Covid-Vaccine-Monitor

Specific Contract No 01 implementing framework contract No EMA/2018/23/PE

Final Study Report for WP3 (electronic health record data) and 4 (Methods)

Deliverable 4.2 – May 2023

Main Authors: Carlos E. Durán, Anna Schultze, Ian Douglas, Sophie Bots, Svetlana Belitser, Davide Messina, Ivonne Martin, Rosa Gini, Fabio Riefolo, Olaf Klungel, Miriam Sturkenboom

Document version: 1
First Release date: 14 May 2023
Updated version July 14, 2023

Disclaimer & acknowledgement
The research leading to these results was conducted as part of the activities of the EU PE&PV (Pharmacoepidemiology and Pharmacovigilance) Research Network (led by Utrecht University) in collaboration with the Vaccine Monitoring Collaboration for Europe network (VAC4EU).
Scientific work for this project was coordinated by the University Medical Center Utrecht. The project has received support from the European Medicines Agency under the Framework service contract nr EMA/2018/23/PE.
This report expresses the opinion of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.

Work in this report is based on:
EU PAS Register No: WP3-4: EUPAS42467

Table of Contents

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

1 Executive Summary/Abstract

1.1 Title

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

1.2 Keywords

Safety; databases; COVID-19; vaccines; adverse events of special interest, methods

1.3 Rationale and objectives

1.3.1 **Rationale**

To complement spontaneous reporting systems for signal detection (routine pharmacovigilance) and other initial safety monitoring activities such as pharmaco-epidemiological studies conducted or planned by different stakeholders, the Agency procured several safety monitoring studies through its framework contracts.
In January 2021 the Agency launched a new tender for safety monitoring of COVID-19 vaccines in the EU. The EU PE & PV and the VAC4EU network received and implemented the tender, which had two objectives, the first was to implement a prospective cohort monitoring in more than 10 countries and the second was signal strengthening. This executive summary is focusing on the second objective which was to conduct signal strengthening activities for potential safety concerns emerging from active surveillance electronic health data.
Based on the technical specifications signal strengthening meant the collection of additional information to further characterise the incidence of the safety concern in comparison to its expected incidence in non- vaccinated populations or suitable comparator populations. This activity should provide additional evidence supporting signal management and regulatory decision-making on the need for a full signal evaluation. The safety concerns for which signal strengthening should be performed could be identified by the Agency, other regulatory authorities, or the consortium itself.

1.3.2 **Objectives**

The request for signal strengthening capacity was translated into three objectives

1) *To create and assess readiness of electronic health record data sources for rapid evaluation of safety signals* by:

    - Providing an overview of the methods for identification of COVID-19 vaccine exposure in the data sources

- Monitoring the number of individuals exposed to any COVID-19 vaccine and to compare this to COVID-19 vaccine exposure (benchmark: ECDC vaccine tracker)[1]
- Generation of updated background rates for AESIs

2) To conduct rapid safety assessment studies using electronic healthcare records and support EMA safety assessment

Moreover, it was planned that this should allow for specific subgroup analyses:

- 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
- patients with a prior history (ever) of that event more than a year before.

3) Methodological assessment

Although not requested in the technical specifications, the consortium proposed to conduct methodological work to test the impact of:

- different comparators in the cohort design, different censoring criteria in the cohort study
- different control periods/duration for the SCRI
- different algorithms to assess events, and covariates
- three different methods to address unmeasured confounding (negative control outcomes, quantitative bias analysis, instrumental variable analysis)
- Misclassification of outcomes

1.4 Methods

1.4.1 Setting

Nine well-known European electronic health record (EHR) data sources in Norway, UK, Italy, the Netherlands and Spain were included (all listed in ENCePP Database Register). The data access providers were members of the EU PE&PV and VAC4EU networks, willing to participate, had access to potential fit for purpose data and had transformed their data already in the ConcePTION CDM for prior studies. From the ACCESS study, which was also conducted by the same consortium it was clear that University Aarhus (Denmark) and University Bordeaux (France) could not get rapid access to required data and that GePARD (DE) did not have access to COVID-19 vaccination data.

1.4.2 Study design

Study designs differed for the three different objectives:

1. Readiness: a retrospective cohort design using data from 2019 to latest data availability
2. Rapid assessment studies: a comparative cohort study and a self-controlled risk interval study;

[1] ECDC vaccine tracker: https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#uptake-tab

3. Methodological studies used several designs to explore assumptions and limitations of the chosen designs including different comparators in the cohort design, comparing SCCS and SCRI, simulations and use of negative controls.

1.4.3 Subjects and study size

Readiness: Study subjects comprised all subjects in the source population of the participating data sources who were in follow-up for at least 365 days during the study period (January 1, 2019, for readiness study) or were born into the cohort during the study period, and for whom vaccination data could be obtained/linked.
Rapid assessment: For self- controlled designs we included only subjects with the outcome of interest and a covid-19 vaccination. For the cohort study, vaccinated subjects and matched comparators were included.
Methodological assessment: for self-controlled designs subjects with the outcome of interest and a covid-19 vaccination. For the cohort study, vaccinated subjects and different matched comparators were included.

1.4.4 Data sources

For the implementation of the readiness study, 10 electronic health care databases in Northern, Southern and Western Europe showed interest to participate. The data sources and the data access providers that were included are:

Italy
- ARS Toscana (Agenzia Regionale di Sanità della Toscana)
- Pedianet (Societa Servizi Informatici)
- Caserta local health database (INSPIRE srl)
- Lazio Regional data source (Pharmacoepidemiology Unit Lazio Region)

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

The United Kingdom
- CPRD: Clinical Practice Research Datalink (University Utrecht)

Norway
- The Norwegian health registers (University of Oslo)

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

Lazio regional data could not be accessed due to changes in data access rules.

For actual rapid assessment studies, choices for data sources were made based on:
- Availability of fit for purpose data
- Sample size and resources
- Ability to commit to timelines.

1.4.5 **Variables**

- 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. The following events were extracted as AESI or potential negative control:

Table ES1. AESI or potential negative control list.

Event	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
Acute 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 (AMI /Ischemic stroke)	✓	✓	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)		✓	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

Event	SCRI	cohort	Naïve period to estimate new onset	Primary Risk period*
Transverse myelitis	✓	✓	365 days	28 days
Bells' palsy	✓	✓	365 days	28 days
Haemophagocytic lymphohistiocytosis	✓	✓	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 events				
Gout	✓	✓	365 days	28 days
Otitis externa	✓	✓	365 days	28 days
Trigeminal neuralgia	✓	✓	365 days	28 days
Acute kidney injury	✓	✓	365 days	28 days
Anaphylaxis (not drug-induced)	✓	✓	365 days	28 days
C. difficile infection	✓	✓	365 days	28 days
Conjunctivitis	✓	✓	365 days	28 days
COVID-19 within 12 days after vaccination	✓	✓	365 days	28 days
Diverticulitis	✓	✓	365 days	28 days
Fractures	✓	✓	365 days	28 days
Gall stones	✓	✓	365 days	28 days
Influenza	✓		365 days	28 days
Liver cirrhosis	✓	✓	365 days	28 days
Organic (secondary) psychosis	✓	✓	365 days	28 days
Osteoarthritis	✓	✓	365 days	28 days
Osteomyelitis	✓	✓	365 days	28 days
Reactive arthritis	✓	✓	365 days	28 days
Renovascular disease	✓	✓	365 days	28 days
Sjögren's syndrome	✓	✓	365 days	28 days
Urinary tract infections	✓	✓	365 days	28 days
Valvular heart disease (non-congenital, not rheumatic)	✓	✓	365 days	28 days

Vaccines: COVID-19 vaccines approved for use by EMA during the study period (Monovalent Pfizer, Moderna, AstraZeneca, Janssen, Novavax).

Specifically, vaccination data were obtained in the following manner:

Exposure to COVID-19 vaccines was based on available recorded prescription, dispensing, or administration of the COVID-19 vaccines. The main exposure of interest for the rapid assessment studies was the receipt of COVID-19 vaccine(s).

- ARS Toscana (IT): ARS identified vaccines from the regional immunization register using the national product code, including batch number.
- Pedianet (IT): Information on COVID-19 vaccine was obtained from the regional immunization register and included the date of immunization, type of vaccine, vaccine batches, dose.
- 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). Those claims data could 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.
- PHARMO (NL): Data on vaccination were obtained from PHARMO's GP database. Information on vaccines include ATC code, brand, and date of administration/recording. Several COVID-19 vaccines have been administered through other routes and information was provided to GP with different lag times.
- 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 is available. This includes, alongside an encrypted unique patient identifier; the name of the vaccine; manufacturing company; dose; and date
- Norwegian health registers (NO): The national, electronic immunization register (SYSVAK) was used. 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 center where the vaccine was administered.
- SIDIAP (ES): SIDIAP has available information on the administration of COVID-19 vaccines to individuals linked to a unique and anonymous identifier. The information originated from electronic medical records. For each patient, SIDIAP had date and center of administration, 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.bifap.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 were 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 were obtained from the Vaccine Information System (VIS), which includes information on vaccine type, manufacturer, batch number, number of doses, location and administration date.

Medicines were used as proxies for co-morbidities or associated with AESIs.

Covariates (medicines or conditions for subgroup analyses)

- Cancer diagnosis or cancer medicines (L01A*, L01B*, L01C*, L01D*, L01X*, L02A*, L02B*, L03*, L04*)
- Chronic kidney disease diagnosis (exclusion criterium for assessment for acute kidney injury)
- Chronic liver disease diagnosis (cirrhosis, non-alcoholic fatty liver disease, alcoholic liver disease, autoimmune hepatitis)
- Chronic respiratory disease diagnosis (chronic obstructive pulmonary disease, bronchiectasis, asthma, interstitial lung disease, cystic fibrosis) or drug proxies (R03*, R07A*)

- Cardio/Cerebrovascular disease (CVD) diagnosis (stroke, transient ischemic attack (TIA), aneurysm, and vascular malformation, coronary artery disease, heart failure or cardiomyopathies) or drug proxies for such disease (C01*, C03*, C07*, C08*, C09*, B01AC*)
- Obesity diagnoses or anti-obesity medicines as proxy (A08AB*, A08AA*)
- Down syndrome diagnoses
- Mental health disease (depression, dementia, and schizophrenia spectrum disorders) or drug proxies (N05A*, N06A*, N06D*)
- Sickle cell disease diagnosis or drug proxies (L01XX05, B06AX01)
- Diabetes (type 1 or 2) or diabetes medicines as proxy (A10B*, A10A*)
- Human immunodeficiency virus diagnoses or drug proxies (J05AE*, J05AR*, J05AF*, J05AG*)
- Immunosuppressants: Use of corticosteroids or other immunosuppressive medications (H02*, L04*)

COVID-19 History

- COVID-19 infection: Covid-19 Dx diagnosis code or positive test further classified by severity:
- Level 1: any recorded COVID-19 diagnosis or positive test
- Level 2: hospitalization for COVID-19 (COVID-19 diagnosis in primary/secondary discharge diagnosis)
- Level 3: ICU admission in those with COVID-19 related admission
- Level 4: death during hospitalization for COVID-19 (any cause)

Prior history of events

- prior VTE (deep venous thromboembolism, Pulmonary embolism, splanchnic) or drug proxies (B01AB*)
- History of anaphylaxis diagnosis or use of injectable epinephrine (C01CA24)
- History of allergic reactions

Comedication that may be associated with any of the AESI, assessed at start of follow-up and at time zero (prescription/dispensing 90 days prior)

- Antithrombotic agents (B01A*)
- Sex hormones (G03*) year prior
- Antibiotics (J01*)
- Antiviral medications (J05*)
- Lipid lowering drugs (C10*)
- Vaccines (J07 not J07BX03)

1.4.6 Data management

This study was conducted in a distributed manner using a common protocol, the ConcePTION common data model (CDM), and a common distributed analytics program. The data pipeline has been developing from the EU-ADR project and was further improved in the IMI-ConcePTION project and used in multiple EMA-tendered and VAC4EU studies. The ConcePTION CDM has been described by Thurin *et al, 2022*.[2]

[2] Thurin, N.H., Pajouheshnia, R., Roberto, G., Dodd, C., Hyeraci, G., Bartolini, C., Paoletti, O., Nordeng, H., Wallach-Kildemoes, H., Ehrenstein, V., Dudukina, E., MacDonald, T., De Paoli, G., Loane, M., Damase-Michel, C., Beau, A.-B., Droz-Perroteau, C., Lassalle, R., Bergman, J., Swart, K., Schink, T., Cavero-Carbonell, C., Barrachina-Bonet, L., Gomez-Lumbreras, A., Giner-Soriano, M., Aragón, M., Neville, A.J., Puccini, A., Pierini, A., Ientile, V., Trifirò, G., Rissmann, A., Leinonen, M.K., Martikainen, V., Jordan, S., Thayer, D., Scanlon, I., Georgiou, M.E., Cunnington, M., Swertz, M., Sturkenboom, M. and Gini, R. (2022), From Inception to ConcePTION: Genesis of a

1.4.7 Statistical analysis

Detailed methodology for summary and statistical analyses of data collected in this study are documented in the statistical analysis plan that was delivered to EMA. All analyses were conducted using R version R-4.0.3 or higher (Foundation for Statistical Computing, Vienna, Austria; or SAS version 9.3 software or higher (Cary, North Carolina, USA; SAS Institute, Inc.).

1.5 Results

1.5.1 Readiness

During the readiness phase, all Data Access Providers (DAP) requested approval to participate in the studies specified in the CVM readiness and rapid assessment protocol (including all potential AESI). The *Extraction, Transformation, and Load* (ETL) design document was updated based on required data. Required data was ETL'ed into the ConcePTION CDM. To assess the quality of the data level 1-3quality checks were conducted. These quality checks were reported in the interim report and comprise assessment of completeness, correctness, plausibility of the data, and accuracy. They were conducted for each data instance, and some data sources conducted these multiple times when data was refreshed (e.g., for updated rapid assessments for myocarditis).
Nine data sources from Italy (ARS, Pedianet, Caserta), Spain (BIFAP, VID, SIDIAP), Netherlands (PHARMO), UK (CPRD) and Norway (national registers) completed this phase. The regional database from Lazio (Italy) could not participate because of administrative issues and data access rules.
The study population on January 1, 2019, included in the readiness assessment comprised a total of 52,862,735 persons, CPRD and BIFAP contributed the largest populations. Data sources had completed data instances with information from December 2021 to June 2022.

Population characteristics

ARS has a relatively old population (8.4% is above 80 years of age) whereas the PEDIANET population is very young since it only captures children 0-14 years of age. The rest of the data sources all had median ages of 40 years of age, with a slightly higher prevalence of women in all data sources. This reflects the national population. The most prevalent co-morbidity at baseline (1/1/2020) was a history of cardio/cerebrovascular disease (28% in ARS and lower in others). Based on the population shapes (level 3 quality checks), the population gender/age trees were similar to national data, and date of birth and gender were available. Some DAPs censored data instances to earlier dates than the extraction dates, to ensure that all databanks would have had the time to be updated.

COVID-19 vaccinations data

COVID-19 vaccination data was available in each of the data sources, and timing of recording as well as uptake percentage was comparable with data from the COVID-19 vaccine tracker at ECDC. The PHARMO data source saw some delays since it was based on GP data, and GPs received the data from the national health agency with delay. All data sources were considered fit 'for' purpose to study COVID-19 vaccination uptake.

In general, more than 70% of persons received Pfizer vaccine in each data source except in UK, followed by Moderna, AstraZeneca and Janssen. In UK the pattern was different, AstraZeneca had a much higher percentage of first dose (48%), Pfizer was first dose for 49% of population, and Janssen vaccine was not used. In Norway, mostly Pfizer and Moderna were used and no Janssen.

Network to Support Better Monitoring and Communication of Medication Safety During Pregnancy and Breastfeeding. Clin. Pharmacol. Ther., 111: 321-331. https://doi.org/10.1002/cpt.2476

For those starting with Pfizer vaccine dose 1, more than 80% had a homologous second Pfizer dose, in PEDIANET second dose was lower, in Norway second dose was frequently Moderna (16.25%). Median distance to second dose differed between regions from 21-63 days (UK) and was much longer when there was a heterologous second dose. In most countries, those vaccinated first with Moderna vaccine had a homologous second dose, in PHARMO and Norway second dose was also frequently Pfizer (14.7% and 12.95% respectively), median distance to second dose was usually 28 days, but there was variation across regions. In persons with AstraZeneca dose 1 a large proportion had a homologous second dose, except in Norway, where 97% used either Pfizer or Moderna as a second dose. The median distance to second dose was between 75-80 days. Boosters after Janssen vaccine were infrequently a Janssen vaccine, the majority had a booster with an mRNA platform vaccine (Pfizer or Moderna).

Strong channeling of different vaccines to certain age groups was observed, which could even change per region within countries. Pfizer was mainly administered to very old people (+80 years old) and children, AstraZeneca was mostly administered between the 50-69 age group, and Moderna was broadly distributed across age categories. AstraZeneca 1st dose users had the highest prevalence of co-morbidities.

AESIs

Age and gender standardized and age-specific incidence rates of AESIs were created for 2019, and 2020 prior to COVID-19 disease, as well as post-COVID-19 disease until vaccination, rates were benchmarked with published data from the ACCESS project[3] and the rates by Li et al.[4], Gubernot et al.[5], mainly. Based on the type of event data that the DAP can access and the setting in which these events are assessed (e.g. in primary care, outpatient specialist and or discharge/emergency) as well as the vocabularies of diagnostic codes, the rates differed, as was described already by Willame et al.[3] The methodological assessment on misclassification shows the impact of the differences of event provenance in studies and this should be considered in the choice of data sources when conducting safety evaluation studies.

Table ES2. AESI list and comparison with ACCESS literature, impact of COVID-19 pandemic and lock down, and heterogeneity by provenance.

AESI	Comparison ACCESS and literature	Effect of lock down	Effect of COVID-19 infection	Heterogeneity by provenance and impact on fitness for purpose
CAD	Consistent	Consistent absolute decrease of 20-40/100,000 PY	1.5-3 fold increase after infection	Underestimation in GP only or hosp. only highest when hosp & outpatient & GP. Norwegian data overestimate due to lack of precise codes
ADEM	Consistently very low (<0.6/100,000)	Not visible, but very rare event	Increased rate after COVID-19	Small data sources do not observe, and neither those with ICPC coding. Hospital data required
ARDS	Lower rates than in ACCESS due to retagging of codes	Lowering of rates	5-80 fold increase	Extreme effect of having hospital data, only data sources with hospital are fit for purpose
AKI	consistent	Decrease of rates	2-10 fold increase	Underestimation in GP only or hosp. only highest when hosp & outpatient & GP. Norwegian data did not have the proper codes in the instance, rates in Caserta and PHARMO too low

[3]Willame C, Dodd C, Durán CE, Elbers R, Gini R, Bartolini C, et al. Background rates of 41 adverse events of special interest for COVID-19 vaccines in 10 European healthcare databases - an ACCESS cohort study. Vaccine. 2023 Jan 4;41(1):251-262. doi: 10.1016/j.vaccine.2022.11.031

[4] Li X, Ostropolets A, Makadia R, Shoaibi A, Rao G, Sena AG, et al. Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. BMJ. 2021 Jun 14;373:n1435. doi: 10.1136/bmj.n1435.

[5] Gubernot D, Jazwa A, Niu M, Baumblatt J, Gee J, Moro P, et al. U.S. Population-Based background incidence rates of medical conditions for use in safety assessment of COVID-19 vaccines. Vaccine. 2021 Jun 23;39(28):3666-3677. doi: 10.1016/j.vaccine.2021.05.016.

AESI	Comparison ACCESS and literature	Effect of lock down	Effect of COVID-19 infection	Heterogeneity by provenance and impact on fitness for purpose
ALI	consistent	Decrease of rates	2-10 fold increase	No adequate data in Pedianet, Caserta and PHARMO. Rest of source fit for purpose
Anaphylaxis	consistent	Decrease of rates	1.5-2 fold increase	No adequate data in the data instance from Norway, more specific ICD10 codes are required. GP data is required
Anosmia, ageusia	consistent	Increase of rates (maybe undetected COVID-19	10-30 fold increase	Hospital data alone are not fit for purpose. GP data are required.
Arrhythmia	consistent	Decrease of rates	2-5 fold increase	All provenances add sensitivity
Arterial thrombosis	Not done in ACCESS	Decrease of rates	2-5 fold increase	GP data alone underestimate, inclusion of hospital data doubles the rate
Bell's Palsy	Not done in ACCESS, but consistent with literature	Small decrease		Caserta and PHARMO data instances not fit for purpose
Chilblain-like lesions	consistent	Small increase	2-5 fold increase	Data from hospital alone not adequate, GP data are required. Instances from Casera, Norway, Caserta are not fit for purpose
Coagulation disorders	Not done as aggregate in ACCESS	decrease	2-10 fold increase	PHARMO, Caserta instances not fit for purpose, hospital & GP data required
Cerebral Venous Sinus Thrombosis (CVST)	consistent	Not much impact	2-5 fold increase	PHARMO, Caserta instances not fit for purpose, hospital & GP data required
Diabetes type 1*	Not comparable against ACCESS. Higher rates than literature	Not much impact	2-5-fold increase, except in IT-ARS where the increase was more than 10 times.	
Disseminated Intravascular Coagulation (DIC)	consistent	Not much impact	Strong increase	GP data alone not fit for purpose for this event
Death (any cause)	Consistent	No impact	Strong increase. 8-23-fold increase.	Instances are fit-for-purpose. No death information in Pedianet and Caserta.
Erythema multiforme	consistent	decrease	Small increase	GP data alone not fit for purpose for this event
Generalized convulsion*				
Guillain Barré Syndrome (GBS)	consistent	decrease	High increase	PHARMO, Norwegian and Caserta instances not fit for purpose
Haemophagocytic lymphohistiocytosis	Not measured in ACCESS	decrease	2-5-fold increase	Hospital data are required, Caserta, Norwegian, PHARMO instance not fit for purpose
Kawasaki´s disease	consistent	No impact	>10 fold (may be MIS)	Caserta, Norwegian and PHARMO instance not fit for purpose
(Meningo) encephalitis	Slightly higher	Decrease in rates	2-5-fold higher	Norwegian data very high. Caserta instance not fit for purpose
Microangiopathy	consistent	Decrease	2-10-fold higher	Data instance from Caserta, CPRD and BIFAP not fit for purpose for this event
Multisystem inflammatory syndrome (MIS)	Lower, since kawasaki was not included anymore	Did not exist as code	Strong increase	ICD9 and ICPC codes do not exist for this condition. Only ICD10 and SNOMED codes. To study MIS & KD should be combined
Myocarditis	consistent	decrease	10-200 fold increase	GP only data underestimate by 50%. PHARMO data not fit for purpose due to lack of specific ICPC
Narcolepsy	consistent	decrease	Increase in some	Hospital only data underestimate. Data instance of PHARMO and Norway not fit for purpose for this event
Pancreatitis	Not measured in ACCESS	Slight decrease	increase	Data sources with ICPC and ICD10 codes had very low rates, code tagging should be inspected
Pericarditis	consistent	decrease	Up to 10-fold increase	PHARMO data not fit for purpose for this event
Rhabdomyolysis	Not measured in ACCESS	decrease	2-fold increase	PHARMO, Norwegian, Pedianet, Caserta data instances not fit for purpose, BIFAP data is low, which may be due to SNOMED codes

AESI	Comparison ACCESS and literature	Effect of lock down	Effect of COVID-19 infection	Heterogeneity by provenance and impact on fitness for purpose
Severe cutaneous adverse reactions to drugs (SCARs)	Not measured in ACCESS	decrease	Up to tenfold increase	ARS, Caserta, PHARMO and Norwegian data sources not fit for purpose for this event.
Sensorineural hearing loss	Not measured in ACCESS	decrease	2-fold increase	Caserta and ARS data not fit for purpose, outpatient data is required
Single organ cutaneous vasculitis (SOCV)	Decrease due to reclassification of narrow codes	decrease	increase	PHARMO, Caserta, ARS, Pedianet and Norwegian data instances not fit for purpose
Stroke haemorrhagic	Lower	decrease	increase	Hospital data are required. Caserta, Pedianet, Norwegian data not fit for purpose. GP only underestimates
Thrombocytopenia	consistent	decrease	2-5-fold increase	Caserta, Norwegian, PHARMO data instances are not fit for purpose
TTS	Slightly higher	No impact	Increase, especially strong in FISABIO, SIDIAP and PHARMO.	Pedianet and NHR instances are not fit for purpose.
Thyroiditis (autoimmune)	Not measured in ACCESS	decrease	increase	Norwegian, ARS, PHARMO, Caserta data not fit for purpose, GP & Hospital data are required
Transverse myelitis	consistent	decrease	increase	Norwegian, PHARMO, Caserta and Pedianet instances not fit for purpose. BIFAP rates lower than expected.
VTE	consistent	decrease	Tenfold increase	Both GP & Hospital data are required, otherwise underestimation, Norwegian data overestimate
*Full analysis is presented in the final report.				

1.5.2 Conduct of electronic healthcare records-based rapid assessment studies

During the 2-year phase of the project, EMA requested 3 rapid evaluation studies to address emerging safety concerns under review by PRAC or research questions important to support regulatory decision-making.

Multi-inflammatory syndrome (MIS)

The request from EMA was to generate incidence rates (IRs) for MIS stratified by COVID-19 and pre-post-vaccination. The analysed study population included more than 6 million persons, with 650,731 children aged between 0 and 17 years old. Since MIS is a condition related to COVID-19 disease, MIS codes were created only at the end of 2020. ARS-IT could not identify MIS codes as this data source makes use of ICD9 codes, which are not updated anymore. In the absence of MIS codes, KD-like disease codes were used by the Italian colleagues due to the reported association between MIS and KD in children. Rates of KD were highest in 0-11 years old individuals, both in males and females, with only one case of MIS effectively occurring after the COVID-19 pandemic, in 2021. An increment of the KD-like disease cases in 0-11 years old children was also observed in 2020, during the COVID-19 pandemic. KD and MIS rates were both very low. No cases of KD & MIS in children post-vaccination were observed, also because very few vaccinated children were present on the April and May 2021 data extractions of BIFAP and ARS, respectively.

For this final report updated Kawasaki and MIS specific incidence rates were calculated. Kawasaki disease rates increased more than 10-fold after COVID-19 diagnosis, and MIS also increased very much, but could only be observed in Norwegian data after COVID-19, which have issues with specificity of the codes.

COVID-19 severity in children

The EMA Pediatric Committee (PDCO) requested an estimation of the incidence rates of serious COVID-19 in children, Data were initially presented to the PDCO in July 2022, and a final report delivered on May 8, 2023. Results have been updated for this final report, including data from Norway since it has fit-for-purpose for this study.

Four COVID-19 severity levels were considered (diagnosis, hospitalization, intensive care unit admission, and death after COVID-19*).* Non hospitalized COVID-19 disease was considered non-severe, and severe disease was hospitalization, ICU or death.

The total study population comprised 6,719,867 under 18 years old individuals (51% women) across the 7 data sources. The median age ranged from 6-10 years old. The at-risk of severe COVID-19 disease population comprised 445,174 (6.6%) children and adolescents with comorbidities. Vaccine uptake in children (mostly Comirnaty) was mainly from July 2021 and September 2021 in Italy and Spain, respectively, whereas in Norway in September 2021 for adolescents. In children and adolescents without risk factors, the highest incidence rates of non-severe COVID-19 across data sources varied between 27 to 143 cases/100 PY in December 2021 and January 2022. Rates were much lower (0 to 1/100 PY) for severe COVID-19 infection. Incidence rates of severe COVID-19 were higher among children and adolescents with at-risk conditions for a severe disease. Overall, mortality cases were almost zero across all databases and cohorts.

Myocarditis and pericarditis

EMA requested to evaluate the signal of COVID-19 vaccines and myocarditis/pericarditis at the end of September 2021. Study results were first reported to EMA and PRAC in November 2021, updates with additional data sources and more follow-up were conducted. Results have been published in a peer-reviewed journal.[6] To include longer follow-up and data sources, an update of the SCRI myocarditis was again presented to PRAC in January 2023. In this report (May 2023), we include a re-analysis taking account of small methodological adjustments.

Key primary results from the May 2023 analysis with fit for purpose data sources confirmed what had been found before: Pfizer dose 2 and Moderna dose 2 were associated with an increased risk of myocarditis in persons below 30 years of age, and not for a booster Pfizer dose, but it persisted when the third dose was Moderna. Analyses by week rather than 28 days, showed that elevations of risk occurred.

Exclusion of subjects with COVID-19 during follow-up resulted in an increase of the IRR (not stratified by age) for second dose of Pfizer, Moderna and AstraZeneca, which were all significantly elevated. After exclusion of persons with COVID-19 disease, third doses were not associated with significant elevation anymore. The negative control sensitivity analysis showed estimates around 1 and an effect towards the 1 when persons with COVID-19 were excluded.

1.5.3 Methodological assessment

The robustness of the methodology and study designs used in EMA requested analyses was assessed using a combination of empirical research and simulation studies.

Choice of comparator

The objective of this methodological study was to explore the following three different approaches to selecting comparator groups for studying the risk of COVID-19 vaccines and myocarditis:

- Approach 1: COVID-19 vaccination vs. no vaccination; time zero of the comparator group is assigned via matching on calendar time to the date of vaccination in the exposed group.

[6] Bots SH, Riera-Arnau J, Belitser SV, Messina D, Aragón M, et al. Myocarditis and pericarditis associated with SARS-CoV-2 vaccines: A population-based descriptive cohort and a nested self-controlled risk interval study using electronic health care data from four European countries. Front Pharmacol. 2022 Nov 24;13:1038043. doi: 10.3389/fphar.2022.1038043.

- Approach 2: COVID-19 vaccination vs. no vaccination; time zero of the comparator group is assigned via random sampling and seasonality is modelled.
- Approach 3: vaccination with one brand vs. vaccination with a different brand; time zero will be the date of vaccination.

The study design comprised the emulation of two hypothetical target randomized controlled trials, comparing the risk of myocarditis and/or pericarditis of A) each COVID-19 vaccine distributed in Europe against no vaccination; and B) each COVID-19 vaccine distributed in Europe against the Pfizer COVID-19 vaccine.

This report presents first data from the CPRD in the UK to emulate a target trial that compares vaccination vs. no vaccination via random sampling of time zero in the comparator group.
A total of 8,417,115 eligible vaccinated individuals were included, and for these, 8,417,115 eligible non-vaccinated individuals were selected at random from the calendar period of December 2020 to March 2021. We found that the application of stabilized inverse probability weights to the eligible population resulted in balancing of the baseline variables, which would have been achieved by design via the more computationally intensive approach of matching.

Sensitivity analysis for SCRI on the association between COVID-19 vaccines and myocarditis

A series of sensitivity analyses were conducted comparing different design options for self-controlled studies and their sensitivity to violations of a core assumption (event-dependency of the exposure) using the association between COVID-19 vaccines and myocarditis. We found no evidence that using control time before the first dose of the vaccination had led to a systematic overestimation of the relative risks in core analyses with good agreement between four different self-controlled study designs (an SCRI using time before vaccination as control time, an SCRI using time after vaccination as control time, a standard SCCS, and an extended SCCS). This is reassuring, and together with a series of graphical assessments indicates that bias due to event-dependency of the exposure was likely limited in our evaluation of the association between COVID-19 vaccines and myocarditis. A simulation experiment supported these conclusions, and showed that using a pre-exposure period in a standard SCCS may introduce bias depending on the nature of the event-dependency of the exposure. This is an interesting finding with implications for the design of future vaccine safety studies and highlights the value of both simulation experiments and sensitivity analyses using different designs when the impact of an adverse event on vaccination probability is uncertain.

Unmeasured confounding in the association between COVID-19 vaccines and myocarditis

To explore the impact of unmeasured (residual) confounding we evaluated three different methods that would deal with/show unmeasured confounding, namely negative control outcomes (NCOs), quantitative bias analyses, and instrumental variable (IV) analyses.

- The NCO analysis repeated the rapid assessment study on myocarditis but replaced the true outcome (myocarditis) with an NCO (otitis externa). Overall, the negative control analyses indeed showed null results across vaccine brands and dosing instances. This analysis confirms earlier work suggesting that the SCRI design is robust to unmeasured confounding and expands that to include a multi-dose setting where the time between the pre-vaccination control window and dose-specific risk windows increases for each subsequent dosing instance.
- The quantitative bias analysis consisted of a simulation SCRI study where we introduce a confounder *U* that is increasingly associated with the outcome and then check whether the effect estimate changes in the presence of this confounder. Our results suggest that unmeasured confounding in the risk window biases the effect estimate more strongly than unmeasured confounding in the control window.

- For the IV instrumental variable, our findings confirm earlier work on IVs in vaccination settings, which showed it was impossible to find valid IVs for vaccine administration.

Misclassification of outcomes

Data sources used heterogeneous provenances and meanings of events to detect AESIs, partly due to the availability of data banks, partly due to choices made by the local experts (e.g. as to in- or exclude secondary discharge diagnoses). We showed empirically that many AESIs may have low sensitivity when narrow diagnostic codes are used.

A literature review of diagnostic codes of AESI informed the simulation study settings. The bias of three different estimates (log of odds ratio, relative risk, and rate difference) was provided for a setting where there is nondifferential misclassification. *For RR estimates* in the non-differential misclassification, we observed the bias towards null as expected.

We considered two settings of differential misclassification, namely (1) when the specificities for both vaccinated and non-vaccinated are the same and (2) when sensitivities and specificities for both vaccinated and unvaccinated groups are different. When specificities for both exposure groups are the same, the EM and naïve estimators perform equally poorly especially for low rate of outcome. The same pattern is also observed in the case of higher sensitivities and sample sizes lower than 1M. When differential misclassification is occurring bias can be large and unpredictable.

1.6 Discussion

The CVM EHR data studies had three objectives, first to create readiness of data sources and assess whether data sources were fit-for-purpose. All 9 data sources were fit for purpose as regards population and COVID-19 vaccinations, but depending on the AESI would not be fit to participate in evaluation studies due to misclassification of the AESI.

Misclassification depended on type of databanks that were available in the data sources (primary care, outpatient specialist and hospitalization), meanings of codes (primary discharge vs. secondary discharge diagnoses) as well as the use of narrow (specific) codes and/or broad codes (sensitive). Most fit for all type of events were data sources that could link GP data to hospital data (e.g. SIDIAP, BIFAP, VID) A review of the existing literature on the PPV of these events showed a range of false positive rates and an impact on the RR which would lead to bias towards the null in case of non-differential misclassification and different directions when there would be differential misclassification in comparative studies.

Confounding may have impacted the results of the evaluation study on COVID-19 vaccines and myocarditis which the EMA requested. We showed considerable channeling of certain COVID-19 vaccines towards specific age groups, which could confound comparative studies. The self-controlled designs automatically adjust for time-fixed confounding factors but are still sensitive to time-varying confounding. COVID-19 disease was a strong time varying confounder that needed to be controlled for (adjustment, restriction) post vaccination follow-up data are not rapidly available during a vaccination campaign because of time lags, and multiple vaccine doses, among other reasons. Design choices such as pre-vaccination control or post-vaccination control period needed to be made. It was shown that using a pre-vaccination control period did not overestimate the effect but rather yielded a more conservative estimate. The SCRI design was less susceptible to time varying confounding than the SCCS design.

1.7 Conclusion

The CVM EHR and methodology studies showed that several data sources are ready to evaluate COVID-19 vaccine-AESI associations, but data sources are not always fit for each type of event.

Depending on the health care setting where such events are diagnosed and treated, and the provenance of the databanks, a data instance may or may not be fit.
Misclassification of the outcome may have a large impact on the absolute and relative estimates and only the 'fit' data should be used. Because of the large channeling of the different vaccines, the designs chosen (SCRI) dealt best with time stable and time varying confounding. Using this design, we were able to estimate the associations between COVID-19 vaccines and myocarditis repeatedly. For myocarditis we showed significant associations between the second dose of mRNA platform vaccines and myocarditis. When we excluded patients diagnosed with COVID-19, the relative risks increased and also showed a significant association for the AstraZeneca vaccine. Other associations can be studied using this design with fit for purpose data sources for the AESI.