	✓	✓	✓	365 days	
microangiopathy	✓	✓	✓	365 days	28 days
CAD	✓	✓	✓	365 days	28 days
Arrhythmia	✓	✓	✓	365 days	28 days
Myocarditis	✓	✓	✓	365 days	28 days
Pericarditis	✓	✓	✓	365 days	28 days
Coagulation disorders, including deep vein thrombosis, pulmonary embolus, cerebrovascular stroke, limb ischaemia, haemorrhagic disease	✓				
VTE (DVT & PE & Splanchnic)	✓	✓	✓	365 days	28 days
CVST	✓	✓	✓	365 days	28 days
Arterial thrombosis	✓	✓	✓	365 days	28 days
TTS (VTE, arterial thrombosis, or CVST with thrombocytopenia in 10 days)	✓	✓	✓	365 days	28 days
Hemorrhagic stroke	✓	✓	✓	365 days	28 days
DIC	✓	✓	✓	365 days	28 days
Generalised convulsion	✓	✓	✓	365 days	14 days
Guillain Barré Syndrome	✓	✓	✓	365 days	42 days
Diabetes (type 1 and unspecified type)	✓		✓	365 days	180 days
Acute kidney injury	✓		✓	365 days	180 days
Acute liver injury	✓		✓	365 days	180 days
Anosmia, ageusia	✓	✓	✓	365 days	28 days
Chilblain-like lesions	✓	✓	✓	365 days	28 days
Single organ cutaneous vasculitis	✓	✓	✓	365 days	28 days
Erythema multiforme	✓	✓	✓	365 days	7 days
Anaphylaxis	✓	✓	✓	30 days	2 days
Death (any cause)** (postvaccination control window)	✓	✓	✓	365 days	7 days
Sudden death (by codes)** (postvaccination control window)	✓	✓	✓	365 days	7 days
Meningoencephalitis	✓	✓	✓	365 days	28 days
Acute disseminated encephalomyelitis (ADEM)	✓	✓	✓	365 days	28 days
Narcolepsy	✓		✓	365 days	180 days
Thrombocytopenia	✓	✓	✓	365 days	28 days
Transverse myelitis	✓	✓	✓	365 days	28 days
Bells' palsy		✓	✓	365 days	28 days
Haemophagocytic lymphohistiocytosis[2]		✓	✓	365 days	180 days
Kawasaki's disease		✓	✓	365 days	28 days

[2] https://primaryimmune.org/disease/hemophagocytic-lymphohistiocytosis-hlh

Event	ACCESS	SCRI	cohort	Naïve period to estimate new onset	Primary Risk period*
Pancreatitis		✓	✓	365 days	28 days
Rhabdomyolysis		✓	✓	365 days	28 days
SCARs		✓	✓	365 days	28 days
Sensorineural hearing loss			✓	365 days	180 days
Thyroiditis			✓	365 days	180 days
Negative control event					
Gout		✓	✓	365 days	28 days
Otitis externa		✓	✓	365 days	28 days
Trigeminal neuralgia		✓	✓	365 days	28 days

**for death the SCRI design will start at date of vaccination and the risk period will be divided in smaller risk windows to compare.[3]
Contingent on time to onset analysis, for any event with median onset> 60 days, a cohort approach is recommended

Data sources:

The study will include data from 10 electronic health care data sources in 5 European countries (Italy, Netherlands, Spain, Norway, United Kingdom) that can link event data to vaccination data. Data sources will capture outcomes from hospitalization and/or general practice.

Study size:

The source population will comprise approximately 45 million individuals.

Data analysis:

In the readiness phase, we will use negative control outcomes to assess the impact of misclassification of outcome (AESI and negative controls) and exposure, unmeasured confounding, and choice of comparators and to further test the assumptions of the design in collaboration with the methods workpackage (WP4).

Relative risks of specific AESI will be estimated for each specific brand of COVID-19 vaccine in comparison to non-vaccinated individuals (prior to vaccination or in non-vaccinated) and between different Covid-19 vaccines.

For the secondary analysis, stratified analyses will be conducted.

5. Amendments and updates

Number	Date	Section of study protocol	Amendment or update	Reason
N/A				

[3] https://academic.oup.com/biostatistics/article/10/1/3/269598

6. Milestones

Milestones and deliverables	Planned date
Contract signature	6 Apr 2021
Start of project	6 Apr 2021
D1 Study plan*	6 May 2021
D2 Study protocol(s)*	7 Jun 2021
Study start	7 July 2021
D3 Monthly interim statistical report 1* on dashboard	30 Sep 2021
D3 Monthly interim statistical report 2* on dashboard	29 Oct 2021
D3 Monthly interim statistical report 3* on dashboard	30 Nov 2021
D3 Monthly interim statistical report 4* on dashboard	31 Dec 2021
D3 Monthly interim statistical report 5* on dashboard	31 Jan 2022
D3 Monthly interim statistical report 6* on dashboard	28 Feb 2022
D3 Monthly interim statistical report 7* on dashboard	31 Mar 2022
D4.1 Interim report*	6 Apr 2022
D3 Monthly interim statistical report 8* on dashboard	29 Apr 2022
D3 Monthly interim statistical report 9* on dashboard	31 May 2022
D3 Monthly interim statistical report 10* on dashboard	30 Jun 2022
D3 Monthly interim statistical report 11* on dashboard	29 Jul 2022
D3 Monthly interim statistical report 12* on dashboard	31 Aug 2022
D3 Monthly interim statistical report 13* on dashboard	30 Sep 2022
D3 Monthly interim statistical report 14* on dashboard	31 Oct 2022
D3 Monthly interim statistical report 15* on dashboard	30 Nov 2022

D3 Monthly interim statistical report 16* on dashboard	30 Dec 2022
D3 Monthly interim statistical report 17* on dashboard	31 Jan 2023
D3 Monthly interim statistical report 18* on dashboard	28 Feb 2023
D3 Monthly interim statistical report 19* on dashboard	31 Mar 2023
D4.2 Final report*	6 Apr 2023
D5 Manuscript*	6 Apr 2023

* Deliverable to be submitted to EMA

Note for rapid assessment study reports: monthly statistical reports will be delivered to EMA if they have been requested

7. Background and current situation

EMA's mission is the protection and promotion of public and animal health, through the evaluation and supervision of medicines for human and veterinary use.

COVID-19 vaccines in the EU are evaluated by EMA via the centralised procedure, based on a rolling review. While a large number of COVID-19 vaccines are still progressing in clinical development, four vaccines (from Pfizer/BioNTech, Moderna, AstraZeneca, Janssen) have been granted conditional marketing authorisation. While more vaccines are expected to be authorised in 2021, large-scale vaccination campaigns are being rolled out across the EU, with tens and perhaps hundreds of millions of EU citizens expected to be vaccinated in 2021 and 2022.

Multiple vaccine products are being used at the national level, many of them based on novel technologies, with safety experience limited to pre-licensure clinical trials. Therefore, there is a public health need for comprehensive safety surveillance. Real-world safety monitoring of COVID-19 vaccines through observational studies should be implemented across Europe in a multi-layer approach by (i) Member States (ii) vaccine manufacturers and (iii) the Agency, to complement its routine pharmacovigilance activities.

8. Goal and objectives

8.1 Goal

In order to complement spontaneous reporting systems for signal detection (routine pharmacovigilance) and other initial safety monitoring activities such as pharmacoepidemiological studies conducted or planned by different stakeholders, the Agency procured an early safety monitoring study through its framework contracts (Early-Covid-Vaccine-Monitor; EUPAS39798) which is conducted by the EU PE&PV research network and VAC4EU.

The **objective** of the COVID-Vaccine Monitor study is to rapidly assess **signals of** potential safety concerns emerging from active surveillance and identified by PRAC. Rapid signal assessment means the collection of additional information in order to further characterise the incidence of the safety concern in comparison to its expected incidence in **non-vaccinated populations or with suitable comparator populations**.

8.2 Objectives

The objectives are divided into two phases, the first phase is the readiness phase. This will be conducted by all 10 DAPs, and be the basis for the selection to participate in real studies as well as the basis for the methods group (WP4) to assess the impact of methodological choices and assumptions using the study designs with negative controls.

1. *Readiness phase conducted wit hall DAPs*

The readiness phase will include the following objectives:

- To provide an overview of the methods and results for identification of COVID-19 vaccine exposure in the data sources
- To monitor the number of individuals exposed to any COVID-19 vaccine and to compare this to COVID-19 vaccine exposure (benchmark: https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#uptake-tab).
- To quantitatively evaluate different algorithms to identify adverse events by provenance in electronic health care data
- To conduct time-to-onset analyses for the AESI with respect to time since vaccination
- To assess the association between the vaccines of interest and negative control using the SCRI to estimate systematic bias (unmeasured confounding), this will be performed by methods WP4
- To test the impact of (by WP4)
    - different comparators in the cohort design, by using the negative control outcomes
    - different censoring criteria in the cohort study
    - different control periods/duration for the SCRI
    - different algorithms to assess vaccine exposure (doses), events, and covariates

based on the analysis of negative control outcomes and quality checks by WP4.

2. *Rapid assessment studies requested by EMA with selected number of DAPs*

Primary objective

The primary objective for this rapid assessment study is to assess the potential association between the occurrence of specific AESIs and vaccination with COVID-19 vaccines within disease-specific risk periods in individuals exposed to the COVID-19 vaccines compared to other COVID-19 vaccine exposed individuals, or compared to a control window within the same individual.

Secondary objectives

The secondary objectives for the rapid assessments studies are:

- To assess the potential association between the occurrence of specific AESIs and vaccination with COVID-19 vaccines in the following subgroups
    - immunocompromised persons
    - persons with the presence of co-morbidities elevating the risk of serious COVID-19
    - persons with a history of diagnosed COVID-19 disease
    - pregnant women
    - age groups (<20 ,and 10-year age categories))
    - patients with a prior history (ever) of that event more than a year before.
    - Gender
- To conduct sensitivity analyses requested by methods group (WP4)

The following VAC4EU and/or EU PE&PV research network data access providers will be invited to participate in the readiness proposal

Table 2: Participating data access providers and datasources.

Country	Data Access Provider	Name Data source	Experience ConcePTION CDM v2.2	AESI experience	Active population	Type of data source	Sources for diagnoses	Pregnancy data	COVID-19 vaccine data	Lag time availability key outcome data
NL	PHARMO / UMCU	PHARMO	Yes	Yes (ACCESS)	6 million	Record linkage	GP, Hospital	Yes (perined) linkage, but requires specific approval	Yes, GP and potentially CIMS	Hospital: 1 year GP <3 month Perined: 1 year
ES	AEMPS	BIFAP	Yes	Yes (ACCESS)	10 million	GP medical records	GP & hospital (larger lag)	yes	Yes	GP: 2-6 months Hosp: 2-12 months
ES	IDIAPJGol	SIDIAP	Yes	Yes (ACCESS)	5.7 million	Record linkage	GP, hospital	Yes	Yes	3-6 months
ES	FISABIO	VID	Yes	Yes (ACCESS)	5 million	Record linkage	GP, hospital, emergency visits, specialist visits	Yes (mother-baby linkage is not available)	Yes	3-6 months
IT	SoSeTe	PEDIANET	Yes	Yes (ACCESS)	0.5 million	Pediatric medical record	Primary care	no	Not yet	< 3 months
IT	ARS Toscana	ARS data	Yes	Yes (ACCESS)	3.6 million	Record linkage	Hospital	Yes (those ending in delivery)	yes	3 months
IT	Lazio	Lazio data	No	No	5.8 million	Record linkage	Hospital, emergency visits	Yes (those ending in delivery)	yes	3 months
IT	INSPIRE srl	Caserta data	No	No	1 million	Record linkage	Hospital, emergency visits	yes (those ending in delivery)	Not yet	< 3 months
UK	Utrecht University	CPRD/HES GOLD	Yes	Yes (ACCESS)	16 million	GP & Hospital medical record	GP, Hospital	Yes	Yes, but no brands	3-6 months
NO	University Oslo	Norwegian	Yes	No	5 million	Record linkage	Hospital, outpatient, and GP	Yes	Yes	>6 months

9. Research methods

9.1 Study Design

The study will comprise a readiness phase, to assess whether the data source is fit for the purpose of vaccine studies. The pool of data sources that are ready can then be utilized for specific rapid assessment studies when requested by EMA.

9.1.1 Readiness phase

The readiness phase study period will start follow-up on 1 January 2019. The primary design will be a cohort study including all subjects with at least one day of follow-up after January 1, 2019, and at least 365 days of availability prior to that date.

In the readiness phase, data sources will

- Prepare the ETL design for the transformation of local data into the ConcePTION CDM (CCDM)
- Run level 1-3 quality checks on data required for all AESI and covariates, aiming at investigating the completeness (level 1), the logic of the converted data (level 2), and subsequently whether the data is fit for purpose, especially as regards vaccine and events data.
    - Level 3 checks include the generation of incidence rates for the events and covariates (2019-2020) by age and gender, using different algorithms. These data will also be utilized to further understand misclassification of outcomes and exposure by the methods group in WP4 task 4.5.
- Conduct the study designs with vaccine-negative control outcome pairs in collaboration with WP4 to:
    - develop and run analytical R-code that will be used for the negative control outcomes and can be re-used for the rapid assessment studies and its sensitivity analyses
    - To assess systematic bias and generate information to `assess the methodological developments by WP4 that may be incorporated in sensitivity analyses of rapid assessment studies
        - risk windows duration and timing (pre-post control window)(exposure misclassification)
        - residual confounding assessment and impact of different comparators such as
            - historic comparators before the COVID-19 pandemic,
            - contemporary unvaccinated comparators,
            - contemporary unvaccinated comparators identified during influenza vaccination,
            - historic influenza vaccinated comparators identified before the COVID-19 pandemic,
            - subjects vaccinated with a different COVID-19 vaccine.
        - outcome misclassification (using different algorithms for events)
        - censoring (left and right)

9.1.2 Overview of Study Design for rapid assessment studies

Rapid assessment of safety concerns will be conducted using a retrospective observational study using electronic health care databases that have gone through the readiness phase. Eligible individuals will be included in the study from the start of vaccination campaigns: 1 December 2020, and the study will end at the last date of data availability in each database.

For specific events of concern, the study design will depend on whether the event is considered acute or non-acute and follow the decision framework described in the ACCESS template protocols (EUPAS 39361)

The primary study design for acute events (events expected to occur within 60 days of vaccination) will be a self-controlled risk interval (SCRI) design and for non-acute events (events expected to occur or be diagnosed with delay, within 180 days) will be a cohort design with contemporary exposed comparators. Acute events may also be studied using the cohort design to address uncertainties around risk windows and limitations of the SCRI design. In the SCRI design, we will use a non-exposed control window.

Subjects start follow-up at time zero (time of vaccination or the start of the pre-vaccination control window for the SCRI) and end follow-up at the earliest of occurrence of latest data availability of the databank, subject exit, or the completion of the period, or death. At least one year of enrollment/ presence prior to time zero (cohort entry) will be required to determine whether individuals meet the study criteria and to define baseline characteristics.

9.1.2.1 Self-controlled Risk Interval Design

The SCRI design will compare the risk of the event of interest in a post-vaccination to a pre-vaccination control window within the same individual. We use a pre-vaccination risk window to allow for rapid assessment, since lag times may occur, we do not want to wait too long after COVID-19 vaccines to be able to analyse.

The implications of using a pre-vaccination control period will be investigated in a simulation study for the methodological development (WP4) of the project, and from these learnings, we will adapt the SAP. Key issues are:

- length and timing of the buffer period (to account for any healthy vaccine effect),
- contra-indication,
- death
- the use of exposed comparators.

Together with the methods group (WP4) in the CVM study, methods will be tested in the readiness phase and implemented in the rapid assessment phase where needed.

The SCRI design will include only individuals who received at least one dose of a COVID-19 vaccine during the study period and who experience the specific event in the control period or after vaccination

(starting date of vaccination). Study subjects will enter the study at the time of the start of the control window, which starts 90 days (as a default) before the date of vaccination with a COVID-19 vaccine. The SCRI design will compare the risk of each outcome during the risk window following dose 1 or dose 2 with the self-matched control interval, used to assess the baseline risk of the outcome. The control period will be 60 days long and is followed by a 30-day buffer/washout period, to account for healthy vaccinee effect, the length of the buffer period may be adapted based on on the assessment of the methods by WP4 and the specific event of interest. Depending on the index date of the case, the occurrence will be calculated in the control or risk window. If a case occurs in this buffer period, it will be kept in the study for sensitivity analyses on the buffer period.

The risk window post-vaccination starts at day 1 and will be divided into dose-specific risk intervals following each dose of the COVID-19 vaccine, except for anaphylaxis for which the risk interval starts at day 0. If a second dose is given within the risk interval of the first dose, the period of follow-up for the first dose will be censored. Sensitivity analyses will be conducted that included day 0 in the risk interval.

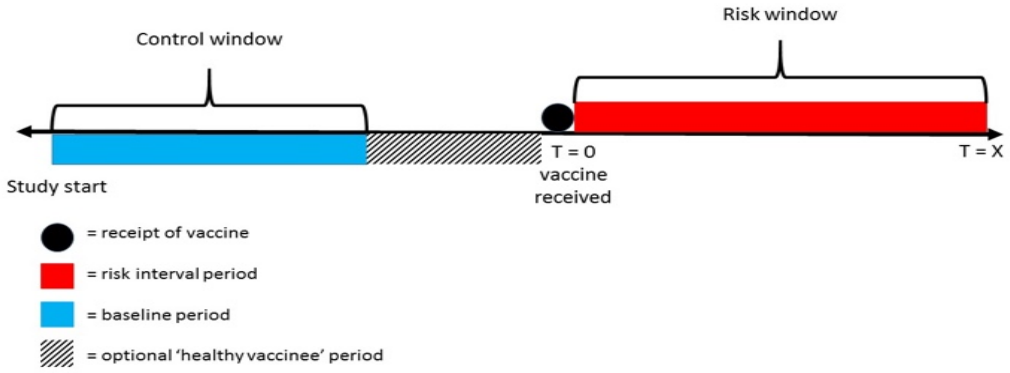


Figure 1*: Self-Controlled Risk Interval Design.*

9.1.2.2 Cohort Design With Concurrent comparator

A retrospective cohort design will be used to estimate the rate of non-acute events of interest after receipt of COVID-19 vaccination dose and compare this incidence primarily with that occurring in a COVID-19 vaccinated matched comparator group. Additional comparators will be tested for the methods work (e.g. non-exposed, influenza, historic) using the negative control events in the readiness phase.

For sensitivity analysis of acute events that initially will be analysed using an SCRI approach, the same cohort design may be chosen, if there is uncertainty about the risk window or direct comparison between different vaccines is needed.

Exposed cohort (index cohort): individuals who have received at least one dose of a specific COVID-19 vaccine.

Concurrently exposed cohort (reference cohort): individuals that have been vaccinated with another type of COVID-19 vaccine.

In this retrospective cohort design, time zero (cohort entry) will be defined as the time at which the exposure status is assigned, when selection criteria are applied, and when study outcomes start to be counted. Time zero (ie, recipients of the vaccine) will be the day the specific COVID-19 vaccination (index cohort) was received for anaphylaxis and date of vaccination +1 for other events of interest.

Persons in the index cohort will be individually matched to one individual in the concurrent cohort on key clinical variables (exact age, sex, and presence of one or more risk factors for severe COVID-19 [eg. cancer, sickle cell, obesity, chronic kidney disease, chronic respiratory disease, human immunodeficiency virus infection]) at time zero. In case there is no balance in several variables these may be included in a propensity score. The effect of propensity score matching or matching on key variables and adjusting for the other non-matched risk factors will be investigated in the methods development group. Individuals will be classified into exposure groups that are compatible with their data at time zero.

9.1.2.3 Rationale for Study Design

The AESIs included initially in this study have been defined and were included in the ACCESS and ECVM study (EUPAS40404), AESI can be divided into acute (<60 days), and non-acute events. For new events, an assessment of the design will be conducted according to the decision framework that was developed in ACCESS.

The SCRI design can only be applied to acute events (ie, AESIs expected to occur within a disease risk period of 60 days). It has the key advantage that it uses each individual as its own control, and thereby avoids potential confounding by factors that do not vary with time. The assumption for the SCRI design is that we can distinguish well between control and risk periods and that we know the risk windows.

Because the SCRI design is not appropriate for non-acute events, the cohort design with contemporary comparators will be used to assess these types of events (ie, AESIs expected to occur within a disease risk period of more than 60 days). Some of the participating data sources may not have complete information on COVID-19 vaccination and therefore misclassification may occur.

The impact of using a non-exposed comparator and an active comparator with incomplete information will be investigated by the methods group (WP4).

Disease-specific risk periods have been defined for all AESIs, however, these risk periods are based on currently available evidence from other vaccines/case reports and may not be strictly applicable to the COVID-19 vaccines. For this reason, sensitivity analyses with extended risk periods will be considered in the SCRI and cohort study for outcomes that do not have a well-defined risk interval (to be investigated in methods (WP4) work, and this will be defined in SAP).

9.2 Setting, Study Population and follow-up

9.2.1 Study Setting

For the implementation of the readiness study, 10 electronic health care databases in Northern, Southern and Western Europe that have shown interest, will be used. The data sources that were included are those who have been working in prior studies (EU PE&PV or VAC4EU) and were interested to participate

Italy
- ARS Toscana (Agenzia Regionale di Sanità della Toscana)
- Lazio region, Department of Epidemiology
- Pedianet (Societa Servizi Informatici)
- Caserta local health database (INSPIRE srl)

Netherlands
- PHARMO Database Network (PHARMO Institute for Drug Outcomes Research) (NL)

United Kingdom
- CPRD (Clinical Practice Research Datalink) & HES data (UK)

Norway
- The Norwegian health registers

Spain
- SIDIAP (Sistema d'Informació per el Desenvolupament de la Investigació en Atenció Primària)
- BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria)
- FISABIO (VID, Valencia health system Integrated Database)

Further information on the data sources used in this study can be found in Section 9.4

All data sources will be participating in the readiness phase.
For actual rapid assessment studies, choices will be made based on:
- Availability of fit for purpose data
- Sample size
- Ability to commit to timelines

9.2.2 Source population

The source population will be made up of all individuals registered in each of the participating healthcare data sources

9.2.3 Study Duration and Follow-Up

Readiness

For the readiness phase study, the study period will start on 1 January 2019 and will end on July 30st, 2021. Subjects will be followed from 1 January 2019 until the earliest of the following dates: death, end of data availability, subject exit, or the completion of the period. If persons have multiple periods within the same data source we will only use the latest.

Rapid assessment study

For the SCRI & cohort design, the study period will start on 1 December 2020 and last until the end of the study period.

- SCRI: Follow-up ends at the earliest of the following: end of data availability, subject withdrawal of the data sources, end of the duration of the risk period. For death and fatal events, specific additional criteria may be posed (pending methods group work)

- Cohort: The cohort design follow-up ends at occurrence of each AESI, death, end of data availability, subject exits the database, date that 6 months of follow-up is reached.

9.2.4 Inclusion Criteria

9.2.4.1 Readiness study

For the readiness study, the person will be included if there is at least one day of follow-up and the person has at least 12 months of data in the data source at the start of follow-up.

9.2.4.2 SCRI Design

For analyses of outcomes assessed with the SCRI design, the following criteria must be met. Note that the study population for each outcome-specific analysis will thus be different.

- Received a dose of COVID-19 vaccine during the study period.

- Have experienced a specific event of interest during the predefined observation periodError! Reference source not found..

- Have at least 12 months of data/registration in the data sources at study entry.

9.2.4.3 Cohort design

Individuals must meet all the following inclusion criteria to be eligible for inclusion in the cohort study:

- At time zero, being in the underlying population of the data source for at least 12 months; or, being born in the previous 12 months in the underlying population.
- Study participants must be eligible (eg, lack of contra-indications) to receive the COVID-19 vaccines at time zero.

9.2.5 Exclusion Criteria

For the readiness study, there will be no exclusion criteria.

Individuals will be excluded from the rapid assessment studies if:

- They have a recorded diagnosis for the specific event in the one year prior to cohort /SCRI entry. Persons with such acute diagnoses more than a year ago will be maintained to allow for subgroup analyses. Upon investigation of one event, we do not exclude any history or prevalence of other groups of events (AESIs).
- They have a contra-indication for one of the COVID-19 vaccines.

9.3 Variables

9.3.1 Exposure Assessment

Exposure will be based on available recorded prescription, dispensing, or administration of the COVID-19 vaccines. Vaccine receipt and date of vaccination will be obtained from all possible sources that capture COVID-19 vaccination, such as dispensing records, general practice records, immunisation registers, vaccination records or other data banks. The main exposure of interest for the rapid assessment studies is the receipt of COVID-19 vaccine.

- ARS Toscana (IT): ARS will identify vaccines from the regional immunization register using the nationally used product code, including batch number.

- Pedianet (IT): Information on COVID-19 vaccine will include date of immunisation, type of vaccine, vaccine batches, dose. Immunization data is not collected yet as children are not vaccinated at this moment.

- Lazio (IT): DEP Lazio will identify vaccines from the regional immunization register using the nationally used product code, including batch number.

- PHARMO (NL): Data on vaccination will be obtained from PHARMO's GP database. Information on vaccines include ATC code, brand, batch, and date of administration/recording. Several COVID-19 vaccines have been administered through other routes and original immunization data are not yet linked with GPs, this may change in the future.

- Caserta LHU database (IT): Caserta LHU record linkage database contains information from all claims databases (e.g. hospitalizations, drug dispensing, etc.) of Caserta province catchment area (around 1 million population). In addition, those claims data can be linked to the local immunization registry which includes name and batch of the vaccine; manufacturing company; dose; administration route; administration location (eg, general practice); date of administration.

- CPRD (UK): The CPRD contains information recorded by National Health Service (NHS) primary care general practitioners (GPs); and information on the administration of COVID-19 vaccines to individuals will become available. This will include, alongside an encrypted unique patient identifier; the name of the vaccine; manufacturing company; dose; stage of the vaccine schedule; administration route; administration location (eg, general practice); batch identifiers/numbers; date of administration; and GP prior to, on, or after the vaccination date. In addition, patient demographic, practice-level, and staff-level information will also be available.

  Furthermore, other CPRD-linked COVID-19 data sets that may provide further follow-up information on AESI include the PHE COVID-19 Hospitalisation in England Surveillance System and the Intensive Care National Audit and Research Center data on COVID-19 intensive care admissions. Standard CPRD-linked data sets will also be obtained including Hospital Episode Statistics (HES) data sets covering hospital secondary care (Accident & Emergency, Admitted

Patient Care, Inpatient and Outpatient), and the Office for National Statistics data sets for Death Registry information, Mother-baby Link.

- Norwegian health registers (NO): The national, electronic immunisation register (SYSVAK) was established in 1995 and records an individual's vaccination status and vaccination coverage in Norway. All vaccinations are subject to notification to SYSVAK and are registered without obtaining patient consent. This applies to all COVID-19 vaccines. In SYSVAK, the following data are registered: individual personal identifier, vaccine name and Anatomical Therapeutic Chemical (ATC) code, vaccine batch number, date of vaccination, reason for vaccination as health care professional versus risk-group patient, and the centre where the vaccine was administered.

- SIDIAP (ES): SIDIAP will have available information on the administration of COVID-19 vaccines to individuals linked to a unique and anonymous identifier. The information will be originated from the electronic medical records. For each patient, SIDIAP will have date and centre of administration, health professional administering the vaccine, dose, brand, reasons for vaccination (eg, risk group), and other information related to vaccination.

- BIFAP (ES): BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atencion Primaria), a computerized database of medical records of primary care (www.bifa p.aemps.es) is a non-profit research project funded by the Spanish Agency for Medicines and Medical Devices (AEMPS). Data on vaccination with COVID-19 vaccines are obtained from the COVID-19 vaccination registries in the participating regions and linked to the primary care medical records in BIFAP. Date of vaccination, brand, batch, and dose are registered.

- FISABIO (ES): Data on vaccine exposure will be obtained from the Vaccine Information System (VIS), which includes information on vaccine type, manufacturer, batch number, number of doses, location and administration date.

The vaccination strategies for the different exposure groups will be defined as follows:

- Subjects who receive a first dose of a specific COVID-19 vaccine will be classified as exposed to D1 for that specific vaccine (if brand is unknown it will be unknown).

- Subjects who receive a second dose of COVID-19 vaccine will be censored for the first dose risk window, and move into the risk window of the second dose for a COVID-19 vaccine, by brand for both the cohort as well as the SCRI design.

The vaccination strategy for the matched reference cohort(s) in the rapid assessment study will be defined as

- Pfizer
- Moderna
- Janssen
- AstraZeneca
- Unknown

depending on the type of index cohort. For subjects who receive more than one type at the same day, we will make a mixed category. For the unknown category we will assess with the methods group whether imputation is possible based on the vaccination role out time, and vaccination group.

In the readiness study on negative control outcomes, different types of comparator cohorts will be tested for assessment of impact (Historic comparators before the COVID-19 pandemic, matched on seasonality or calendar month (EHRs), Contemporary unvaccinated comparators, matched on calendar time (EHRs), Contemporary unvaccinated comparators identified during influenza vaccination, may be challenging given the current vaccination periods (EHRs), Historic influenza vaccinated comparators identified before the COVID-19 pandemic, matched on relevant characteristics (EHRs), subjects vaccinated with a different COVID-19 vaccine (EHRs or from event monitoring cohort).

For the rapid assessment SCRI design, person-time in the risk interval will be considered "exposed" while person-time in the control interval will be considered "unexposed." Risk intervals are specific to the outcome of interest.

9.3.2 Study Outcomes

AESIs, as listed below (Table 3) and in line with the definitions and code lists that have been created for the ACCESS project (further defined in the SAP), will be identified, with a date of diagnosis, using algorithms (as created in readiness phase), based on diagnosis codes (with the procedure and/or pharmacy dispensing codes and/or limited to specific medical care settings if applicable to the outcome). During the readiness phase, the impact of the provenance of information on outcomes, as well as different algorithms, will be assessed by WP4. In case a new signal arises the protocol will be amended and the event will be included. For new events, level 3 checks (including incidence rates) will be conducted.

Table 3: List of AESI and the negative control events, design and primary risk period duration

Event	ACCESS	SCRI	cohort	Naïve period to estimate new onset	Primary Risk period*
Multisystem inflammatory syndrome	✓	✓	✓	365 days	28 days
Acute respiratory distress syndrome	✓	✓	✓	365 days	28 days
Acute cardiovascular injury	✓	✓	✓	365 days	
microangiopathy	✓	✓	✓	365 days	28 days
CAD	✓	✓	✓	365 days	28 days
Arrhythmia	✓	✓	✓	365 days	28 days
Myocarditis	✓	✓	✓	365 days	28 days
Pericarditis	✓	✓	✓	365 days	28 days
Coagulation disorders, including deep vein thrombosis, pulmonary embolus, cerebrovascular stroke, limb ischaemia, haemorrhagic disease	✓				
VTE (DVT & PE & Splanchnic)	✓	✓	✓	365 days	28 days
CVST	✓	✓	✓	365 days	28 days
Arterial thrombosis	✓	✓	✓	365 days	28 days

Event	ACCESS	SCRI	cohort	Naïve period to estimate new onset	Primary Risk period*
TTS (VTE, arterial thrombosis, or CVST with thrombocytopenia in 10 days)	✓	✓	✓	365 days	28 days
Hemorrhagic stroke	✓	✓	✓	365 days	28 days
DIC	✓	✓	✓	365 days	28 days
Generalised convulsion	✓	✓	✓	365 days	14 days
Guillain Barré Syndrome	✓	✓	✓	365 days	42 days
Diabetes (type 1 and unspecified type)	✓		✓	365 days	180 days
Acute kidney injury	✓		✓	365 days	180 days
Acute liver injury	✓		✓	365 days	180 days
Anosmia, ageusia	✓	✓	✓	365 days	28 days
Chilblain-like lesions	✓	✓	✓	365 days	28 days
Single organ cutaneous vasculitis	✓	✓	✓	365 days	28 days
Erythema multiforme	✓	✓	✓	365 days	7 days
Anaphylaxis	✓	✓	✓	30 days	2 days
Death (any cause)** (postvaccination control window)	✓	✓	✓	365 days	7 days
Sudden death (by codes)** (postvaccination control window)	✓	✓	✓	365 days	7 days
Meningoencephalitis	✓	✓	✓	365 days	28 days
Acute disseminated encephalomyelitis (ADEM)	✓	✓	✓	365 days	28 days
Narcolepsy	✓		✓	365 days	180 days
Thrombocytopenia	✓	✓	✓	365 days	28 days
Transverse myelitis	✓	✓	✓	365 days	28 days
Bells' palsy		✓	✓	365 days	28 days
Haemophagocytic lymphohistiocytosis[4]		✓	✓	365 days	180 days
Kawasaki's disease		✓	✓	365 days	28 days
Pancreatitis		✓	✓	365 days	28 days
Rhabdomyolysis		✓	✓	365 days	28 days
SCARs		✓	✓	365 days	28 days
Sensorineural hearing loss			✓	365 days	180 days
Thyroiditis			✓	365 days	180 days
Negative control event					
Gout		✓	✓	365 days	28 days
Otitis externa		✓	✓	365 days	28 days
Trigeminal neuralgia		✓	✓	365 days	28 days

*For death we may conduct different SCRI analyses

[4] https://primaryimmune.org/disease/hemophagocytic-lymphohistiocytosis-hlh

9.3.3 Covariate Definition

Readiness study

In the readiness study covariates (as listed below for the rapid assessment study) will be extracted and inspected for algorithms and for methodological analysis.

Rapid assessment study

Time-varying variables for the SCRI design will be measured at the start of the risk period and control window for time-varying factors. For the cohort design and SCRI, covariate status for others factors will be measured at time zero. All covariates will be assessed in specific periods, default is during the one year prior. For short-term time varying confounders (e.g. antibiotics) we will address whether smaller exposure periods would be required in the WP4 methods group. Some of these covariates may be intermediates for certain events, with WP4 we will describe how to assess these in the SAP

Population characteristics will be identified based on diagnoses, medicines, laboratory data, survey observation or medical observations, and observation period information.

Demographic characteristics

- Age
- Sex

Pregnancy

- Pregnancy status and pregnancy trimester at time zero (if this can be measured in the datasources)

Comorbidities

- Cancer
- Chronic kidney disease (exclusion for acute kidney injury)
- Coronary artery disease (exclusion for cardiac/cardiovascular events)
- Chronic respiratory disease (chronic obstructive pulmonary disease, asthma)
- Obesity: BMI>30 kg/m^2
- Down syndrome
- Type 1 (excluded for Type 1 when AESI) or Type 2 diabetes
- prior VTE
- Morbidity index: number of different ATC codes (level 7) dispensed in year prior to cohort entry
- History of anaphylaxis
- History of any type of allergic reaction
- Immunocompromising conditions (will be used to define subgroups for secondary analyses)
    - Immunodeficiencies
    - Systemic Immunosuppressant medication use
    - Human immunodeficiency virus and other immunosuppressing conditions

Covid-19 History

- Prior recorded COVID-19 infection by severity (non-hospitalized, hospitalized)

Comedication (dispensed/prescribed)

- Analgesics (N02, month prior, timevarying for SCRI)
- Systemic Corticosteroids (H02, month prior, timevarying for SCRI)
- Antithrombotic agenst (B01A month prior, timevarying for SCRI)
- Aspirin (month prior, timevarying for SCRI)
- Sex hormones (G03) year prior
- Immunosuppressants (L04A*, H02*, timevarying for SCRI)
- Antihypertensive meds (C02*, C03*, C07*, C08*, C09*, year prior)
- Use of diabetes medications (year prior)
- Antibiotics (month prior, timevarying for SCRI) (J01)
- Antiviral medications (month prior timevarying for SCRI) (J05)
- Non-steroidal anti-inflammatory drugs (month prior timevarying for SCRI) (M01)
- Psychotropics (N03, N04, N05, N06, year prior)
- Lipid lowering drugs (year prior, C10)
- Obesity meds (A08A*, year prior)
- Cardiovascular meds (C01B*, C01C*, C01D* C01E*, year prior)
- Sickle cell meds (L01XX05, B06AX*, year prior),
- bronchodilating meds (R03*, R07AA*, R07AB*, year prior),
- diabetes meds (A10B*, A10A*, year prior)
- cancer meds (L01*, L02*, L03*, L04*, year prior),

The AESI may have different sets of risk factors, and outcome-specific analyses may contain different covariate sets. Potential covariates may include the following information, as available in each data source: For the SCRI design, covariate status will be measured at the start of the risk period and control window for time-varying factors, impact smaller periods for rapidly changing covariates will be investigated by the methods group.

For subgroup analyses, we will use the following groups

- immunocompromised persons (yes/no)
- persons with the presence of co-morbidities elevating the risk of serious COVID-19 (yes/no)
- persons with a history of diagnosed COVID-19 disease (yes/no)
- pregnant women
- age groups (<20 ,and 10-year age caregories))
- patients with a prior history (ever) of that event more than a year before.
- gender

9.4 Data Sources

The study will use data from secondary electronic health record databases that are population-based. All data sources will have the ability to provide data on COVID-19 vaccines, outcomes (diagnoses, procedures, and treatments), and important covariates. It is not currently known the extent to which COVID-19 vaccines, product types, and batch numbers will be captured well in the data sources.

9.4.1 PHARMO (NL)

The PHARMO Database Network, which is maintained by the PHARMO Institute for Drug Outcomes Research, is a population-based network of electronic health record databases that combines anonymous data from different primary and secondary health care settings in the Netherlands. These different data banks—including data from general practices, in- and outpatient pharmacies, clinical laboratories, hospitals, the cancer register, the pathology register, and the perinatal register—are linked on a patient level through validated algorithms. To ensure data privacy in the PHARMO Database Network, the collection, processing, linkage, and anonymisation of the data are performed by STIZON, which is an independent, ISO/IEC 27001 certified foundation that acts as a trusted third party between the data sources and the PHARMO Institute. The longitudinal nature of the PHARMO Database Network enables the follow-up of more than 9 million individuals of a well-defined population in the Netherlands for an average of 12 years. Currently, the PHARMO Database Network covers over 6 million active individuals out of 17 million inhabitants of the Netherlands. Data collection period, catchment area, and overlap between data banks differ. Therefore, the final cohort size for any study will depend on the data banks that are required. All electronic patient records in the PHARMO Database Network include information on age, sex, socioeconomic status, and mortality. Other available information depends on the data banks. A detailed description of the different data banks is given in subsequent sections. The PHARMO Institute is always seeking new opportunities to link with additional databanks, currently it is exploring linkage with the COVID-19 immunization register that is collected by RIVM. The PHARMO Database Network is listed under the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database.

The General Practitioner databank comprises data from electronic patient records registered by GPs. The records include information on diagnoses and symptoms, laboratory test results, referrals to specialists, and health care product/drug prescriptions. The prescription records include information on type of product, prescription date, strength, dosage regimen, quantity, and route of administration. Drug prescriptions are coded according to the WHO ATC coding system. Diagnoses and symptoms are coded according to the International Classification of Primary Care (ICPC) [www.nhg.org], which can be mapped to the International Classification of Diseases (ICD) codes but can also be entered as free text. General practitioner data cover a catchment area representing 3.2 million residents (~20% of the Dutch population). PHARMO GP databank captures vaccinations supplied by the GP (influenza, zoster, COVID-19).

The Netherlands Perinatal Registry is maintained by Perined and comprises data on pregnancies, births and neonatal outcomes of births in the Netherlands, voluntarily collected by perinatal caregivers, mainly for benchmarking. For the current study permission has been obtained from PHARMO as well as Perined to link the data with the PHARMO Database Network via the TTP and use the PHARMO Perinatal Research Network (PPRN).

9.4.2 Clinical Practice Research Datalink and Hospital Episode Statistics (UK)

The 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 are 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 when CPRD GOLD (General Practitioner Online Database) and CPRD Aurum versions are used. There are currently approximately 59 million individuals (acceptable for research purposes) -16 million 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 (eg, 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 the Drug Safety Research Unit (DSRU). The HES database contains details of all admissions to NHS hospitals in England (Accident & Emergency, Admitted Patient Care, Outpatients); approximately 44.6 million individuals in the CPRD are linked to the HES database. Not all patients in the CPRD have linked data (eg, if they live outside England, if their GP has not agreed that their data should be used in this way). As with standard CPRD patients, HES data are limited to patients who are research standard. CPRD records are linked to the HES using a combination of the patient's NHS number, sex, and date of birth. Additional CPRD-linked data sets include Death Registration data from the Office of National Statistics, which includes information on the official date and causes of death (using ICD codes), Mother-baby Link, and an algorithm-based Pregnancy Register. In addition, other CPRD-linked COVID-19 data sets, which may provide further follow-up information on AESI, include the Public Health England (PHE) Second Generation Surveillance System (SGSS) COVID-19 positive virology test pillar 1 tests, PHE COVID-19 Hospitalisation in England Surveillance System, and the Intensive Care National Audit and Research Centre data on COVID-19 intensive care admissions.

9.4.3 Norwegian Health Registers (NO)

The Norwegian data sources in this project are several national health registers, ie, the Medical Birth Registry of Norway (MBRN), the National Patient Register (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Immunisation Registry (SYSVAK), the National Prescription Registry, and Statistics Norway. The source population will be identified using the Norwegian Institute of Health's (NIPH) copy of the Norwegian population data file from the National Registry. The NPR and KUHR (and the MBRN for the pregnant population) provide data on inpatient and outpatient diagnostic codes. Information on population background data is derived from Statistics Norway (eg, education, occupation status, sex, age). Data on vaccination status are derived from SYSVAK and the Norwegian Prescription Database. The latter register includes data on filled prescriptions for possible co-medications and other prescription drug use.

Norwegian Immunisation Registry

The SYSVAK is the national electronic immunisation register that records an individual's vaccination status and vaccination coverage in Norway. It became nationwide in 1995, and includes information such as personal identity number, the vaccine code, disease vaccinated against, and vaccination date.

The Norwegian Patient Registry

The NPR is an administrative database of records reported by all government-owned hospitals and outpatient clinics and by all private health clinics that receive governmental reimbursement. The NPR contains information on admission to hospitals and specialist health care on an individual level from 2008. The data include date of admission and discharge as well as primary and secondary diagnosis. The NPR has included Norwegian national identification numbers since 2008. Consequently, person-specific data from 2008 onwards are available. Diagnostic codes in the NPR follow ICD-10.

Norway Control and Payment of Health Reimbursement

The KUHR is an administrative database based on electronically submitted reimbursement claims from physicians to the Norwegian Health Economics Administration. It contains information from primary health care, GP, and emergency services on morbidity, utilisation of health care services, and health care use. Person-specific data are available since 2006 . Diagnostic codes in the KUHR follow ICD-10, but the ICPC is more frequently used by GPs.

The Norwegian Prescription Database

Since January 2004, all pharmacies in Norway have been obliged to send data electronically to the Norwegian Institute of Public Health regarding all prescribed drugs (irrespective of reimbursement) dispensed to individuals in ambulatory care. Relevant variables for this project include detailed information on drugs dispensed and date of dispensing.

The Medical Birth Registry of Norway

The MBRN is a population-based register containing information on all births in Norway since 1967 (more than 2.3 million births). The MBRN is based on mandatory notification of all births or late abortions occurring at 12 weeks of gestation onwards. The MBRN includes identification of the mother and father, including national identification numbers, parental demographic information, the mother's health before and during pregnancy, complications during pregnancy and delivery, and length of pregnancy, as well as information on the infant, including congenital malformations and other perinatal outcomes.

Statistics Norway

Statistics Norway provides microdata for research projects and includes information on population characteristics, housing conditions, education, income, and welfare benefits. These data are potential important confounders.

The National Registry

The National Registry (Folkeregisteret) holds information about all inhabitants in Norway. The NIPH holds a copy of the Norwegian population data file from the National Registry that will be used to identify the source population in Norway.

Norwegian Surveillance System for Communicable Diseases

Notification of infectious diseases to the Norwegian Surveillance System for Communicable Diseases (MSIS) is an important part in the surveillance of infectious diseases in Norway. Microbiological laboratories analysing specimens from humans, and all doctors in Norway, are required by law to notify cases of certain diseases (71 in total including SARS-CoV-2) to the MSIS central unit at the Norwegian

Institute of Public Health. The following variables are available since 1977: notifiable disease, month and year of diagnosis, age groups, county of residence, and place of infection. Data on positive COVID-19 tests are updated continuously.

9.4.4 SIDIAP (ES)

The Information System for the Improvement of Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària' [SIDIAP]; www.sidiap.org) was created in 2010 by the Catalan Health Institute and the IDIAPJGol Institute. It includes information collected since 01 January 2006 during routine visits at 278 primary care centres pertaining to the Catalan Health Institute in Catalonia (North-East Spain) with 3,414 participating GPs. SIDIAP has pseudo-anonymised records for 5.7 million people (80% of the Catalan population) and is highly representative of the Catalan population. The SIDIAP data comprise the clinical and referral events registered by primary care health professionals (eg, 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. The SIDIAP data can also be linked to other data sources, such as the hospital discharge database, on a project-by-project basis. Health professionals gather this information using ICD-10 codes, ATC codes, and structured forms designed for the collection of variables relevant for 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 and urine test results. Regarding vaccinations, SIDIAP includes all routine childhood and adult immunisations, including the antigen and the number of administered doses. Encoding personal and clinic identifiers ensures the confidentiality of the information in the SIDIAP database. Currently, with the COVID-19 pandemic, there is the possibility to have shorter term updates in order to monitor the evolution of the pandemic. Recent reports have shown the SIDIAP data to be useful for epidemiological research. SIDIAP is listed under the ENCePP resources database.

9.4.5 BIFAP database (ES)

BIFAP (Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria) is a longitudinal population-based database of EMRs from patients attended in primary care facilities of the SNS (Sistema Nacional de Salud), the Spanish National Health System, and located in one of the participating regions throughout Spain. Since 2001, this database has been progressively and increasingly collecting health data, with annual updates, and the current complete version of the database with information until December 2019 includes clinical data of 10.153 Primary Care Practitioners (PCPs) and pediatricians. Nine participant Autonomous Region send their data to BIFAP every year. BIFAP database currently includes anonymized clinical and prescription/dispensing data from more than 13.7 million (9.4 active population) patients representing 85% of all patients of those regions participating in the database, and 29% of the Spanish population. Mean duration of follow-up in the database is 8.7 years. Information collected by PCPs includes administrative data, socio-demographic data, lifestyle, and other general data, clinical diagnosis and health problems, results of diagnostic procedures, interventions, and prescriptions/dispensations. Diagnoses are classified according to the International Classification of Primary Care (ICPC)-2 and ICD-9 code system, and a variable proportion of clinical information is registered in "medical notes" in free text fields in the EMR. Additionally, information on hospital discharge diagnoses coded in ICD-10 terminology is linked to patients included in BIFAP 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.

BIFAP has now been updated until 2021, February 28th including information on COVID-19 diagnosis and vaccines for several regions, and it will progressively and regularly be updated.

The BIFAP database was characterized in the ADVANCE project and considered fit for purpose for vaccine coverage, benefits and risk assessment (Sturkenboom et al. 2020). The BIFAP program currently participates in several European projects financed by the EMA, the main objective of some of them is to contribute to the surveillance of vaccine safety against COVID-19: ACCESS ("VACcine Covid-19 tracking readinESS") and "Early-Covid-Vaccine-Monitoring".

9.4.6 FISABIO, VID database (ES)

The VID is a set of population-wide electronic databases covering residents of the Valencia region in Spain, representing approximately 5 million individuals (Garcia-Sempere et al 2020). All the information in the VID databases can be linked at the individual level through a single personal identification. The data sets in the VID are as follows:

The Population Information System (SIP) is a database that provides basic information on health system coverage (eg, dates and causes of Valencia health system entitlement or disentitlement, insurance modality, pharmaceutical copayment status, assigned Healthcare Department) as well as some sociodemographic data (eg, sex, date of birth, nationality, employment status, geographic location). Importantly, the SIP database includes the date of death captured from the Mortality Registry. The SIP database is paramount to the VID, as it is the source of the individual, exclusive, and permanent identifier number associated with each individual (the SIP number), which is then used throughout the rest of the databases, thereby allowing data linkage across the multiple databases in the network.

The Ambulatory Medical Record (ABUCASIS) is the electronic medical record for primary and specialised outpatient activity, with 96% population coverage since 2009. ABUCASIS is integrated by two main modules: the Ambulatory Information System (SIA) and the Pharmaceutical Module (GAIA), including paediatric and adult primary care, mental health care, prenatal care, and specialist outpatient services, as well as providing information about dates, visits, procedures, laboratory test results, diagnoses, and clinical and lifestyle information. It also includes information on several health programmes (eg, healthy children, vaccines, pregnancy, notifiable diseases), the primary care nurse clinical record, and the health-related social assistance record. The SIA module uses the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) for coding diagnoses (and, partially, ICD-10-ES from 2019). The SIA also uses the Clinical Risk Groups system to stratify the morbidity of the entire population.

The GAIA Pharmaceutical module stores data on all outpatient pharmaceutical prescriptions and dispensings, including both primary care and outpatient hospital departments, using the Anatomical Therapeutic Chemical (ATC) classification system and the National Pharmaceutical Catalogue, which allow the identification of the exact content of each dispensing. GAIA does not include in-hospital medication or medication administered in the Accident and Emergency Department (AED). GAIA provides detailed information on prescriptions issued by physicians, such as the duration of treatment and dosage.

The Hospital Medical Record (ORION) provides comprehensive information covering all areas of specialised care, from admission, outpatient consultations, hospitalisation, emergencies, diagnostic services (eg, laboratory tests, imaging, microbiology, pathology), pharmacy, surgical block including day

surgery, critical care, prevention and safety, social work, at-home hospitalisation, and day hospitalisation. ORION is currently in the process of being integrated for the whole region, with several databases already fully integrated and available for all hospitals, including the Minimum Basic Data Set at Hospital Discharge (MBDS) and the AED clinical record.

The MBDS is a synopsis of clinical and administrative information on all hospital admissions and major ambulatory surgery in the Valencia health system hospitals, including public-private partnership hospitals (approximately 450,000 admissions per year in the region). The MBDS includes admission and discharge dates, age, sex, geographic area and zone of residence, main diagnosis at discharge, up to 30 secondary diagnoses (comorbidities or complications), clinical procedures performed during the hospital episode, and the diagnosis-related group(s) assigned at discharge. The MBDS used the ICD-9-CM system for coding through December 2015 and ICD-10-ES afterwards. The MBDS was extended in 2015 to include the "present on admission" diagnosis marker and information on tumour morphology.

The AED clinical record was launched in 2008 and collects triage data, diagnoses, tests, and procedures performed in public emergency departments. As with the MBDS, the coding system used the ICD-9-CM until December 2015 and the ICD-10-ES thereafter. Diagnosis codification has been increasing from approximately 45% of all emergency department visits between 2008 and 2014 up to approximately 75% in 2017, largely due to the progressive incorporation of hospital coding.

Data on vaccine exposure may be obtained from the Vaccine Information System (VIS), which includes information on vaccine type, manufacturer, batch number, number of doses, location and administration date, adverse reactions related to vaccines, and if applicable, risk groups. Information in the VIS is updated daily.

All databases included in the VID are updated frequently (every 1 to 3 months), except the MBDS database, which is updated every 6 months.

9.4.7 ARS Toscana Database (IT)

The Italian National Healthcare System is organised at the regional level: the national government sets standards of assistance and tax-based funding for each region, which regional governments are responsible for providing to all their inhabitants. Tuscany is an Italian region, with approximately 3.6 million inhabitants. The Agenzia Regionale di Sanità della Toscana (ARS Toscana) is a research institute of the Tuscany region. The ARS Toscana database comprises all information collected by the Tuscany region to account for the health care delivered to its inhabitants. Moreover, ARS Toscana collects data from regional initiatives. All data banks in the ARS Toscana data source can be linked at the individual level through a pseudo-anonymous identifier. Two data banks collect dispensings of reimbursed medicines from, respectively, community pharmacies and hospital pharmacies. In the latter data bank, dispensings for outpatient and ambulatory use are complete, and dispensings for inpatient use are partial. Other data banks include hospital discharges, emergency care admissions, records of exemptions from copayment, diagnostic tests and procedures, causes of death, the mental health services register, the birth register, the spontaneous abortion register, and the induced terminations register. A pathology register is available, mostly recorded in free text, but with morphology and topographic SNOMED codes. A COVID-19 registry including all positive cases with clinical follow up is also available. Mother-child linkage is possible through the birth register. Vaccination data are available for children since 2016 and for adults since 2019. All the data banks can be linked at the individual level

through a pseudonymous identifier. Data banks are updated approximately every 2 months. Some of them are updated at the date of transmission (eg, vaccines, COVID-19 registry, access to emergency room), others (eg medicines dispensings and hospital discharge records) have a delay of approximately 4 months.

9.4.8 Lazio regional database (IT)

Lazio is an Italian region, with approximately 5.8 million inhabitants. The Department of Epidemiology, ASL Roma1 (DEP Lazio) is a department of the Local Health Authority ASL Roma1, recognised as a regional reference center for epidemiological services and research 35ort he Lazio Regional Health Service.

DEP Lazio has access to data collected in the regional administrative healthcare databases referring to mortality, hospital discharge records, emergency room visits, co-payment exemptions, drug claims for outpatients from community and hospital pharmacies, and ambulatory specialist visits. A COVID-19 case registry and COVID-19 vaccine registry are also available.

All data are collected at patient level and can be linked between databases through a pseudo-anonymous identifier.

Data are updated with different lag times, and delays vary between 2 weeks and 6 months.

9.4.9 Caserta LHU database (IT)

The Caserta database is a claims database containing patient-level data from the city of Caserta, in the Campania region. The coverage of this database is very high: from 2005-2020 the catchment area population in Caserta consists of more than 1 million persons (15% of the Campania regional population). The Caserta linkage databases consists of several databases which are linked through a unique patient identifier: a demographic registry, pharmacy claims database with information on concerning all dispensed drugs reimbursed by the Italian NHS, a as well as hospital discharge diagnose databases, emergency department admissions database, claims for diagnostic and laboratory tests ordered, and a registry of patients exempt from reasons for healthcare service co-payment exemptions (e.g. diabetes mellitus, dementia, and other chronic diseases), emergency department visit diagnoses and diagnostic tests. Patient level data from these claims databases, including other drugs reimbursed by the NHS and dispensed by community pharmacies, can be linked together, using a unique patient identifier. The healthcare information in the databases is coded using international coding systems, such as International Classification of Diseases, 9th Edition (ICD 9 CM) for diagnoses and Anatomic Therapeutic and Chemical (ATC) classification for drugs.

A COVID-19 registry including all positive cases with clinical follow up is also available.

9.4.10 PEDIANET (IT)

PEDIANET, a pediatric general practice research database, contains reason for accessing healthcare, health status (according to the Guidelines of Health Supervision of the American Academy of Pediatrics), demographic data, diagnosis and clinical details (free text or coded using the ICD-9 CM), prescriptions (pharmaceutical prescriptions identified by the ATC code), specialist appointments, diagnostic procedures, hospital admissions, growth parameters and outcome data of the children habitually seen by about 140 family pediatricians (FPs) distributed throughout Italy.

PEDIANET can link to other databases using unique patient identifiers. In the first database, information on routine childhood vaccination are captured including vaccine brand and dose. In the second database, information on patient hospitalization date, reason for hospitalization, days of hospitalizations and discharge diagnosis (up to six diagnosis) are captured. The FPs participation in the database is voluntary and patients and their parents provide consent for use of their data for research purposes. In Italy each child is assigned to a FP, who is the referral for any health visit or any drug prescription, thus the database contains a very detailed personal medical history. The data, generated during routine practice care using common software (JuniorBit®), are anonymized and sent monthly to a centralized database in Padua for validation. The PEDIANET database can be linked to regional vaccination data which was successfully tested in the ADVANCE project where it was characterized and deemed fit for purpose for pediatric routine vaccines (Sturkenboom et al., 2020).

9.5 Study Size

The study will be conducted in a source population of 45 million individuals.

Cohort

Table 4 shows the statistical power that can be obtained for a range of relative risks and a range of population sizes for a matching ratio index/reference 1:1. For example, 100,000 individuals in the index cohort and 100,000 individuals in the reference cohort will allow the detection of a relative risk equal to or greater than 2 with 80% statistical power for diseases with a background incidence rate of ≥100 per 100,000 person-years.

Table 4: Statistical Power for Cohort Design Based on Incidence and Relative Risk.

Number of exposed individuals*	Incidence rate in reference cohort (cases per 100,000 personyears)	Relative risk						
		1.5	2	3	4	5	7	10
50,000	1	3.89	4.88	6.21	7.17	7.99	9.63	12.57
	10	6.17	10.68	21.69	34.84	49.10	75.34	95.86
	50	12.78	30.99	71.98	93.98	99.34	100.00	100.00
	100	20.20	52.57	94.59	99.87	100.00	100.00	100.00
100,000	1	4.26	5.73	8.19	10.42	12.70	17.82	27.59
	10	7.99	16.03	37.24	60.29	79.26	97.06	99.98
	50	20.20	52.55	94.58	99.87	100.00	100.00	100.00
	100	34.22	80.42	99.89	100.00	100.00	100.00	100.00
200,000	1	4.84	7.14	11.74	16.68	22.18	35.07	57.36
	10	11.23	26.14	62.59	88.12	97.73	99.98	100.00
	50	34.20	80.40	99.89	100.00	100.00	100.00	100.00
	100	57.87	97.60	100.00	100.00	100.00	100.00	100.00

Source: Rothman, 2015
*Assuming each individual contributes a 60-day risk window

SCRI

Table 5 shows the statistical power that can be obtained for a range of relative risks and a range of sample sizes. For example, a sample size of 100 cases in the risk or control period will allow the detection of a relative risk equal to or greater than 2 with 93% statistical power. The methods group will assess the impact of the length of the control period on the power.

Table 5: Detectable Relative Risk and Statistical Power for SCRI Design.

Relative Risk	Sample Size*	Power
1.5	20	0.142
2	20	0.320
2.5	20	0.495
3	20	0.638
1.5	50	0.292
2	50	0.667
2.5	50	0.881
3	50	0.963
1.5	100	0.519
2	100	0.926
2.5	100	0.994
3	100	1.000
1.5	150	0.692
2	150	0.987
2.5	150	1.000
3	150	1.000

1.5	200	0.812
2	200	0.998
2.5	200	1.000
3	200	1.000

*sample size = number of events in risk and control period

Table 6 shows the number of vaccinated subjects assuming the same time in control and risk period to obtain an 80% statistical power for a range of relative risks. For example, a sample size of 69 vaccinated individuals with an AE of interest will allow the detection of a relative risk equal to or greater than 2 with 80% statistical power.

Table 6: Detectable Relative Risk and Sample Size for 80% Statistical Power for SCRI Design

Relative Risk	Subjects with AE of Interest	Power
1.5	195	0.802
2	69	0.805
2.5	41	0.809
3	29	0.804

9.6 Data Management

This study is conducted in a distributed manner using a common protocol, common data model (CDM), and common analytics programs (Figure 2). The data pipeline has been developing from the EU-ADR project and was further improved in the IMI-ConcePTION project (https://www.imiconception.eu/) and used to generate background rates in the ACCESS project. This process maximizes the involvement of the data providers in the study by utilizing their knowledge on the characteristics and the process underlying the data collection which makes analysis more efficient.

9.6.1 Data Extraction

Each database access provider (DAP) creates extraction, transform, and load (ETL) specifications using the standard ConcePTION ETL design template (accessible via this link: https://docs.google.com/document/d/1SWi31tnNJL7u5jJLbBHmoZa7AvfcVaqX7jiXgL9uAWg/edit) and upload it to the VAC4EU FAIR Catalogue.

The most recent version of the ConcePTION CDM will be used for this analysis.

Following completion of this template and review with study statisticians and principal investigators, each DAP extracts the relevant study data locally using their software (eg, Stata, SAS, R, Oracle). This data is loaded into the CDM structure in csv format. These data remain local (Error! Reference source not found.~~Figure 2~~).

Description of Data Transformation & Analysis Pipeline

This study uses data that is already collected for analysis and available in electronic health care data sources in 5 European countries and follows the following principles.

First, to harmonize the structure of the data sets held by each partner, a shared syntactic foundation is utilized, we will use the CDM that was developed in the IMI-ConcePTION project. In this CDM, data is represented in a common structure but the content of the data remains in their original format. The ETL design is made available on paper and later on in the VAC4EU FAIR catalogue. The validity of the ETL will be assessed using Level 1 (completeness) and Level 2 (logical consistency) R-scripts that have been developed as part of the IMI-ConcePTION project.

Second, to reconcile differences across terminologies a shared semantic foundation is built for the definition of events under study by collecting relevant concepts in a structured fashion using a standardized event definition template. The Codemapper tool (https://vac4eu.org/codemapper/) was used to create diagnosis code lists based upon completed event definition templates for each event and comorbid risk condition in the ACCESS-BGR protocol https://drive.google.com/drive/folders/1Y_3cuGRN1g-jBv2ec1fC0aYcpxEjtrY9?usp=sharing.

Based on the relevant diagnostic medical codes, as well as other relevant concepts (eg, medications), algorithms will be constructed to operationalize the identification and measurement of each event. These algorithms may differ per data source, as the components that go into the study variable may differ. Wherever possible the event definition sheet specifies prior validation of algorithms and codes for benchmarking. Scripts for semantic harmonization are created centrally and provided in R and distributed to data access providers for local deployment. This will result in a set of study variables that are both semantically and syntactically harmonized. The quality of the semantic harmonisation will be assessed using Level 3 checks and compared against published rates and between databases (benchmarking). Limited to outcomes, components will be analysed in each data source, to assess the unique contribution of each data bank, therefore providing evidence on sensitivity and accuracy of the algorithms. Results from validation activities will be used to inform the calculation of validity of the composite algorithms.

Third, following conversion to harmonized study variable sets, additional R- or SAS-scripts for calculation of analytical datasets are distributed to data access providers for local deployment. The output datasets produced by these scripts are then be uploaded to the Digital Research Environment (DRE) for pooled analysis of incidence and visualization (Figure 2). The DRE is made available through UMCU/VAC4EU (https://www.andrea-consortium.org/). The DRE is a cloud-based, globally available research environment where data is stored and organized securely and where researchers can collaborate (https://www.andrea-consortium.org/azure-dre/).

All final statistical computations are performed on the DRE using R and SAS. Data access providers have access to the project workspace for verification of the results.

9.6.2 Data Access

Within the DRE, each project-specific area consists of a separate secure folder called a “workspace.” Each workspace is completely secure, and researchers are in full control of their data. Each workspace has its own list of users, which can be managed by its administrators. The DRE architecture allows researchers to use a solution within the boundaries of data management rules and regulations. Although General Data Protection Regulation and Good (Clinical) Research Practice still apply to researchers, the DRE offers tools to more easily control and monitor which activities take place within

projects. All researchers who need access to the DRE are granted access to study specific secure workspaces.

Access to this workspace is only possible with double authentication using an identification code and password together with the user's mobile phone for authentication. Upload of files is possible for all researchers with access to the workspace within the DRE. The Download of files is only possible after requesting and receiving permission from a workspace member with an "owner" role.

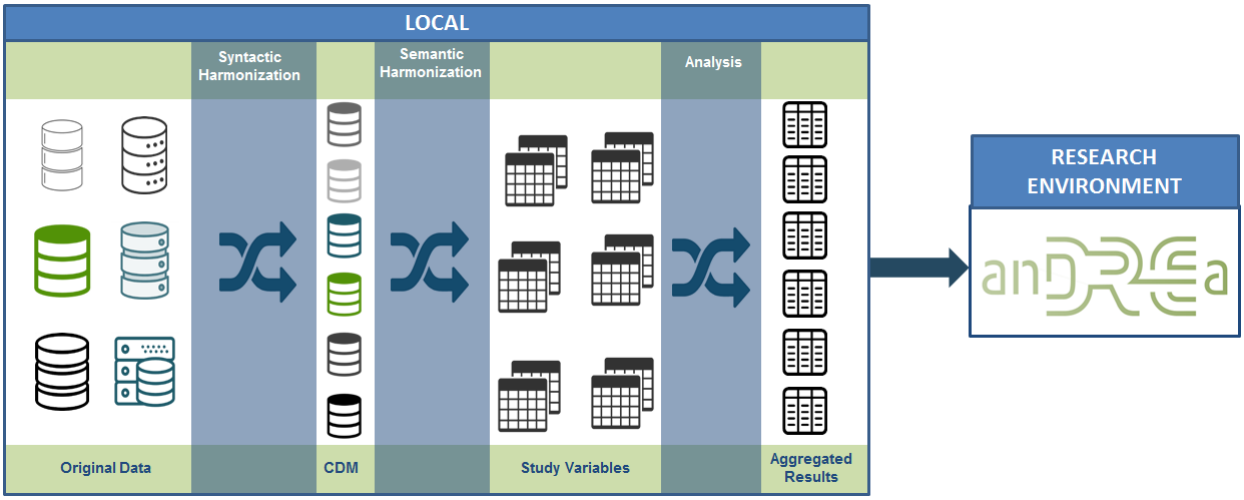


Figure 2: Data transformation and flow

9.6.3 Data Processing

Due to the nature of the study, a repeated data processing procedure is envisioned, based on the pipeline described in the previous section. This allows optimising the data processing timelines and archiving procedures.

At the study start, a baseline data extraction will be made by each of the DAPs. This creates a baseline instance of the data source. This is ETL'ed into the ConcePTION CDM and forms the baseline instance of the CDM. The data pipeline will be run for the first time on the baseline instance of the CDM of each DAP, and produce a baseline set of analytic datasets that will be centrally analysed for the baseline assessment.

9.6.4 Record Retention

DAPs are responsible locally to archive each data source instance that is used for the study. The meta-data table in the CDM allows for storing of details on the data source instance. The DAP has the obligation to archive the data source instances, the ETL scripts, the R-scripts that were used, and the results that were uploaded to the DRE, locally.

Aggregated results from DAPs, ETL design documents, and a repository of study scripts will be stored in the DRE for inspection by the study sponsor for at least five years. The final study aggregated results sets and statistical programs to pool and visualize will be archived and stored on the DRE for five years.

Documents that individually and collectively permit evaluation of the conduct of a study and the quality of the data produced will be retained for a period of 5 years in accordance with Good

Pharmacoepidemiology Practices (GPP) guidelines. Study records or documents may also include the analyses files, syntaxes (usually stored at the site of the database), ETL specifications, and output of data quality checks.

After 5 years all materials from the DRE will be retained for at least 15 years on a UMCU secure drive. The final study protocol and possible amendments, the final statistical report, statistical programs and output files will be archived on the UMCU secure drive according to Julius Clinical standard operating procedures.

9.7 Data Analysis

Detailed methodology for summary and statistical analyses of data collected in this study will be documented in an SAP for the readiness study and the respective SCRI and cohort analyses. All analyses will be conducted using R version R-4.0.3 or higher (Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org) or SAS version 9.3 software or higher (Cary, North Carolina, USA; SAS Institute, Inc.). The SAP will contain more detail of the analysis and data pooling and may modify the plans outlined in the protocol; any major modifications of primary endpoint definitions or their analyses would be reflected in a protocol amendment.

9.7.1 Main Summary Measures

For the readiness phase analysis, we will characterise and monitor over time the utilisation patterns of COVID-19 vaccines. Description of demographics and clinical characteristics will be reported for different groups of vaccine recipients and non-recipients, overall and among subcohorts of interest—such as individuals who are immunocompromised, pregnant, or have specific comorbidities. Negative control outcomes will be used using the SCRI and cohort design to estimate the presence of systematic bias.

The rapid assessment analysis will focus on the calculation and comparison of the incidence rates of each AESI between exposed individuals and reference individuals (cohort) or between the risk period and control period (SCRI). All analyses will be conducted within each data source and pooled across data sources using a random-effects model.

Readiness phase

Demographic and Baseline Characteristics/confounding

The distributions of baseline characteristics at the start of COVID-19 vaccination for each COVID-19 vaccine exposure group and a time-matched non-exposed group will be calculated to describe differences between the groups. For continuous variables, means, standard deviations, medians, and ~~other~~ quartiles will be estimated. For categorical variables, counts and proportions plus their confidence intervals will be estimated. The missingness of variables will also be described. To describe the relative imbalance of characteristics between different exposed groups, absolute standardised differences will be calculated for each baseline characteristic. Multilevel categorical variables will calculate an overall standardised difference across all levels. The larger the absolute standardised difference values, the greater the imbalance between baseline characteristics, this calculation will be done before and after matching.

Vaccine uptake

For every data source, the number of administered doses per vaccine brand within the primary series (dose 1 and dose 2) will be calculated by calendar time (in months) over the follow-up period. We will also calculate the distance between different doses and describe so-called heterologous vaccine schedules whereby patients receive different vaccine types for their first and second dose

For persons with a given year of birth, we will calculate vaccination uptake by dose 1 and 1+2 over time. The coverage at month i for birth year j is calculated by dividing the number of vaccinated subjects n_ij by the total number of subjects under follow-up at month i (N_ij), expressed as a percentage. Estimates of vaccination uptake will be compared with data provided by ECDC.

Algorithms for outcome definitions

For each outcome, we will count the number of cases by diagnosis code, and we will assess the impact on the case count when we use confirmatory medications /procedures or laboratory tests in algorithms. All such counts will be stratified by setting and recording type (GP record, hospitalization primary diagnosis, hospitalization secondary diagnosis). The impact of different algorithms and differential misclassification will be investigated with task 4.5 (methods, we will use component analysis: estimating which cases remain and disappear.

Time to onset

Temporal scan statistics and/or descriptive analyses of time to onset of an event (up to 6 months after vaccination) will be calculated to verify the selection of risk intervals.

Comparators

Comparators from population-based databases will be used to support signal refinement resulting from findings from the event monitoring and EHR-based cohort studies (see section 9.9 for more details).

Potential comparator groups for consideration, depending of the event of interest, may include the following:

- Historic comparators before the COVID-19 pandemic matched on seasonality or calendar month (EHRs)
- Contemporary unvaccinated comparators, matched on calendar time (EHRs)
- Contemporary unvaccinated comparators identified during influenza vaccination, may be challenging given the current vaccination periods (EHRs)
- Historic influenza vaccinated comparators identified before the COVID-19 pandemic, matched on relevant characteristics (EHRs)
- Subjects vaccinated with a different COVID-19 vaccine (EHRs or from event monitoring cohort)

Cohort Design With Concurrent Comparator

Demographic and Baseline Characteristics

The distributions of baseline characteristics at time zero by exposure group will be calculated to describe the study cohort and illustrate differences between the groups. For continuous variables, means, standard deviations, medians, and other quartiles will be estimated. For categorical variables, counts and proportions will be estimated. To describe the relative imbalance of characteristics between exposed groups, absolute standardised differences will be calculated for each baseline characteristic.

Multilevel categorical variables will calculate an overall standardised difference across all levels. The larger the absolute standardised difference values, the greater the imbalance between baseline characteristics.

Measures of Occurrence and Association

Incidence rates for each AESI will be calculated by dividing the number of cases by the follow-up person-time. Poisson regression models will estimate incidence rate ratios and 95% CIs. Robust variance estimation will account for an individual's possibility to be included in the index and reference cohorts. The Poisson model for each AESI will include the number of events in each cohort as the dependent variable, the exposure status as the binary independent variable, and the natural logarithm of the person-time as the offset variable. The attributable risk will be computed as the difference between the incidence rate in the index cohort and the reference cohort in the cohort design only since the SCRI does not allow for a proper incidence estimation, only the relative incidence.

Adjustment for Baseline Imbalances

Individuals following in the index and reference groups may have different characteristics that may determine their risk of AESI. To account for such potential confounding, we will match on variables such as age and sex and adjust for independent risk factors for the outcome (Poisson regression). If the number of outcome events is limited (<10/covariate) we will develop a propensity score (PS) and match index and reference group on the PS. The impact of matching or propensity score methods will be explored by the methods group.

Self-controlled Risk Interval Design

Descriptive Statistics

The number of cases of each AESI will be reported, with summary measures of age and sex.

Measures of Association

Conditional Poisson regression will be used to estimate incidence rate ratios and 95% confidence intervals (CIs). Time-varying confounders (eg, medications) may be included as covariates in regression models. Subgroup analyses will be conducted by groups defined by demographic and clinical characteristics and other covariates of interest.

9.7.2 Main Statistical Methods

Matching

For the cohort designs, matching will be used to ensure comparability between index and reference cohort on observed covariates. We will assess the impact of the match on the following potential key variables at time zero:

- Age (year of birth)
- Gender
- Pregnancy status
- Immunocompromised state
- Calendar month
- Prior recorded COVID-19 infection (yes/no)

We decided to focus on these covariates to ensure that no residual confounding for these variables remains. The impact will be tested in the methods group based on the readiness phase data.

For the SCRI design, a person is matched with itself and therefore controlled for all of the stable variables.

Stratified Analyses

- Analyses may be stratified by the following clinically relevant subgroups:
- Select comorbidities (by presence or absence of comorbidity of the specific event),
- Age ([<20], 10-year age categories)
- Gender
- Pregnancy status
- Immunocompromised state

Missing Values

Further details of the analysis, including handling of missing data and imputations will be described in the SAP.

9.7.3 Sensitivity Analyses by WP4 in the readiness phase and potentially (if needed) for rapid assessment studies

9.7.3.1 Choice of comparator (task 4.2)

In this section, we address considerations related to task WP4.2, specifically regarding selecting appropriate comparators for potential use in sensitivity analyses and selecting appropriate negative controls for use in the self-controlled and cohort approaches. We discuss the key factors that drive the appropriate selection of comparators and of negative controls, propose potential comparators and negative control outcomes, and describe the circumstances when each proposal is appropriate.

9.7.3.1.1 Selection of Comparators

The primary comparison proposed in the protocol consists of the comparison of recipients of an individual COVID-19 vaccine to contemporary recipients of other COVID-19 vaccines. Numerous additional comparisons are being considered, including within-person self-controlled comparisons, and comparisons using contemporary (i.e., during the vaccination period) or historical (i.e., prior to the COVID-19 pandemic) active or unvaccinated comparators. The suitability of a particular comparator differs depending on the specific outcome being examined.

Historical comparators

The COVID-19 pandemic resulted in profound changes in healthcare utilisation[5] as care was delayed or forgone during periods of lockdown or social distancing. These decreases have applied to inpatient and emergency department visits, as well[6] Therefore, comparisons of vaccinated individuals during the

[5] Giannouchos TV, et al. Trends in outpatient emergency department visits during the COVID-19 pandemic at a large, urban, academic hospital system. American Journal of Emergency Medicine 40 (2021) 20–26.

[6] Xu. J Med Internet Res 2021;23(4):326558.

vaccination period with comparators during the pre-pandemic period may identify non-null associations unrelated to confounding due to decreased incidence of all diagnoses during the vaccination period relative to the pre-pandemic period. However, research in the United States has demonstrated that while patterns of healthcare may have changed, increases in telehealth visits offset decreases in outpatient visits relatively early in the pandemic, and many other measures of healthcare utilisation have neared pre-pandemic levels by later in 2020, as well[7]. Before using a historical comparator for a particular outcome, trends in incidence before the pandemic and during the vaccination period should be examined to assess the suitability, particularly for outcomes with minor presentations that may have delayed diagnosis. Beyond healthcare utilization, the pandemic had a widespread impact on most routine aspects of life such as in-person school and work attendance that may have affected the occurrence of outcomes. These impacts are subsiding as lockdowns have ended or restrictions have been eased during the vaccination period, meaning the use of historical comparators may be adequate but trends in occurrence should be investigated before use. Historical comparators may not be suitable for outcomes that are associated with SARS-CoV-2 infection as the incidence of these outcomes will not be comparable before and during the pandemic. Table 7 lists outcomes for which occurrence may have been impacted by changes in healthcare utilization or exposures to causes of these outcomes during the pandemic.

Table 7. Outcomes for which occurrence may be impacted by the pandemic and temporal changes in occurrence should be examined prior to use of historical comparators.

Event
Microangiopathy
Coronary artery disease (CAD)
Arrhythmia
Myocarditis/pericarditis
Diabetes (type 1 and unspecified type)
Acute kidney injury
Acute liver injury
Anosmia, ageusia
Chilblain-like lesions
Single organ cutaneous vasculitis
Erythema multiforme
Transverse myelitis

Contemporary comparators

For comparisons in which vaccinated and comparator patients are drawn from the same time period (e.g., vaccinated vs. contemporary unvaccinated, COVID-19 vaccines vs. other COVID-19 vaccines), the issue of changing healthcare utilisation would not a factor, assuming the comparison is matched on time or exposure groups are balanced within time periods. As the COVID-19 vaccination period continues, it will be important to match on calendar time as temporal changes in healthcare utilisation will continue.

When using contemporary unvaccinated patients as comparators, it may be challenging to assess long-term outcomes (i.e., >90 days) because most of these patients will likely become vaccinated and thus

7

be censored from the analysis. Long-term outcomes include diabetes, acute kidney injury, acute liver injury, and narcolepsy.

When comparing one COVID-19 vaccine to others, temporality effects should be considered. Vaccines were not released at the same time so there may be periods of non-overlap for head-to-head comparisons. To reduce effects of temporality, patients can be matched on calendar time so non-overlapping periods would be discarded. Discarding non-overlapping time periods may also help balance patient characteristics as the timing of vaccine eligibility is related to age, underlying conditions, and profession. At the time of writing, there are two distinct platforms for the COVID-19 vaccines, mRNA and adenovirus. The platform should be considered when choosing an active comparator as certain outcomes may be platform-specific. There is a known or suspected association of the adenovirus vaccines with VTE, CVST, arterial thrombosis, and TTS, and the mRNA vaccines with Bells' palsy; there have been concerns of myocarditis/pericarditis with mRNA vaccines in some age and gender groups. As vaccines with similar platforms may share many characteristics, including potential safety concerns, it may be most appropriate to compare recipients of a specific COVID-19 vaccine to recipients of other COVID-19 vaccines of another platform to avoid comparing within the same platform.

Influenza Vaccination Comparators

Recipients of influenza vaccinations have been proposed as a potential historical or contemporary comparison group since persons who receive COVID-19 vaccines might be healthier or less healthy than those who do not receive any COVID-19 vaccine. Assuming that the factors leading to influenza vaccination and to COVID-19 vaccination are similar, a comparison of COVID-19 vaccinees to influenza vaccinees could potentially be less subject to confounding bias than a comparison of COVID-19 vaccinees and unvaccinated individuals. Influenza vaccinations are typically given in relatively narrow, seasonal periods of time (late summer and fall) which may maximize their susceptibility to the influence of seasonal variation in outcomes. Additionally, the period of primary influenza vaccination may not overlap with the temporal distribution of receipt of COVID-19 vaccines. For example, in the 2020/2021 influenza season, most influenza vaccines were distributed before the end of December 2020, while COVID-19 vaccination had barely begun by the end of 2020, and vaccination has continued throughout the duration of 2021. Therefore, comparisons of COVID-19 vaccinated individuals in 2021 to influenza vaccinated individuals in 2020 would be subject both to issues of the seasonality of outcomes (discussed in the next section) and potential changes in healthcare utilization. One way to enable contemporary comparisons while accounting for the healthy user bias intended to be addressed using influenza vaccination comparators might be to restrict the study population to individuals who have received the influenza vaccine in the recent past, and compare COVID-19 vaccinees to unexposed contemporary comparators. However, restricting the population to persons who have received the influenza vaccine in previous years may potentially lead to a selected population that would limit generalisability.

Furthermore, as COVID-19 vaccination campaigns continue, the temporality of COVID-19 and influenza vaccination may overlap in future influenza seasons (e.g., 2021/2020 season), and it may be more appropriate to perform comparisons with contemporary influenza-vaccinated patients. However, given the current rate of vaccination campaigns, influenza-vaccinated patients who have not received a COVID-19 vaccine by late summer/fall may be a selected population (e.g., younger age groups not eligible for COVID-19 vaccination yet) that would limit the generalisability of results.

Special Considerations for Outcomes with Seasonality

The occurrence of some outcomes may be dependent on seasonal patterns (Table 8). For example, many respiratory illnesses and their sequelae have increased incidence during the fall and winter months. For both contemporary and historical comparators, patients should be matched on calendar time when examining outcomes with seasonality. The exact temporality of these seasonal patterns can vary from year to year (e.g., the beginning of the influenza season can vary by months from late fall to late winter). Comparisons using historical comparators, even those matched on calendar week or month, may be subject to residual confounding by temporal changes.

Table 8. Outcomes that may have seasonal patterns and temporal changes in occurrence should be examined prior to analysis.

Event
Multisystem inflammatory syndrome
Acute respiratory distress syndrome
Coronary artery disease (CAD)
Arrhythmia
Myocarditis/pericarditis
Coagulation disorders
VTE (DVT & PE)
Hemorrhagic stroke
Generalised convulsion
Guillain Barré Syndrome
Diabetes (type 1 and unspecified type)
Acute kidney injury
Anosmia, ageusia
Chilblain-like lesions
Anaphylaxis
Death (any cause)
Sudden death (by codes)
Meningoencephalitis
Acute disseminated encephalomyelitis (ADEM)
Narcolepsy
Bells' palsy

Summary of Key Points on Selection of Appropriate Comparators

The primary cohort approach of comparing individuals receiving a COVID-19 vaccine to a contemporary comparator group of recipients of other COVID-19 vaccine types should be a viable option for all outcomes, with the following caveats:

- The two vaccine types should be available during the same period of calendar time (or the analysis should be restricted to periods of overlap) to reduce the influence of seasonal variation in outcomes.

- These comparisons may be particularly appropriate for outcomes hypothesized with one of the vaccine platforms but not others.

- Direct comparisons of different vaccines answer different research questions than comparisons of vaccinated and unvaccinated individuals. While direct comparisons of different vaccines answer questions related to comparative safety/effectiveness, comparisons of vaccinated vs. unvaccinated

address whether the risk of the outcome is different between vaccinated persons and persons not receiving any COVID-19 vaccine. A null result from a comparison of two vaccines may not indicate a lack of overall risk associated with a vaccine as compared to no vaccination, as a null estimate may be received for two vaccines with equally elevated risks.

Contemporary unvaccinated comparators may be viable for most outcomes. Caution may be needed for outcomes with long risk periods.

Historical comparator groups of unvaccinated patients matched on calendar time may be appropriate for outcomes without strong and variable seasonal variation or strong changes in coding due to changes in healthcare utilization during the pandemic.

Historical comparator groups of influenza vaccine recipients may be viable only if the COVID-19 vaccination groups are restricted to time periods similar to those of the influenza vaccination period, and there are not strong time trends in diagnosis recording between autumn/winter 2020 (or earlier years) when influenza vaccines were distributed, and the COVID-19 vaccination period.

Contemporary comparator groups of influenza vaccine recipients may be identified during the 2021/2022 influenza season which may overlap with the continued distribution of COVID-19 vaccines to reduce the influence of seasonal variation in outcomes. However, those individuals who remain unvaccinated by the 2021/2022 influenza season may be a selected population that would limit generalisability. As an alternative, one could consider restricting the study population to individuals who have received the influenza vaccine in the recent past and compare COVID-19 vaccinees to unexposed contemporary comparators with the unexposed index date matched on calendar time to the COVID-19 vaccinees. However, restricting the population to persons who have received the influenza vaccine in previous years may potentially lead to a selected population that would limit generalisability.

9.7.3.1.2Selection of Negative Controls

Negative control outcomes will be used in both the self-controlled and cohort approaches to ensure adequate control of systematic biases before implementing final analyses using the study outcomes. Negative control outcome analyses investigate the association between the exposure of interest and an outcome known not to be caused by the exposure. If a non-null association is observed, the result is assumed to be caused by bias rather than a causal effect of the exposure.

In order to be a valid negative control outcome, the potential event must meet the following criteria[8]

1. The set of common causes of receiving a COVID-19 vaccine (or the specific COVID-19 vaccine being evaluated) and the safety event(s) of interest should be as identical as possible to the set of common causes of receiving the COVID-19 vaccine and the negative control (e.g., the negative control outcome and safety outcomes of interest share common unmeasured confounders with vaccine receipt)
2. The COVID19 vaccine cannot cause a negative control outcome.

[8] Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010 May;21(3):383-8.

Considerations for Selecting an Appropriate Negative Control for COVID-19 Vaccine Studies

The COVID-19 pandemic resulted in profound changes in healthcare utilisation [9] as care was delayed or forgone during periods of lockdown or social distancing. These decreases are applied to inpatient and emergency department visits. Therefore, comparisons of vaccinated individuals during the vaccination period with comparators during the pre-pandemic period may identify non-null associations with a negative control outcome unrelated to confounding due to decreased incidence of all diagnoses during the vaccination period relative to the pre-pandemic period. Additionally, in self-controlled designs, comparisons between different time periods within the same individuals may lead to non-null associations with a negative control outcome due to changes in healthcare utilization overtime during the pandemic.

However, research in the United States has demonstrated that while patterns of healthcare may have changed, increases in telehealth visits offset decreases in outpatient visits relatively early in the pandemic[10] , and many other measures of healthcare utilisation have neared pre-pandemic levels by later in 2020, as well[11]. While comparisons of the vaccination period to the early pandemic period (i.e. March-May 2020) may result in substantial differences in healthcare utilisation, reduced healthcare utilisation may be less of an issue when comparing the vaccination period to the pre-pandemic period. However, more current data about healthcare utilisation during the vaccine administration period is needed to inform the suitability of these negative control outcomes.

Comparator-Specific Considerations for Identifying Appropriate Negative Controls

Numerous comparisons are being considered, including comparisons using self-controls, contemporary or historical active or unvaccinated comparators. The suitability of a particular negative control outcome may depend on the choice of comparators, as outlined below:

Contemporary Comparisons

For comparisons in which vaccinated and comparator patients are drawn from the same time period (e.g. vaccinated vs. contemporary unvaccinated, COVID-19 vaccines vs. other COVID-19 vaccines), the issue of changing healthcare utilisation would not be a factor in determining an appropriate negative control outcome, assuming the comparison is matched on calendar time or exposure groups are balanced within time periods. However, if both exposure groups are drawn from the same larger time period, but temporality exists (e.g. an earlier-released vaccine is compared with a later-released vaccine within a broad study time period), there may still be temporality effects in diagnoses of the negative control.

Historical Comparators

The suitability of using a particular outcome as a negative control when using a historical comparator depends on whether the outcome is seasonal. Many respiratory illnesses and their sequelae have increased incidence during the fall and winter months. However, the exact temporality of these seasonal

[9] Giannouchos TV, et al. Trends in outpatient emergency department visits during the COVID-19 pandemic at a large, urban, academic hospital system. American Journal of Emergency Medicine 40 (2021) 20–26.

[10] Xu. J Med Internet Res 2021;23(4):e26558

[11] Giannouchos TV, et al. Trends in outpatient emergency department visits during the COVID-19 pandemic at a large, urban, academic hospital system. American Journal of Emergency Medicine 40 (2021) 20–26.

patterns varies from year to year (e.g. the beginning of the influenza season can vary by months from late fall to late winter). Comparisons using historical comparators, even those matched on calendar week or month, may be subject to confounding by temporal changes in incidence if negative control outcomes with seasonal patterns are used.

Historical Comparators, Influenza Vaccination

Recipients of influenza vaccinations have been proposed as a potential comparison group. Influenza vaccinations are typically given in relatively narrow, seasonal periods of time (fall and early winter) which may not overlap with the temporal distribution of receipt of COVID-19 vaccines. As described previously, comparisons of COVID-19 vaccinated individuals in 2021 to influenza vaccinated individuals in 2020 would be subject both to issues of the seasonality of outcomes and potential changes in healthcare utilization. Negative control outcomes with seasonal patterns or changes in temporal patterns due to disrupted healthcare utilisation would not be appropriate for this comparison.

As COVID-19 vaccination campaigns continue, the timing of COVID-19 and influenza vaccination may overlap in future influenza seasons, and it may be more appropriate to perform comparisons with contemporary influenza-vaccinated patients, assuming that the factors that lead to COVID-19 vaccination and to influenza vaccination are similar.

Self-controlled approaches

For self-controlled approaches, each vaccinated serves as his/her own control by comparing two time periods within the same person. Negative control outcomes with seasonal patterns or changes in temporal patterns due to disrupted healthcare utilisation are not appropriate for use in self-controlled cohort approaches.

Testing Assumptions

The occurrence of potential negative control outcomes can be evaluated in the data sources to describe if disruption of healthcare utilisation has influenced the coding of outcomes during the study period relative to proposed historical control periods or led to changes in coding throughout the study period (relevant for self-controls). Weekly or monthly rates of recorded diagnoses can be plotted across the study period.

Additionally, negative control analyses can be performed comparing incidence rates of the negative control outcomes in unvaccinated patients during the vaccination period to unvaccinated patients in historical periods to evaluate changes in incidence rates. However, it is important to note that non-null associations in contemporary unvaccinated patients vs. historically unvaccinated patients may be due to either change in healthcare utilization or due to unmeasured confounding.

Potential Negative Control Outcomes

Suitable for Contemporary Comparisons

- Gout
    - Gout exacerbations may have seasonal patterns making it less suitable for historical comparisons or self-controlled cohort approaches.
- COVID-19 within 14 days of index date
    - COVID-19 vaccines are not expected to demonstrate efficacy until 14 days after vaccination as recipients need to mount an immune response. Observed differences in COVID-19 infections within 14 days of vaccination (or the comparator index date,

depending on the choice of comparator) would be due to confounding rather than vaccine effect. Similar negative controls with influenza have been used previously.
    - Obviously, historical comparisons of COVID-19 would not be feasible, as COVID-19 did not exist in the pre-pandemic historical period. Even comparisons within the pandemic period should be temporally close enough that changing incidence over time wouldn't influence the results. COVID-19 disease in the 14 days after vaccination is not a suitable negative control for self-controlled approaches because the event is time-limited and is therefore only appropriate for use with external comparators.
- Otitis externa
    - Seasonal variation in otitis externa has been noted, corresponding with seasonal variation in viral respiratory infections[12]. Therefore, otitis externa is not suitable to be used as a negative control for self-controlled cohort approaches. Additionally, with reductions in influenza and other viral infections during the pandemic period, comparison of the pandemic period to pre-pandemic periods may not be appropriate.

Suitable for Historical Comparisons and Self-Controlled Approaches, Pending Evaluation of Changing Patterns Over Time

- Automobile accident
    - Driving patterns were severely disrupted early in the pandemic and during period of national or regional lockdown. These periods have varied by geography and time. It is possible that during portions of the vaccination period, however, driving patterns in many geographies have returned to normal.
    - Automobile accidents may follow a predictable variation (more accidents in the summer), though without the variability of other outcomes (e.g. viral respiratory infections).
    - Patterns of diagnoses of injuries due to automobile accidents over time should be evaluated to determine if it could be a suitable negative control.
- Trigeminal neuralgia
    - Trigeminal neuralgia, a rare pain disorder, does not have noted seasonal trends.
    - Being a reasonably severe condition requiring treatment with medications or surgery but without emergency hospitalisations or emergency department visits, diagnoses of trigeminal neuralgia may remain constant during the vaccination period and pre-pandemic periods.
- Foreign object in ear canal
    - The majority, but not all, of cases of foreign objects in the ear occur in children. Seasonal variation in diagnoses have not been reported, and the incidence of cases requiring medical intervention may be constant in the vaccination and pre-pandemic periods.
- Poisoning
    - Poisoning accidents have been associated with adherence to health promoting behaviours in previous work[13] and may represent an important negative control when

[12] Stockman C. Seasonality of Acute Otitis Media and the Role of Respiratory Viral Activity in Children. Pediatr Infect Dis J. 2013 Apr; 32(4): 314–319.

[13] Dormuth 2009 STATIN ADHERENCE AND RISK OF ACCIDENTS: A CAUTIONARY TALE. Circulation. 2009 Apr 21; 119(15): 2051–2057.

evaluating potential confounding by adherence to guidelines or recommendations, such as receipt of vaccination.

- Poisoning accidents can be emergency events requiring intervention, and routine seasonal variation has not been noted.

9.7.3.3 SCRI sensitivity analyses (task 4.3)

Readiness phase

In the readiness SCRI studies on negative controls, several sensitivity analyses will be conducted to test assumptions and results from simulations of the WP4 methods group:

- excluding all patients who die within the observation period;
- beginning observation at the date of vaccination for the outcome of death
- Length of the buffer period
- Length of the control window and the impact of the lockdown
- A simple test to determine whether censoring may have biased effect estimates is to fit an interaction term between the main exposure and an indicator marking whether a patient had their theoretical observation period censored (for any reason). If there is evidence of an interaction, further discussion will determine whether a) the SCCS/SCRI method is inappropriate for investigating the outcome of interest or b) extensions of the SCCS/SCRI method can be applied to avoid any possible bias.

Two sensitivity analyses will be performed on the negative control events. 1) excluding all patients who die within the observation period; 2) A simple test to determine whether censoring may have biased effect estimates by fitting an interaction term between the main exposure and an indicator marking whether a patient had their theoretical observation period censored (for any reason). If there is evidence of an interaction, further discussion will determine whether a) the SCRI method is inappropriate for investigating the outcome of interest or b) extensions of the SCRI method can be applied to avoid any possible bias.

Rapid assessment

Signals being evaluated in WP3 work packages will be evaluated. An initial triage will determine the potential suitability of the SCRI design, using the following steps to assess how well the assumptions of the method are met:

1. Is the outcome acute and with good accuracy of dating in the relevant datasets? If yes, move to step 2. If no, SCRI is not a valid design for the signal.

2. Are outcomes independent within person? If not, consider the first occurrence only.

4. Event-dependent censoring; the SCRI design assumes the timing of outcome events is independent of the observation period. This assumption is not met for events that are frequently fatal and can lead to biased estimates of the risk ratio for exposure, especially if

censoring is common. This will be investigated using a) clinical knowledge of the outcome of interest and b) a histogram of the time between the outcome and the end of the observation period.

5. Assess for an event (outcome) dependent exposures; independence of the exposure from the outcome is assumed, but this assumption is frequently not completely met because outcomes often lead to a short termshort-term change in the risk of exposure, usually a temporary avoidance of exposure. This will be investigated using event centred plots of the time between the outcome and the vaccination date. This histogram will inform the length of any pre-exposure period to be removed from baseline time.

After triage, if SCRI is considered an appropriate signal strengthening design, the next step is to define the observation period and control/risk windows (see Fig 1). The observation period will be bound within the time period since the vaccine of interest became available in the EU and the last date of data availability defining both the vaccine exposure and the outcome event. Within this window, an observation period comprising 1) a control window, 2) an optional pre-exposure window to account for possible event-depending exposure delay and 3) a risk widow. The precise length of these windows will vary between outcomes and will need to be decided using clinical knowledge of the likely time to onset and empirical evidence of any delays in vaccination following the outcome of interest (see above).

If outcome-dependent observation censoring is considered a potential problem two further sensitivity analyses will be performed. 1) excluding all patients who die within the observation period; 2) A simple test to determine whether censoring may have biased effect estimates is to fit an interaction term between the main exposure and an indicator marking whether a patient had their theoretical observation period censored (for any reason). If there is evidence of an interaction, further discussion will determine whether a) the SCCS/SCRI method is inappropriate for investigating the outcome of interest or b) extensions of the SCCS/SCRI method can be applied to avoid any possible bias.

9.7.3.4 Bias analysis (task 4.4)

We will perform an array approach for unmeasured confounding and calculate E-values[14] [15]. Both approaches are similar in presenting how large an unmeasured confounder, or group of confounders, would have to be to explain the observed association. The array approach yields a full overview of different values for the confounders in terms of the strength of the association, the prevalence of the confounder, and the association of the confounder with the exposure. The E-value consists of one number that embodies these variables into one value. The strength of the E-value lies in the possibility to compare E-values to those in other studies. The strength of the array approach is that one gets a better impression of what the characteristics of a confounder should have to explain the association and can then compare this to the values of confounders that were found in the study. We will calculate E-values to identify the minimum strength of association that an unmeasured confounder would need to have with either vaccine exposure and the AE of interest, conditional on the measured confounders, to explain away the observed associations, a quantitative bias analysis will be conducted.

[14] Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf. 2006 May;15(5):291-303. doi: 10.1002/pds.1200. PMID: 16447304.

[15] VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017 Aug 15;167(4):268-274. doi: 10.7326/M16-2607. Epub 2017 Jul 11. PMID: 28693043.

Negative or falsification endpoints are outcomes that are not supposed to be associated with the treatment of interest, but with potential unmeasured confounders. In the readiness phase, we will use 3 negative endpoints. If a falsification endpoint is analysed in the same way as the actual outcomes are analysed, and also have the same confounders, the falsification endpoints are supposed to yield a neutral association with the treatment[16] [17]. However, if the falsification endpoint shows a significant association with any of the treatments it indicates that there is residual confounding. If the association with the falsification endpoint is in the same direction as the actual association of interest, this means that the found association of interest is probably (partly) explained by this residual confounding. We will test the association between vaccine exposure and negative control events in the readiness phase. As vaccine exposure should not affect the risk of these three events, these could serve as falsification endpoint.

Differences between countries in vaccination strategies (prioritization of subgroups, difference in vaccine brands, different time between 1st and 2nd dose) give the opportunity to evaluate the impact on safety and effectiveness of these strategies. Account should be taken of different non-pharmaceutical government interventions against COVID-19[18]. Although direct comparisons may not be valid, an instrumental variable (IV) approach using the country as an IV may provide valid comparisons between the intensity of different vaccination strategies (including different brands) and safety and effectiveness. The validity of IV analysis depends on 3 major assumptions:

1) (strong) association between instrument (country) and exposure (vaccine brand/strategy),
2) IV only affects outcome through exposure (exclusion restriction) and
3) effect on outcome is unconfounded (exchangeability).

Several falsification strategies and tools are available in addition to subject matter knowledge to assess the plausibility of these main assumptions and will be considered to assess the suitability of IV analyses[19].
We will apply the IV method to the negative control events in the readiness phase.

9.7.5 Misclassification of outcomes (task 4.5)

We will examine and, where necessary, adjust for the effects of misclassification (measurement error) To accomplish this, we will (1) develop several statistical (Bayesian) models of increasing complexity that incorporate misclassification into the analysis; (2) support a literature review of extant knowledge on the degree of error that serves as input for the modelling efforts; and (3) provide user-friendly open-source software that allows third parties to examine the likely sensitivity of study estimates to measurement error.

A particular challenge in modelling the effect of outcome misclassification on further analyses is the (plausible) possibility that detection probabilities of events differ for patients with different exposure status. To tackle this problem, we will take the approach of first simply quantifying the possible sensitivity of estimates under various scenarios, without a strong modelling effort. In the second instance, we will produce model-based estimates of the effects of differentially misclassified events; Because such assumptions are more tenuous, we will present them with due care.

We will

[16] Prasad V, Jena AB. Prespecified Falsification End Points: Can They Validate True Observational Associations?*JAMA.* 2013;309(3):241–242. doi:10.1001/jama.2012.96867
[17] Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Statistics in Medicine* 2014; 33:209–218.

[18] Brauner JM, et al. Science 10.1126/science.abd9338(2020)
[19] Labrecque J, Swanson S. Epidemiologic Methods 2018;5:214-220, Ali M, et al. Epidemiology 2014;25:770-2

- Conduct a further literature review on validation and misclassification in the various events, based on the event definitions as defined in ACCESS;
- Develop models that describe the effect of the above misclassification on estimates:
- A simple Bayesian model per event type using plug-in estimates, accounting for uncertainties about these as resulting from the review;
- Hierarchical version of (a) to increase the precision of estimates;
- Sensitivity analysis of model accounting for differential misclassification w.r.t. exposure;
- Estimation of model accounting for differential misclassification under various assumptions.
- Development of user-friendly software (e.g. R package, Shiny app) to support the interpretation of estimates in the rapid assessment studies.

9.8 Quality Control

Rigorous quality control (QC) will be applied to all deliverables. Data transformation into the CDM will be conducted by each subcontracted research partner in its associated database, with processes as described in the following corresponding sections. 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.

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.

10.1 Patient Information

This study mainly involves data that exist in anonymized/pseudonimized structured format and contain no patient personal information. Data remain local and only aggregated data that is the result of the analytical script implementing the study design will be transferred to the DRE for postprocessing an pooling. When there are privacy concerns around anonymized SCRI data sharing, postprocessing can be done locally and only coefficients can be shared.

10.1.1 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.

10.1.2 Institutional Review Board /Independent Ethics Committee

Each DAP will be following the local country and institutional requirements to apply for access and analysis of the data for this study. At the coordinating centre, UMCU will ask approval for exemption

from review by the UMCU International IRB. All correspondence with the IRB or independent ethics committee and applicable documentation will be retained as part of the study materials.

10.1.3 Ethical Conduct of the Study

This study will adhere to the Guidelines for GPP and has been designed in line with the ENCePP Guide on Methodological Standards in Pharmacoepidemiology. The ENCePP Checklist for Study Protocols will be completed. The study is a post-authorisation study of vaccine safety and will comply with the definition of the non-interventional (observational) study referred to in the International Conference on Harmonisation tripartite guideline Pharmacovigilance Planning E21 and provided in the EMA Guideline on Good Pharmacovigilance Practices (GVP) Module VIII: Post-Authorisation Safety Studies and with the 2012 EU pharmacovigilance legislation, adopted 19 June 2012. The study will be registered in the EU PAS Register The research team should adhere to the general principles of transparency and independence in the ENCePP Code of Conduct

The study will be conducted in accordance with legal and regulatory requirements, as well as with scientific purpose, value, and rigour and follow generally accepted research practices described in Guidelines for GPP issued by the International Society for Pharmacoepidemiology, and Good Epidemiological Practice guidelines issued by the International Epidemiological Association. An independent scientific advisory committee will be installed, comprising experts in vaccine safety studies.

11. Management and reporting of adverse events/adverse reactions

For studies in which the research team uses only data from automated health care 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 summarised 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-authorisation 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 summarised in the interim safety analysis and in the final study report."

Module VIII – Post-Authorisation 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.

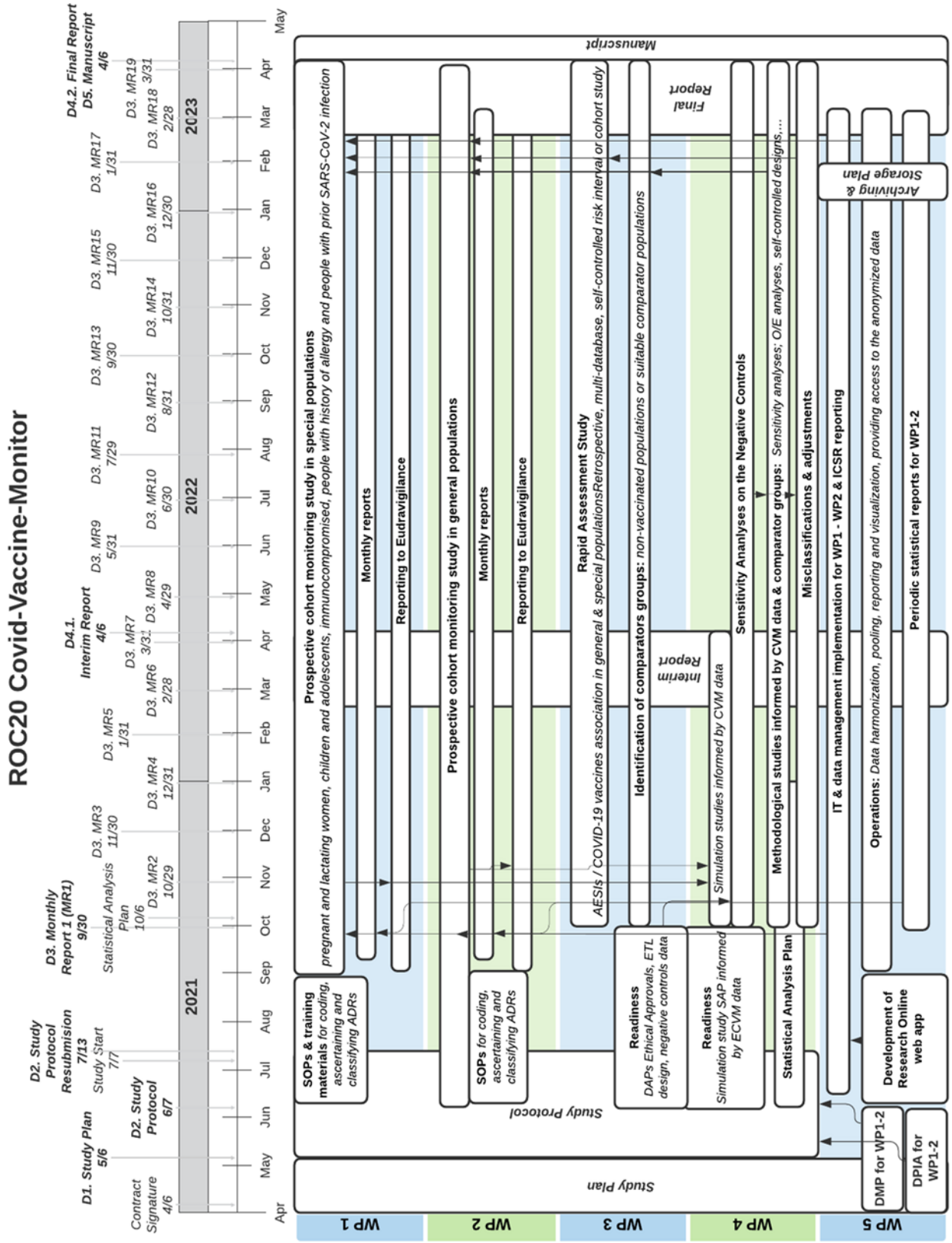
12. Plans for disseminating and communicating study results

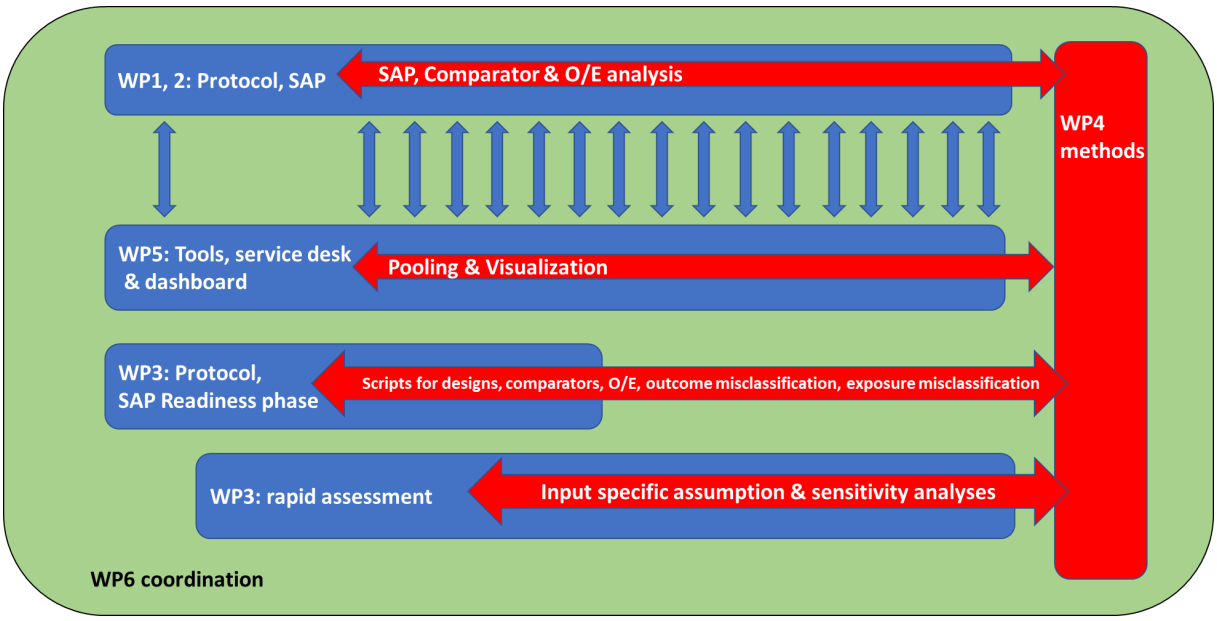
Results of analysis and interpretation will be delivered in the form of a report for the EMA.

Study results will be published following guidelines, including those for authorship, established by the International Committee of Medical Journal Editors. When reporting results of this study, the appropriate Strengthening the Reporting of Observational Studies in Epidemiology checklist 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 GVP Module VIII: Post-Authorisation Safety Studies.

Communication via appropriate scientific venues will be made according the ENCePP CoC.

Annex 1: WPs collaboration and timelines scheme




WP1, 2: Protocol, SAP
SAP, Comparator & O/E analysis
WP4 methods
WP5: Tools, service desk & dashboard
Pooling & Visualization
WP3: Protocol, SAP Readiness phase
Scripts for designs, comparators, O/E, outcome misclassification, exposure misclassification
WP3: rapid assessment
Input specific assumption & sensitivity analyses
WP6 coordination