WP1. COVID-19 infection and medicines in pregnancy – a multinational registry-based study
Protocol Amendment Plan[1]
E. Hurley/H. Nordeng, 28th May 2022

Amendment or update	Reason	Section of the protocol
DAPs: The Leibniz Institute for Prevention Research and Epidemiology - BIPS group in Germany will not participate in the project.	They are not able to provide necessary data during pandemic period due to long lag time	9.4. Data sources
Use of historical controls: We will present prevalence estimates of adverse pregnancy and neonatal outcomes for women whose pregnancy was completed prior to the start of the pandemic (2018/2019). This will not be a matched analyses, but provide relevant baseline rates that will be valuable to discussed in the final report.	Change discussed and agreed with EMA. Note: FDA/Sentinel are not presenting historical baseline rates of adverse pregnancy outcomes. Thus, these data from WP1 will not be useful for the meta-analyses. FDA restricts to a matched analysis between women+-COVID-19 whose pregnancies aligned time wise during the pandemic.	9.7.5 Analytical approaches – objective 3 Adverse pregnancy and neonatal outcomes in pregnant women with COVID-19
Type and level of Medication examined: Medication use will be examined at ATC level-2 listed in the protocol, with some further additions: (antineoplastic agents (L01) antigout preparations (M04), antiprotozoals (P01), antihelminthics (P02)). Furthermore, we have pre-specified those medications of special interest in COVID-19 that will be examined at the ATC- level-3, ATC-level-4 and ATC level-5; conditional on sufficient sample size *(≥5 cases per cell)*	Accommodation of request from the EMA. Following review of the literature on medications of special interest, this comprehensive list of medications was derived.	9.3.2. Exposure to medication And 9.3.4 Outcomes
COVID-19 severity classification: Instead of classifying severity of COVID-19 disease according to the original WHO 5 point scale, it will be categorized as severe and non-severe as follows: Non-severe COVID-19: Recording of a positive COVID-19 test (PCR or antigen test) or diagnosis with no subsequent hospital admission with a recording (primary or secondary) of COVID-19. Severe COVID-19: A hospitalisation with a primary or secondary discharge diagnosis of COVID-19.	Following recognition that distinctions in severity of COVID-19 during hospitalization would not be accurately measurable by all Data Access Providers, this new categorization was agreed upon with EMA, post protocol approvement.	9.3.4. Outcomes Objective 2 Severity of COVID-19
Analyses of Major congenital anomalies:	Due to expected sample size. Note: Aligned with FDA.	9.3.4. Outcomes

[1] Includes specifications outlined in the Statistical Analysis Plan v0.9, 18 April 2022.

Major congenital anomalies will be grouped together, and not reported upon individually given the small numbers.		Objective 3: pregnancy and neonatal outcomes
Approach to analysing severity of COVID-19: Obj. 2: Instead of assessing the impact of medication use on COVID-19 severity, we will stratify by COVID-19 severity: Severity of COVID-19 will be used as a stratification factor in all analyses in Objective 1 (prevalence of medication use) and in Objective 3 (impact of COVID-19 on pregnancy and neonatal outcomes). Obj.2 is thus embedded in Obj. 1 and Obj. 3. Obj.3 Research question 3b (ii) will be restricted to pregnant women with non-severe COVID-19 alone, to avoid misclassification of medicines exposure during hospitalization.	Revision based on results from WP2. Severity of COVID-19 is so highly correlated with the type and number of medications used, that this analysis is less informative than expected. Furthermore, WP1 is based on outpatient dispensing, as we do not capture in hospital use, which is done in WP2 and 3. This decision was reached in agreement with EMA post protocol approval.	9.7.4 Analytical approaches – objective 2 Severity and clinical outcomes of COVID-19 disease in pregnant women. 9.7.5 Analytical approaches – objective 3 Adverse pregnancy and neonatal outcomes in pregnant women with COVID-19
Modelling of associations: Due to small sample size, we will focus on descriptive results for medication use for all objectives, and will in Obj. 3 only model the association between COVID-19 and adverse maternal and neonatal outcomes. The prevalence of individual pregnancy and neonatal outcomes will be presented per level of medication exposure (for each medication class), and separately for trimester of infection: 1. Exposed to medication X in the 30 days pre COVID-19 positive test/diagnosis 2. Exposed to medication X in the 30 days post COVID-19 positive test/diagnosis *3.* Not-exposed to medication X in 30 days post COVID-19 positive test/diagnosis	This decision was reached in agreement with EMA post protocol approval.	9.7.5 Analytical approaches – objective 3 Adverse pregnancy and neonatal outcomes in pregnant women with COVID-19 Research question 3a (descriptive) and 3bi (analytical), have been combined for ease of presentation.

Protocol

COVID-19 infection and medicines in pregnancy – a multinational registry based study

Version 1.0
31.01.2021

Title	COVID-19 infection and medicines in pregnancy – a multinational registry based study
Protocol version identifier	1.0
Date of last version of protocol	January 31nd 2021
EU PAS register number	Study will be registered in the EU PAS register.
Active substance	**NA**
Medicinal product	**NA**
Product reference	**NA**
Procedure number	EMA/2017/09/PE/11 (Lot 4). Study on impact of COVID-19 infection and medicines in pregnancy.
Marketing authorisation holder(s)	None
Research question and objectives	This study is based on electronic health care registry data from 9 health care data bases in 8 European countries between 2018 and 2020. The primary objectives are: **1)** To estimate the prevalence of medicines used, by trimester of pregnancy and compare this between pregnant women with COVID-19, pregnant women without COVID-19 and non-pregnant women of reproductive age with COVID-19. **2)** To describe severity and clinical outcomes of COVID-19 in pregnant and non-pregnant women. **3)** To compare the rates of adverse maternal and neonatal outcomes in pregnant women with and without COVID-19, using different medicines.
Country(-ies) of study	Participating electronic health care databases among CONSIGN partners: Denmark – nationwide registries France – nationwide registries Germany – 20% population coverage Italy – regional registry (Tuscany) Norway – nationwide registries Spain – regional registries (Valencia and Aragon) Sweden – nationwide registries UK – national registry (Wales)
Authors	**Professor Hedvig Nordeng** PharmacoEpidemiology and Drug Safety Research Group, Faculty of Mathematics and Natural Sciences, University of Oslo, Norway. Post box 1068 Blindern, 0316 OSLO, Norway Telephone: +47 450 33 984 E-mail: h.m.e.nordeng@farmasi.uio.no **Professor Miriam Sturkenboom** University Medical Center Utrecht, Julius Global Health, Julius Center Heidelberglaan 100, 3584 CX Utrecht, The Netherlands Telephone: +31657831983

Contributors	Andrea Margulis, Bradley Layton, Alejandro Arana, Estel Plana, Elena Rivero (RTI Health Solutions) Angela Lupattelli (University of Oslo, UiO) Caitlin Dodd (University Medical Center Utrecht) Satu Siiskonen (University Utrecht, UU) Vera Ehrenstein (Aarhus University, AUH) Rosa Gini (Agenzia Regionale di Sanita, ARS) Ulrike Haug, Tania Schink (Leibniz Institute for Prevention Research and Epidemiology - BIPS) Carolyn Cesta, Helle Kieler, Marie Linder (Karolinska Institute, KI) Gabriel Sanfelix-Gimeno, Isabel Hurtado, Clara Rodriguez, Anibal Garcia Sempere, Salvador Peiro (The Foundation for the Promotion of Health and Biomedical Research of Valencia Region, FISABIO) Mònica Sabaté, Xavier Vidal, Judit Riera, Cristina Aguilera, Elena Ballarin (Catalan Institute of Pharmacology Foundation, FICF) Beatriz Poblador-Plou, Alexandra Prados-Torres, Antonio Gimeno-Miguel, Antonio Poncel-Falcó, Aida Moreno-Juste, Jonás Carmona-Pirez (Instituto Aragones de Ciencias de la Salud, IACS) Cécile Droz-Perroteau (Université de Bordeaux, BPE) Nicolas Thurin (Université de Bordeaux, BPE) Sue Jordan (Swansea University, USWANSEA)

Table of contents

List of 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
CEPI	Coalition for Epidemic Preparedness Innovations
CI	Confidence interval
DAP	Data Access Provider
DDD	Daily defined dose
DNA	Desoxyribonucleic acid
DRE	Digital Research Environment
ECDC	European Centre for Disease Prevention and Control
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
ICD	International Classification of Diseases
ICMJE	International Committee of Medical Journal Editors
ICU	Intensive Care Unit
IMI	Innovative Medicines Initiative
LMP	Last menstrual period
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
TOPFA	Termination of Pregnancy for Fetal Anomaly
VAC4EU	Vaccine monitoring Collaboration for Europe

1. Title

COVID-19 infection and medicines in pregnancy – a multinational registry-based study

2. Marketing authorisation holder(s)

Not applicable

3. Responsible parties

Utrecht University (UU), The Netherlands
Olaf Klungel, PhD, professor, lead of the EU PE&PV research network
University Medical Center Utrecht (UMCU), The Netherlands
Miriam Sturkenboom, PharmD, PhD, MSc, professor & CONSIGN project PI
University of Oslo (UIO), Norway
Hedvig Nordeng, PhD, Professor & Study lead

Collaborating institutions

Country	Coordinating organization	Contact person
Denmark	Aarhus University Hospital (AUH) University of Copenhagen (UCPH)	Vera Ehrenstein Morten Andersen
France	Université de Bordeaux, Bordeaux PharmacoEpi (BPE)	Cécile Droz-Perroteau
Germany	Leibniz Institute for Prevention Research and Epidemiology - BIPS	Ulrike Haug
Italy, Tuscany	Agenzia Regionale di Sanita Toscana (ARS)	Rosa Gini
Norway	University of Oslo (UIO)	Hedvig Nordeng
Sweden	Karolinska Institute (KI)	Carolyn Cesta
Spain, Valencia	The Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO) Catalan Institute of Pharmacology Foundation (FICF)	Clara Rodriguez Mònica Sabaté
Spain, Aragon	Instituto Aragones de Ciencias de la Salud (IACS)	Beatriz Poblador-Plou
Spain	RTI Health Solutions (RTI)	Andrea Margulis
UK, Wales	Swansea University (SWANSEA)	Sue Jordan

See also **Annex 1** with list of all project participants.

4. Abstract

Title: CONSIGN: Covid-19 infectiOn aNd medicineS In pregnancy – a multinational observational cohort study based on electronic health care registries

Main authors:
Prof. dr. Hedvig Nordeng, University of Oslo, Norway
Prof. dr. Miriam Sturkenboom, University Medical Center Utrecht, The Netherlands
Dr. Caitlin Dodd, University Medical Center Utrecht, The Netherlands

Rationale and background: More than 100 million pregnant women around the world may be exposed to the new coronavirus (SARS-CoV-2). Currently, we do not know how pregnant women in general are cared for during the pandemic, whether they are more easily infected with SARS-CoV-2 or more likely to have severe illness than the general population, whether COVID-19 can directly or indirectly harm the unborn child, and how pharmacological treatments modify the above.

Research question and objectives:
The primary objectives are:
1) To estimate the prevalence of medicines used, by trimester of pregnancy, and compare this among pregnant women with COVID-19, pregnant women without COVID-19, and non-pregnant women with COVID-19.

a. To estimate the prevalence of medication use in pregnant women with COVID-19, by age and trimester of pregnancy.

b. To compare these data with those collected for pregnant women without COVID-19, by age and trimester of pregnancy.

c. To compare these data with those collected for non-pregnant women of reproductive age with COVID-19, by age.

2) To describe severity and clinical outcomes of COVID-19 disease in pregnant women with COVID-19, according to treatments received during pregnancy, and compare these data with those of non-pregnant women of reproductive age with COVID-19.

a. To describe the severity of COVID-19 disease in pregnant women according to treatments received during pregnancy (including 'no treatment'), by age, and trimester of pregnancy at diagnosis.

b. To compare these data with those of non-pregnant women of reproductive age with COVID-19, by age.

3) To assess and compare the rates of adverse maternal and neonatal outcomes in pregnant women with and without COVID-19, using different medicines.

a. To assess and compare the rate of relevant adverse pregnancy and neonatal outcomes in different treatment groups of pregnant women for COVID-19, by age group and trimester of pregnancy at diagnosis

b. To compare estimated rates in pregnant women with COVID-19 to those estimated for pregnant women without COVID-19 during the pandemic period, by age group and trimester of pregnancy at diagnosis.

Study design: A retrospective multi-database dynamic cohort study, conducted during the years 2018 to 2020, including a period of SARS-CoV-2 circulation in Europe, or until the date of last data availability for each data source. Within the cohort we will conduct nested case control analyses

Population: The study population will include women of reproductive age (12-55 years), pregnant women and their children observed in one of the participating data sources for at least one day during the study period (01.01.18 – 31.12.20/last data availability).

Variables:
COVID-19: identified by diagnostic codes in surveillance systems, health care records and/or laboratory test results.
Severity of COVID-19: classified according to WHO[1] scale as follows:
Level 1: any recorded diagnosis;
Level 2: hospitalization for COVID-19 (confirmed or suspected);
Level 3: ICU admission in those with COVID-19 related admission;
Level 4: Acute respiratory distress requiring ventilation (ARDS) during a hospitalization for COVID-19;
Level 5: death during a hospitalization for COVID-19 (any cause)
Exposures: All medications, identified by outpatient prescription/dispensing codes will be described by therapeutic class (ATC-level 2) and when sufficient exposures, on an individual drug level (ATC-level 5).
Risk/protective factors for COVID-19: trimester of infection, age at the start of pregnancy, obesity, smoking, parity, co-morbidity (e.g. cardiovascular, respiratory, diabetes, rheumatic diseases, cancer, mental disorder), vaccinations (e.g. influenza), timing of pregnancy during the pandemic.
Pregnancy outcomes: congenital anomalies (incl. microcephaly), preterm birth, low birth weight, small for gestational age/ intrauterine growth restriction, spontaneous abortions, stillbirth, type of delivery, induced terminations of pregnancy for fetal anomaly (TOPFA).
Neonatal outcomes: Infection (e.g. SARS-CoV-2 infection), low APGAR score at 5 min, neonatal complications, neonatal death.
Available variable information will vary across data sources.

Data sources: The study will include data from 9 electronic health care registries in 8 European countries (Denmark (DK), Germany (DE), France (FR), Italy (IT), Norway (NO), Spain (ES), Sweden (SE), United Kingdom (UK)). Data sources include general practice databases (e.g. UK, ES), claims databases (FR, DE) and record linkage of demographic data, registers and dispensing (SE, NO, DK, IT, ES).

Study size: The source population will include data from 9 data sources in 8 European countries, comprising 113 million individuals and approximately 1 million pregnancies per year, depending on the in- and exclusion criteria that will differ for each objective the study cohorts.

Data analysis:
Descriptive analysis will focus on 3-monthly prevalence rates of medication use, incidence rates of COVID-19 outcomes and prevalence of pregnancy outcomes. Log-binomial regression model and Cox proportional hazards regression (survival analysis) will be used as appropriate. Advanced confounder adjustment methods, including propensity score methods, will be used to mitigate measured

confounding. Timing of COVID-19 infection in pregnancy as well as medication use and risk factors for the outcome will be taken into account.
Data will be accessed in a distributed manner using a common protocol, the ConcePTION common data model and common analytics developed through the ConcePTION collaboration. Data will be transformed locally to the ConcePTION CDM and analysed using R-scripts that are generated centrally, results will be sent to the digital research environment for pooling. ConcePTION data quality indicators will be used.
Results will be presented separately for each data source and aggregated across data sources.

5. Amendments and updates

None

6. Deliverables and Milestones

Deliverable	**Date**
D2a. Protocol submitted to the EMA	16 October 2020
D3. Interim report, objective 1	16 July 2021
D4. Final report of study results	16 July 2022

Milestone	**Planned date**
Protocol submitted to the EMA	16 October 2020
Extract, Transform, and Load (ETL) design finalized	February 2020
Statistical analysis plan finalized	March 2021
Data extraction & ETL	March/ April 2021
Running quality & first analysis	April 2021
Interim report, objective 1	16 July 2021
Data extraction, objective 2 and 3	March 2022
Updated report of study results, objective 2-3	16 July 2022

7. Rationale and background

7.1 Background

Over 100 million pregnant women worldwide, including more than 5 million pregnancies in Europe in 2020, are potentially at risk of infection with SARS-CoV-2. The new coronavirus leaves pregnant women, health care providers and authorities with numerous questions, but with little data to guide them. Although several case series, small cohort studies and meta-analyses[2-15] of COVID-19 and pregnancies are appearing, we still do not know the full consequences of COVID-19 infection and treatment on maternal and neonatal health and early infant development. Foregoing clinical care may also impact pregnancy outcomes. It is unknown how the pandemic impacts fertility, birth rates, pregnancy outcomes and pregnancy termination. Most of the evidence available today is based on third trimester exposures to COVID-19 and use of medicines. Key questions remain about 1st and 2nd trimester exposures.

There is no information about the impact of compassionate medication use for SARS-CoV-2 in pregnancy. Very little data exists on the fetal safety of SARS-CoV-2 medications, and what we do know is gleaned from the use of these drugs for other indications. For example, a case series of six pregnant women with Ebola treated with remdesivir in pregnancy did not show any adverse pregnancy outcomes[16]. For lopinavir/ritonavir, data on about 300 pregnancies did not suggest an increased risk of malformations, but preterm delivery, low birth weight and stillbirth have been observed[17]. Longstanding concerns regarding the risk of retinopathy in infants exposed to (hydroxy)chloroquine *in utero* remains unresolved[18]. Management of COVID-19 during pregnancy is currently based on clinical "best guess" and includes aggressive infection control procedures, testing for SARS-CoV-2 and coinfection, antibiotics (for secondary bacterial infection risk) and fetal monitoring[19]. Advice on delivery mode was lacking in the early part of the pandemic[20]. WHO considered that COVID-19 positive status alone is not an indication for caesarean section.[21]

7.2 Rationale

This study is funded by the European Medicines Agency (EMA) as part of the CONSIGN project (specific contract 04 implementing framework contract No EMA//28/PE, Lot 4).

In the current pandemic, pregnant women are becoming infected with the SARS-CoV-2 virus and being treated for COVID-19 disease and its complications. Most of this treatment is off-label when administered in pregnancy. EMA wishes to be prepared for situations where questions arise regarding the impact of medicine use for COVID-19 during pregnancy on the unborn child. To answer such questions, insight will be needed into the natural history of COVID-19 disease, including disease severity in pregnant women, pregnancy outcomes, impact of COVID-19 in different gestational ages, and drug utilisation in affected pregnancies compared with other pregnancies. The regulatory network will be better placed to judge the appropriateness of use in pregnancy, as well as the appropriateness of any proposed pregnancy-related risk minimisation measures, with better insight into how the disease and currently used treatments impact on the pregnant woman and her infant.

8. Research question and objectives

8.1 Primary objectives

The three primary objectives have been defined by EMA. The primary objectives are:

1. To estimate the prevalence of medicines and other treatments used in pregnant women with COVID-19, and compare these data with those collected for:
a) pregnant women without COVID-19
b) non-pregnant women of reproductive age with COVID-19.

2. To describe the severity and clinical outcomes of COVID-19 disease in pregnant women with COVID-19 according to treatments received during pregnancy (including 'no treatment') and compare these data with those of non-pregnant women of reproductive age with COVID-19.

These comparisons will take into account the trimester of pregnancy (in the pregnant women) and other predictors of clinical outcomes such as underlying health conditions and obesity.

3. To assess and compare the rate of relevant adverse pregnancy and neonatal outcomes in different treatment groups of pregnant women for COVID-19 and its clinical sequelae in the first, second and third trimester of pregnancy (including 'no treatment').

The rates estimated in pregnant women with COVID-19 will be compared to those estimated for pregnant women without COVID-19 during the pandemic period.

8.2 Specific objectives

1. Assess and compare the prevalence of medication use used in pregnant women with COVID-19
The specific objectives are:
a. To estimate the prevalence of medication use in pregnant women with COVID-19
b. To compare prevalence of medicine use between pregnant women with COVID-19 with those for pregnant women without COVID-19
c. To compare prevalence of medication use between pregnant women with COVID-19 and non-pregnant women of reproductive age with COVID-19.

2. Severity and clinical outcomes of COVID-19 disease in pregnant women and the impact of medications
The specific objectives are:
a. To describe the severity of COVID-19 disease in pregnant women and non-pregnant women with COVID-19
b. To compare medication use between severe and non-severe clinical outcomes and the interaction with pregnancy

3. Rates of relevant pregnancy and neonatal outcomes
The specific objectives are:

a. To describe and compare the rate of pregnancy and neonatal outcomes between women with and without COVID-19
b. To assess the impact of COVID-19 on pregnancy and neonatal outcomes
c. To assess the impact of medication use on pregnancy and neonatal outcomes in COVID-19 and non-COVID-19 women and assess the interaction between COVID-19 and medication use.

All medication use refers to outpatient prescribed/dispensed medications as recorded in the electronic health care registries participating in the project.

9. Research methods

9.1. Study design

The study will be a retrospective, multi-database, dynamic cohort study, conducted during the years 2018 to 2020, including a period of SARS-CoV-2 circulation in Europe until the date of last data availability for each data source.

The source population will include women of reproductive age (12-55 years), pregnant women and their children observed in one of the participating data sources for at least one day during the study period (01.01.18 – 31.12.20/last data availability).

Studying the outcomes of pregnancy and identifying pregnancies is subjected to left and right censoring. Several time varying variables need to be considered, pregnancies last 9 months before outcomes can be measured and many data sources identify pregnancies based on end of pregnancy.

Figure 1 shows how the left and right censoring operate.
Left censoring is caused by the study start date, 1/1/2018, to avoid left censoring bias for pregnancy analyses we will only include pregnancies that start after 1/1/2018 (B, C, D) and the source population should have data available in 2017.
Right censoring of pregnancies is a more important concern, any data cut that does not capture full information up until October 2021, will suffer from right censoring. Since data sources identify pregnancies based on pregnancy end, shorter follow-ups would select preterm births. Therefore, the first analysis will have bias on all pregnancies for which the end cannot be observed prior to the data cut.

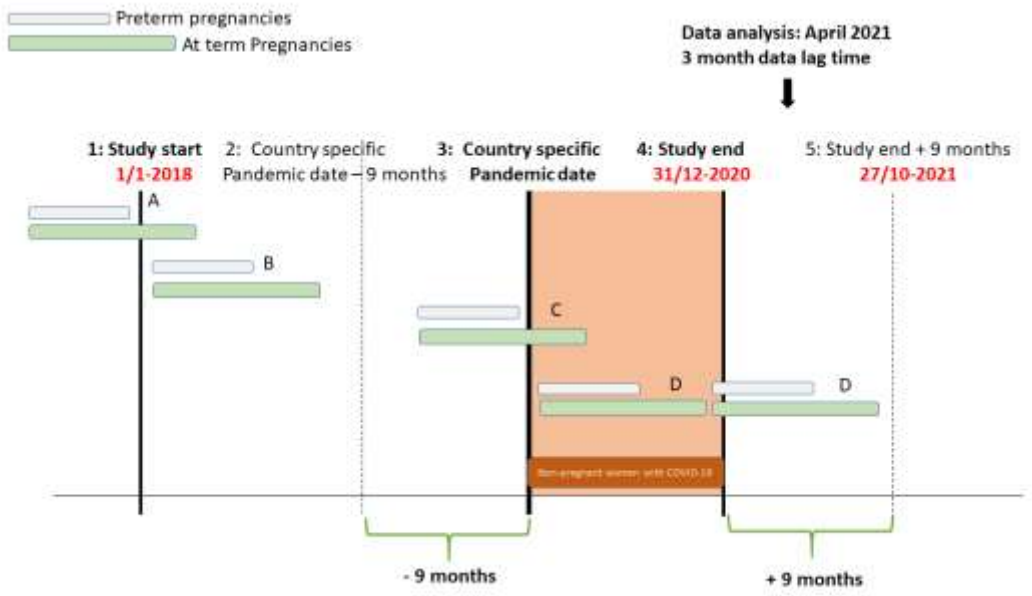


Figure 1. Schematic illustration of pregnancy cohorts considering SARS-COV-2 circulation. In brown: the group of non-pregnant women with COVID-19.

9.2. Setting

The study results will include data from 8 data sources in 7 European countries, comprising 113 million individuals (Table 1). Data sources are described in section 9.4.

Table 1. Overview of data sources to be used for the study

Country	Data Access Provider	Name Data source	Active population	Type of data source	Diagnoses recordings	Medical birth registry data	Mortality registry
Denmark	Aarhus University	Danish Registries	5.8 million	Record linkage	Hospital	Yes	Yes
Germany	BIPS	GePaRD	16 million	Health insurance	GP, Hospital	Yes	No
France	BPE*	SNDS	6.7 million	Health insurance	Hospital	No*	Yes
Italy	ARS	ARS	3.6 million	Record linkage	Hospital**	Yes	No
Norway	University of Oslo	Norwegian registries	5.4 million	Record linkage	GP, Hospital	Yes	Yes
Spain-Valencia	FISABIO	VID	5 million	Record linkage	GP, specialist Hospital	Yes	Yes
Spain- Aragon	Aragon Health System	PRECOVID Study and EpiChron Cohort	1.3 million	Cohort	GP, Hospital	Yes	No
Sweden	Karolinska Institutet	Swedish registries	10.2 million	Record linkage	Specialists, Hospital, Pregnancy Registry	Yes	Yes

Country	Data Access Provider	Name Data source	Active population	Type of data source	Diagnoses recordings	Medical birth registry data	Mortality registry
Wales, UK	USWAN	SAIL	3 million	Record linkage databank	GP, Hospital	Yes	Yes
Total			**57 mill**				

GP: General practitioner. * Given the very broad inclusion criteria, a representative 1/10th sample of the full SNDS will be used. Birth related outcomes are available through hospital discharge summary database, but there is not specific birth registry. **also emergency admissions, exemptions from copayment for chronic conditions, access to mental healthcare.

9.3. Variables

9.3.1 SARS-CoV-2 and COVID-19

Exposures of main interest are SARS-CoV-2 infections and COVID-19, which will be identified by recordings in surveillance systems or diagnostic codes health care records and/or laboratory results.

Diagnostic coding system	COVID-19 diagnostic code
Primary care (ICPC)	**R991:** COVID-19, suspected **R992:** COVID-19**,** confirmed by laboratory testing
Secondary care (ICD-10/ICD-11)	**U07.1:** COVID-19 diagnosis confirmed by laboratory testing **U07.2**: COVID-19 clinical diagnosis without laboratory confirmation **RA01.0:** confirmed diagnosis of COVID-19 **RA01.2:** clinical diagnosis (suspected or probable) of COVID-19

Validity of measurement: COVID-19 was not widely tested in the first wave, and has increased over time. Therefore, misclassification may be highest during January-June 2020, especially for non-severe COVID-19. Regulations for coding in ICD9 may have been specified later in time. Different algorithms defining COVID-19 disease and severity of disease as proposed in the ACCESS[22] project will be utilized in sensitivity analysis.

9.3.2 Exposure to medication

Medications are exposures of interest as they may impact pregnancy outcomes, be used to treat symptoms of COVID-19 infections, or be proxies for maternal medical conditions (risk factors for adverse outcomes). Consequently, timing of medication use relative to COVID-19 as well as timing of medication use in pregnancy (trimester), is relevant to understand drug utilization patterns (objective 1) as well as for COVID-19 outcomes (objective 2) and pregnancy and neonatal outcomes (objective 3).

We will describe all medications use in Objective 1. In Objective 2 and 3 we will specifically assess medication group being explored in COVID-19 trials (see list 9.3.4. Outcomes).

Medications, identified by outpatient prescription/dispensing codes and dates will be used (therapeutic classes (ATC-level 2) and if possible, individual drugs (ATC-level 5). In this study, we will

not be able to capture in-hospital medications and/or treatments or OTC medications validly.

Information on medication exposure will be retrieved from dispensing/prescription records. For each prescription/dispensing record we will assume the prescription/dispensing date to reflect timing of actual medication use.

Timing of medication use (trimester, time since COVID-19), administration form, strength of medication will be recorded.

Medication use may also be used as proxies for maternal underlying illness.

9.3.3. At-Risk Medical Conditions to develop severe COVID-19

At-risk medical conditions for developing severe COVID-19 have been defined based on scientific evidence available on Centres for Disease Control and Prevention (CDC website, July 2020) and National Health Services (NHS website, July 2020) websites.

At-risk groups: table 2 lists the at-risk medical conditions, diagnosis codes and proxies will be identified in line with the ACCESS project[22].

Table 2. Comorbid conditions with evidence of increased COVID-19 severity.

Defined Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines (ACCESS-BGR)[22] EU PAS Register Number: EUPAS37273 www.encepp.eu/encepp/viewResource.htm?id=37274

At-risk medical conditions (diagnostic codes)	Medicinal product proxy(ies) (ATC code)
Cardiovascular incl. blood	
Cardiovascular disease/ Serious heart conditions including heart failure, coronary artery disease, cardiomyopathies	Antiarrhythmics, class I and III (C01B) Cardiac stimulants excl. Cardiac glycosides (C01C) Vasodilators used in cardiac diseases (C01D) Other cardiac preparations (C01E) Antithrombotic agents (B01A)
Sickle Cell Disease	Hydroxyurea (L01XX05) Other hematological agents (B06AX)
Respiratory	
Chronic lung disease including COPD, cystic fibrosis, severe asthma	Drugs for obstructive airway diseases (R03) Lung surfactants (R07AA) Respiratory stimulants (R07AB)
Endocrine	
Type 1 & 2 Diabetes	Blood glucose lowering drugs (A10B)
Obesity diagnosis or having a BMI $\geq$30 kg/m^2	Peripherally acting antiobesity products (A08AB) Centrally acting antiobesity products (A08AA)
Renal	

At-risk medical conditions (diagnostic codes)	Medicinal product proxy(ies) (ATC code)
Chronic kidney disease	Erythropoietin (B03XA01)
Immunological	
HIV	Protease inhibitors (J05AE) Combinations to treat HIV (J05AR) NRTI (J05AF) NNRTI (J05AG)
Immunosuppression	Immunosuppressants (L04A) Corticosteroids (H02)
Cancer	
Cancer	Alkylating agents (L01A) Antimetabolites (L01B) Plant alkaloids and other natural products (L01C) Cytotoxic antibiotics and related substances (L01D) Other antineoplastic agents (L01X) Hormones and related agents (L02A) Hormone antagonists and related agents (L02B) Immunostimulants (L03) Immunosuppressants (L04)

9.3.4. Outcomes

Objective 1 Use of medication:

We will describe all recorded medication use (ATC code). Pre-hypothesized groups of special relevance for COVID-19 include:

- Antihypertensives (C02, C03, C04, C07, C08, C09)
- Anticoagulants/platelet inhibitors (B01),
- Antivirals (J05),
- Antibacterials (J01),
- Antimycotics (J02),
- Antimycobacterials (J04),
- Immune sera and globulins (J06),
- Vaccinations (J07),
- Analgesics (N02),
- Psycholeptics (N05),
- Psychoanaleptics (N06),
- Diabetes (A10),
- Corticoisteroids (H02),
- Immunostimulants (L03),
- Immunosuppressants (L04),
- anti-inflammatory drugs (M01) (especially ibuprofen),

- Nasal preparations (R01), Medicines for obstructive airway disease (R03),
- Cough and cold medications (R04).

We will not assess OTC-medications and in-hospital treatments in this project.

Objective 2 Severity of COVID-19

Severity of COVID-19 has been re-categorized according to WHO[1] scale as follows:

- Level 1: any recorded diagnosis;
- Level 2: hospitalization for COVID-19 (confirmed or suspected);
- Level 3: ICU admission in those with COVID-19 related admission;
- Level 4: Acute respiratory distress requiring ventilation (ARDS) during a hospitalization for COVID-19;
- Level 5: death during a hospitalization for COVID-19 (any cause)

To ensure harmonization across studies, we will use the 8-point WHO progression scale:

The Lancet Infectious Diseases 2020 20e192-e197DOI: (10.1016/S1473-3099(20)30483-7)

Patient State	Descriptor	Score
Uninfected	Uninfected; no viral RNA detected	0
Ambulatory mild disease	Asymptomatic; viral RNA detected	1
	Symptomatic; independent	2
	Symptomatic; assistance needed	3
Hospitalised: moderate disease	Hospitalised; no oxygen therapy*	4
	Hospitalised; oxygen by mask or nasal prongs	5
Hospitalised: severe diseases	Hospitalised; oxygen by NIV or high flow	6
	Intubation and mechanical ventilation, pO_2/FiO_2 ≥150 or SpO_2/FiO_2 ≥200	7
	Mechanical ventilation pO_2/FIO_2 <150 (SpO_2/FiO_2 <200) or vasopressors	8
	Mechanical ventilation pO_2/FiO_2 <150 and vasopressors, dialysis, or ECMO	9
Dead	Dead	10

Clinical outcomes of COVID-19: duration of hospitalization.

Validity of measurements:

Data on the severity of COVID-19 will be extracted from electronic health records; codes and algorithms for identification will be made available. Different algorithms defining COVID-19 disease and severity of disease as proposed in the ACCESS[22] project, be utilized in sensitivity analysis.

Objective 3: pregnancy and neonatal outcomes

Pregnancy and neonatal health indicators: include all pregnancy complications, pregnancy outcomes and newborn complications listed below, as defined by Euro-PERISTAT[23] and EMA guidelines[24].

They are as follows:
Pregnancy outcomes (alphabetical order)

- Congenital anomalies/Major congenital anomalies
- Delivery type (spontaneous, induced, Caesarian section) and location (maternity ward, outside institution)
- Low birth weight
- Microcephaly
- Neonatal death
- Preterm birth
- Small-for-gestational age (SGA)/ Intrauterine growth restriction (IUGR)
- Spontaneous abortions
- Stillbirth
- Termination of Pregnancy for Fetal Anomaly (TOPFA)

Definitions of these events are given in **Annex 5.**

Note on congenital anomalies: We will describe major groups of anomalies individually (e.g. cardiac anomalies), and explore whether there could be patterns of malformations. Only analyses for which we have sufficient a priori power (80%) to detect a two-fold increase in the risk of the outcome in question will be performed, and analyses for groups of rare congenital anomalies will be essentially descriptive. Minor congenital anomalies will be explored as a part of malformation patterns.

Newborn outcomes/complications:

- Infection
  - SARS-CoV-2 infection
  - Other infections
- Neonatal complications
  - Transfer to NICU
  - Low Apgar score at 5 minutes.

We will classify 5-minute Apgar score of 7–10 as normal, a score of 4–6 as moderately abnormal, and a score of 0–3 as low in the term infant and in the late-preterm infant.

- Neonatal mortality

Definitions of these events are given in **Annex 5.**

Validity of measurements:
Data on each of the outcomes will be extracted from electronic health records, codes and algorithms for identification will be made available.
These events will be characterized in the data sources as part of the ConcePTION data characterization task, or through the ACCESS project (ACCESS-BGR[22]). For DAPs not participating in ConcePTION or ACCESS, pregnancy back ground rates (BGR) (**Annex 5)** will be calculated in CONSIGN.

9.3.5. Covariates

A broad range of covariates will be used. Covariates will be used for stratification and adjustment, and only covariates determined to be likely confounders will be considered for adjustment. Trimester of pregnancy, age, calendar time and prior co-morbidity will be important variables for stratification or matching as they may confound the effects.

For the comparative studies (objectives 2-3), we define six groups of covariates that may be used for confounding adjustment/matching:

1. Demographic baseline characteristics
2. Maternal characteristics at start of pregnancy
3. Maternal underlying conditions (may also be proxied by medications)
4. Maternal history of obstetric conditions
5. General morbidity scores: Bateman score, number of different drugs
6. Vaccination status

1. Demographic characteristics: dates of birth and death, geographic region (Denmark, France, Germany, Tuscany, Norway, Valencia/Spain, Aragon/Spain, Sweden, Wales/UK).

2. Maternal characteristics at start of pregnancy: will be identified on the basis of diagnosis codes recorded in patients' hospital discharge records, general practice records and/or information in the medical birth registry and include:

- Age at start of pregnancy
- Parity
- Reproductive history (e.g. previous spontaneous abortions, malformations, stillbirths)
- Folic acid use
- Smoking status (i.e. smoking prior to pregnancy)
- Alcohol use (start of pregnancy)
- Educational level (high, medium, low)
- Obesity before pregnancy (BMI≥30), BMI and/or its components
- Socio-economic status and/or proxies of SES

Availability and completeness of these variables will vary across data bases and will be described.

3. Maternal medical conditions

This includes all "At-risk medical conditions for developing severe COVID-19" (Table 2), but also chronic maternal conditions that might be affected by the COVID-19 pandemic, such as maternal mental disorders. They will be identified on the basis of diagnosis codes recorded in patients' hospital discharge records, general practice records, other sources, drug utilization patterns and/or information in the medical birth registry at any time prior to or during the pregnancy and include:

- Cardiovascular diseases (e.g. chronic hypertension, pregnancy hypertension)
- Respiratory diseases (i.e. asthma, COPD)
- Diabetes (i.e. pre-gestational diabetes, gestational diabetes)

- Rheumatic diseases
- Cancer
- Obesity
- Mental disorders

Data sources where only hospital records are available for disease diagnosis (e.g. SNDS), will rely on proxies to identify underlying conditions (e.g. SNDS). Such proxies can be primary care prescriptions or drug dispensing and/or codes for long term diseases in hospital charts.

4. Maternal obstetric conditions
These are variables typically registered in the medical birth registries including gestational diabetes, gestational hypertensive disorders (i.e. gestational hypertension, preeclampsia, HELLP) and previous adverse reproductive history (i.e. prior still birth, prior small for gestational age (SGA) child or fetal growth restriction (FGR), congenital anomaly). Identification will be based on the ConcePTION definitions and codes

5. Proxies for maternal morbidity
- An obstetric comorbidity index will be calculated where possible, such as the one suggested by Bateman *et al*[25] that has been shown to predict maternal morbidity[26].

- Proxy based on medication use: number of different drugs (ATC codes) in year prior to pregnancy or index date (for non-pregnant women)

6. Woman's vaccination status during the study time period will comprise of the following vaccines:
- Influenza A-H1N1
- Pneumococcal disease
- Pertussis

ATC code J07 will be used in addition to diagnostic codes and procedural codes.

9.4. Data sources

Data sources are population-based health care datasets (see Table 1 and 3) that will be linked at person level. Together they capture data on a source population of more than 100 million individuals with 1 million births per year. The data sources include general practice databases (e.g. UK, ES), claims databases (FR, IT, DE) and record linkage of demographic data, registers covering primary & secondary care and prescription registries (NO, SE, DK, IT). Infected pregnant women and their infants are identified through Surveillance System for Communicable Diseases (e.g. MSIS in Norway), hospital contact (e.g. IT, DK, UK, ES-Valencia) and through a Microbiological Surveillance System and Metabolic Diseases Registry (e.g. FISABIO-SPAIN). Laboratory data are included in some registries (e.g. ES-Valencia).

The project will use data from the data source described below. Collectively we will access data from 8 countries from northern, southern, western and more eastern Europe. The following have/are being mapped onto the ConcePTION CDM and obtained/will obtain approvals for data

characterization if needed: UK (Swansea), Norway (UiO), Denmark (AUH), Germany (GePaRD), Spain (FISABIO/Aragon), Italy (ARS), France (SNDS), Sweden (KI).

A description of data sources participating in this project is given below (alphabetical order).

Denmark: Danish linked registries

Denmark has a tax-funded health care system ensuring easy and equal access to health care for all its citizens, and with this system all contacts are recorded in administrative and medical registers (Schmidt et al., 2019). The records carry a unique personal identification number, called the CPR-number, assigned to every Danish citizen. Linkage between registers at an individual level is possible because this CPR-number is used in all Danish registers and assigned by the Danish Civil Registration System. All registers have a nationwide coverage and an almost 100% capture of contacts covering information on currently 5.8 million inhabitants plus historical information. For the purpose of the study we will obtain information from the following registries: *The Danish National Prescription Registry (DNPR*) includes data on all outpatient dispensing of medications and vaccines at Danish pharmacies from 1995 and onwards, including dispensing date, ATC code, product code and amount. From the *Danish Civil Registration System,* data on demographics (sex, date of birth) and censoring (migration, vital status). *The Danish National Patient Registry* contains diagnoses and procedures from all hospitalizations since 1977 and contacts to hospital specialist outpatient clinics since 1995. The Danish National Health Service Register contains information on referral for vaccine administration from GPs. **COVID-19:** COVID and SARS-COV-2 data cannot be promised by the data cut date. **Aarhus University** will be Data Access Provider for the Danish registries, and receive consultancy support from the **University of Copenhagen.**

France: Système National des Données de Santé (SNDS)

The SNDS (Système National des Données de Santé) is the French nationwide healthcare database. It currently covers the overall French population (about 67 million persons) from birth (or immigration) to death (or emigration), even if a subject changes occupation or retires. Using a unique pseudonymized identifier, the SNDS merges all reimbursed outpatient claims from all French health care insurance schemes (SNIIRAM database), hospital-discharge summaries from French public and private hospitals (PMSI database), and the national death registry. SNDS data are available since 2006 and contains information on: *General characteristics:* gender, year of birth, area of residence, etc. *Death:* month, year and cause. Long-Term Disease registration associated with an ICD-10 diagnostic codes. *Outpatient reimbursed healthcare expenditures with dates and codes (but not the medical indication nor result):* visits, medical procedures, nursing acts, physiotherapy, lab tests, dispensed drugs and medical devices, etc. For each expenditure, associated costs, prescriber and caregiver information (specialty, private/public practice) and the corresponding dates are provided. *Inpatients details:* primary, related and associated ICD-10 diagnostic codes resulting from hospital discharge summaries with the date and duration of the hospital stay, the performed medical procedures, and the related costs. Drugs included in the diagnosis related group cost are not captured. However, expansive drugs (i.e. the one charged in addition to the group cost) are. **COVID-19:** To date, only hospitalized COVID-19 cases are captured by the SNDS through the ICD-10 hospital discharge diagnoses codes U07.1. Testing for SARS-CoV2 in outpatient settings can be tracked through French CCAM procedures codes but results are not recorded. **Bordeaux PharmacoEpi (BPE),**

a research platform of the University of Bordeaux specialized in real world studies, will be Data Access Provider for SNDS data. Given the very broad inclusion criteria, a representative 1/10th sample of the full SNDS will be used (around 6.7 million patients).

Germany: GePaRD

GePaRD is based on claims data from four statutory health insurance providers in Germany and currently includes information on approximately 25 million persons who have been insured with one of the participating providers since 2004 or later. Per data year, there is information on approximately 20% of the general population and all geographical regions of Germany are represented. In addition to demographic data, GePaRD contains information on dispensations of reimbursable prescription drugs as well as outpatient (i.e., from general practitioners and specialists) and inpatient services and diagnoses. GePaRD also contains information on influenza vaccinations. **COVID-19:** In February 2020 the two codes for COVID 19 infection (U07.1 and U07.2) were added to the ICD-10-GM classification and further codes were added later, e.g. regarding referral to a test centre. It is not clear yet if these tests were used adequately in the outpatient setting, i.e. there may be an underreporting of mild (non-hospitalized) cases. There are also codes for the test itself but it is not clear yet to which extent these tests are coded on an individual vs. an institutional level. Linkage studies aiming to assess suitability of German claims data to address research questions on COVID-19 are planned. **The Leibniz Institute for Prevention Research and Epidemiology – BIPS** will be Data Access Provider for the GePaRD data. GePaRD data have been used for vaccine safety studies and pregnancy studies. GePaRD is listed under the ENCePP resources database.

Italy: ARS database

The Italian National Healthcare System is organized at regional level: the national government sets standards of assistance and a tax-based funding for each region, and regional governments are responsible to provide to all their inhabitants. Tuscany is an Italian region, with around 3.6 million inhabitants. **The Agenzia Regionale di Sanita' della Toscana (ARS)** is a research institute of the Tuscany Region. The ARS database comprises all information that are collected by the Tuscany Region to account for the healthcare delivered to its inhabitants. Moreover, ARS collects data from regional initiatives. All the data in the ARS data source can be linked with each other at the individual level, through a pseudo-anonymous identifier. The ARS database routinely collects primary care and secondary care prescriptions of drugs for outpatient use, and is able to link them at the individual level with hospital admissions, admissions to emergency care, records of exemptions from copayment, diagnostic tests and procedures, causes of death, mental health services registry, birth registry, spontaneous abortion registry, induced terminations registry. Mother-child linkage is possible through the birth registry. Vaccine data is available since 2016 for children and since 2019 for adults. **COVID-19:** COVID-19 is identified via the Tuscan section of the national COVID surveillance registry, which records and follows all persons with a positive test result. This can be integrated with other sources to improve quality of follow-up study variables.

Norway: Norwegian linked registries

The core data that UIO has access to is the health care administrative databases of the entire Norwegian population, which amounts to approximately 5.3 million inhabitants. Norway has a universal public health care system, consisting of primary health care services and specialist health

care services. Many population-based health registries were established in the 1960s, with use of unique personal identifiers facilitating linkage between registries. The mandatory national health registries were established to maintain national functions. They are used for health analysis, health statistics, improving the quality of healthcare, research, administration and emergency preparedness. **The Norwegian data sources**[9], in this project are the national, mandatory Norwegian Surveillance System for Communicable Diseases (MSIS), which will be linked to five national health registries, i.e. the Medical Birth Registry, the National Patient Register, Norway Control and Payment of Health Reimbursement, the Norwegian Immunisation Registry, and the National Prescription Registry. Information about all Norwegian National Registries can be found here: www.fhi.no/en/more/access-to-data/about-the-national-health-registries2/ **University of Oslo** is Data Access Provider for Norwegian linked registry data in CONSIGN.

Spain: The PRECOVID Study EpiChron Research Group

The source of data for this study will be the PRECOVID study and the EpiChron Cohort. The PRECOVID study includes all individuals with laboratory-confirmed infection by SARS-CoV-2 from an ad hoc registry developed by the Aragon Health System for monitoring the evolution of the COVID-19 disease pandemic in the region of Aragon. This registry links, at a patient level and in a pseudo-anonymized way, the information contained in the users' database, primary care EHRs and primary care and hospital pharmaceutical billing records. The EpiChron Cohort Study links socio-demographic and clinical anonymized information of all the inhabitants of Aragon, build from the BIGAN platform. Aragon BIGAN platform integrates a technical infrastructure and a data lake gathering individual patient data from the regional health service information systems, including primary care, specialized care, hospitalizations, ER episodes, drug prescription, image diagnosis, laboratory analytical determinations, diagnostics, vaccination, anamnesis and demographics from the whole population of Aragon, about 2M subjects with historic data, and 1.3 M active population subjects. Mother-child linkage, and some information regarding labour outcomes are possible through hospital birth registry. The EpiChron Research Group at the **Instituto Aragones de Ciencias de la Salud (IACS)** will provide these data. It has not been mapped onto the ConcePTION CDM or ACCESS project.

Spain: Valencia Integrated Databases (VID)

The Valencia health system integrated database (VID) is a set of multiple, public, population-wide electronic databases for the Valencia Region, the fourth most populated Spanish region, with ≈5 million inhabitants and an annual birth cohort of 48 000 newborns, representing 10.7% of the Spanish population and around 1% of the European population. The VID provides exhaustive longitudinal information including sociodemographic and administrative data (sex, age, nationality, etc.), clinical (diagnoses, procedures, diagnostic tests, imaging, etc.), pharmaceutical (prescription, dispensation) and healthcare utilization data from hospital care, emergency departments, specialized care (including mental and obstetrics care), primary care and other public health services. It also includes a set of associated population databases and registries of significant care areas such as cancer, rare diseases, vaccines, congenital anomalies, microbiology and others, and also public health databases from the population screening programmes. All electronic health systems in the VID use the ICD-9-CM and the ICD-10-CM. All the information in the VID databases can be linked at the individual level through a single personal identification code. The databases were initiated at different moments in time, but all in all the VID provides comprehensive individual-level data fed by all the databases from 2008 to date. Information on PCR test results as well as serological/antibody

tests results for the whole population of the Valencia region is available and linkable from the Microbiological Surveillance Network (RedMIVA). **The Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO)** is Data Access Provider for Valencia Integrated Databases (VID).

Sweden: Swedish registries

Each Swedish citizen is assigned a unique personal identification (PID), at birth or immigration, kept unchanged throughout life with few exceptions. Linkage between registers at an individual level is possible since the PID is used as key in all Swedish national registers. All registers have nationwide coverage, currently on 10.2 million inhabitants and also including historical information. For the Covid-19 study we will obtain information from the Swedish prescribed drug register (PDR), the Swedish national patient register (NPR), the cause of death register (CDR) and the medical birth register (MBR). The Swedish Prescribed Drug Register (PDR) includes information on all prescribed drugs sold in Sweden since July 1, 2005. The register contains patient level data on dispensed medicines for the entire Swedish population, including information on the dispensed drug (product, quantity, price) as well as the dates of prescription and dispensing. Data on total and reimbursed expenditure and certain characteristics of the prescriber and the workplace of the prescriber are also recorded. All drugs are classified according to the World Health Organization Anatomical Therapeutic Chemical (ATC) classification and the register is updated monthly. The National Patient Register (NPR) includes information on patient data (personal identification number, sex, age, and place of residence), hospital data (county council, hospital, and department), administrative data (e.g. date of admission and discharge, acute/planned admission) and medical data (main diagnosis, secondary diagnoses, external cause of injury and poisoning and surgical procedures). Information to the register is delivered once a year to the National Board of Health and Welfare from each of the 21 county councils in Sweden. Cause of Death Register (CDR) contains data from 1961 and is updated annually (Brooke et al 2017). The register includes all deaths that occurs in Sweden. This includes also deaths where the person was not registered in Sweden at the time of death. Stillbirths are not included in the register. Medical birth register (MBR) started in 1973 and records all pregnancies lasting 22 weeks or longer. The information includes the woman's previous pregnancies, smoking, maternity clinic, length of pregnancy, pain relief, method of delivery, diagnoses in mother and child, surgeries, the child's gender, weight, height and head circumference and the baby's condition at birth. The Swedish Pregnancy Register is a national quality register which started in 2013 and collects data on pregnancy and childbirth, starting at the first visit to antenatal care and ending at the follow-up visit to the antenatal care. The majority of data is collected directly from the standardized electronic medical records. Data included in the Pregnancy Register is similar to that of the Medical Birth Register, but is available without a lag-time from the register holder and will be used for this study for pregnancies in 2019 and 2020. **COVID-19:** COVID-19 will be identified as registered diagnoses of Covid-19 from the international classification of diseases, tenth revision (ICD-10) codes: Covid-19, virus identified (U071); Covid-19, virus not identified (U072) and/or procedure code procedures related to Covid-19 (ZV100) in the Swedish national patient register. Additional information about Covid-19 infection will be retrieved from the National Quality Registry for Intensive Care. Sweden will not be able to retrieve data on negative Covid-19 test results. **Centre for Pharmacoepidemiology, Karolinska Institutet** will be Data Access Provider for the Swedish registries in this project.

United Kingdom: Wales, Swansea University

The Secure Anonymised Information Linkage (SAIL) Databank sources, accesses, links and analyses prospectively collected routine health and population data, within a governed infrastructure that is safe and secure. All datasets are anonymised and encrypted by a third party, and returned to SAIL for linkage. Data are held on 5, 400,000 people, since 1998. Data are available within 3 months of events. SAIL holds linkable, anonymised national datasets, including: Accident and emergency care from 2009, Critical care from 2016, Congenital Anomaly Register and Information Service for Wales (CARIS), In-patient and out-patient PEDW records, Maternity dataset from 2015 for additional data on childbirth, National Community Child Health Database (NCCHD, includes gestation (ultrasound), birth centiles, childbirth, infant feeding, developmental screening and vaccinations), National Pupil Database Wales (education attainment to 16), ONS births and deaths (compulsory registration), Primary care data (including all prescriptions and diagnoses) from ~75% of Welsh GP practices. **COVID-19:** SAIL receives a daily extract from the national laboratory system containing COVID-19 test results from all sources (both public and private testing). This dataset is available in SAIL and previously linked to other datasets in the databank, as above. **Swansea University** will be Data Access Provider for the SAIL data in this project.

9.4.1 Data Availability

Data delivery is scheduled in March 2021, and data analysis is scheduled in April 2021 using the most up-to-date data available for each data source. Table 3 displays estimated end dates of available data for analysis in March/April 2021; data which is not available for a data source is indicated by NA.

Table 3. Estimated end date of data availability for each data source given data extraction in March 2021 (analysis March/April 2021). Study period will be 01.01.2018 to 31.12.2020. Some database access providers (DAPs) will not have access to data covering all of 2020.

	Data Access Provider (DAP)	Data Source (DS)	End date of data availability/lag times (LT)						
			Diagnoses, Signs, and Symptoms						
			Hospital	ER	Primary care	Specialist care	Rx	COVID-19 test SARS-CoV-2 +/-	Birth Registry
Denmark	AUH	Nationwide linked registires	2018	2018 (from hospitals)	NA	2018 (from hospitals)	2018	NA	2018
France	BPE	SNDS	31.12.19	NA (visit only)	NA (visit only)	NA (visit only)	31.12.19 Dispensing	NA*	31.12.19**
Germany	BIPS	GePaRD	31.12.18	31.12.18 (2 year LT)	31.12.18 (2 year LT)	31.12.18 (2 year LT)	31.12.18 (2 year LT)	NA	31.12.18 (2 year LT)
Italy, Tuscany	ARS	ARS database	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)	-	.-	31.12.20 (<3m. LT)	01.03.21 (no LT) #	31.12.20 (<3m. LT)
Norway	UIO	Nationwide linked registires	31.12.20 (3m. LT)	01.03.21 (no LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)	01.03.21 (no LT)	01.06.20 (9m. LT)
Spain, Aragon	IACS	PRECOVID study	31.12.20 (2m. LT)	31.12.20 (1m. LT)	31.12.20 (1m. LT)	NA	31.12.20 (1m. LT)	31.12.20 (1m. LT)	31.12.20 (2m. LT)
Spain-Valencia	FISABIO	VID	31.12.20 (3m. LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)	31.12.20 (3m. LT)

	Data Access Provider (DAP)	Data Source (DS)	End date of data availability/lag times (LT)						
			Diagnoses, Signs, and Symptoms				Rx	COVID-19 test SARS-CoV-2 +/-	Birth Registry
			Hospital	ER	Primary care	Specialist care			
Sweden	KI	Nationwide linked registers	31.12.20 (2m. LT)	31.12.20 (2m. LT)	NA	31.12.20 (2m. LT)	31.12.20 (1m. LT)	NA	31.12.20 (Pregnancy register) (3m. LT)
UK, Wales	USWAN	SAIL	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)	31.12.20 (<3m. LT)

m: months. NA: Not available: Emergency ward. Rx: Prescription and/or Dispensing. Abbreviations for DAPs and DS – see annex 1.

* Procedure without result will be available in 2020 ** Birth and related outcomes in hospital discharge summaries # negative results not available.

Based on the need to have data up till October 2021 to observe all pregnancy outcomes from pregnancies starting in 2020, we suggest to perform a second data extraction in March 2022 to deliver an updated report covering data from more data sources and a higher number of pregnancies starting in 2020 in July -August 2022.

Table 4. Estimated births per year and estimated time of data availability of complete data for pregnancies starting in 2020 (ending before October2021), based on current lag times.

Country	Data source	Estimated births per year	Time when complete 2020 data from the Medical Birth Registry is available	Proportion of pregnancies starting in 2020, available by March 2022
Italy – Tuscany	ARS database	25 000	Q1 2022	All
Spain – Aragon	PRECOVID Study and EpiChron Cohort	10 000	Q1 2022	All
Spain – Valencia	Valencia Integrated Database (VID)	45 000	Q1 2022	All
Sweden	Several linked national registries	100 000	Q1 2022	All
UK – Wales	SAIL database	33 000	Q2 2022	All
Norway	Several linked national registries	60 000	Q2 2023	~ 50%
France	Système National des Données de Santé (SNDS)	70 000	Q2 2023	~ 50%
Denmark	Several linked national registries	60 000	Q3 2023	unsure
Germany	The German Pharmacoepidemiological Research Database (GePaRD)	100 000	Q3 2023	unsure

*Conditional on delivery times from the data custodians.

9.5. Study size

Overall, the source population will give rise to approximately 1 million pregnancies per year covering countries and regions with highly variable COVID-19 infection rates (Table 4). Per objective the size of the study population will differ due to restrictions and matching that will deal with confounding.

9.6. Data management

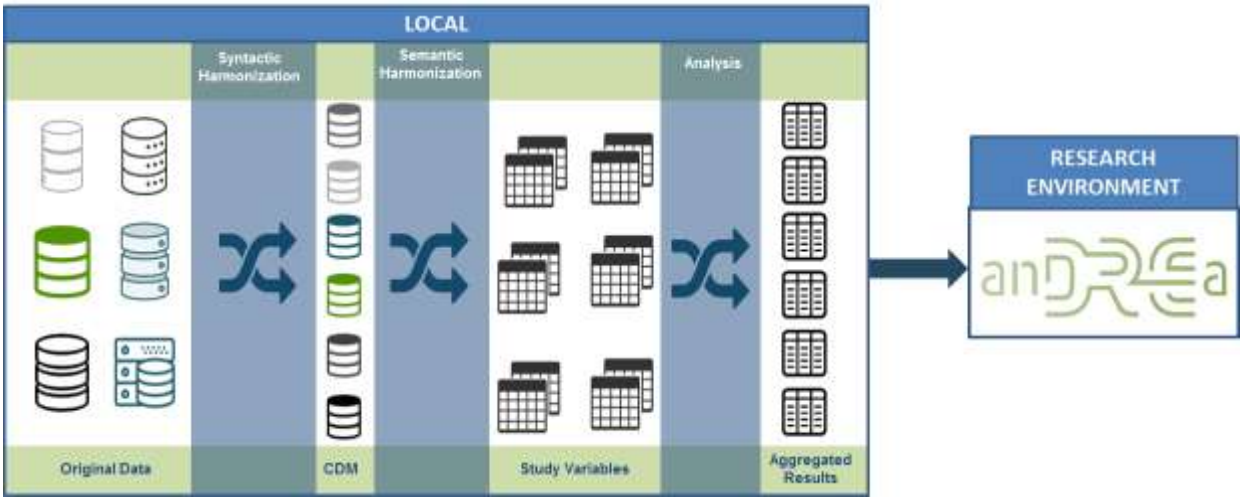
This study will be conducted in a distributed manner using a common protocol, the ConcePTION common data model (CDM), and common analytics programs. **Data will remain local, and effect estimates will be uploaded and aggregated within the secure platform.**

The general principles are:

1) Study specific data will be extracted by data access providers from their local data source, based on instructions
2) These data will be transformed and loaded in the common data model.
3) The extraction, transform, and load (ETL) script will be shared by data access providers for transparency.
4) The data loaded in the CDM will undergo quality control (see section data quality) using a common procedure prior to being processed
5) The data loaded in the CDM will then be processed to create the required study variables and analytical datasets.
6) Data access providers are fully involved in the study by utilizing their knowledge on the characteristics and the process underlying the data collection. This makes analysis rapid, transparent and more efficient.

This process was used successfully in several other European multi-database projects[27] (**Figure 2**). The data pipeline has been further specified in theoretical terms by Gini et al. [28] and further improved in the IMI-ConcePTION project (www.imi-conception.eu/ Deliverable D7.5).

Figure 2 Data management plan



First, to harmonize the structure of the data sets held by each partner, a shared syntactic foundation is utilized. Syntactic foundation is described in **Annex 3** and refers to the syntactically harmonized CDM. In this common data model, data is represented in a common structure, but the content of the data remain in their original format, and the context where the data was originated (e.g. whether a code was collected during inpatient care, or during a primary care visit) is captured along with the data itself. The extraction, transform, and load (ETL) design will be shared on a searchable FAIR catalogue. The ConcePTION FAIR data catalogue is a meta-data management tool designed to contain searchable meta-data describing organizations that can provide access to specific data sources. FAIR means: findable, accessible, interoperable and re-usable. The ConcePTION catalogue but will be populated with new and additional data sources to capture SARS-CoV-2 testing results

and COVID-19 diagnosis. Data quality checks will be conducted to measure the integrity of the ETL as well as internal consistency within the instance of the CDM (see **section 10. Quality Control**).

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. Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g. medications), one or more algorithms will be constructed to operationalize the identification and measurement of each event. Typically a sensitive, or broad, set of codes and one specific, or narrow, set of codes, will be combined with the context where the medical codes are captured (e.g., inpatient care, emergency care, primary care), in a standardised yet flexible approach called *component strategy* [29, 30].
These algorithms may differ per database, as the components that go into the study variable may differ[29,30]. No validation will be done for this study, as there are no resources for this within the budget of the EMA tender. Wherever possible, the event definition sheet will specify prior validated algorithms and codes. Scripts for semantic harmonization will be developed in R, distributed to data access providers for local deployment, and shared on the catalogue. The impact of choices of different algorithms will be assessed quantitatively. This will result in a set of study variables which are both semantically and syntactically harmonized. An attempt at estimating validation indices for outcomes will be performed leveraging on the component strategy [30]

Third, following conversion to harmonized study variable sets, R programs for generation of analytical datasets will be distributed to data access providers for local deployment.

The output of these scripts will then be uploaded to the Digital Research Environment (DRE) for aggregated analysis and visualization. The DRE is made available through UMCU (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 (www.andrea-consortium.org/azure-dre).

9.6.1 Data extraction

Each database access provider (DAP) will create ETL specifications using the standard ConcePTION ETL design template (accessible via this link: https://docs.google.com/document/d/1SWi31tnNJL7u5jJLbBHmoZa7AvfcVaqX7jiXgL9uAWg/edit).
Following completion of this template and review with study statisticians, each DAP will extract the relevant study data locally using their software (e.g. Stata, SAS, R, Oracle): this data is called *instance* of the datasource. An ETL program compliant with the specifications will be developed by the DAPs in their preferred software. Using this program, the instance will be loaded into the CDM structure (see **Annex 3**) in csv format. These data, called *instance* of the CDM, remain local (see **Figure 2**). If the DAP has already participated in studies using the ConcePTION CDM, both ETL design and program may be based on previous experiences. The instance of the CDM contains a table, named INSTANCE, where the characteristics of the instance of the data source are specified for transparency.

9.6.2 Data Processing and transformation

Data processing, level 1-3 quality checks and analysis will be conducted using R code against the syntactically harmonized CDM. The R scripts will first transform the data in the syntactically harmonized CDM to semantically harmonized study variables (see **Figure 2**). Following creation of study variables, the data will be characterized. This characterization will include calculation of code counts and incidence rates, as well as benchmarking within data source (over time), between data sources, and externally (against published estimates). Subsequently, R code to create analytical datasets against semantically harmonized study variables will be distributed and run locally to produce aggregated results. The R scripts for these processing and analysis steps will be developed and tested centrally and sent to the DAPs. The R scripts will be structured in modular form in such a way that transparency is ensured. Functions to be used in the modules will be either standard R packages or packages designed, developed and tested on purpose for multi-database studies. As a result, scripts will be thoroughly documented, and this will allow verification. The DAPs will run the R code locally and send results to the anDREa digital research environment using a secure file transfer protocol. In the anDREa DRE, results will be further plotted, inspected (for quality assessment) and pooled and analysed (if needed) for final reporting.

9.6.3 Software and Hardware

All final statistical computations will be performed on the DRE using R and/or STATA or SAS. Data access providers will have access to the workspace for verification of the scripts.

9.6.4 Storage

Aggregated results, ETL specifications, and a repository of study scripts will be stored in the DRE.

9.6.5 Access

Within the DRE, each project-specific area consists of a separate, secure folder, called a 'workspace'. Each workspace is completely secure, so researchers are in full control of their data. Each workspace has its own list of users, which can be managed by its administrators.

The architecture of the DRE allows researchers to use a solution within the boundaries of data management rules and regulations. Although General Data Protection Regulation (GDPR) and Good (Clinical) Research Practice still rely on researchers, the DRE offers tools to more easily control and monitor which activities take place within projects.

All researchers who need access to DRE are granted access to study-specific secure workspaces. Access to this workspace is only possible with double authentication using an ID 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. Download of files is only possible after requesting and receiving permission from a workspace member with an

'owner' role.

9.6.6 Archiving and record retention

The final study aggregated results sets and statistical programs will be archived and stored on the DRE Sharepoint. The validation of the quality control (QC) of the statistical analysis will be documented. The final study protocol and possible amendments, the final statistical report, statistical programs and output files will be archived on a specific and secured drive centrally.

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 GPP guidelines. These documents could be retained for a longer period, however, if required by the applicable regulatory requirements or by an agreement between study partners. It is the responsibility of the principal investigator to inform the other investigators/institutions as to when these documents no longer need to be retained. 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.

9.7 Data analysis

Table 5 given an overview of data analysis according to objective.

Table 5. Overview of data analysis according to objective

Objective	Study design	Cohorts	Outcome	Exposure	Stratification	Statistical method	Estimator
1a	Cohort	Pregnant women with COVID-19	Medication groups	-	Trimester of pregnancy, age groups, and COVID-19 severity	Descriptive	3-monthly (trimester) prevalence and 95% CI
1b	Matched cohort	Pregnant women	Medication groups	COVID-19 (Yes/No)	Trimester of pregnancy, age groups, and presence of COVID-19 and its severity	Log-binomial regression model	Prevalence ratio (PR) and 95% CI
1c	Matched cohort	Women with COVID-19	Medication groups	Pregnant (Yes/No)	Age groups, presence of chronic at-risk conditions and obesity	Log-binomial regression model	Prevalence ratio (PR) and 95% CI
2a	Cohort	Pregnant women with COVID-19	Severity of COVID-19	Medication groups	Trimester of pregnancy, age groups, presence of chronic at-risk conditions and obesity	Descriptive	Prevalence and 95% CI
2b	Matched cohort	Women with	Severity of COVID-19	Medication groups	Pregnant (Yes/No), age groups	Kaplan-Meier and	Hazard Ratio (HR) and 95% CI

Objective	Study design	Cohorts	Outcome	Exposure	Stratification	Statistical method	Estimator
		COVID-19				proportional-hazard Cox regression model	
3a	Cohort	Pregnant women with COVID-19	Adverse pregnancy and neonatal outcomes	Medication groups	Trimester of pregnancy at diagnosis, age groups, severity of disease, at-risk conditions and obesity	Descriptive	Prevalence and 95% CI
3b	Matched cohort	Pregnant women	Adverse pregnancy outcomes Adverse neonatal outcomes	COVID-19	Trimester of pregnancy at diagnosis, age groups, medication groups	Log-binomial regression model Kaplan-Meier and proportional-hazard Cox regression model	Prevalence ratio (PR) and 95% CI Hazard Ratio (HR) and 95% CI

9.7.1 Definitions and categorization of key variables

Age groups will be defined based upon age at the start date of the pregnancy. Primary grouping will follow

- 12-24 years of age;
- 25-39 years of age;
- 40-55 years of age

Pregnancy start and end dates:

Pregnancy start and end dates will be assessed from medical birth registers for those data sources with access to a registry, while existing algorithms for defining start and end of pregnancy will be utilized in those data sources with an existing algorithm, novel algorithms will be developed if they do not exist to limit bias. Start of pregnancy is date of *last menstrual period* (LMP), end of pregnancy is the date of delivery or abortion (elective/spontaneous). The period of pregnancy will be divided into trimesters, start of trimesters will also be labelled with the calendar month during the pandemic time period. We will try to develop algorithms to predict condition of pregnancy in women whose pregnancy has not ended.

Across all analyses, the following definition of timing in pregnancy will be used:

- Trimesters will be defined as first (T1= days 0-90 since pregnancy start), second (T2 = days 91-180) and third (T3 = 181 days until pregnancy end).
- The first 20 weeks of pregnancy: LMP to LMP+140 days
- LMP: The first day of the LMP is estimated by subtracting the gestational age at delivery from the pregnancy end date. Due date, and thus, gestational length and LMP, is estimated by ultrasound, and only if unavailable, by the woman's recall of LMP.

At-risk groups are women with one or more of the following conditions

- Cardiovascular (including blood disorders)
- Respiratory
- Endocrine
- Immunological
- Cancer
- Women with a diagnostic code of obesity or with a BMI of ≥30 will also be defined as belonging to the at-risk groups.

Hospitalizations: will be categorized as follows based on the primary discharge diagnosis[12]:

- Hospitalizations due to COVID-19 (no obstetric reason)
- Hospitalizations with obstetric reasons and COVID-19
- Hospitalizations due to obstetric reasons without COVID-19

The most common obstetric reasons for hospitalization (not delivery in itself) include threatening premature delivery (incl. placental complications), cardiovascular disorders (incl. hypertensive disorders, coagulation disorders, HELLP) and hyperemesis gravidarum.

9.7.2 Descriptive analysis

Maternal baseline characteristics (e.g. age, smoking status, parity), at-risk medical conditions and pregnancy history will be summarized for each data source and for each cohort using descriptive statistics.

Frequency tables including numbers and proportions will be generated for categorical variables (e.g. maternal age in categories, and at-risk medical conditions).

Mean, standard error, median and interquartile range will be provided for continuous variables (e.g. maternal age).

9.7.3 Analytical approaches – objective 1 Prevalence of medicines use in pregnant women with COVID-19

Descriptive research question 1a: *To what extent do women with COVID-19 in pregnancy use medication (overall and by type) during pregnancy?*

Objective 1a: utilization of medicines will be described by presenting counts and monthly (calendar month) prevalence of medicine use during pregnancy. Analyses will be further stratified by age group, presence of COVID-19, presence of chronic at-risk conditions at start of pregnancy, calendar month/year and trimester.

The numerator will be the number of women having one or more dispensing/prescription (date of prescription/dispensing) of the specific medication within the given time-period. The denominator will be the number of women contributing person-time to that month.

Medications, identified by outpatient prescription/dispensing codes will be evaluated by therapeutic classes (ATC-level 2) and if possible, individual drugs (ATC-level 5).

Inclusion:
Study population includes women who are pregnant with LMP after 1/1/2018 and had a diagnosis of COVID-19 during pregnancy.

Analytical research question 1b: *Is medication use (overall and by type) among women with COVID-19 different from women without COVID-19 in pregnancy?*

Objective 1b: Comparison of exposure to medicines between pregnant women with COVID-19 and those without COVID-19, by age group and trimester of pregnancy at diagnosis. This comparison will be conducted using a matched analysis of cohorts of pregnant women with and without COVID-19 from groups C and D **(Figure 1):**

Population: Pregnant women diagnosed with COVID-19 with LMP after 1/1/2018 will be matched to pregnant women not diagnosed with COVID-19 on age group, at risk conditions (yes/no) trimester of pregnancy at the time of COVID-19 diagnosis and week of healthcare encounter in pregnancies without COVID-19. The index date for the non-COVID pregnant woman will be a healthcare encounter occurring in the week of the diagnosis for the COVID-19 case, to allow for changes in healthcare utilization during lock downs (**Figure 3).** Women with start of follow-up after LMP are not eligible.

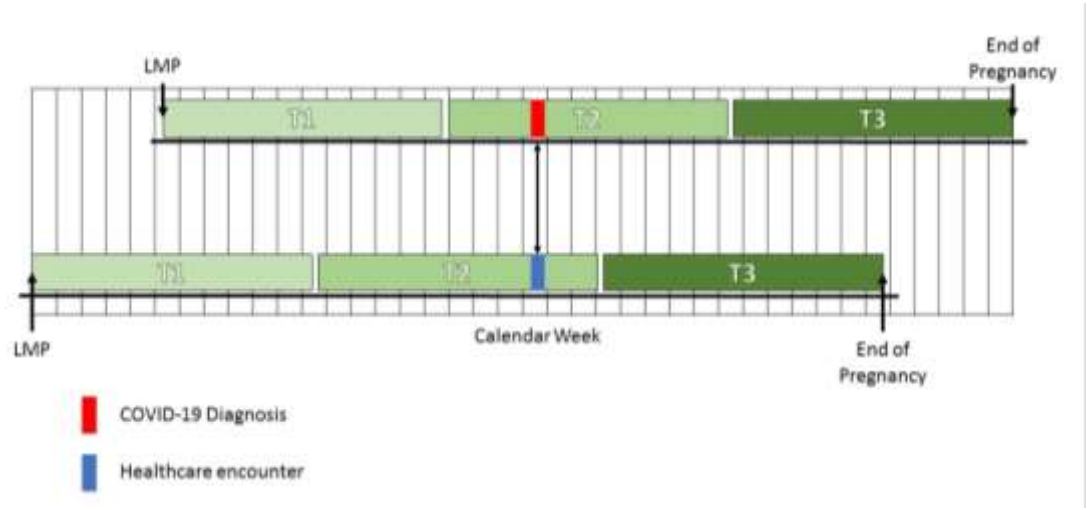


Figure 3. Matching schematic for comparing pregnant women with and without COVID-19 (objectives 1b)

Monthly prevalence of medication use in the 3 months prior to the COVID-19 infection and in the months after the index date of COVID-19 will be computed in the sub-cohort of pregnant women diagnosed with COVID-19 and in the sub-cohort of pregnant women not diagnosed with COVID-19.
Numerator will be the prescription/dispensing of a specific medicine in the monthly periods prior to index date and in the months after index date. Denominator will be the number of persons.
Any medication and treatment prevalence rates in these groups with non-overlapping 95% confidence intervals will be formally compared using log-binomial regression adjusting for baseline at-risk conditions, exact age and important maternal background characteristics.

Analytical research question 1c: *Is medication use (overall, by type) among women with COVID-19 different compared to medication use among non-pregnant women of reproductive age with COVID-19 when taking into account calendar time, severity of infection and key confounders?*

Objective 1c: Comparison of exposure to medicines between pregnant women with COVID-19 and non-pregnant women with COVID-19.

Population: This comparison will be conducted using a matched cohort of pregnant women with COVID-19 during pregnancy (biased if data cut does not capture pregnancy outcomes as far as October 2021) and non-pregnant women with COVID-19.
Pregnant women diagnosed with COVID-19 will be matched to non-pregnant women diagnosed with COVID-19 on age group and calendar week of diagnosis (**Figure 4**).

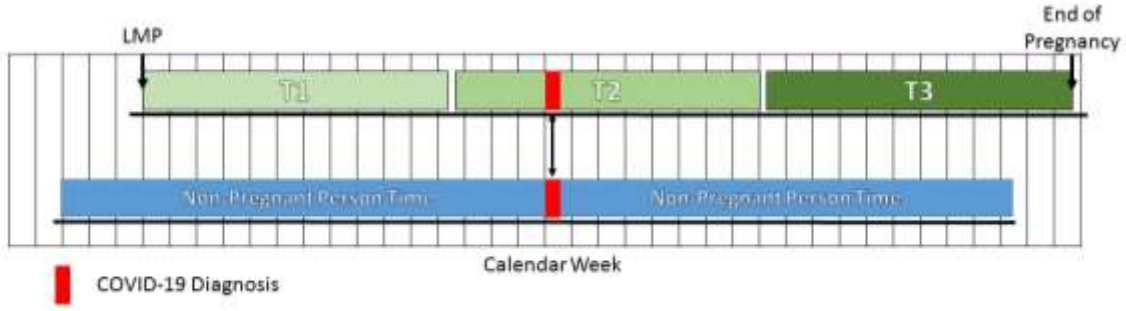


Figure 4. Matching schematic for comparing pregnant and non-pregnant women with COVID-19 (objective 1c, 2b)

Monthly prevalence of medication use in the 3 months prior to the COVID-19 infection and in the months after the index date of COVID-19 will be computed in the sub-cohort of pregnant women diagnosed with COVID-19 and in the sub-cohort of non-pregnant women diagnosed with COVID-19.
Numerator will be the prescription/dispensing of a specific medicine in the monthly periods prior to index date and in the months after index date. Denominator will be the number of persons.
Any medication and treatment prevalence rates in these groups with non-overlapping 95% confidence intervals will be formally compared using log-binomial regression adjusting for at-risk conditions, exact age and important background characteristics.

9.7.4 Analytical approaches – objective 2 Severity and clinical outcomes of COVID-19 disease in pregnant women.

Descriptive research question 2a: *To what extent do women with COVID-19 in pregnancy have severe COVID-19 disease and how is this affected by calendar time and trimester of pregnancy?*

Objective 2a severity of COVID-19 disease in pregnant women and non-pregnant women will be described by presenting counts and prevalence of COVID-19 cases by severity level, by age group, trimester of pregnancy at diagnosis, month of first COVID diagnosis and at-risk conditions among pregnant and non-pregnant women diagnosed with COVID-19 (in groups C and D & E during 2020, D may be biased when last data cut is before October 2021). Duration of hospitalizations will also be described.

Population included will be women pregnant after 1/1/2018 and having COVID-19 during pregnancy (C, D).
For COVID-19 cases, monthly treatment received in the three month before first diagnosis will be described by severity of disease (highest class prevails) and stratified by trimester of pregnancy and calendar period (January -June 2020, vs July-December 2020).
Numerator will be the prescription/dispensing of a specific medicine in the monthly periods prior to index date (date of first diagnosis). Denominator will be the number of COVID-19 cases.

Analytical research questions 2b:
**Do women with COVID-19 in pregnancy have increased risk of severe COVID-19 disease and adverse clinical outcomes of COVID-19 compared to non-pregnant women with COVID-19?*
**Does medication use (type, time of use) impact the association between being pregnant and risk of severe COVID-19 disease?*

Objective 2b: Comparison of COVID-19 severity levels in pregnant women with those of non-pregnant women of reproductive age with COVID-19 (see **Figure 4**). The key question to be answered is whether medication has an impact on severity of COVID-19, and whether this is modified by pregnancy.

Population included: Severe COVID-19 cases in cohorts C-E (D may be biased if data cut is before October 2021) will be matched to up to 10 non-severe COVID-19 cases based on month of COVID-19 diagnosis (month), age and presence of at-risk conditions in the year prior to COVID-19 diagnosis.
To study whether pregnancy status and medication use prior to COVID-19 diagnosis may impact severity of COVID-19 this association, we create a) dummy variables for pregnancy trimester of COVID-19 diagnosis (reference level no pregnancy) and b) medication use (only for medications with at least 5 exposed controls) and additionally test whether there is a significant interaction term between medication (exposure) and pregnancy status. Conditional logistical regression adjusting for co-variates that are related to the severity of COVID-19 will be utilized.

9.7.5 Analytical approaches – objective 3 Adverse pregnancy and neonatal outcomes in pregnant women with COVID-19

Descriptive research question 3a: *To what extent do women with COVID-19 in pregnancy have adverse pregnancy and neonatal outcomes?*

Objective 3a (adverse pregnancy and neonatal outcomes in pregnant women with COVID-19) will be described by presenting counts and incidence of pregnancy and neonatal outcomes overall, by age group and trimester of pregnancy at COVID-19 diagnosis and by COVID-19 severity as well as at risk conditions. Risk factors for adverse pregnancy outcomes will be assessed using binomial regression.

Population included: This will be conducted in cohorts B, C & D (D may be biased when data cuts do not capture data until October 2021).

Analytical research questions 3b:
**Do women with COVID-19 in pregnancy have increased risk of adverse pregnancy and neonatal outcomes compared to pregnant women without COVID-19?*
**Does medication use (type, timing) impact the association between COVID-19 in pregnancy and risk of adverse pregnancy and neonatal outcomes?*

Objective 3b: Comparison of adverse pregnancy and neonatal outcomes in pregnant women with COVID-19 with those of pregnant women without COVID-19. This comparison will be conducted using a matched analysis of cohorts of pregnant women with and without COVID-19 from groups B, C and D**.**

Pregnant women from cohorts C and D diagnosed with COVID-19 during pregnancy and for whom pregnancy has ended, will be matched to pregnant women in cohort C & D not diagnosed with COVID-19 during pregnancy, for whom pregnancy has ended on age group, by week of LMP and absence/presence for at risk conditions.

The risk of different adverse pregnancy and neonatal outcomes will be assessed using log-binomial regression models. To study whether medication use may modify the association between COVID-19 exposed and non-exposed pregnancies, we will first assess the impact of medication use on the pregnancy outcome within each of the COVID-19 and non-COVID-19 cohorts, we will also assess medications use in pregnancy between cohorts and subsequently assess whether there is a significant interaction term between COVID-19 (exposure) and the medication of interest.

9.7.6 Missing data

Since the underlying data represent attended medical care, we generally assume that absence of information of clinical events means absence of medical care for that condition. No imputation will be done for missing data.

9.7.7 Sensitivity analysis

COVID-19 is not always tested, e.g. because of shortage of tests and capacity issues. Our strategy to take this into account will be to a) match on time, b) use different definitions: based on laboratory tests, on clinical symptoms and/or different diagnostic codes, and assess the impact on results, and c) use Probabilistic bias analysis (PBA) to quantify the misclassification bias due to low specificity of infectious disease status (i.e., asymptomatic but COVID-19 positive patients) under a range of scenarios. Diagnostic algorithms for COVID-19 as developed in the ACCESS project[22] will be used.

Bias will occur in cohort D whenever analyses are done that are right censored prior to October 2021. Sensitivity analyses will be conducted excluding all pregnancies that have an LMP < 40 weeks prior to last data cut.

In the analysis of the pregnancy outcomes (objective 3b), we will also consider more fine-tuned matching on gestational week instead of pregnancy trimester. This will address the potential concern that the impact of COVID-19 on pregnancy may be different within a trimester (12-week period).

Sensitivity analysis restricting the calculation according to timing in pregnancy to pregnancies with

ultrasound determined pregnancy start, for those data sources where pregnancy start is determined by maternal recall of LMP in more than 10% of pregnancies, will be performed.

9.8 Quality control

ConcePTION quality indicators will be used. Quality checks in the pre-pandemic and pandemic periods will be used to identify data quality issues related to the pandemic.

To investigate potential changes in health care behaviours during the pandemic and associated lockdown periods on COVID-19 infection, sensitivity analyses will be conducted on:
*A pre-SARS-CoV2 period (overall period starting from 2018 until the start of SARS-CoV2 circulation period) and SARS-CoV2 circulation period. Periods will be defined for each country (see **Annex 4**).
*A pre-lockdown (overall period starting from 2018 until the imposition of lockdown policies limiting face-to-face healthcare encounters), lockdown, and post-lockdown (dependent upon availability of data) periods. Periods will be defined for each country (see **Annex 4)**.

9.8.1 Quality management

The study will be conducted according to the guidelines for Good Pharmacoepidemiology Practice (GPP) (International Society for Pharmacoepidemiology 2008) and according to the ENCePP code of conduct (European Medicines Agency 2018). All data access providers have experience in conducting pharmacoepidemiological research and research is done by researchers trained in pharmacoepidemiology. All programs will be developed according to agreed coding standards and will be validated by double programming or source code review with second programmer involvement. Only validated software (Stata, R and/or SAS version 9.4, SAS Institute Inc., Cary, NC) will be used for statistical analyses.

9.8.2 Data Quality

Data quality will be characterized in a transparent manner according to the procedures developed in the IMI-ConcePTION project on the syntactically harmonized data (www.imi-conception.eu/). This process will proceed iteratively and in collaboration with each data access provider.

Level 1 data checks review the completeness and content of each variable in each table of the CDM to ensure that the required variables contain data and conform to the formats specified by the CDM specifications (e.g., data types, variable lengths, formats, acceptable values, etc.).

This is a check conducted in collaboration with DAPs to verify that the extract, transform, and load (ETL) procedure to convert from source data to the syntactically harmonized CDM has been completed as expected. Formats for all values will be assessed and compared to a list of acceptable formats. Frequency tables of variables with finite allowable values will be created to identify unacceptable values. Distributions of date variables to assess any rounding will be constructed.

The Level 1 checks proceed as follows for each table of interest in the CDM:

1. Within the METADATA table of the CDM, check for presence of the table of interest in the instance.
2. Verify that the table is present in the directory specified by the DAP. If the table is not present, print a notification of its absence to the report.
3. Verify that mandatory variables are present and contain data. If a mandatory variable is absent or contains only missing data, print a notification of this to the report.
4. Check that all conventions for the table of interest have been adhered to. If a convention is not adhered to, print a notification of this to the report.
5. Check consistency between listed allowable values in the METADATA table and data in the table of interest.
6. Tabulate missingness in all variables, overall and by calendar year.
7. Construct distributions of date variables.
8. Construct frequency tables of categorical variables, overall and by calendar year.

Each DAP will be responsible for running the script to complete the Level 1 checks. An R Markdown report describing results of the checks for each table of the CDM will be produced. After addressing any issues identified in Level 1 checks, DAPs may rerun the script and inspect the results. This may proceed iteratively until the DAP declares the ETL to be sufficiently complete and correct. An example R Markdown report produced using simulated data will be included as an annex to the study report.

Level 2 data checks assess the logical relationship and integrity of data values within a variable or between two or more variables within and between tables.

In the level 2 checks, we assess records occurring outside of recorded person time (i.e. before birth, after death, or outside of recorded observation periods). We will identify persons listed in the PERSONS table who do not have any associated records in the other tables of the CDM and verify that persons identified as the mother of an infant in the PERSON_RELATIONSHIPS table of the CDM have a birth date at least twelve years prior to the birth date of their identified child.

Each DAP will be responsible for running the script to complete the Level 2 checks. An R Markdown report describing results of the checks for each table of the CDM will be produced. After addressing any issues identified in Level 2 checks, DAPs may rerun the script and inspect the results. This may proceed iteratively until the DAP declares the ETL to be sufficiently complete and correct. An example R Markdown report produced using simulated data will be included as an annex to the study report.

Level 3 data checks produces incidence and prevalence rates or proportions and trends over time within a data source (by examining output by age and year) for benchmarking between data sources and against external sources.

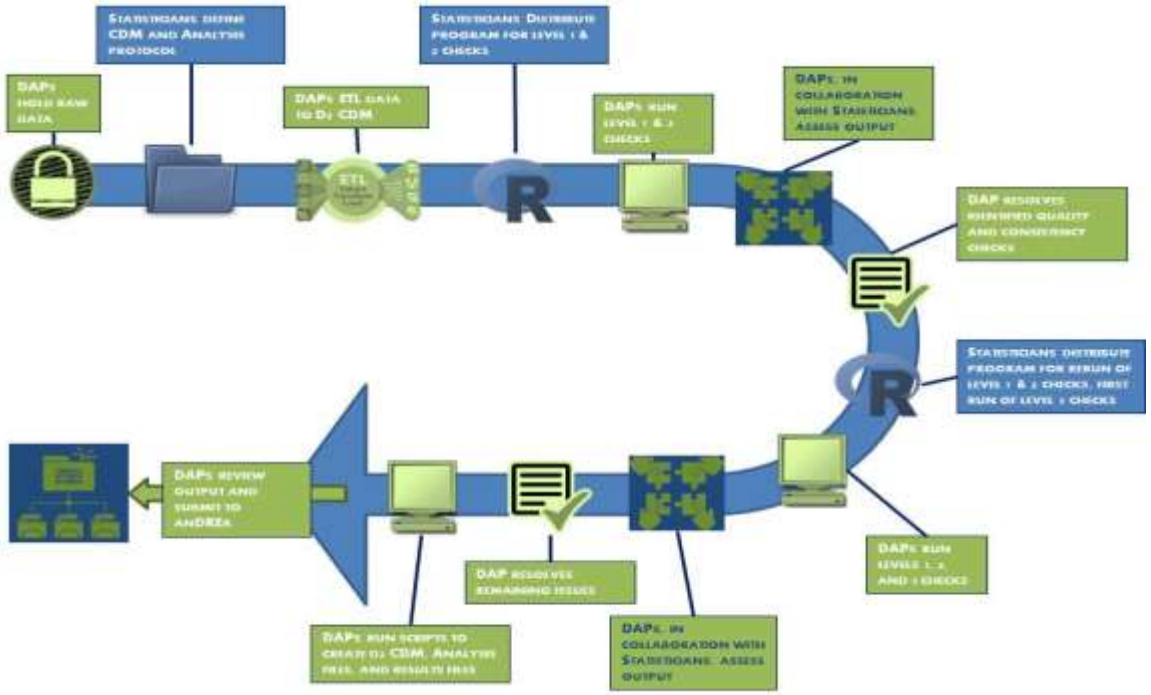
For the current study, Level 3 checks will quantify person time in each data source for the study population as a whole as well as for subpopulations of interest. These will be calculated overall and by calendar year. Codes will be grouped into concept sets based upon Unified Medical Language System (UMLS) Concept Unique Identifiers (CUIs) as identified using the Codemapper tool[31]. Counts

of codes in each concept found in the datasource set will be calculated overall and by calendar year, stratified by the context of code recording. Counts and rates of each concept set will be calculated overall and by calendar year, and stratified by the context of concept recording. Characterization summaries based upon level 3 checks will be included as an annex to the study report.

External benchmark data will be incidence rates of disease that have been obtained from the literature and are listed in the event definition form. Incidence rates from literature will be presented together with incidence rates estimated for the current study in the final study report. Discrepancies will be identified and interpreted based upon descriptions of the data source(s), algorithms for identification of events, and design choices including in and exclusion criteria in published studies vs. those employed for this protocol. A standardised analysis of components will also be performed [30].

In order to identify data capture and recording issues associated with the SARS-CoV2 pandemic period, counts and rates of codes and concept sets will be calculated for the pandemic and lockdown periods of 2020 for each data source and compared graphically against the same period in the three prior years (2017-2019).

Figure 6 Data Quality Pipeline



9.9 Limitations of the research methods

9.9.1 Limitations related to the data sources

1) The key challenge in this project is the issue of left and right censoring, and the fact that analyses are needed, but also requiring time to pass to be able to say anything about pregnancy outcomes. Moreover, in many databases pregnancy is identified upon the end of the pregnancy which may

create addition selection bias. We will mitigate this by 1) Trying to reduce data lag time through support letter from EMA; 2) Sensitivity analysis excluding cohort D; 3) Re-analysis in 2022.

2) The major challenge in this project is the complexity data across the different data sources. This challenge will be overcome by capitalizing on the prior work and collaboration within prior projects including the ConcePTION project (www.imi-conception.eu), the ACCESS project and the ADVANCE project[32]. This study will include 9 different data sources in 8 countries which will be used based on available and permissions in March 2021. These data sources were chosen based on availability, ability to run multisite studies and experience in using common data models as well as capacity. These data sources contain various type of data which are either representative of the national population (eg. Nordic registries), or have a regional/multiregional scope (eg. ARS, SIDIAP, SAIL). Some data are collected at hospital level including or not emergency department or at GPs level only, others are collected at both hospital and GPs level. As the data sources rely on dispensed/filled prescriptions from prescription registries, in-hospital medication use and OTC medications will not be validly captured. These will be captured in the other CONSIGN work packages (INOSS and COVI-PREG).

Given the heterogeneity in the type of encounters recorded, our analyses will be computed per data sources. Only when deemed appropriate will pooled estimates be generated. The CONSIGN consortium is now describing the data sources in this protocol, but a wide assessment of additional data sources and their capacity in conducting near real time monitoring, and vaccine registries is being collected and delivered by 29th January 2021 (Deliverable 1).

3) Some of the participating data sources in this protocol have long lag times, which means that they cannot contribute with data for the entire 2020. This limitation may be overcome by including a second data collection in February 2022, and delivering an additional report to EMA by July 2022.

3) Some of the data sources that have information on SARS-CoV-2 test will only capture positive tests (e.g. SE, DK, ARS), this should not constitute a bias.

4) Some of the data sources do not encompass information on induced terminations and/or spontaneous abortions. For objective 3, we will perform sensitivity analysis taking into considering the live birth bias. Quality of information on the pregnancy start and end dates and pregnancy outcome is conditional on this availability. For most data sources, pregnancy start is calculated by subtracting the date of birth by the pregnancy length (determined preferable by ultrasound examination). If ultrasound is not available (e.g. <3% of pregnancies in the Norwegian medical birth registry), the gestational length is calculated from the last menstrual period as reported by the woman.

9.2 Limitations in the methodology

Other limitations are:

1) In order to provide information rapidly, we need to wait for pregnancies that occur to have an outcome, this means we have a ‘waiting time’ The trade-off between waiting time and rapid

answers means we limited the pandemic study period to 2020. We suggest a second data collection in February/March 2022 to capture a larger proportion of pregnancies starting in 2020 to overcome this limitation.

2) Currently it is not known which treatments are used for COVID-19, we therefore cannot pre-plan the treatment groups. We will assess the outcomes of the most frequently used treatments and will have a minimum of 5 exposed controls to study the treatment group

3) Only medications prescribed/dispensed outside the hospital setting will be captured validly. Treatments of COVID-19 during hospitalization will not be captured and therefore analyses in objective censor on COVID-19 outcome. Misclassification of exposure may occur for objective 2b and 3b.

4) Time to recovery will only be available for hospitalized patients. We use this as a descriptive only in objective 2.

5) COVID-19 is not always tested, especially not because of shortage of tests and capacity issues. We have therefore used different definitions: based on tests, based on clinical symptoms and based on circulation. This means we may misclassify COVID-19 status, but we gain generalizability and power. Any comparison is assumed to be attenuated when using clinical symptoms and periods of circulation. We also have the opposite challenge; screening, resulting in asymptomatic positive patients who would have not been identified if not for screening.

7) Another challenge is the study power, and this limitation is recognized: even in a source population of over 1 million pregnancies, we will have limited power to study rare outcomes among SARS-CoV-2 infected pregnant women and the effect of medications. For rare outcomes specifically related to severe COVID-19 disease, however, we will have challenges, including low power/precision and the limitations of sharing data with few cases/small counts. Only analyses for which we have high a priori power (80%) to detect a two-fold increase in the risk of outcome in question will be performed, and analyses for very rare outcome will be essentially descriptive.

8) It is expected that healthcare behaviours will be impacted during the SARS-CoV2 circulation period due to lockdown situations in most countries. To better take into account this period, sensitivity analyses may be conducted dividing the year 2020 into the periods prior to lockdown measures limiting face-to-face healthcare contacts, during the lockdown, and post-lockdown (dependent upon data availability) (see **Annex 4** for dates of circulation and lockdowns).

10. Protection of human subjects

The study will be conducted in accordance with all applicable regulatory requirements, including all applicable subject privacy requirements and the guiding principles of the Declaration of Helsinki.
The study is part of the CONSIGN project which follows the framework contract stipulations of the EMA and EU PE&PV research network.
The project will follow the EU General Data Protection Regulation as well as all ethical and

institutional regulations relevant for each partner/data source in the project. Governance approval will be obtained from DAPs to conduct the study prior to data extraction.
The output files are stored securely on the anDREa platform of UMCU. These output files do not contain any data that allow identification of subjects included in the study. In fact, each record is completely anonymous and does not contain any identifier key.

11. Management and reporting of adverse events/adverse reactions

Not applicable.

As this is a non-interventional study is based on secondary use of data, safety monitoring and safety reporting, where there is a safety relevant result, is provided on an aggregate level only; no reporting on an individual case level is required. In studies based on secondary use of data with a safety relevant result, reports of adverse events/adverse reactions should be summarized in the study report, i.e. the overall association between an exposure and an outcome. Relevant findings from the study report will be included in the periodic aggregated regulatory reports submitted to Health Authorities.

12. Plans for disseminating and communicating study results

The study protocol will be posted on the EU PAS register. Upon study completion and finalization of the study report, the results of this non-interventional study will be submitted for publication and posted in the EU PAS publicly accessible database of results. Publications will comply with the International Committee of Medical Journal Editors (ICMJE) guidelines.

13. References

1. World Health Organization (WHO). Coronavirus disease (COVID-19) outbreak situation. Accessed 20.05.2020.
2. Allotey J, Stallings E, Bonet M, et al. Clinical manifestations, risk factors, and maternal and perinatal outcomes of coronavirus disease 2019 in pregnancy: living systematic review and meta-analysis. *BMJ.* 2020;370:m3320.
3. Kc A, Gurung R, Kinney MV, et al. Effect of the COVID-19 pandemic response on intrapartum care, stillbirth, and neonatal mortality outcomes in Nepal: a prospective observational study. *Lancet Glob Health.* 2020;8(10):e1273-e1281.
4. Chen H, Guo J, Wang C, et al. Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *Lancet.* 2020;395(10226):809-815.
5. Elshafeey F, Magdi R, Hindi N, et al. A systematic scoping review of COVID-19 during pregnancy and childbirth. *Int J Gynaecol Obstet.* 2020.
6. Liu D, Li L, Wu X, et al. Pregnancy and Perinatal Outcomes of

Women With Coronavirus Disease (COVID-19) Pneumonia: A Preliminary Analysis. *AJR Am J Roentgenol.* 2020:1-6.
7. Mehan A, Venkatesh A, Girish M. COVID-19 in pregnancy: risk of adverse neonatal outcomes. *J Med Virol.* 2020.
8. Schwartz DA. An Analysis of 38 Pregnant Women with COVID-19, Their Newborn Infants, and Maternal-Fetal Transmission of SARS-CoV-2: Maternal Coronavirus Infections and Pregnancy Outcomes. *Arch Pathol Lab Med.* 2020.
9. Yang Z, Wang M, Zhu Z, Liu Y. Coronavirus disease 2019 (COVID-19) and pregnancy (review). *J Matern Fetal Neonatal Med.* 2020:1-4.
10. Zaigham M, Andersson O. Maternal and perinatal outcomes with COVID-19: A systematic review of 108 pregnancies. *Acta Obstet Gynecol Scand.* 2020.
11. Hedermann G, Hedley PL, Baekvad-Hansen M, et al. Danish premature birth rates during the COVID-19 lockdown. *Arch Dis Child Fetal Neonatal Ed.* 2020.
12. Panagiotakopoulos L, Myers TR, Gee J, et al. SARS-CoV-2 Infection Among Hospitalized Pregnant Women: Reasons for Admission and Pregnancy Characteristics - Eight U.S. Health Care Centers, March 1-May 30, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(38):1355-1359.
13. Delahoy MJ, Whitaker M, O'Halloran A, et al. Characteristics and Maternal and Birth Outcomes of Hospitalized Pregnant Women with Laboratory-Confirmed COVID-19 - COVID-NET, 13 States, March 1-August 22, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(38):1347-1354.
14. Ellington S, Strid P, Tong VT, et al. Characteristics of Women of Reproductive Age with Laboratory-Confirmed SARS-CoV-2 Infection by Pregnancy Status - United States, January 22-June 7, 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(25):769-775.
15. Ahlberg M, Neovius M, Saltvedt S, et al. Association of SARS-CoV-2 Test Status and Pregnancy Outcomes. *JAMA.* 2020.
16. Mulangu S, Dodd LE, Davey RT, Jr., et al. A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *N Engl J Med.* 2019;381(24):2293-2303.
17. Pasley MV, Martinez M, Hermes A, d'Amico R, Nilius A. Safety and efficacy of lopinavir/ritonavir during pregnancy: a systematic review. *AIDS Rev.* 2013;15(1):38-48.
18. Osadchy A, Ratnapalan T, Koren G. Ocular toxicity in children exposed in utero to antimalarial drugs: review of the literature. *J Rheumatol.* 2011;38(12):2504-2508.
19. Rasmussen SA, Smulian JC, Lednicky JA, Wen TS, Jamieson DJ. Coronavirus Disease 2019 (COVID-19) and pregnancy: what obstetricians need to know. *Am J Obstet Gynecol.* 2020;222(5):415-426.
20. American College of Obstetricians and Gynecologists. Outpatient Assessment and Management for Pregnant Women With Suspected or Confirmed Novel Coronavirus (COVID-19). In:24.04.2020.
21. *Clinical management of COVID-19.* Vol WHO/2019-nCoV/clinical/2020.5: World Health Organization; 2020.
22. Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines (ACCESS-BGR). http://www.encepp.eu/encepp/viewResource.htm?id=37274.
23. EURO-PERISTAT T. Indicators of Perinatal health. https://www.europeristat.com/our-indicators/indicators-of-perinatal-health.html. Published 2020. Accessed 14.05.2020.

24. EMA/653036/2019. Guideline on good pharmacovigilance practices (GVP). Product- or Population-Specific Considerations III: Pregnant and breastfeeding women. EMA/653036/2019 DRAFT FOR PUBLIC CONSULTATION, 4 December 2019 In:2019.
25. Bateman BT, Mhyre JM, Hernandez-Diaz S, et al. Development of a comorbidity index for use in obstetric patients. *Obstet Gynecol.* 2013;122(5):957-965.
26. Bateman BT, Gagne JJ. The Obstetric Comorbidity Index predicts severe maternal morbidity. *BJOG.* 2015;122(13):1756.
27. Gini R, Sturkenboom MCJ, Sultana J, et al. Different Strategies to Execute Multi-Database Studies for Medicines Surveillance in Real-World Setting: A Reflection on the European Model. *Clin Pharmacol Ther.* 2020;108(2):228-235.
28. Gini R, Schuemie M, Brown J, et al. Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies. *EGEMS (Wash DC).* 2016;4(1):1189.
29. Roberto G, Leal I, Sattar N, et al. Identifying Cases of Type 2 Diabetes in Heterogeneous Data Sources: Strategy from the EMIF Project. *PLoS One.* 2016;11(8):e0160648.
30. Gini R, Dodd CN, Bollaerts K, et al. Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project. *Vaccine.* 2019.
31. Becker BFH, Avillach P, Romio S, et al. CodeMapper: semiautomatic coding of case definitions. A contribution from the ADVANCE project. *Pharmacoepidemiol Drug Saf.* 2017;26(8):998-1005.
32. Sturkenboom M, Braeye T, van der Aa L, et al. ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations. *Vaccine.* 2020.
33. McIntosh JJ. Corticosteroid Guidance for Pregnancy during COVID-19 Pandemic. *Am J Perinatol.* 2020.
34. Dolk H, Loane M, Garne E. The prevalence of congenital anomalies in Europe. *Adv Exp Med Biol.* 2010;686:349-364.

14. Annexes
Annex 1. List of stand-alone documents

Number	Document reference number	Date	Title
1	Number	22.01.2021	List of investigators
2	*Number*	*Date*	
…	*Number*	*Date*	*text*

Stand-alone documents

1. List of all investigators

Collaborating Institutions	Country	Key persons
Data access provider (DAPs)		
Aarhus University (AUH)	Denmark	Vera Ehrenstein
Agenzia Regionale di Sanita Toscana (ARS)	Tuscany, Italy	Rosa Gini, Claudia Bartolini
Leibniz Institute for Prevention Research and Epidemiology - BIPS	Germany	Ulrike Haug, Tania Schink
Karolinska Institute (KI)	Sweden	Carolyn Cesta, Marie Linder, Helle Kieler
The Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO) Catalan Institute of Pharmacology Foundation (FICF)	Valencia, Spain	Gabriel Sanfelix-Gimeno, Isabel Hurtado, Clara Rodriguez, Anibal Garcia Sempere, Salvador Peiro. Mònica Sabaté, Xavier Vidal, Cristina Aguilera, Judit Riera, Elena Ballarín
Instituto Aragones de Ciencias de la Salud (IACS)	Aragon, Spain	Beatriz Poblador-Plou, Antonio Gimeno-Miguel, Alexandra Prados-Torres, Antonio Poncel-Falcó, Aida Moreno-Juste
Université de Bordeaux, Bordeaux PharmacoEpi (BPE)	France	Cécile Droz-Perroteau, Nicolas Thurin
Swansea University (SWANSEA)	Wales, UK	Sue Jordan, Daniel Thayer
University of Oslo (UIO)	Norway	Hedvig Nordeng, Angela Lupattelli
Non-DAPs		
University Medical Center Utrecht (UMCU)	The Netherlands	Miriam Sturkenboom, Viola Hoxha
Utrecht University (UU)	The Netherlands	Olaf Klungel, Satu Siiskonen
RTI-Health Solutions (RTI)	Spain	Andrea Margulis, Bradley Layton, Alejandro Arana, Estel Plana, Elena Rivero
University of Copenhagen (UCPH)	Denmark	Morten Andersen

Annex 2. ENCePP checklist for study protocols

Adopted by the ENCePP Steering Group on 15/10/2018

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) welcomes innovative designs and new methods of research. This Checklist has been developed by ENCePP to stimulate consideration of important principles when designing and writing a pharmacoepidemiological or pharmacovigilance study protocol. The Checklist is intended to promote the quality of such studies, not their uniformity. The user is also referred to the ENCePP Guide on Methodological Standards in Pharmacoepidemiology, which reviews and gives direct electronic access to guidance for research in pharmacoepidemiology and pharmacovigilance.

For each question of the Checklist, the investigator should indicate whether or not it has been addressed in the study protocol. If the answer is "Yes", the section number of the protocol where this issue has been discussed should be specified. It is possible that some questions do not apply to a particular study (for example, in the case of an innovative study design). In this case, the answer 'N/A' (Not Applicable) can be checked and the "Comments" field included for each section should be used to explain why. The "Comments" field can also be used to elaborate on a "No" answer.

This Checklist should be included as an Annex by marketing authorisation holders when submitting the protocol of a non-interventional post-authorisation safety study (PASS) to a regulatory authority (see the Guidance on the format and content of the protocol of non-interventional post-authorisation safety studies). The Checklist is a supporting document and does not replace the format of the protocol for PASS presented in the Guidance and Module VIII of the Good pharmacovigilance practices (GVP).

Study title: COVID-19 infection and medicines in pregnancy – a multinational registry based study

EU PAS Register® number: **Study reference number (if applicable):**

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	X	☐	☐	9.1
1.1.2 End of data collection[2]	X	☐	☐	9.1
1.1.3 Progress report(s)	☐	☐	X	
1.1.4 Interim report(s)	☐	☐	X	
1.1.5 Registration in the EU PAS Register®	X	☐	☐	12
1.1.6 Final report of study results.	X	☐	☐	6

Comments:

[1] Date from which information on the first study is first recorded in the study dataset or, in the case of secondary use of data, the date from which data extraction starts.
[2] Date from which the analytical dataset is completely available.

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☐	☐	☐	
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	X	☐	☐	7.1, 7.2
2.1.2 The objective(s) of the study?	X	☐	☐	8
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	X	☐	☐	9.1
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	X	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	X	

Comments:

The study will calculate incidence and prevalence rates and is not a hypothesis testing study.

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	X	☐	☐	9.1
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	X	☐	☐	9.1, 9.2, 9.4
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	X	☐	☐	8, 9.7.2
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	X	☐	☐	9.7.2
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	X	

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	X	☐	☐	9.2
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	X	☐	☐	9.1
4.2.2 Age and sex	X	☐	☐	9.1
4.2.3 Country of origin	☐	X	☐	
4.2.4 Disease/indication	☐	☐	X	
4.2.5 Duration of follow-up	X	☐	☐	9.1

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	X	☐	☐	9.1, 9.3.1

Comments:

The source population is not described according to country of origin, but according to country in which they are registered in a participating data source.

Women are not selected into the study on the basis of a disease or indication. All women of reproductive age (12-55 years), pregnant women and their children observed in one of the participating data sources for at least one day during the study period (01.01.18 – 31.12.20/last data availability) are included.

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	X	☐	☐	9.3.1, 9.3.2
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	X	☐	☐	9.3.1.
5.3 Is exposure categorised according to time windows?	X	☐	☐	9.3.1, 9.3.2, 9.7.1
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	X	☐	☐	9.3.1, 9.3.2
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	X	☐	☐	9.3.2, 9.7.1
5.6 Is (are) (an) appropriate comparator(s) identified?	X	☐	☐	9.1, fig. 1

Comments:

5.4 Intensity of COVID-19 = severity. Intensity of medication use = amount prescribed/dispensed.

5.5 Trimester of exposure

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	X	☐	☐	8, 9.3.4
6.2 Does the protocol describe how the outcomes are defined and measured?	X	☐	☐	9.2.3, Annex 5
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	X	☐	☐	9.3.1, 9.3.4
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	X	

Comments:

The study does not address outcomes relevant for Health Technology Assessment

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	X	☐	☐	9.3.5
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☐	☐	X	9.7.7
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	X	☐	☐	9.7.7, 9.6, 9.7, 10, 11.2

Comments:

All women of reproductive age (12-55 years), pregnant women and their children observed in one of the participating data sources for at least one day during the study period (01.01.18 – 31.12.20/last data availability) will be selected into the study population.

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	X	☐	☐	9.3.2

Comments:

Effect modification by medication use

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	X	☐	☐	9.3
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	X	☐	☐	9.3, 9.4
9.1.3 Covariates and other characteristics?	X	☐	☐	9.3, 9.4
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	X	☐	☐	9.3, 9.4
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	X	☐	☐	9.3, 9.4
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	X	☐	☐	9.3, 9.4
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	X	☐	☐	9.3

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	X	☐	☐	9.3
9.3.3 Covariates and other characteristics?	X	☐	☐	9.3
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	X	☐	☐	9.4

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	X	☐	☐	9.7
10.2 Is study size and/or statistical precision estimated?	X	☐	☐	9.2
10.3 Are descriptive analyses included?	X	☐	☐	9.7
10.4 Are stratified analyses included?	X	☐	☐	9.7
10.5 Does the plan describe methods for analytic control of confounding?	X	☐	☐	9.7
10.6 Does the plan describe methods for analytic control of outcome misclassification?	X	☐	☐	9.3.4, 9.6
10.7 Does the plan describe methods for handling missing data?	X	☐	☐	9.7.6
10.8 Are relevant sensitivity analyses described?	X	☐	☐	9.7.7

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	X	☐	☐	9.6
11.2 Are methods of quality assurance described?	X	☐	☐	10.2
11.3 Is there a system in place for independent review of study results?	☐	X	☐	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☐	☐	X	
12.1.2 Information bias?	X	☐	☐	9.9

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	X	☐	☐	
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	X	☐	☐	9.4 Table 3

Comments:

All women of reproductive age (12-55 years), pregnant women and their children observed in one of the participating data sources for at least one day during the study period (01.01.18 – 31.12.20/last data availability) will be selected into the study population.

Section 13: Ethical/data protection issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	X	☐	☐	10
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	X	
13.3 Have data protection requirements been described?	X	☐	☐	9.6

Comments:

No ethical review procedure have been conducted.

Section 14: Amendments and deviations	**Yes**	**No**	**N/A**	**Section Number**
14.1 Does the protocol include a section to document amendments and deviations?	X	☐	☐	5

Section 15: Plans for communication of study results	**Yes**	**No**	**N/A**	**Section Number**
15.1 Are plans described for communicating study results (e.g. to regulatory authorities)?	X	☐	☐	12
15.2 Are plans described for disseminating study results externally, including publication?	X	☐	☐	12

Name of the main author of the protocol: Hedvig Nordeng

Date: 14 December 2020

Signatur: Hedvig Nordeng (sign.)

Annex 3. Syntactically Harmonized Common Data Model

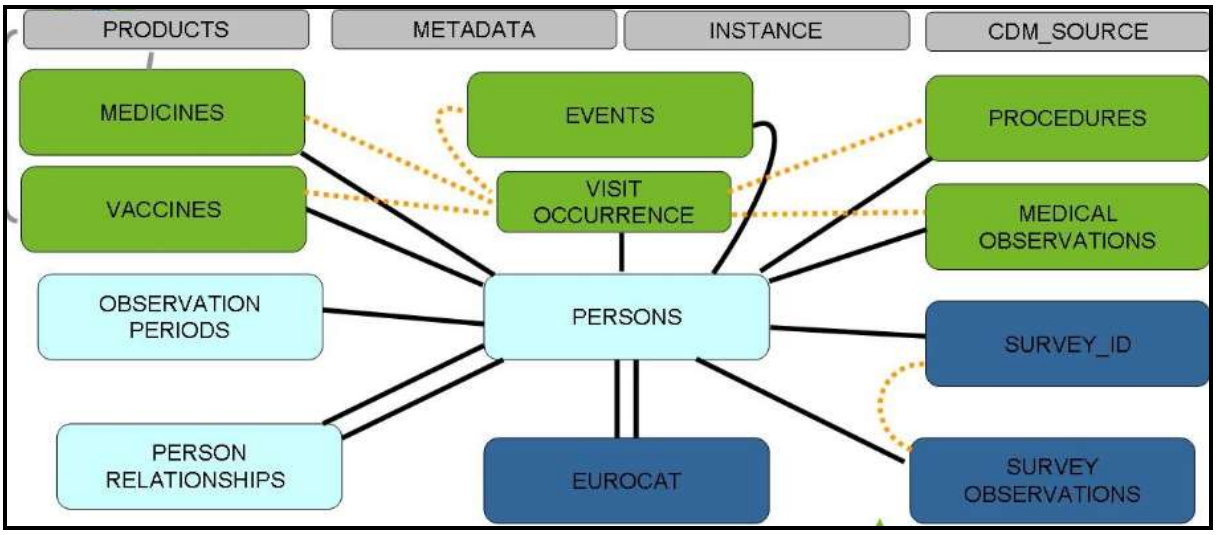


Figure. Schematic representation of the ConcePTION CDMv2.0

METADATA TABLES

The metadata tables contain data in a machine readable format which allows for processing of the data in the CDM. The CDM includes 4 tables in total:

PRODUCTS

Listing of national product codes for medicinal products. Contains a product ID foreign key to the DRUGS and VACCINES table. The PRODUCT_CODE table contains detailed data on products at the package level.

METADATA

The metadata table contains indicators which can act as machine readable guides for code written against the CDM. For instance, whether data in the drug table represents prescription or dispensing.

INSTANCE

The instance table contains data on the specific instance of the ConcePTION CDM, such as tables and columns from source data which have been included.

CDM_SOURCE

Contains high-level meta data describing the source data for the current instance such as the name of the source, data access provider, and date of last update.

CURATED TABLES

Curated tables differ from the other tables of the CDM in that data access providers are asked to

create these tables using rule-based algorithms. These tables therefore represent a *syntactic* and *semantic* harmonization. The CDM includes 3 tables in total:

PERSON
One row of data per subject present in the data and meeting inclusion criteria for the CDM instance at any point during the study period. Data on each subject includes sex at the date of the instance creation, one date of birth, and one date of death (these may be derived using DAP-specific rules)

OBSERVATION_PERIODS
One row per period during which a subject is present in the data source. This may be based upon registration in a geographical area, registration in a GP practice, presence in a registry, etc.

PERSON_RELATIONSHIPS
Contains one row of data for each child present in the data and meeting inclusion for the CDM instance at any point during the study period, together with an identifier for the mother of the child and the father of the child if available.

ROUTINE HEALTH DATA TABLES

Routine health care data tables capture data observed in the course of routine health care in hospitals, GP offices, pharmacies, outpatient clinics, etc. The CDM includes 6 tables in total:

VISIT_OCCURRENCE
Contains an identifier of a visit to allow for linkage of diagnoses, procedures, dispensings, etc in the same visit if this information is available in a data source.

EVENTS
Contains data on events indicated by a diagnosis code or free text. It contains one row per diagnosed event.

MEDICINES
One record per prescription or dispensing. Contains data required to estimate duration of exposure. Linkage to PRODUCT_CODE table to access data on drugs at the package level.

PROCEDURES
Contains data on procedures ordered or completed. For those procedures with an associated result, results and units are recorded. It contains one row per procedure.

VACCINES
Contains data on vaccinations with one row per vaccine. Data on dose number for childhood vaccines and manufacturer are accommodated by this table.

MEDICAL_OBSERVATIONS
Contains observations recorded during routine healthcare. Can be a result from a laboratory test, or

physical measurement, but also level of education, or sex, or a pathology report.
SARS-CoV-2 test results will be mapped under this Table.

SURVEILLANCE TABLES

Surveillance tables contain data collected for purposes beyond routine health care either for surveillance of specific events or for recording of detailed information related to a unit of observation such as a pregnancy or chronic illness. The CDM includes 3 tables in total:

EUROCAT

Contains the EUROCAT or EUROmediCAT (a subset of the EUROCAT) table for those data access providers which have access to this standard table.

SURVEY_ID

Contains metadata on observations contained in the SURVEY_OBSERVATION table and allows for linkage between mothers and infants captured in a medical birth registry.

SURVEY_OBSERVATION

Contains one row per observation in any survey or registry data table – such as a medical birth registry, well child program database, cancer registry, etc.

Full CDM specifications can be accessed here:
https://drive.google.com/file/d/1hc-TBOfEzRBthGP78ZWIa13C0RdhU7bK/view?usp=sharing

Associated CDM vocabularies can be accessed here:
https://docs.google.com/spreadsheets/d/1idAEKC440rkIYIxCSRmEVgEPj_UouUI-I3kxNCpJt3U/edit?usp=sharing

Annex 4. Dates of COVID-19 pandemic response measures per country

Country	Date of country's first detected case of COVID-19 in 2020$	Date at which 1/10 of the highest daily number of cases was reached$	Start date of lockdown measures limiting face-to-face healthcare encounters	End date of lockdown measures limiting face-to-face healthcare encounters
Denmark	27 February	10 March	12 March	Healthcare encounters are still limited for people with respiratory tract infection symptoms. Patients with symptoms of possible COVID-19 have to call the GP asking for authorization to go to the medical centre to be tested.
France	24 January	14 March	17 March	11 May
Germany	28 January	13 March	March (exact date varies between federal states)	June (exact date varies between federal states)
Italy	31 January	06 March	07 March	18 May
Norway	27 February	11 March	12 March	20 April
Spain	01 February	11 March	14 March	Government lockdown measures were eased on 21 June. Healthcare encounters are still limited. Patients have to call the GP asking for authorization to go to the medical centre (as of 11 August 2020)
Sweden	31 January	7 March	10 March	Healthcare encounters are still limited for people with respiratory tract infection symptoms. Patients with symptoms of possible COVID-19 have to call an information number before seeking health care.
United Kingdom	31 January	22 March	26 March	01 June

$ Data extracted from ECDC website (https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases)

Annex 5. Pregnancy & neonatal outcome definitions

Terms for defining pregnancy and neonatal outcomes will be aligned with recommendations and current recognized standards:

- Guideline on good pharmacovigilance practices (GVP). Product- or Population-Specific Considerations III: Pregnant and breastfeeding women. EMA/653036/2019 DRAFT FOR PUBLIC CONSULTATION, 4 December 2019 (www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-good-pharmacovigilance-practices-product-population-specific-considerations-iii_en.pdf):
- EuroPeriStat definitions and recommendations (www.europeristat.com)
- EUROCAT Classification (EUROCAT Subgroups of Congenital Anomalies).

The pregnancy events in CONSIGN are aligned with the definitions and are operationalized as in the IMI ConcePTION project. A brief overview of the definitions is given below.

Pregnancy outcome: End result of pregnancy, which includes ectopic pregnancy, miscarriage, foetal death, termination of pregnancy and live birth.

Delivery: Deliveries will be classified according to initiation type: Spontaneous, induction or Caesarian section. Place of delivery will be classified as at the maternity ward, at home (planned, unplanned) or at other place of delivery outside institution.

Foetal death (intrauterine death, *in utero* death): Death prior to complete expulsion or extraction from the mother of a foetus, irrespective of the duration of pregnancy. Early foetal death (before 22 completed weeks of gestation) is known as miscarriage, whereas late foetal death (after 22 completed weeks of gestation) is known as stillbirth.

Spontaneous abortion/Miscarriage: Spontaneous abortions or miscarriages are foetal losses before the gestational age or birthweight threshold for defining stillbirth. The definition of this indicator varies by country/data sources. According to EUROCAT, spontaneous abortions/miscarriages are foetal deaths <20 weeks. The WHO and the EMA define spontaneous abortion pregnancy ending spontaneously before 22 weeks of gestation (i.e. up to and including 21 6/7 weeks of gestation) (European Medicines Agency (EMA). Guideline on the exposure to medicinal products during pregnancy: Need for post-authorisation data. EMA, 2005. Accessed May 26, 2020. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-exposuremedicinal-products-during-pregnancy-need-post-authorisation-data_en.pdf).
The upcoming ICD-11 guidelines will likely define stillbirth from 22 weeks onwards with spontaneous abortion/miscarriage referring to deaths at <22 weeks.
Given that spontaneous abortions or miscarriages tend to occur early in pregnancy often these are often not recorded in register-based data sources. Women might not declare the occurrence of spontaneous abortion or this might be managed in emergency room or in primary health care. This leads to underestimation in pregnancy studies using register data.

Termination of pregnancy (induced abortion, elective abortion): Artificial interruption of pregnancy for any reason.

Trimesters: Trimester are usually defined as three months period where the first trimester covered the first 12 weeks of gestation. Trimesters can be defined as first (T1= days 0-90 since LMP), second (T2 = days 91-180 since LMP) and third (T3 = LMP + 181 days until pregnancy end).

Live birth: Complete expulsion or extraction from the mother of a foetus, irrespective of the duration of the pregnancy, that, after such separation, breathes or shows any evidence of life.

Gestational age: Measure of the age of a pregnancy calculated from the first day of a woman's last menstrual period or as estimated by a more accurate method such as ultrasound. The method used needs to be clearly stated in any reporting. Gestational age is expressed in completed days or completed weeks (e.g. events occurring 280 to 286 days after the onset of the last menstrual period are considered to have occurred at 40 weeks of gestation).

Birth weight: Initial weight of the infant at birth.

Pre-term birth (premature birth): Birth at less than 37 completed weeks (less than 259 days) of gestation.

Term birth: Birth at any time from 37 to less than 42 completed weeks (259 to 293 days) of gestation.

Post-term birth: Birth after 42 completed weeks of gestation or more (294 days or more).

Low birth weight: Body weight of the newborn at birth of less than 2,500 grams (up to and including 2,499 g).

Intrauterine growth retardation (IUGR) (small-for-gestational age): Observed weight of a live born infant or size of a foetus lower than expected, usually below the tenth percentile, on the basis of gestational age.

Neonatal death / mortality:
Neonatal mortality refers to death of a live-born baby within the first 28 days of life. Early neonatal mortality refers to the death of a live-born baby within the first seven days of life, while late neonatal mortality refers to death after 7 days until before 28 days.

Perinatal mortality
The total number of stillbirths and deaths within the first 7 days after birth according to the inclusion criteria (specified below) divided by the total number of births according to the inclusion criteria.
As a result of the general development within neonatal medicine most official institutions have lowered the inclusion criteria for gestational age and birth weight from 28 weeks/1000 grams to 22 weeks/500 grams.

*Perinatal mortality***:** Euro-Peristat: Gestational age ≥ 22 weeks. If gestational age is unknown, birth weight ≥ 500 grams is used as a reasonable indicator for inclusion.
Perinatal mortality: WHO's primary definition Birth weight ≥ 500 grams. If birth weight is unknown, gestational age ≥ 22 weeks is considered a reasonable indicator for inclusion. If both birth weight and gestational age are unknown, length (crown-heel) ≥ 25 cm is used as an indicator for inclusion.
Perinatal mortality: WHO's secondary definition Birth weight ≥ 1000 grams. If birth weight is unknown, gestational age ≥ 28 weeks is considered as an indicator for inclusion. If both birth weight and gestational age are unknown, length (crown-heel) ≥ 35 cm is used as an indicator for inclusion.

Maternal death / mortality:
Maternal death or maternal mortality is defined by the World Health Organization (WHO) as "the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes.

Terms for defining congenital anomalies (birth defects)
Definitions and classifications are in line with EuroCat definitions and classifications (https://eu-rd-platform.jrc.ec.europa.eu/sites/default/files/EUROCAT-Guide-1.3.pdf)**:**

Congenital anomaly: Morphological, functional and/or biochemical developmental disturbance in the embryo or foetus whether detected at birth or not. The term congenital anomaly is broad and includes congenital abnormalities, foetopathies, genetic diseases with early onset, developmental delay. Both onset and diagnosis of congenital anomalies can be delayed.

Major anomaly: A life-threatening structural anomaly or one likely to cause significant impairment of health or functional capacity and which needs medical or surgical treatment. The prevalence of major abnormalities recognised at birth among live-born infants is 2%-4% in most series published.

Minor anomaly: Relatively frequent structural anomaly not likely to cause any medical or cosmetic problems.

Major congenital anomalies:
The prevalence of major congenital anomalies among live-born infants is generally considered to be 2%-4%[33]. Prevalence of major congenital anomalies (including chromosomal anomalies) recorded by EUROCAT was 2.4% (2003-2007)[34]. All congenital anomalies will be classified and analysed according to the EUROCAT Classification (EUROCAT Subgroups of Congenital Anomalies). This includes diagnosis in the Q chapter of ICD-10 (and equivalent ICD-9 codes), but excludes a recognized list of minor anomalies as specified by EUROCAT.

As recommended by European regulatory guidelines (GVP III)[24], congenital anomalies should be considered in both live and non-live births (i.e. miscarriage > 12 gestational week, stillbirths).
As recommended by these guidelines we will use the following definitions:

$$\text{Live birth prevalence rate} = \frac{\text{Number of cases among live born infants}}{\text{Total number of live born infants}} * 1000$$

$$\text{Birth prevalence rate} = \frac{\text{Number of cases among live and stillborn infants}}{\text{Total number of (live + still) born infants}} * 1000$$

The numerator is the number of cases of the subject of interest. The denominator is the population from which the numerator came.