



NON-INTERVENTIONAL (NI) STUDY PROTOCOL

PASS information

Title	Post Conditional Approval Active Surveillance Study Among Individuals in Europe Receiving the Pfizer-BioNTech Coronavirus Disease 2019 (COVID-19) Vaccine
Protocol number	C4591021
Protocol version identifier	Version 2 (20 May 2021)
Date	20 May 2021
EU Post Authorization Study (PAS) register number	To be registered before the start of data collection
Active substance	BNT162b2
Medicinal product	COVID-19 messenger ribonucleic acid (mRNA) vaccine is a nucleoside-modified ribonucleic acid (modRNA) encoding the viral spike glycoprotein S of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Marketing Authorization Holder(s) (MAH)	BioNTech Manufacturing GmbH
Joint PASS	No
Research question and objectives	The research question addressed by this study is: Is there an increased risk of select adverse events of special interest (AESI) after being vaccinated with the Pfizer-BioNTech COVID-19 vaccine?

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 1

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

	<p>Objectives</p> <p><u>Primary study objective</u></p> <p>To determine whether an increased risk of prespecified AESI exists following the administration of at least one dose the Pfizer-BioNTech COVID-19 vaccine using two approaches: (a) a cohort design comparing risk in vaccinated and non-vaccinated individuals and (b) a self-controlled risk interval (SCRI) design.</p> <p><u>Secondary study objectives</u></p> <ul style="list-style-type: none">• To estimate the incidence rates of prespecified AESI among individuals who receive at least one dose of the Pfizer-BioNTech COVID-19 vaccine using a cohort study design.• To describe the incidence rates and determine whether an increased risk of prespecified AESI exists following the administration of at least one dose the Pfizer-BioNTech COVID-19 vaccine compared with a matched comparator group with no COVID-19 vaccination within subcohorts of interest (i.e., individuals who are immunocompromised, individuals who are frail and have comorbidities, individuals diagnosed with previous COVID-19 infection, and age-specific groups) in Europe using a cohort study design and/or a SCRI design.• To determine whether an increased risk of prespecified AESI exists following the administration of at least one dose of the Pfizer-BioNTech COVID-19 vaccine compared with no COVID-19 vaccination, in pregnant people and their neonates using a cohort study design.
--	---

	<ul style="list-style-type: none"> To characterise utilisation patterns of Pfizer-BioNTech COVID-19 vaccine among individuals within Europe, including estimating the proportion of individuals receiving the vaccine; two-dose vaccine completion rate and distribution of time gaps between the first and second doses; and demographics and clinical characteristics of recipients, overall and among subcohorts of interest, such as individuals who are immunocompromised, elderly, or have specific comorbidities.
Country(-ies) of study	The Netherlands (NL), Italy (IT), Spain (ES), United Kingdom (UK) of Great Britain, Norway (NO)
Author	<p>Alejandro Arana Senior Director Epidemiology RTI Health Solutions, in collaboration with University Medical Center Utrecht on behalf of the Vaccine monitoring Collaboration for Europe (VAC4EU) Consortium research team Av. Diagonal, 605, 9-1, 08028 Barcelona SPAIN</p>

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Marketing Authorization Holder(s)

Marketing Authorization Holder(s)	BioNTech Manufacturing GmbH An der Goldgrube 12 55131 Mainz Germany
MAH contact person	Constanze Blume

This document contains confidential information belonging to Pfizer. Except as otherwise agreed to in writing, by accepting or reviewing this document, you agree to hold this information in confidence and not copy or disclose it to others (except where required by applicable law) or use it for unauthorized purposes. In the event of any actual or suspected breach of this obligation, Pfizer must be promptly notified.

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

1. TABLE OF CONTENTS

1. TABLE OF CONTENTS.....	5
2. LIST OF ABBREVIATIONS.....	9
3. RESPONSIBLE PARTIES.....	12
4. ABSTRACT.....	14
5. AMENDMENTS AND UPDATES.....	17
6. MILESTONES.....	18
7. RATIONALE AND BACKGROUND.....	18
8. RESEARCH QUESTION AND OBJECTIVES	19
8.1. Objectives.....	19
9. RESEARCH METHODS	20
9.1. Study design	20
9.1.1. Retrospective cohort design.....	20
9.1.1.1. Matching process.....	21
9.1.2. Self-controlled risk interval design.....	22
9.2. Setting.....	24
9.2.1. Inclusion criteria	24
9.2.1.1. Cohort design	24
9.2.1.2. Self-controlled risk interval design	25
9.2.2. Exclusion criteria	25
9.2.2.1. Cohort design	25
9.2.3. Source population	25
9.2.4. Study period.....	25
9.3. Variables.....	26
9.3.1. Exposure definition, by data source.....	26
9.3.1.1. Cohort design	27
9.3.1.2. Self-controlled risk interval design	28
9.3.2. Outcomes definition.....	28
9.3.2.1. Safety outcomes	28
9.3.3. Covariate definition	32
9.4. Data sources	36

9.4.1. PHARMO (NL) (6 million active individuals).....	37
9.4.1.1. Vaccine exposure	38
9.4.2. ARS Toscana database (IT) (3.6 million active individuals)	39
9.4.2.1. Vaccine exposure	39
9.4.3. Pedianet/Health Search Database (IT) (1 million active individuals)	39
9.4.3.1. Vaccine Exposure.....	41
9.4.4. EpiChron – Aragon data sources (ES) (1.3 million active individuals)	41
9.4.4.1. Vaccine exposure	42
9.4.5. Clinical Practice Research Datalink and Hospital Episode Statistics (UK) (16 million active individuals).....	42
9.4.5.1. Vaccine exposure	44
9.4.6. Norwegian health registers (NO) (5.3 million active individuals)	44
9.4.6.1. Norwegian Immunisation Registry	44
9.4.6.2. The Norwegian Patient Registry	44
9.4.6.3. Norway Control and Payment of Health Reimbursement.....	45
9.4.6.4. The Norwegian Prescription Database	45
9.4.6.5. The Medical Birth Registry of Norway.....	45
9.4.6.6. Statistics Norway.....	45
9.4.6.7. The National Registry	45
9.4.6.8. Norwegian Surveillance System for Communicable Diseases	45
9.4.6.9. Vaccine Exposure.....	46
9.4.7. SIDIAP (ES) (5.7 million active individuals).....	46
9.4.8. Cohort design.....	52
9.4.8.1. Exposure assignment and follow-up	52
9.4.8.2. Descriptive statistics.....	53
9.4.8.3. Description of vaccination categories	54
9.4.8.4. Crude outcome measures	54
9.4.8.5. Adjustment for baseline imbalances	55
9.4.8.6. Adjustment for adherence to recommended vaccination schedule	55
9.4.8.7. Meta-analysis	55

9.4.8.8. Vaccine exposure	47
9.5. Study size	47
9.6. Data management.....	48
9.6.1. Case report forms (CRFs)/Data collection tools (DCTs)/Electronic data record.....	49
9.6.2. Record retention.....	50
9.6.3. Data extraction.....	51
9.6.4. Data processing and transformation	51
9.6.5. Data access.....	51
9.7. Data analysis	52
9.7.1. Self-controlled risk interval	55
9.7.1.1. Descriptive statistics.....	55
9.7.1.2. Measures of association	55
9.8. Quality control.....	56
9.8.1. PHARMO (NL)	56
9.8.2. ARS Toscana (IT).....	56
9.8.3. Pedianet/HSD (IT).....	57
9.8.4. EpiChron - Aragon data sources (ES)	57
9.8.5. CPRD (UK)	58
9.8.6. SIDIAP (ES)	58
9.9. Limitations of the research methods	58
9.10. Other aspects	60
10. PROTECTION OF HUMAN SUBJECTS	60
10.1. Patient information.....	60
10.2. Patient consent.....	61
10.3. Institutional review board (IRB)/Independent ethics committee (IEC).....	61
10.4. Ethical conduct of the study.....	61
11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS	61
12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS.....	63
13. REFERENCES	64
14. LIST OF TABLES.....	68

15. LIST OF FIGURES68
ANNEX 1. LIST OF STAND ALONE DOCUMENTS68
ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS68
ANNEX 3. ADDITIONAL INFORMATION.....73

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
ACCESS project	vACcine Covid-19 monitoring readinESS
AESI	adverse events of special interest
ARS Toscana	Agenzia Regionale di Sanita' della Toscana (a research institute of the Tuscany region of Italy)
ATC	Anatomical Therapeutic Chemical (classification system)
BDU	user database at EpiChron
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en Atención Primària (a data resource for pharmacoepidemiology in Spain)
CDM	common data model
CHESS	COVID-19 Hospitalisation in England Surveillance System (UK)
CI	confidence interval
COVID-19	coronavirus disease 2019
CPRD	Clinical Practice Research Datalink
DAP	database access provider
DRE	Digital Research Environment (NL)
DSRU	Drug Safety Research Unit (UK)
DTP	diphtheria, tetanus, and pertussis vaccine
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
EpiChron	EpiChron Research Group on Chronic Diseases at the Aragon Health Sciences Institute (Spain)
ES	Spain
ETL	extraction, transformation, and loading (a process for putting data into a common data model)
EU PAS Register	European Union electronic register of post-authorisation studies
EU	European Union

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 9

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

GOLD	General Practitioner Online Database (of the CPRD)
GP	general practitioner
GPP	Good Pharmacoepidemiology Practices
GVP	Good Pharmacovigilance Practices
HES	Hospital Episode Statistics
HSD	Health Search Database (Italy)
ICD	International Classification of Diseases
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
ICD-10	International Classification of Diseases, 10th Revision
ICPC	International Classification of Primary Care
ISPE	International Society for Pharmacoepidemiology
IT	Italy
KUHR	Norway Control and Payment of Health Reimbursement
MAH	marketing authorisation holder
MBRN	Medical Birth Registry of Norway
mRNA	messenger RNA
MSIS	Norwegian Surveillance System for Communicable Diseases
NHS	National Health Service (UK)
NIPH	Norwegian Institute of Health
NL	Netherlands
NO	Norway
NPR	National Patient Register (Norway)
ONS	Office for National Statistics
PASS	post-authorisation safety study
PHARMO	PHARMO Institute for Drug Outcomes Research or PHARMO Database Network (Netherlands)
PHE	Public Health England
QC	quality control

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 10

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

RTI-HS	RTI Health Solutions
SAP	statistical analysis plan
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2 (cause of COVID-19 disease)
SCRI	self-controlled risk interval (study design)
SIDIAP	Sistema d'Informació per el Desenvolupament de la Investigació en Atenció Primària [Information System for the Improvement of Research in Primary Care] (Spain)
SQL	Structured Query Language
SSB	Statistics Norway
SYSVAK	national, electronic immunisation register
UK	United Kingdom
UMCU	University Medical Center Utrecht
USA	United States of America
VAC4EU	Vaccine monitoring Collaboration for Europe
VAED	vaccine-associated enhanced disease
VV	Varicella zoster virus
WHO	World Health Organization

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

3. RESPONSIBLE PARTIES

Principal Investigator(s) of the Protocol

Name, degree(s)	Job Title	Affiliation	Address
Heather Rubino	Director, Epidemiology	Pfizer	235 E 42nd St., New York, NY 10017
Daniel Weibel	Assistant Professor	University Medical Center Utrecht	Internal mail no Str 6.131 PO Box 85500 3508 GA UTRECHT
Alejandro Arana	Senior Director, Epidemiology	RTI Health Solutions	Av. Diagonal, 605, 9-1, 08028 Barcelona, Spain
Miriam Sturkenboom	Professor	University Medical Center Utrecht	Internal mail no Str 6.131 PO Box 85500 3508 GA UTRECHT
Xabier Garcia de Albeniz	Director, Epidemiology	RTI Health Solutions	Av. Diagonal, 605, 9-1, 08028 Barcelona, Spain
Estel Plana	Director, Biostatistics	RTI Health Solutions	Av. Diagonal, 605, 9-1, 08028 Barcelona, Spain
Alison Kawai	Senior Research Epidemiologist	RTI Health Solutions	307 Waverley Oaks Road, Suite 101, Waltham, MA 02452-8413 USA
Bradley Layton	Senior Research Epidemiologist	RTI Health Solutions	3040 East Cornwallis Rd, PO Box 12194 Research Triangle Park, NC 27709-2194 USA
Rachel Weinrib	Research Epidemiologist	RTI Health Solutions	Av. Diagonal, 605, 9-1, 08028 Barcelona, Spain

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Country Coordinating Investigators

Name, degree(s)	Job Title	Affiliation	Address
Rosa Gini	Head Pharmacoepidemiology Unit	Agenzia regionale di sanità della Toscana	Via Pietro Dazzi 1, 50141 Firenze, Italy
Ron Herings	Director	PHARMO Institute for Drug Outcomes Research	Van Deventerlaan 30-40 3528 AE Utrecht The Netherlands
Carlo Giaquinto	President	PENTA foundation	Corso Stati Uniti 4 35127 Padova Italy
Saad Shakir	Director	Drug Safety Research Unit (DSRU)	Drug Safety Research Unit Bursledon Hall Blundell Lane Southampton, Hampshire SO31 1AA United Kingdom
Alexandra Prados Torres	National Health Service (NHS) Senior Researcher	EpiChron Research Group. Instituto Aragonés de Ciencias de la Salud	Hospital Universitario Miguel Servet Paseo Isabel la Católica 1-3 50009, Zaragoza, Spain
Antonio Gimeno Miguel	Researcher		
Beatriz Poblador-Plou	Researcher		
Boni Bolibar	Scientific Director	IDIAP-Jordi Gol	Gran Via Corts Catalanes 587 08007, Barcelona, Spain
Felipe Villalobos	Researcher		
Angela Lupattelli	Researcher	University of Oslo	PO 1068 Blindern 0316 Oslo, Norway

Note: Country Coordinating investigators have reviewed and contributed to this protocol.

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

4. ABSTRACT

Title: Post Conditional Approval Active Surveillance Study Among Individuals in Europe Receiving the Pfizer-BioNTech Coronavirus Disease 2019 (COVID-19) Vaccine; Version 2 (20 May 2021); Main author. Alejandro Arana, Senior Director Epidemiology, RTI Health Solutions, in collaboration with University Medical Center Utrecht on behalf of VAC4EU Consortium research team

Rationale and background: The novel coronavirus SARS-CoV-2, the cause of COVID-19, has resulted in a global pandemic. The Pfizer-BioNTech COVID-19 vaccine, tozinameran (Comirnaty®) a novel mRNA-based vaccine, has been authorised for use in the European Union (EU), for the prevention of COVID-19. Efficient and timely monitoring of the safety of the vaccine is needed in European countries.

Research question and objectives:

Is there an increased risk of select adverse events of special interest (AESI) after being vaccinated with the Pfizer-BioNTech COVID-19 vaccine?

Objectives

Primary study objective

To determine whether an increased risk of prespecified AESI exists following the administration of at least one dose the Pfizer-BioNTech COVID-19 vaccine using two approaches: (a) a cohort design comparing risk in vaccinated and non-vaccinated individuals and (b) a self-controlled risk interval (SCRI) design.

Secondary study objectives

- To estimate the incidence rates of prespecified AESI among individuals who receive at least one dose of the Pfizer-BioNTech COVID-19 vaccine using a cohort study design.
- To describe the incidence rates and determine whether an increased risk of prespecified AESI exists following the administration of at least one dose the Pfizer-BioNTech COVID-19 vaccine compared with a matched comparator group with no COVID-19 vaccination within subcohorts of interest (i.e., individuals who are immunocompromised, individuals who are frail and have comorbidities, individuals diagnosed with previous COVID-19 infection, and age-specific groups) in Europe using a cohort study design and/or a SCRI design.
- To determine whether an increased risk of prespecified AESI exists following the administration of at least one dose of the Pfizer-BioNTech COVID-19 vaccine compared with no COVID-19 vaccination, in pregnant people and their neonates using a cohort study design.

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 14

- To characterise utilisation patterns of Pfizer-BioNTech COVID-19 vaccine among individuals within Europe, including estimating the proportion of individuals receiving the vaccine; two-dose vaccine completion rate and distribution of time gaps between the first and second doses; and demographics and clinical characteristics of recipients, overall and among subcohorts of interest, such as individuals who are immunocompromised, elderly, or have specific comorbidities.

Study design: A retrospective cohort design will be used to estimate the incidence of AESI after receiving vaccine doses and compare this incidence with that occurring in an unvaccinated comparator group matched on relevant individual characteristics (e.g., age, comorbidities). Where appropriate, the study will also use a SCRI design.

Population: The source population will comprise all individuals registered in each of the health care data sources who are eligible to receive the Pfizer-BioNTech COVID-19 vaccine. The study period will start on the date of launch of the Pfizer-BioNTech COVID-19 vaccine and will end on the date of the latest data availability or 31 Dec 2023. It is expected that follow-up will last for 2 years for AESI. People who are pregnant at time of vaccination or who become pregnant within two years of study start and their live born infants will be followed for an additional 12 months to collect information about birth outcomes and linked infant outcomes.

Variables: Exposure will be based on recorded prescription, dispensing, or administration of the Pfizer-BioNTech COVID-19 vaccine. Vaccine administration and date of vaccination should be obtained from all possible sources that capture COVID-19 vaccination. The outcomes will be based on the AESI proposed by the European Medicines Agency (EMA)-sponsored ACCESS project (vACcine COVID-19 monitoring readinESS). AESI will be identified based on patient profile review of electronic records by health care professionals. In addition, manual review of patient charts conducted by clinicians blinded to COVID-19 vaccine exposure will be performed. Confirmation of an event diagnosis will be classified against existing definitions of the Brighton Collaboration and those currently being developed. Key covariates include demographics; COVID-19 history, as available in each data source (will be used to define a subgroup of interest); personal lifestyle characteristics; comorbidities; immunocompromising conditions (will be used to define subgroups for secondary analyses); comedication use during the year before time zero (prescriptions or dispensing, no over-the-counter medication use); health care utilisation descriptors; other vaccinations; and surrogates of frailty.

Data sources: The study will be performed within the following selected data sources: PHARMO (PHARMO Institute for Drug Outcomes Research) (NL), ARS Toscana (Agenzia Regionale di Sanita' della Toscana) (IT), Pedianet/Health Search Database (HSD) (IT), EpiChron (EpiChron Research Group on Chronic Diseases at the Aragon Health Sciences Institute) (ES), CPRD (Clinical Practice Research Datalink) (UK), the Norwegian health registers (NO), and SIDIAP (Sistema d'Informació per el Desenvolupament de la Investigació en Atenció Primària) [Information System for the Improvement of Research in Primary Care] (ES).

Data analysis: The distributions of baseline characteristics at time zero by exposure group will be calculated to describe the study cohort and illustrate differences between the groups. For safety outcomes, the risk over specific time period(s), incidence rates and their corresponding 95% confidence intervals (CIs) will be computed after the receipt of a first dose and similarly after the receipt of a second dose. Crude risks, cumulative incidence over different time periods, and measures of association (risk differences and risk ratios) for each AESI after vaccination will be estimated in the entire population overall across both doses and separately by dose. Subgroup analyses will be conducted by subgroups defined by demographic and clinical characteristics as well as other covariates of interest. Individuals following each vaccination category under study (vaccination with at least one dose of the Pfizer-BioNTech vaccine vs. no vaccination) may have different characteristics that may determine their risk of AESI. To account for such potential confounding, propensity score methods will be used to estimate the adjusted risk ratios and 95% CIs. Using the main estimates from each data source, appropriate random-effects meta-analytic methods will be used to obtain a combined effect estimate. Where appropriate, the study will also use a SCRI design.

Milestones

Milestone	Planned Date
Registration in the European Union electronic register of post-authorisation studies (EU PAS Register)	Before the start of data collection
Start of data collection	30 September 2021
End of data collection	31 December 2023
Progress report ¹	30 September 2021
Interim report 1	31 March 2022
Interim report 2	30 September 2022
Interim report 3	31 March 2023
Interim report 4	30 September 2023
Interim report 5	31 March 2024
Final study report	30 Sept 2024

¹ Data will not be provided in the progress report

5. AMENDMENTS AND UPDATES

None

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

6. MILESTONES

Milestone	Planned Date
Start of data collection	30 September 2021
End of data collection	31 December 2023
Progress report ¹	30 September 2021
Interim report 1	31 March 2022
Interim report 2	30 September 2022
Interim report 3	31 March 2023
Interim report 4	30 September 2023
Interim report 5	31 March 2024
Final study report	30 Sept 2024

¹ Data will not be provided in the progress report

7. RATIONALE AND BACKGROUND

The novel coronavirus SARS-CoV-2, the cause of COVID-19, has resulted in a global pandemic. The Pfizer-BioNTech COVID-19 vaccine, tozinameran (Comirnaty®) a novel mRNA-based vaccine, has been authorised for use in several countries, including those in the EU, for the prevention of COVID-19. Rapid uptake of the vaccine is expected. Because of the relatively short prelicensure period and limited number of participants in clinical studies, efficient and timely monitoring of the safety of the vaccine will be needed in European countries.

The safety of the Pfizer-BioNTech COVID-19 vaccine has been investigated in clinical studies conducted in the United States, Europe, Turkey, South Africa, and South America and included over 43,000 patients aged 16 years and older. The overall safety profile of the vaccine was found to be favourable in the trial setting. Reported adverse reactions from unblinded data (i.e., from the overall trial population) on participants aged 16 years and older who received two doses of Pfizer-BioNTech COVID-19 vaccine 21 days apart after 2 months of follow-up included pain at the injection site (84.1%), fatigue (62.9%), headache (55.1%), muscle pain (38.3%), chills (31.9%), joint pain (23.6%), fever (14.2%), injection site swelling (10.5%), injection site redness (9.5%), nausea (1.1%), malaise (0.5%), and lymphadenopathy (0.3%). The safety database revealed an imbalance of cases of Bell's palsy (four in the vaccine group and none in the placebo group) ^[1]. Severe allergic reactions have been reported following receipt of the Pfizer-BioNTech COVID-19 vaccine in mass vaccination campaigns outside clinical trials in various countries. Additional safety events may become evident with more widespread use in the general population.

The mRNA vaccine requires careful storage in ultra-low temperature freezers at temperatures between -80°C and -60°C and must be protected from light and ultraviolet radiation before use [1]. Before administration, thawing and dilution of the vaccine are required. These requirements may restrict vaccination events to larger medical centres with appropriate storage capabilities.

The Pfizer-BioNTech COVID-19 vaccine was investigated and is currently recommended as a two-dose vaccine series, with two doses of 0.3 mL each, administered intramuscularly 21 days apart [1]. Little or no data are currently available on the safety and effectiveness of incomplete vaccine series, series mixed with other potentially available vaccines, or altered dosing schedules.

Public health authorities have identified priority populations for vaccination based on health care or essential worker status, comorbidities, and age [2]. Early distribution of the vaccine may be limited to vulnerable groups at higher risk for COVID-19 infection and COVID-19 complications. As recommendations for vaccination are updated over time, the characteristics of vaccine recipients are expected to vary considerably. Approaches for investigating vaccine safety must flexibly account for changing vaccine distribution, which may vary by country or jurisdiction in Europe.

This non-interventional study is designated as a post-authorisation safety study (PASS) and is a commitment to the EMA.

8. RESEARCH QUESTION AND OBJECTIVES

Research question: Is there an increased risk of select adverse events of special interest (AESI) after being vaccinated with the Pfizer-BioNTech COVID-19 vaccine?

8.1. Objectives

Primary study objective

- To determine whether an increased risk of prespecified AESI exists following the administration of at least one dose the Pfizer-BioNTech COVID-19 vaccine using two approaches: (a) a cohort design comparing risk in vaccinated and non-vaccinated individuals and (b) a SCRI design.

Secondary study objectives

- To estimate the incidence rates of prespecified AESI among individuals who receive at least one dose of the Pfizer-BioNTech COVID-19 vaccine using a cohort study design.
- To describe the incidence rates and determine whether an increased risk of prespecified AESI exists following the administration of at least one dose the Pfizer-BioNTech COVID-19 vaccine compared with a matched comparator group with no COVID-19 vaccination within subcohorts of interest (i.e., individuals who are immunocompromised, individuals who are frail and have comorbidities, individuals diagnosed with previous

COVID-19 infection, and age-specific groups) in Europe using a cohort study design and/or a SCRI design.

- To determine whether an increased risk of prespecified AESI exists following the administration of at least one dose of the Pfizer-BioNTech COVID-19 vaccine compared with no COVID-19 vaccination, in pregnant people and their neonates using a cohort study design.
- To characterise utilisation patterns of Pfizer-BioNTech COVID-19 vaccine among individuals within Europe, including estimating the proportion of individuals receiving the vaccine; two-dose vaccine completion rate and distribution of time gaps between the first and second doses; and demographics and clinical characteristics of recipients, overall and among subcohorts of interest, such as individuals who are immunocompromised, elderly, or have specific comorbidities.

9. RESEARCH METHODS

9.1. Study design

This post-authorisation active surveillance study of safety events of interest associated with the Pfizer-BioNTech COVID-19 vaccine will use a retrospective cohort design involving multiple databases.

In addition to the cohort analysis, for a subset of the study endpoints (see [Table 1](#)), the SCRI design will also be used to assess risk. The SCRI design will be used to sequentially monitor the occurrence of AESI while controlling for time-invariant confounders (such as sex, race, chronic illness, and health state).

9.1.1. Retrospective cohort design

A retrospective cohort design will be used to estimate the incidence of AESI after receipt of the vaccine. Incidence rates of prespecified AESI among individuals who receive at least one dose of the Pfizer-BioNTech COVID 19 vaccine will be calculated.

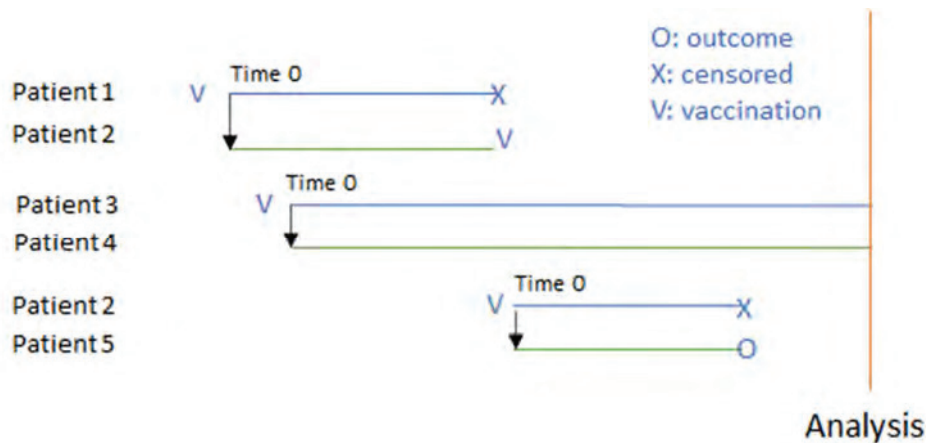
The primary objective will be addressed in a comparative analysis of this incidence with that occurring in an unvaccinated matched comparator group.

In this retrospective cohort design, time zero will be defined as the time at which the exposure status is assigned, when inclusion and exclusion criteria are applied and when study outcomes start to be counted ^[1-5]. Time zero in the *exposed* groups (i.e., recipients of the vaccine) will be the day the first vaccination dose was received. Time zero in the *unexposed* group will be a day when they did not receive a Pfizer-BioNTech COVID-19 vaccine dose. This day will be chosen by calendar matching to the time zero of the corresponding exposed group; at each calendar day when an individual is vaccinated, those individuals who were not vaccinated that same day (time zero) or before will be assigned to the unexposed group, matching them to the vaccinated individual by important clinical variables (e.g., age,

indicated and recommended characteristics to be vaccinated at the time, stratification variables) at time zero.

9.1.1.1. Matching process

As in prior applications of real-world data studies of the Pfizer-BioNTech COVID-19 vaccine^[6], we propose to perform a 1:1 matching without replacement using a “rolling cohort” design. Starting from the first day of the study period, for each day, we will attempt to match newly vaccinated individuals who meet the eligibility criteria that day, even if they had previously been included as an unvaccinated control. Newly vaccinated individuals will be matched 1:1 with unvaccinated controls meeting the eligibility criteria that day who were not previously matched. If at a later date an unvaccinated control is vaccinated, they and their vaccinated match will be censored from this comparative analysis. The newly vaccinated individual (Patient 2 in the figure below) will be censored from the unvaccinated group and a new matched unvaccinated control will be sought so that this newly vaccinated individual and the new pair will be included in the analyses, if eligibility criteria are still met on the day of vaccination.



Patients will be matched on the following variables, which have shown good control of confounding for vaccine effectiveness in a prior study^[6]:

- Age – of 2-year age groups (consecutive years)
- Sex (male, female) – exact matching
- Previous COVID-19 infection (yes/no) – exact matching
- Place of residence – exact matching, at the level of clinical practice, neighbourhood, or small town or proxy as available (specific for data source)
- Influenza vaccines in the past 5 years (0, 1-2, 3-4, 5) – exact matching
- Pregnancy (yes, no) – exact matching

- Immunocompromised (yes/no) – exact matching
- Number of pre-existing conditions considered by the Centers for Disease Control (CDC) as risk criteria (0, 1, 2, 3, 4+)^[7,8] – exact matching

This selection of variables for matching was based on a prior real-world data study^[1]. The selection of variables will be tailored based on the variable availability of each data source. Conditional exchangeability will be evaluated by studying the difference in the risk of outcomes that depend on the antibody-mediated immunogenicity of the COVID-19 disease, in the first 14 days after the first dose, which should be close to 0. If conditional exchangeability is not achieved after implementing all statistical adjustments, the variables for matching may be revisited to improve it. To ensure exchangeability, additional health history measures, such as comorbidities associated with an increased risk of AESI will be explored. Additional details on the matching process, criteria for establishing exchangeability, and analyses to account for the potential autocorrelation introduced by this situation will be specified in the statistical analysis plan (SAP).

A single individual may contribute to both exposed and unexposed groups at different time points (details will be described in the SAP). The causal contrast of interest will be the observational analogue of a per-protocol effect, that is, the event rate difference that would be observed if all individuals received at least one dose of the Pfizer-BioNTech vaccine vs. if no individuals received it. Individuals will be classified into exposure groups that are compatible with their data at time zero. Follow-up under unexposed status and its corresponding exposed pair is censored if an individual receives a COVID-19 vaccine.

Unmatched vaccinated individuals will not be included in the retrospective cohort analysis. They will be described in the descriptive analysis.

9.1.2. Self-controlled risk interval design

As an additional and complementary approach for a subset of study outcomes that are acute and meet other necessary assumptions, a SCRI design will be used. These assumptions include that the outcome must have acute onset and short latency and must have relatively well-known risk intervals; the design is less suited to study outcomes that affect the probability of exposure, but this potential bias can be reduced by the use of a post-vaccination control interval.

Vaccine exposure is known to be challenging to measure, particularly in a pandemic setting where vaccines may be administered outside the usual healthcare system. Often, this results in underascertainment of exposure and the inclusion of exposed persons in the unexposed cohort. This underascertainment of exposure could result in a bias towards the null if the vaccine does increase the risk of an event. As the SCRI design includes only people with known vaccine exposure, it is not subject to this bias.

The SCRI design will compare the risk of each outcome during a prespecified period following each dose during which there is a hypothesised increased risk of the outcome (“risk

interval”) with a self-matched control interval, used to assess the baseline risk of the outcome.

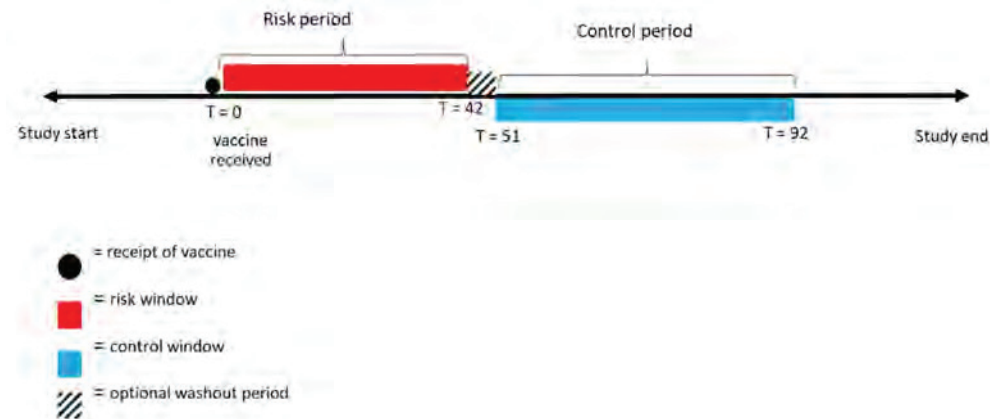
The SCRI design will be performed in the overall vaccinated population, including among vaccinated individuals not included in the retrospective cohort analysis because a matching comparator could not be found. This design will serve as a sensitivity analysis and will allow the evaluation of the exclusion of unmatched pairs from the analysis.

Table 1 defines the risk windows proposed for each AESI and indicates for which AESI a SCRI analysis would be a valid approach.

A prespecified post-vaccination control interval will be used for each outcome. This approach avoids bias because of outcomes affecting the probability of exposure (e.g., the outcome is a contraindication for exposure or delayed exposure). For individuals who receive two doses of the vaccine, the risk interval will extend beyond each vaccine dose.

For outcomes with short risk intervals, for each dose, the control interval will occur close in time to the risk interval associated with that dose and before the second dose is given. For outcomes with risk intervals longer than the gap between doses, among individuals receiving two doses, the control interval for each dose will occur after the risk interval of the second dose (see Figure 1).

Figure 1. Self-controlled risk interval design



T = time measured in days.

Note: Example with a risk period of 42 days and a control period of 42 days.

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

9.2. Setting

For the implementation of this study, we will use electronic health care databases in Europe. The selected data sources and two-letter country codes are as follows:

- PHARMO (PHARMO Institute for Drug Outcomes Research) (NL)
- ARS Toscana (Agenzia Regionale di Sanita' della Toscana) [a research institute of the Tuscany region of Italy] (IT)]
- Pedianet/Health Search Database (HSD) (IT)
- EpiChron (EpiChron Research Group on Chronic Diseases at the Aragon Health Sciences Institute) (ES)
- CPRD (Clinical Practice Research Datalink) (UK)
- The Norwegian health registers (NO)
- SIDIAP (Sistema d'Informació per el Desenvolupament de la Investigació en Atenció Primària) [Information System for the Improvement of Research in Primary Care] (ES)

9.2.1. Inclusion criteria

9.2.1.1. Cohort design

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

- Have a minimum of 12 months (or from birth if enrolled in the data source at birth) of active enrolment and history in one of the selected data sources to ensure adequate characterisation of medical history; this criterion may be met after the start of the study period.
- No history of vaccination with a non-Pfizer-BioNTech COVID-19 vaccine before time zero.

At any point in time, vaccinated individuals may differ from the remaining population in characteristics that may determine their risk of AESI. Measured baseline differences will be adjusted for analytically (Section 9.7).

For the study of pregnancy outcomes, the cohort will be restricted to pregnant women. Details of the differences from the main cohort approach will be described in the SAP.

9.2.1.2. Self-controlled risk interval design

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

- Have received at least one dose of the Pfizer-BioNTech COVID-19 vaccine.
- Have experienced an event during the risk or control interval.
- Have full accrual of data used to define the event in the risk and control intervals combined, taking into account the data lag and timing of data extraction.

9.2.2. Exclusion criteria

9.2.2.1. Cohort and SCRI designs

There are no exclusion criteria. Individuals having any specified contraindication to vaccination or being part of a group not recommended for vaccination in the jurisdiction of the study will be analysed separately

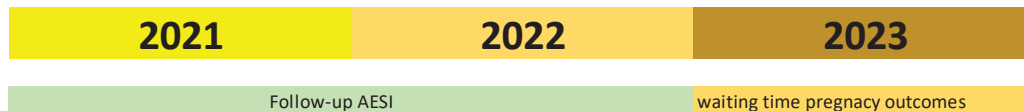
9.2.3. Source population

The source population for both cohort and SCRI designs will be composed of all individuals registered in each of the health care data sources. The selected European populations are the populations underlying the data sources listed in [Section 9.2](#).

9.2.4. Study period

The study period for both cohort and SCRI designs will start on the date of launch of the Pfizer-BioNTech COVID-19 vaccine in each country participating in the study and will end on the date of the latest data availability. Follow-up will last for 2 years for AESI. Differences in follow-up for acute and non-acute events will be described in the SAP. An additional year will accrue for pregnancy outcomes to occur in pregnancies occurring during the 2 years of follow-up (see Figure 2).

Figure 2. Study period and follow-up periods



AESI = adverse events of special interest.

9.3. Variables

9.3.1. Exposure definition, by data source

Exposure will be based on recorded prescription, dispensing, or administration of the Pfizer-BioNTech COVID-19 vaccine. Vaccine receipt and date of vaccination should be obtained from all possible sources that capture COVID-19 vaccination, such as pharmacy dispensing records, general practice records, immunisation registers, vaccination records, medical records, or other secondary data sources. Depending on the data source, vaccines may be identified via nationally used product codes—including batch numbers—where possible. The main exposure of interest is in the receipt of at least one dose of the Pfizer-BioNTech COVID-19 vaccine. Other exposure groups will also be described.

PHARMO (NL): Data on vaccination will be included in PHARMO’s General Practitioner (GP) database. Information on vaccines includes Anatomical Therapeutic Chemical (ATC) code, brand, batch, and date of application.

ARS Toscana (IT) will identify vaccines using the nationally used product code, including batch number.

Pedianet/HSD (IT): Information on COVID-19 vaccine will include date of immunisation, type of vaccine, vaccine batches, dose. They will be collected by the paediatrician at each contact with the patient.

EpiChron – Aragon data sources (ES): The Aragon Health System (Aragon, Spain) has implemented a specific vaccination register embedded in the electronic health record (EHR) system. The COVID-19 vaccine is being systematically registered in this register by health care professionals. This register can collect all the relevant information regarding the vaccination process, such as patient’s identifier; date of administration and due date for next dose, if applicable; centre of administration; part of the body where vaccine is administered; name of the vaccine; brand (laboratory); batch number; dose; and vaccination criteria (risk group to which the patient belongs). There is also a free-text section in which health professionals can include their observations (e.g., presence or not of an allergic reaction).

Clinical Practice Research Datalink (UK): The CPRD contains information recorded by National Health Service (NHS) primary care GPs; and information on the administration of COVID-19 vaccines to individuals will be available. This will include, alongside an encrypted unique patient identifier; the name of the vaccine; manufacturing company; dose; stage of the vaccine schedule; administration route; administration location (e.g.; general practice); batch identifiers/numbers; date of administration; and medical observations, events, referrals, test results, and prescribed medications recorded by the GP prior to, on, or after the vaccination date. Free-text medical notes may also be available if recorded; however, this is dependent on patient anonymity being maintained. In addition, patient demographic, practice-level, and staff-level information is also available.

Furthermore, other CPRD-linked COVID-19 data sets that may provide further follow-up information on AESI include the Public Health England (PHE) Second Generation

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 26

Surveillance System (SGSS) COVID-19 positive virology test data, PHE COVID-19 Hospitalisation in England Surveillance System (CHES), and the Intensive Care National Audit and Research Centre (ICNARC) data on COVID-19 intensive care admissions. Standard CPRD-linked data sets can also be obtained including Hospital Episode Statistics (HES) data sets covering hospital secondary care (Accident & Emergency, Admitted Patient Care, Inpatient and Outpatient), Office for National Statistics (ONS) data sets for Death Registry information, mother-baby link, and an algorithm-based Pregnancy Register.

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

SIDIAP (ES): For all 5.8 million individuals of the Catalan Institute of Health–Primary Care teams, SIDIAP will have available information on the administration of COVID-19 vaccines to individuals linked to a unique and anonymous identifier. The information will be originated from the electronic medical records. For each patient, SIDIAP will have date and centre of administration, health professional administering the vaccine, dose, brand, reasons for vaccination (e.g., risk of group), and other information related to vaccination. As the Pfizer-BioNTech COVID-19 vaccine is indicated as a two-dose vaccine series, multiple vaccinations per person will be identified.

9.3.1.1. Cohort design

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

1. Receipt of at least one dose of the Pfizer-BioNTech COVID-19 vaccine, followed or not by a second dose of the Pfizer-BioNTech COVID-19 vaccine. Individuals will be censored if and when they receive a non–Pfizer-BioNTech COVID-19 vaccine during follow-up.
2. The vaccination category for the matched unexposed group will be defined as not receiving a COVID-19 vaccine of any brand during the study period. Individuals will be censored when they receive a dose of any COVID-19–directed vaccine during follow-up.

As a sensitivity analysis, a vaccination category consisting of the receipt of the two vaccination doses per the recommended schedule will be studied (i.e., receipt of a first dose of the Pfizer-BioNTech COVID-19 vaccine, followed by a second dose by week 4 after the first dose in the absence of an adverse event, and never receiving a non–Pfizer-BioNTech COVID-19 vaccine). For this specific sensitivity analysis, and not for the main analysis, individuals will be censored if they do not receive the second dose of the Pfizer-BioNTech

COVID-19 vaccine by week 6 after the first dose in the absence of an adverse event and if/when they receive a non-Pfizer-BioNTech COVID-19 vaccine during follow-up.

The operationalisation of these exposure strategies is described in 9.7.

9.3.1.2. Self-controlled risk interval design

For the SCRI design, for each dose, person-time in the risk interval will be considered “exposed”, while person-time in the control interval will be considered “unexposed”. Risk intervals will be specific to the outcome of interest and are defined to reflect the duration of time post-vaccine exposure that an incident vaccine-induced event would be expected to occur. Events known to have a risk window limited to a defined period after vaccination are not well known for COVID-19 vaccines, but have been defined based on prior post-marketing studies of other vaccines (where applicable), clinical trial data (where applicable), and passive post-marketing surveillance activities (when they become available). An acute event, while time-limited in duration, does not necessarily have a defined risk window if there is not a known time-limited window post vaccine exposure that the acute event would be expected to occur post-vaccination.

Outcome-specific control intervals will also be defined. For outcomes with short risk intervals, the control interval will occur relatively close in time to the risk interval of each dose. For outcomes with long risk intervals, among individuals receiving two doses, the control interval for both the first and second doses will occur after the risk interval of the second dose. A sensitivity analysis will be performed, where the exposed group of vaccinees is restricted to those who receive vaccine per the recommended schedule, (i.e., two doses of the Pfizer-BioNTech COVID-19 vaccine per the Pfizer-BioNTech recommended dosing schedule).

9.3.2. Outcomes definition

9.3.2.1. Safety outcomes

Outcomes will be defined homogeneously across the data sources to the fullest extent possible. Selected AESI currently planned for inclusion in the study are listed in Table 1 and are based on those proposed by the ACCESS project (vACCine COVID-19 monitoring readinESS), which has been funded by the EMA to ensure that a European infrastructure will be in place to effectively monitor COVID-19 vaccines in the real world, once the vaccines are authorised in the EU (<http://www.encepp.eu/encepp/viewResource.htm?id=37274>).

Table 1. List of Selected Adverse Events of Special Interest

Body system/ classification	Adverse event of special interest	Estimated risk window (days)	Analytic Approach
Autoimmune diseases	Guillain-Barré syndrome ^a	1-42 ³⁰	Cohort/SCRI
	Acute disseminated encephalomyelitis	1-42 ³⁰	Cohort/SCRI

Table 1. List of Selected Adverse Events of Special Interest

Body system/ classification	Adverse event of special interest	Estimated risk window (days)	Analytic Approach
	Narcolepsy ^a	1-42 ^b	Cohort/SCRI
	Acute aseptic arthritis	1-42 ^d	Cohort
	Diabetes (type 1 and broader)	Any	Cohort
	(Idiopathic) thrombocytopenia ^a	1-42 ³¹	Cohort/SCRI
	Heparin-induced thrombocytopenia (HIT)–like event ^a	15 ³⁰	Cohort/SCRI
Cardiovascular system	Acute cardiovascular injury including microangiopathy, heart failure, stress cardiomyopathy, coronary artery disease, arrhythmia, myocarditis	Any ^e	Cohort
Circulatory system	Coagulation disorders: thromboembolism, haemorrhage	1-28 ³⁰	Cohort/SCRI
	Single organ cutaneous vasculitis	1-28 ^f	Cohort/SCRI
Hepato- gastrointestinal and renal system	Acute liver injury	1-42 ^h	Cohort
	Acute kidney injury	1-42 ^h	Cohort
	Acute pancreatitis	1-42 ^h	Cohort
	Rhabdomyolysis	Any	Cohort
Nerves and central nervous system	Generalised convulsion	1-42 ³⁰	Cohort/SCRI
	Meningoencephalitis	1-42 ³⁰	Cohort/SCRI
	Transverse myelitis ^a	1-42 ³⁰	Cohort/SCRI
	Bell’s palsy	1-42 ³⁰	Cohort/SCRI
Respiratory system	Acute respiratory distress syndrome	Any	Cohort
Skin and mucous membrane, bone and joints system	Erythema multiforme	1-42 ^g	Cohort
	Chilblain-like lesions	1-42 ^f	Cohort
Other system	Anosmia, ageusia	1-42	Cohort
	Anaphylaxis ^a	1 ³⁰	Cohort

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Table 1. List of Selected Adverse Events of Special Interest

Body system/ classification	Adverse event of special interest	Estimated risk window (days)	Analytic Approach
	Multisystem inflammatory syndrome	1-42 ^c	Cohort
	Death (any causes)	Any	Cohort
	Subacute thyroiditis	1-42 ^d	Cohort
	Sudden death	Any	Cohort
Pregnancy outcome, maternal	Gestational diabetes	Any time pregnancy	Sub-cohort
	Preeclampsia	Any time pregnancy	Sub-cohort
	Maternal death	Any time pregnancy	Sub-cohort
Pregnancy outcome, neonates. Define design taking trimester into account	Foetal growth restriction	Any time pregnancy	Sub-cohort
	Spontaneous abortions	After vaccination	Sub-cohort
	Stillbirth	After vaccination	Sub-cohort
	Preterm birth	At preterm birth	Sub-cohort
	Major congenital anomalies ^a	1 year after birth	Sub-cohort
	Microcephaly	At birth	Sub-cohort
	Neonatal death	At birth	Sub-cohort
	Termination of pregnancy for foetal anomaly	At termination	Sub-cohort
Any	COVID-19 Disease	Any	Cohort
	Vaccine-associated enhanced disease (VAED) ^a	Any	Cohort

AESI = adverse events of special interest; VAED = vaccine-associated enhanced disease.

Notes:

a For this AESI clinical validation will occur.

b Published risk and control intervals for demyelinating diseases and cranial disorders were applied to TM and narcolepsy/cataplexy.

c As severe COVID-19 ranges from severe pneumonia, acute respiratory distress syndrome, and multisystem organ failure/MIS-A, a 1-42 day risk interval was applied in order to capture the 14-day incubation period of the disease and 4-5 day period from exposure to symptom onset.

d Published risk and control intervals for autoimmune disorders were applied to similar autoimmune rheumatic conditions (i.e., fibromyalgia and autoimmune thyroiditis).

e Published risk and control intervals for myocarditis and pericarditis were applied to other cardiovascular conditions (i.e., heart failure and cardiogenic shock, stress cardiomyopathy, CAD, arrhythmia, AMI).

f Similar risk and control intervals were applied to all cardiovascular and hematological disorders characterized by damage to the blood vessels and/or arteries and clotting (i.e., microangiopathy, DVT, pulmonary embolus, limb ischemia, hemorrhagic disease, DIC, chilblain-like lesions). The published risk and control intervals for KD were applied to vasculitides given that KD is a type of medium and small-vessel vasculitis.

Table 1. List of Selected Adverse Events of Special Interest

Body system/ classification	Adverse event of special interest	Estimated risk window (days)	Analytic Approach
--------------------------------	--------------------------------------	---------------------------------	-------------------

g Published risk and control intervals for non-anaphylactic allergic reactions were applied to hypersensitivity disorders (i.e., erythema multiforme).

h Risk intervals of 42 days were applied for acute kidney injury and liver injury to be consistent with other COVID-19 related safety events of interest.

Outcomes will be identified in EHR databases with algorithms based on codes for diagnoses, procedures, and treatments. Definitions, codes, and proposed algorithms for all AESI will incorporate definitions developed by the ACCESS project (https://drive.google.com/drive/folders/1Y_3cuGRN1g-jBv2ec1fC0aYcpxEjtrY9) and will be described in more detail in the SAP.

9.3.2.1.1. Outcome identification and validation, by data source

AESI will be identified based on patient profile review of electronic records by health care professionals. In addition, for selected outcomes mentioned in Table 1 and others (if considered necessary in a future evaluation of results), manual review of patient charts conducted by clinicians blinded to COVID-19 vaccine exposure will be performed when possible and will be based on data source structure. Confirmation of an event diagnosis will be classified against existing definitions of the Brighton Collaboration and those currently being developed.

Standard algorithms for each outcome definition will be applied to participant data sources, based on the results of the ACCESS project. Algorithms will be tailored to the data source and will consider the nature of the records that have identified the outcome, e.g., primary care, access to hospital care, access to emergency care [9]. Multiple algorithms for the same outcome may be included in the analysis, to assess the potential impact of differential misclassification.

Potential outcomes will be identified based on patient profile review of electronic records by health care professionals.

PHARMO (NL): For the validation study, information on selected endpoints from patient medical records will be abstracted by local medical professionals or PHARMO employees, provided that medical chart review is approved by ethics committees and other local and/or national governing bodies.

Pedianet/HSD (IT): A validation mechanism including an individual linkage with the electronic regional immunisation register will be in place. Furthermore, the validation process includes the review by clinicians of the individuals' electronic medical records, which contain information from primary care reports.

EpiChron (ES): In Aragon (EpiChron) data sources, the proposed validation process is based on the review of the individuals' electronic medical records by clinicians from the research team who are blinded to COVID-19 vaccination status. These records include information from primary care reports, hospital discharge reports (including hospital emergency rooms), and results of diagnostic tests and laboratory tests.

Clinical Practice Research Datalink (UK): In the United Kingdom (UK), validation will be conducted by review of electronic medical record information for selected endpoints by an adjudication committee who will be blinded to COVID-19 vaccination status.

Norwegian health registers (NO): In Norway, the validation process is based on the manual review of hospital charts for a subsample of individuals with the adverse event of interest, compared with registered diagnoses in the Patient Registry of Norway. Validation studies are already available for selected health outcomes (e.g., intracranial haemorrhage, hip fractures, cancer). Depending on the adverse event of interest, validation is possible by comparing the registered diagnosis in two separate registers (e.g., the Norwegian Patient Registry versus the Norwegian Stroke Register).

SIDIAP (ES): In SIDIAP, the validation process is part of data quality control. Validation will be based on the review of the electronic medical record information (ECAP) by members of the SIDIAP research group who will be blinded to COVID-19 vaccination status.

9.3.3. Covariate definition

The following variables will be assessed at time zero (for the cohort design) or the date of initial vaccine dose (for the SCRI design) to be used to define patient populations of special interest or priority vaccination groups, to define subgroups of interest for secondary analyses, or to control for confounding. The AESI may have different sets of risk factors, and outcome-specific analyses may contain different covariate sets. Potential covariates may include the following information, as available in each data source:

- Demographics
 - Age at time zero (will be used to define subgroups for secondary analyses)
 - Age will be categorised as age categories in line with published background incidence rates from ACCESS (0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
 - Sex
 - Pregnancy status and pregnancy trimester at time zero
 - Race and/or ethnicity, as appropriate in each country
 - Geographic region, as appropriate in each country

- Socioeconomic status, as available in each country (including housing, employment, and income, if available)
- Residency in a long-term care facility
- Health care worker or essential worker status, if available
- Date of vaccination (categorised as appropriate, e.g., by year or month)
- Batch of vaccine received
- COVID-19 history, as available in each data source (will be used to define a subgroup of interest)
 - Previous diagnosis of COVID-19
 - Positive test result for COVID-19
- Personal lifestyle characteristics
 - Smoking status (if available)
 - Body mass index (if available)
- Comorbidities
 - History of anaphylaxis
 - History of allergies
 - Diabetes mellitus (types 1 and 2)
 - Hypertension
 - Cardiovascular disease
 - Cerebrovascular disease
 - Chronic respiratory disease
 - Chronic kidney disease
 - Chronic liver disease
 - Cancer

- Autoimmune disorders
- Influenza infection or other respiratory infections
- Charlson Comorbidity Index (may be included as the composite scale, or the scale components may be included as individual terms)
- Immunocompromising conditions (will be used to define subgroups for secondary analyses)
 - Immunodeficiencies
 - Immunosuppressant medication use
 - Human immunodeficiency virus and other immunosuppressing conditions
- Comedication use during the year before time zero (prescriptions or dispensing, no over-the-counter medication use)
 - Analgesics
 - Antibiotics
 - Antiviral medications
 - Corticosteroids
 - Non-steroidal anti-inflammatory drugs
 - Psychotropics
 - Statins
 - Novel oral anticoagulants
 - Warfarin
- Health care utilisation in the year before time zero and in the 2 weeks before time zero
 - Number of hospitalisations
 - Number of emergency department visits
 - Skilled nursing facility, nursing home, or extended care facility stay
 - Primary care utilisation

- Cancer screening
- Other preventive health services, as appropriate
- COVID-19 tests
- Other vaccinations
 - Influenza
 - Pneumococcal
 - DTP (diphtheria, tetanus, and pertussis)
 - TPV (polio)
 - TV (MMR) (measles, mumps and rubella)
 - Hib (Haemophilus influenzae type b)
 - HB (hepatitis B virus)
 - VV (varicella zoster virus)
 - HZ (herpes-zoster virus)
 - HPV (human papillomavirus)
 - Meningococcal
 - Rotavirus
- Surrogates of frailty
 - Wheelchair use
 - Home hospital bed
 - Paralysis
 - Parkinson's disease
 - Skin ulcer
 - Weakness

- Stroke/brain injury
- Ambulance transport
- Dementia
- Difficulty walking
- Home oxygen
- Rehabilitation care
- Psychiatric illness
- Sepsis
- Heart failure
- Podiatric care
- Bladder incontinence
- Diabetes complications
- Arthritis
- Coagulation deficiencies
- Vertigo
- Lipid abnormalities

9.4. Data sources

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

At the proposal stage, members of VAC4EU (Vaccine monitoring Collaboration for Europe) (<https://vac4eu.org/>) were offered the option to participate in the study. Several data sources have indicated the ability to participate in the study and are described in the following subsections. As vaccine delivery and registration are not fully determined, we are exploring additional options in Portugal and Norway (we include here the Norwegian data sources as they will be able to contribute with one data extraction per year).

Data availability for each institution might be affected by third parties or external circumstances that are independent from the institution involved in the study as described below in Section 9.9.

9.4.1. PHARMO (NL) (6 million active individuals)

The PHARMO Database Network, which is maintained by the PHARMO Institute for Drug Outcomes Research, is a population-based network of EHR databases that combines anonymous data from different primary and secondary health care settings in the Netherlands. These different data sources—including data from general practices, in- and outpatient pharmacies, clinical laboratories, hospitals, the cancer register, the pathology register, and the perinatal register—are linked on a patient level through validated algorithms. To ensure data privacy in the PHARMO Database Network, the collection, processing, linkage, and anonymisation of the data are performed by STIZON, which is an independent, ISO/IEC 27001 certified foundation that acts as a trusted third party between the data sources and the PHARMO Institute. The longitudinal nature of the PHARMO Database Network enables the follow-up of more than 9 million individuals of a well-defined population in the Netherlands for an average of 12 years. Currently, the PHARMO Database Network covers over 6 million active individuals out of 17 million inhabitants of the Netherlands ^[10]. Data collection period, catchment area, and overlap between data sources differ. Therefore, the final cohort size for any study will depend on the data sources included. All electronic patient records in the PHARMO Database Network include information on age, sex, socioeconomic status, and mortality. Other available information depends on the data source. A detailed description of the different data sources is given in subsequent sections. The PHARMO Institute is always seeking new opportunities to link with health care databases. Furthermore, it is possible to link additional data collections, such as data from chart reviews, patient-reported outcomes, or general practice trials.

The General Practitioner database comprises data from electronic patient records registered by GPs. The records include information on diagnoses and symptoms, laboratory test results, referrals to specialists, and health care product/drug prescriptions. The prescription records include information on type of product, prescription date, strength, dosage regimen, quantity, and route of administration. Drug prescriptions are coded according to the World Health Organization (WHO) ATC classification system [www.whocc.no]. Diagnoses and symptoms are coded according to the International Classification of Primary Care (ICPC) [www.nhg.org], which can be mapped to the International Classification of Diseases (ICD) codes but can also be entered as free text. General practitioner data cover a catchment area representing 3.2 million residents (~20% of the Dutch population).

The Out-patient Pharmacy Database comprises GP- or specialist-prescribed health care products dispensed by the outpatient pharmacy. The dispensing records include information on type of product, date, strength, dosage regimen, quantity, route of administration, prescriber specialty, and costs. Drug dispensings are coded according to the WHO ATC classification system. Outpatient pharmacy data cover a catchment area representing 4.2 million residents (~25% of the Dutch population). The PHARMO Database Network is listed under the European Network of Centres for Pharmacoepidemiology and

Pharmacovigilance (ENCePP) resources database. PHARMO data capture influenza vaccination and may be linked to the PRAEVENTIS database maintained by RIVM, based on specific permissions.

The GP Database contains vaccinations administered by GPs and by the public health service, as GPs receive an automated notification when a patient has a positive corona test or has been vaccinated via the public health service (provided that individuals have given their consent).

The Dutch government wants everyone from the age of 18 years to have had at least one COVID-19 vaccination by the beginning of July 2021. This vaccination schedule depends on many factors (e.g., approval and effectiveness, delivery and distribution of vaccines to injection sites, such as hospitals and GPs, new developments and advice from, for instance, the Health Council of the Netherlands [i.e., de Gezondheidsraad]).

The Netherlands Perinatal Registry is maintained by Perined and comprises data on pregnancies, births, and neonatal outcomes of births in the Netherlands, voluntarily collected by perinatal caregivers mainly for benchmarking. For research purposes, the data are linked with the PHARMO Database Network via the TTP, resulting in the PHARMO Perinatal Research Network (PPRN)^[11]. Records include information on mothers (e.g., maternal age, obstetric history, parity), pregnancy (e.g., mode of conception, mode of delivery), and children (e.g., birth weight, gestational age, Apgar score). Diagnoses and symptoms are coded according to the Perinatal Registry code lists. For more information: www.perined.nl

Permission to obtain these data on a by-project basis is needed from PHARMO as well as from Perined.

PHARMO acknowledges that the data source they have access to includes data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.1.1. Vaccine exposure

Currently, in the Netherlands, different health care providers administer the COVID-19 vaccines (i.e., GPs, the public health service, and health care institutions). The vaccination data are recorded in a central register (if people have given permission beforehand). This is the COVID-19 vaccination Information and Monitoring System (CIMS). PHARMO is currently exploring the possibilities of linking with this register. Until then, the GP Database will be the basis for the vaccination data and thus vaccination exposure may be underascertained.

As of 3 May 2021, approximately 3.6 million doses of the Pfizer-BioNTech COVID-19 vaccine have been administered in the Netherlands. Assuming that PHARMO covers approximately 20% of the Dutch population, this should be 700,000 doses in PHARMO's catchment area.

Therefore it is difficult to estimate how many individuals will be vaccinated with the Pfizer-BioNTech COVID-19 vaccine during the study period in the Netherlands. Also, the administration to the Pfizer-BioNTech COVID-19 vaccine depends on different factors, such as type of work (care workers), home living (yes or no), year of birth, and existing comorbidities.

9.4.2. ARS Toscana database (IT) (3.6 million active individuals)

The Italian National Healthcare System is organised at the regional level: the national government sets standards of assistance and tax-based funding for each region, which regional governments are responsible for providing to all their inhabitants. Tuscany is an Italian region, with approximately 3.6 million inhabitants. The Agenzia Regionale di Sanita' della Toscana is a research institute of the Tuscany region. The ARS Toscana database comprises all information collected by the Tuscany region to account for the health care delivered to its inhabitants. Moreover, ARS Toscana collects data from regional initiatives. All data in the ARS Toscana data source can be linked at the individual level through a pseudo-anonymous identifier. The ARS Toscana database routinely collects primary care and secondary care drug prescriptions for outpatient use and is able to link them at the individual level with hospital admissions, emergency care admissions, records of exemptions from copayment, diagnostic tests and procedures, causes of death, the mental health services register, the birth register, the spontaneous abortion register, and the induced terminations register. A pathology register is available, mostly recorded in free text, but with morphology and topographic SNOMED codes. Mother-child linkage is possible through the birth register. Vaccination data since 2016 are available for children and since 2019 for adults. However, to date, 2019 vaccination data for adults may still be incomplete resulting in an underascertainment of vaccine exposure. The ARS Toscana database was characterised in the ADVANCE project and considered fit for purpose for vaccine coverage, benefits, and risk assessment when using the new vaccine register (from 2019) ^[12].

ARS Toscana acknowledges that the data source they have access to includes data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.2.1. Vaccine exposure

Using data from the European Centers for Disease control^[13], as of 08 May 2021, 17,801,550 doses of the Pfizer-BioNTech COVID-19 vaccine have been administered. If the distribution of the vaccine were uniform across Italian regions, it is expected that 1 million doses would have been administered in Toscana.

9.4.3. Pedianet/Health Search Database (IT) (1 million active individuals)

Pedianet, a paediatric general practice research database, was set up in 2000. It contains reason for accessing health care, health status (according to the Guidelines of Health Supervision of the American Academy of Pediatrics), demographic data, diagnosis and clinical details (free text or coded using the ICD-9-CM [International Classification of Diseases, Ninth Revision, Clinical Modification]), prescriptions (pharmaceutical

prescriptions identified by the ATC code), specialist appointments, diagnostic procedures, hospital admissions, growth parameters, and outcome data of the children habitually seen by approximately 140 family paediatricians distributed throughout Italy.

Pedianet can link to other databases using unique patient identifiers. In the first database, information on routine childhood vaccination is captured, including vaccine brand and dose. In the second database, information on patient hospitalisation date, reason for hospitalisation, days of hospitalisations, and discharge diagnosis (up to six diagnoses) is captured. The family paediatricians' participation in the database is voluntary, and individuals and their parents provide consent for use of their data for research purposes. In Italy, each child is assigned to a family paediatrician, who is the referral for any health visit or any drug prescription; thus, the database contains a very detailed personal medical history. The data, generated during routine practice care using common software (JuniorBit®), are anonymised and sent monthly to a centralised database in Padua, Italy, for validation. The Pedianet database can be linked to regional vaccination data, which was successfully tested in several large European projects (e.g., ADVANCE, GRIP, EMIF, EU Alliance) where it was characterised and deemed fit for purpose to evaluate prescriptions including paediatric routine vaccines ^[12].

Children aged younger than 12 years will likely start receiving the Pfizer-BioNTech COVID-19 vaccine soon. This vaccine is expected to be the first COVID-19 vaccine rolled out among children, and most children in Italy are likely to receive it. We will be able to capture most of these individuals in Pedianet, as it is expected that approximately 10,000 vaccinated children aged 12 to 14 years will have data available in Pedianet.

The HSD, an Italian general practice data source in place since 1998, comprises data from computer-based patient records registered by a selected group of GPs uniformly distributed throughout Italy. The individuals in the database are representative of the entire Italian population. In HSD, patient demographic details are linked through an encrypted patient code with medical records (e.g., diagnoses, tests performed, test results, hospital admissions), drug prescription information (trade name, dosage form, ATC code, ministerial code, active substance, date of filled prescription, number of days' supply), risk factors and determinants of health (blood pressure, body mass index, smoking habits), and date of death. Diseases are classified according to ICD-9-CM. Ambulatory procedures are encoded in accordance with the Nomenclatore Tariffario, a list of all outpatient specialist medical services and related tariffs, instituted by Ministerial Decree in 1996. Currently, almost 900 GPs are caring for approximately 1 million individuals (almost 20% of whom are aged younger than 19 years). Pedianet will have an individual patient linkage with the Immunization administrative database, which will allow to have all the required information.

Pedianet/HSD acknowledges that the data source they have access to includes data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.3.1. Vaccine Exposure

Using data from the European Center for Disease control, as of 08 May 2021, 17,801,550 doses of the Pfizer-BioNTech COVID-19 vaccine have been distributed among 60.36 million Italians^[13]. If the distribution of the vaccine were uniform across Italian regions, it is expected that 295,000 doses would have been administered in HSD.

In Italy, the COVID-19 vaccination campaign started in December 2020. Every region ($n = 20$) has adopted different vaccination strategies involving hubs and/or general practices. The primary care setting was actively involved in the vaccination campaign only at the beginning of April 2021, and only certain age categories and/or types of vaccines were available for direct administration by GPs. Thus, for the period between January and March 2021, Italian GPs have likely recorded vaccine injections according to three main pathways: a) some regions automatically informed GPs regarding their patients' COVID-19 vaccination; b) GPs referred patients to a specific hub and registered their vaccination status there; and c) patients autonomously reported their vaccination to their GPs. For the first semester of 2021, HSD expects to find complete data for certain age categories, while for the first trimester and for some other age categories, we could find only incomplete data for some regions. In HSD, after preliminary evaluation of data completeness, the study design (e.g., self-controlled or cohort design) will be chosen for the specific objectives.

9.4.4. EpiChron – Aragon data sources (ES) (1.3 million active individuals)

The Spanish National Health System is organised at a regional level. Aragon is one of the regions, with approximately 1.3 million inhabitants. The Aragon data sources to be used in this project, which cover approximately 98% of the reference population, are the following:

- The user database (BDU) with sociodemographic information
- Individuals' electronic medical records from primary care (OMI-AP) and hospital care (Minimum Basic Data Set, CMBD, with data on hospital discharges, and PCH database with data on visits to the emergency room)
- Individuals' pharmacotherapeutic history with prescriptions and dispensation of drugs in community pharmacies (Receta Electrónica database) and hospitals (for hospitalised patients and outpatients)
- Aragon COVID-19 Registry

Furthermore, additional databases and registers at the local (i.e., hospital or primary care health care centre) and national (e.g., Base de Datos para la Investigación Farmacoepidemiológica en Atención Primària [BIFAP] database and its CIAP dictionary; SINASP) level, as well as new potential databases or registers that could be developed for the vaccination process during the project, will be explored and used if appropriate. All the information contained in these data sources is linked at the patient level through a pseudonymisation process and then anonymised for research purposes. The group's researchers have broad experience in the use of these databases for research on chronic

diseases and COVID-19. From the Aragon Health Department, the following key persons will be directly involved in the project: Antonio Poncel-Falcó, Maria Isabel Cano del Pozo, Cristina Navarro Pemán, Ana Cristina Bandrés Liso, Mercedes Aza Pascual Salcedo, and Francisca González Rubio. The group developed the EpiChron Cohort Study [14] for the analysis of the clinical epidemiology of chronic diseases, multimorbidity, and polypharmacy using real-world data from some of the aforementioned databases during the period 2010-2020; more than 30 scientific publications have resulted from this study. The group has also developed the PRECOVID study [15], for the demographic and clinical characterisation of all COVID-19 patients in the Aragon region and for identifying variables associated with increased mortality risk. Diagnoses are coded initially according to the ICPC or ICD and are subsequently grouped into diagnostic clusters, if needed, using open software (i.e., Clinical Classifications Software). Drug prescriptions and dispensations are coded according to the WHO ATC classification system. Once the aforementioned data sources have been gathered and linked at the patient level, data undergo continuous quality control checks that ensure their accuracy and reliability for research purposes.

Information on pregnancy, pregnancy outcomes, and mother-baby linkage from women who give birth in at least the two most relevant public hospitals in the Aragon region in which approximately 70% of births in the region occur is expected to be available in the EpiChron database. The mother-baby linkage is possible using the Neosoft software at the hospital level, in which all information about the mother and baby is recorded. This information will be completed using information from the mother's electronic health records.

EpiChron acknowledges that the data source they have access to includes data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.4.1. Vaccine exposure

Approximately 0.25 million first doses and 0.14 million second doses of the Pfizer-BioNTech COVID-19 vaccine have been administered in the Aragon region from 27 December 2020 to 30 April 2021. During May 2021, Aragon will receive a 0.25 million dose batch, although the number of individuals who may potentially receive at least one dose of the vaccine in Aragon during the study period is currently difficult to estimate and will depend on the availability of the vaccine in Spain in later stages.

9.4.5. Clinical Practice Research Datalink and Hospital Episode Statistics (UK) (16 million active individuals)

The CPRD from the UK collates the computerised medical records of GPs in the UK who act as the gatekeepers of health care and maintain patients' life-long EHRs. Accordingly, GPs are responsible for primary health care and specialist referrals, and they also store information about specialist referrals and hospitalisations. General practitioners act as the first point of contact for any non-emergency health-related issues, which may then be managed within primary care and/or referred to secondary care, as necessary. Secondary care teams also provide information to GPs about their patients, including key diagnoses. The data recorded in the CPRD include demographic information, prescription details, clinical events,

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 Non-Interventional Study Protocol Template For Secondary Data Collection Study

01-Jun-2020

Page 42

preventive care, specialist referrals, hospital admissions, and major outcomes, including death. Most of the data are coded using Read or SNOMED codes. Data validation with original records (specialist letters) is also available.

The data set is generalisable to the UK population based on age, sex, socioeconomic class, and national geographic coverage when CPRD GOLD (General Practitioner Online Database) and CPRD Aurum versions are used.

There are currently approximately 59 million individuals (acceptable for research purposes)—16 million of whom are active (i.e., still alive and registered with the GP practice)—in over 2,000 primary care practices (<https://cprd.com/Data>). Data include demographics, all GP/health care professional consultations (e.g., phone calls, letters, emails, in surgery, at home), diagnoses and symptoms, laboratory test results, treatments (including all prescriptions), all data referrals to other care providers, hospital discharge summary (date and Read/SNOMED codes), hospital clinic summary, preventive treatment and immunisations, and death (date and cause). For a proportion of the CPRD panel practices (> 80%), the GPs have agreed to permit the CPRD to link at the patient level to HES data. The CPRD is listed under the ENCePP resources database, and access will be provided by the Drug Safety Research Unit (DSRU). The CPRD was not yet characterised in the ADVANCE project, for which the UK THIN and RCGP databases were used, but has been largely used in vaccine studies.

The HES database contains details of all admissions to NHS hospitals in England (Accident & Emergency, Admitted Patient Care, Outpatients); approximately 44.6 million individuals in the CPRD are linked to the HES database. Not all patients in the CPRD have linked data (e.g., if they live outside England, if their GP has not agreed that their data should be used in this way). As with standard CPRD patients, HES data are limited to patients who are research standard. CPRD records are linked to the HES using a combination of the patient's NHS number, sex, and date of birth ^[16]. Additional CPRD-linked data sets include Death Registration data from the ONS, which includes information on the official date and causes of death (using ICD codes), mother-baby link, and an algorithm-based Pregnancy Register.

In addition, other CPRD-linked COVID-19 data sets, which may provide further follow-up information on AESI, include the PHE Second Generation Surveillance System (SGSS) COVID-19 positive virology test data, PHE COVID-19 Hospitalisation in England Surveillance System (CHESS), and the ICNARC data on COVID-19 intensive care admissions.

The mother-baby link (which uses a probabilistic algorithm, based on data in the primary care medical records) and the Pregnancy Register are linked data sets available with the CPRD GOLD database. For patients identified in the CPRD Aurum database, the mother-baby link and Pregnancy Register information is not available. However, information on pregnancy status and pregnancy outcomes is likely to be available in both CPRD databases as events reported by the GP in the primary care medical records.

The DSRU acknowledges that the CPRD data sources they have access to include data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.5.1. Vaccine exposure

Approximately 1.2 million patients were identified in the CPRD Aurum database (March 2021 database release) who had received at least one dose of the Pfizer-BioNTech COVID-19 vaccine. Based on the current UK COVID-19 vaccine delivery strategy, with younger age groups more likely to receive the Pfizer-BioNTech COVID-19 vaccine, we would estimate the ability to identify approximately a further 2 to 3 million patients in the CPRD databases who may potentially receive at least one dose of the Pfizer-BioNTech COVID-19 vaccine during the study period. It is expected that, in the near future, the CPRD GOLD database will be able to contribute to the overall CPRD study, although it is not possible to estimate the number of patients that will be suitable for participation in the study.

9.4.6. Norwegian health registers (NO) (5.3 million active individuals)

The Norwegian data sources included in this project, accessed through a partnership with the University of Oslo are several national health registers, i.e., the Medical Birth Registry of Norway (MBRN), the National Patient Register (NPR), Norway Control and Payment of Health Reimbursement (KUHR), the Norwegian Immunisation Registry (SYSVAK), the National Prescription Registry (NoPD), and Statistics Norway (SSB).

The source population will be identified using the Norwegian Institute of Health's (NIPH) copy of the Norwegian population data file from the National Registry. The NPR and KUHR (and the MBRN for the pregnant population) provide data on inpatient and outpatient diagnostic codes. Information on population background data is derived from SSB (e.g., education, occupation status, sex, age). Data on vaccination status are derived from SYSVAK and the Norwegian Prescription Database. The latter register includes data on filled prescriptions for possible co-medications and other prescription drug use.

9.4.6.1. Norwegian Immunisation Registry

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

9.4.6.2. The Norwegian Patient Registry

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

onwards are available. Diagnostic codes in the NPR follow the *International Classification of Diseases, 10th Revision* (ICD-10).

9.4.6.3. Norway Control and Payment of Health Reimbursement

The KUHR is an administrative database based on electronically submitted reimbursement claims from physicians to the Norwegian Health Economics Administration (HELFO). It contains information from primary health care, GP, and emergency services on morbidity, utilisation of health care services, and health care use. Person-specific data are available for the years 2010 through 2018. Diagnostic codes in the KUHR follow ICD-10, but the ICPC is more frequently used by GPs.

9.4.6.4. The Norwegian Prescription Database

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

9.4.6.5. The Medical Birth Registry of Norway

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

9.4.6.6. Statistics Norway

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

9.4.6.7. The National Registry

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

9.4.6.8. Norwegian Surveillance System for Communicable Diseases

Notification of infectious diseases to the Norwegian Surveillance System for Communicable Diseases (MSIS) is an important part in the surveillance of infectious diseases in Norway. Microbiological laboratories analysing specimens from humans, and all doctors in Norway, are required by law to notify cases of certain diseases (71 in total, including SARS-CoV-2) to the MSIS central unit at the Norwegian Institute of Public Health. The following variables are available since 1977: notifiable disease, month and year of diagnosis,

age groups, county of residence, and place of infection. Data on positive COVID-19 tests are updated continuously.

The University of Oslo acknowledges that the data sources they have access to (described above) include data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.6.9. Vaccine Exposure

According to the Vaccination calendar (version 30 April 2021), in Norway, 7,280,000 doses of the Pfizer-BioNTech COVID-19 vaccine will be distributed by the end of September 2021.

9.4.7. SIDIAP (ES) (5.7 million active individuals)

The Information System for the Improvement of Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària' [SIDIAP]) was created in 2010 by the Catalan Health Institute and the IDIAPJGol Institute. It includes information collected since 01 January 2006 during routine visits at 278 primary care centres pertaining to the Catalan Health Institute in Catalonia (North-East Spain) with 3,414 participating GPs. SIDIAP has pseudo-anonymised records for 5.7 million people (80% of the Catalan population) and is highly representative of the Catalan population.

The SIDIAP data comprise the clinical and referral events registered by primary care health professionals (e.g., GPs, paediatricians, and nurses) and administrative staff in electronic medical records, comprehensive demographic information, community pharmacy invoicing data, specialist referrals, and primary care laboratory test results. The SIDIAP data can also be linked to other data sources, such as the hospital discharge database, on a project-by-project basis. Health professionals gather this information using ICD-10 codes, ATC codes, and structured forms designed for the collection of variables relevant for primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, and blood and urine test results. Regarding vaccinations, SIDIAP includes all routine childhood and adult immunisations, including the antigen and the number of administered doses. Encoding personal and clinic identifiers ensures the confidentiality of the information in the SIDIAP database. The SIDIAP database is updated annually at the start of each year.

Currently, with the COVID-19 pandemic, there is the possibility to have shorter term updates in order to monitor the evolution of the pandemic. Recent reports have shown the SIDIAP data to be useful for epidemiological research. SIDIAP is listed under the ENCePP resources database (www.encepp.eu/encepp/resourcesDatabase.jsp). The SIDIAP database was characterised in the ADVANCE project and considered fit for purpose for vaccine coverage, benefits, and risk assessment ^[12].

Information on pregnancy, pregnancy outcomes, and mother-baby linkage will be available in the SIDIAP database.

SIDIAP acknowledges that the data source they have access to includes data on vaccine delivery and registration and undertakes to cooperate on addressing the study objectives by contributing to providing reports based on such data.

9.4.7.1. Vaccine exposure

Approximately 1.7 million doses of the Pfizer-BioNTech COVID-19 vaccine have been administered in Catalonia from 27 December 2020 through 28 April 2021 (<https://dadescovid.cat/>). Based on the National Strategic Vaccination Plan, 600 million doses of the Pfizer-BioNTech COVID-19 vaccine are expected to be administered in Spain. The number of doses for Catalonia will depend on the availability of the vaccine in later stages.

9.5. Study size

The study will be conducted in a source population of 38.9 million individuals captured in the electronic healthcare data sources.

Table 2 shows the sample size calculations for AESI and different risk ratios assumed. As examples, assuming a two-sided alpha = 0.95, power of 80%, and a ratio of 1 to 4 exposed to unexposed, to detect a risk ratio of 3 for vaccine-associated enhanced disease (VAED), we will need to include 1,209 exposed individuals and 4,836 unexposed individuals; and assuming a two-sided alpha = 0.95, power of 80%, and a ratio of 1 to 4 exposed to unexposed, to detect a risk ratio of 2 for Guillain-Barré syndrome, we will need to include 1,700,582 exposed individuals and 6,802,328 unexposed individuals.

Table 2. Number of individuals needed to detect different risk ratios for select AESI^a with a range of background rates

Disease	Background proportion during risk window	Risk ratio	Sample size	
			Exposed	Unexposed
Anaphylaxis	1/40,000	1.5	2,147,100	8,588,400
Anaphylaxis	1/40,000	2	680,233	2,720,932
Anaphylaxis	1/40,000	2.5	365,976	1,463,904
Anaphylaxis	1/40,000	3	241,643	966,572
Guillain-Barré syndrome	1/100,000	1.5	5,367,750	21,471,000
Guillain-Barré syndrome	1/100,000	2	1,700,582	6,802,328
Guillain-Barré syndrome	1/100,000	2.5	914,940	3,659,760
Guillain-Barré syndrome	1/100,000	3	604,106	2,416,424
VAED	1/5,000	1.5	10,736	42,944
VAED	1/5,000	2	3,402	13,608

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 Non-Interventional Study Protocol Template For Secondary Data Collection Study

01-Jun-2020

Page 47

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Table 2. Number of individuals needed to detect different risk ratios for select AESI^a with a range of background rates

Disease	Background proportion during risk window	Risk ratio	Sample size	
			Exposed	Unexposed
VAED	1/5,000	2.5	1,830	7,320
VAED	1/5,000	3	1,209	4,836

AESI = adverse events of special interest; VAED = vaccine-associated enhanced disease.
 a. Assuming a two-sided alpha = 0.95, power of 80%, and a ratio of 1:4 exposed to unexposed.

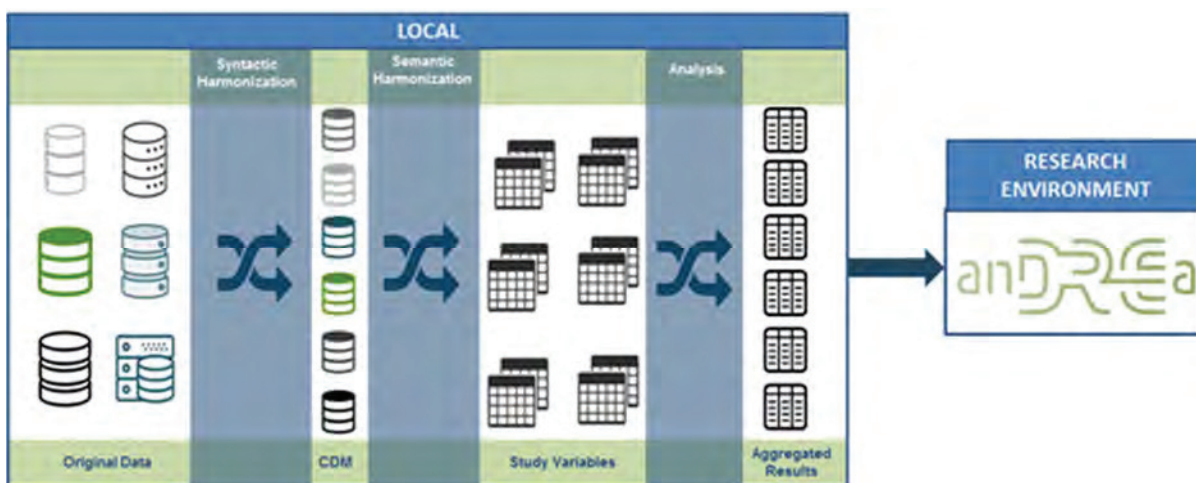
9.6. Data management

This study will be conducted in a distributed manner using a common protocol, common data model (CDM), and common analytics programmes based on existing health data. The following steps will be implemented:

1. Extraction, transformation, and loading (ETL) of data to a CDM. To harmonise the structure of the data sets stored and maintained by each data partner, a shared syntactic foundation is used. The CDM that will be used has been developed during the IMI-ConcePTION project (<https://www.imi-conception.eu/wp-content/uploads/2020/10/ConcePTION-D7.5-Report-on-existing-common-data-models-and-proposals-for-ConcePTION.pdf>). In this CDM, data are represented in a common structure, but the content of the data remain in their original format. The ETL design for each study is shared in a searchable FAIR catalogue. The VAC4EU FAIR data catalogue is a meta-data management tool designed to contain searchable meta-data describing organisations that can provide access to specific data sources. FAIR is defined as findable, accessible, interoperable, and re-usable. Data quality checks will be conducted to measure the integrity of the ETL as well as internal consistency within the context of the CDM (see [Section 9.8](#)).
2. 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 standardised event definition template. The Codemapper tool was used to create diagnosis code lists based on completed event definition templates for each AESI and comorbid risk condition in the ACCESS project. Based on the relevant diagnostic medical codes and keywords, as well as other relevant concepts (e.g., medications), one or more algorithms are constructed (typically one sensitive, or broad, algorithm and one specific, or narrow, algorithm) to operationalise the identification and measurement of each event. These algorithms may differ by database, as the components involved in the study variables may differ. Manual review of electronic records will be conducted for a sample of the events. Specifications for both ETL and semantic harmonisation will be shared in the catalogue.

- Third, following conversion to harmonised study variable sets, R and SAS programs for the calculation of incidence and prevalence will be distributed to data access providers for local deployment. The aggregated results produced by these scripts will then be uploaded to the Digital Research Environment (DRE) for pooled analysis and visualisation (see Figure 3). The DRE is made available through UMCU (University Medical Center Utrecht)/VAC4EU (<https://www.andrea-consortium.org/>). The DRE is a cloud-based, globally available research environment where data are stored and organised securely and where researchers can collaborate (<https://www.andrea-consortium.org/azure-dre/>).

Figure 3. Data management plan



CDM = common data model.

9.6.1. Case report forms (CRFs)/Data collection tools (DCTs)/Electronic data record

This study will use secondary data collected in EHR databases. For the purpose of validating selected study endpoints, special forms will be developed and securely saved in environments assuring data protection and patient confidentiality according to the requirements of each country and database access provider (DAP).

As used in this protocol, the term CRF should be understood to refer to either a paper form or an electronic data record or both, depending on the data collection method used in this study.

A CRF is required and will be completed for each patient who is subject to an event/case verification/validation procedure. The completed original CRFs are the sole property of the DAPs and will not be made available in any form to third parties, except for authorised representatives of Pfizer or appropriate regulatory authorities. The DAPs will ensure that the CRFs are securely stored at the study site in encrypted electronic form and will be password protected to prevent access by unauthorised third parties.

The DAPs have ultimate responsibility for the collection and reporting of all clinical and laboratory data entered on the CRFs for the procedure of event/case verification/validation

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 49

and ensuring that they are accurate, authentic/original, attributable, complete, consistent, legible, timely (contemporaneous), enduring, and available when required. The CRFs must be signed by an authorised staff member of the DAPs to attest that the data contained on the CRFs are true. Any corrections to entries made in the CRFs or source documents must be dated, initialled, and explained (if necessary) and should not obscure the original entry.

The source documents are the hospital's or physician's charts. In these cases, data collected on the CRFs must match those charts.

9.6.2. Record retention

The investigators must obtain Pfizer's written permission before disposing of any records, even if retention requirements have been met.

The final study aggregated results sets and statistical programmes will be archived and stored on the DRE and the VAC4EU SharePoint site. 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 programmes, and output files will be archived on a specific and secured central drive.

It is the responsibility of the principal investigator to inform the other investigators or institutions regarding 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.

To enable evaluations and/or inspections/audits from regulatory authorities or Pfizer, DAPs agree to keep all study-related records, including the identity of all participating patients (sufficient information to link records, e.g., CRFs, hospital records), copies of all CRFs, safety reporting forms, source documents, detailed records of treatment disposition, and adequate documentation of relevant correspondence (e.g., letters, meeting minutes, telephone call reports). The records should be retained by DAPs according to local regulations or as specified in the vendor contract, whichever has a longer retention time. DAPs must ensure that the records continue to be stored securely for as long as they are retained.

If UMCU becomes unable for any reason to continue retaining study records for the required period, Pfizer should be prospectively notified. In this case, the study records must be transferred to a designee acceptable to Pfizer.

Study records must be kept for a minimum of 15 years after completion or discontinuation of the study, unless UMCU and Penta and Pfizer have expressly agreed to a different retention via a separate written agreement. Records must be retained for longer than 15 years if required by applicable local regulations.

UMCU must obtain Pfizer's written permission before disposing of any records, even if retention requirements have been met.

9.6.3. 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 by study statisticians, each DAP will extract the relevant study data locally using their software (e.g., Stata, SAS, R, Oracle). These data will be loaded into the CDM structure in csv format. These data remain local (see [Figure 3](#)).

9.6.4. Data processing and transformation

Data processing and transformation will be conducted using R and SAS code against the syntactically harmonised CDM. The R and SAS scripts will first transform the data in the syntactically harmonised CDM to semantically harmonised study variables (see [Figure 3](#)). Following creation of study variables, the data will be characterised. This characterisation will include calculation of code counts and incidence rates, as well as benchmarking within the data source (over time), between data sources and externally (against published estimates). Subsequently, R and SAS code to conduct analysis against semantically harmonised study variables will be distributed and run locally to produce aggregated results. The R and SAS scripts for these processing and analysis steps will be developed and tested centrally and sent to the DAPs.

The R and SAS scripts are structured in modular form to ensure transparency. Functions to be used in the modules will be either standard R and SAS packages or packages specifically designed, developed, and tested for multidatabase studies. Scripts will be double coded in SAS and R and quality checks will be thoroughly documented.

The DAPs will run the R and SAS code locally and send aggregated analysis results to the DRE using a secure file transfer protocol. In the DRE, results will be further plotted, inspected (for quality assessment), and pooled (if needed) for final reporting.

All final statistical computations will be performed on the DRE using R and/or SAS (SAS Institute; Cary, North Carolina). Data access providers will have access to the workspace for script verification.

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

9.6.5. Data access

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

The DRE architecture allows researchers to use a solution within the boundaries of data management rules and regulations. Although General Data Protection Regulation and Good

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 51

(Clinical) Research Practice still apply to researchers, the DRE offers tools to more easily control and monitor which activities take place within projects.

All researchers who need access to the DRE are granted access to study-specific secure workspaces. Access to this workspace is only possible with double authentication using an identification code and password together with the user's mobile phone for authentication.

Upload of files is possible for all researchers with access to the workspace within the DRE. Download of files is only possible after requesting and receiving permission from a workspace member with an "owner" role.

9.7. Data analysis

Detailed methodology for summary and statistical analyses of data collected in this study will be documented in a SAP, which will be dated, filed, and maintained by the sponsor. The SAP will also provide additional detail regarding the evaluation of a threshold of excess risk for each of the safety events of interest. This will be determined based on background incidences for each event (e.g., based on a historical influenza vaccinated active comparator cohort data to be determined during the study), in addition to prespecified significance level (e.g., $\alpha = 0.01$ or 0.05) and power. All analyses will be conducted using R version R-4.0.3 or higher (Foundation for Statistical Computing, Vienna, Austria; <https://www.R-project.org>) or SAS version 9.3 software or higher (Cary, North Carolina, United States of America (USA); SAS Institute, Inc.).

The SAP will contain more detail of the analysis and data pooling and may modify the plans outlined in the protocol; any major modifications of primary endpoint definitions or their analyses would be reflected in a protocol amendment.

Data extraction, descriptive analyses of AESI that are not pregnancy outcomes, incidence rates, and comparative analyses (when the data exist and are available to support those analyses with sufficient precision) are planned to occur every 6 months during the first 2 years of the study and will be reported in the interim and the final reports. An additional data extraction and analysis will take place at the end of year 3 of follow-up and will include all outcomes, including pregnancy outcomes. Pfizer-BioNTech proposes to include in the interim reports descriptive results.

9.7.1. Cohort design

9.7.1.1. Exposure assignment and follow-up

The main exposure of interest is in the receipt of at least one dose of the Pfizer-BioNTech COVID-19 vaccine. Other exposure groups will also be described. Individuals will be assigned to each vaccination category (see [Section 9.1.1](#)) at time zero, as outlined below.

- A. Vaccination category "Receipt of a first dose of the Pfizer-BioNTech COVID-19 vaccine, followed or not by a second dose": Individuals will be assigned to this exposure category if they receive a first dose of the Pfizer-BioNTech COVID-19 vaccine. Individuals will be censored if and when they receive a non-Pfizer COVID-19 vaccine during follow-up.

PFIZER CONFIDENTIAL

For the matched comparative analysis individuals will be censored when their matched pair is censored.

- B. No vaccination category. Individuals will be assigned to this exposure group if they do not receive a vaccination at time zero (see [Section 9.1.1](#)). Individuals will be censored when they receive a dose of any COVID-19–directed vaccine during follow-up. Time zero in the unexposed group will be a day when they did not receive a Pfizer-BioNTech COVID-19 vaccine dose. This day will be chosen by calendar matching to the time zero of the corresponding exposed group.

Censoring will only apply to the sensitivity analysis for a vaccination category consisting of the receipt of the two vaccination doses per the recommended schedule. For the sensitivity analysis studying the receipt of a full vaccination regimen of the Pfizer-BioNTech COVID-19 vaccine, individuals will be assigned to this exposure category if they receive a first dose of the Pfizer-BioNTech COVID-19 vaccine. Individuals will be censored if they do not receive a second dose of the Pfizer-BioNTech COVID-19 vaccine by week 4 after the first administration, in the absence of an adverse event that contraindicates the second dose. Individuals will also be censored if the second dose of Pfizer-BioNTech COVID-19 vaccine is received within 2 weeks of the first dose or if they receive a non–Pfizer-BioNTech COVID-19 vaccine during follow-up.

Individuals will be followed from time zero (see [Section 9.1.1](#)) until the censoring described above, death, or the administrative end of follow-up, whichever occurs first. For analyses of AESI with known risk windows, follow-up will be truncated at the end of the risk window.

9.7.1.2. Descriptive statistics

The distributions of baseline characteristics at time zero by exposure group will be calculated to describe the study cohort and illustrate differences between the groups. For continuous variables, means, standard deviations, medians, and other quartiles will be estimated. For categorical variables, counts and proportions will be estimated. The missingness of variables will also be described. Interim reports will be limited to descriptive analyses. Comparative analyses will be included in the final report. Further details will be described in the SAP.

To describe the relative imbalance of characteristics between exposed and unexposed groups, absolute standardised differences will be calculated for each baseline characteristic ^[15, 16]. Multilevel categorical variables will calculate an overall standardised difference across all levels ^[18]. The larger the absolute standardised difference values, the greater the imbalance between baseline characteristics. Balance will also be checked after propensity score methods are applied to control for confounding.

9.7.1.3. Description of vaccination categories

The counts and proportions of administered doses with the following characteristics will be reported:

- Receipt of a second dose of the Pfizer-BioNTech COVID-19 vaccine outside the Pfizer-BioNTech recommended COVID-19 vaccination schedule
- Receipt of a dose of the Pfizer-BioNTech COVID-19 vaccine after receipt of a dose of a different vaccine
- Receipt of the Pfizer-BioNTech COVID-19 vaccine in individuals with contraindications to the vaccine or in groups not recommended to receive vaccination (i.e., adults who are not health care workers in January 2021)
- To characterise utilisation patterns of Pfizer-BioNTech COVID-19 vaccine, the absolute and relative frequency of individuals receiving at least one dose of the vaccine and the two-dose vaccine completion rate will be calculated. The distribution of time gaps in weeks between the first and second dose will be described by categories (< 2, 2-4, 5-8, 9-12, 13-18, > 18 weeks) and by median, other quantiles, and minimum and maximum.

9.7.1.4. Crude outcome measures

For safety outcomes with a known short risk window (e.g., anaphylaxis), the risk (number of events/number of vaccinated persons) and the corresponding 95% CIs will be computed baseline. Effect estimates will be calculated both as risk differences and as risk ratios, along with their corresponding 95% CIs.

For safety outcomes with unknown risk windows or those that require long follow-up (e.g., death), the cumulative incidence will be computed, which will be estimated with a 1 – Kaplan-Meier survival curve, as well as with adjusted parametric incidence curves ^[19]. Time to outcome will be defined as the time from the baseline date (time zero) until the occurrence of the outcome or censoring. For individuals without outcomes, the censoring date is defined as the earlier of date of death, censoring of the match pair and the end of follow-up (as described in Section 9.7). The variance will be computed using approaches that account for autocorrelation (e.g., the robust estimator or via bootstrapping). Risk differences and risk ratios (and their corresponding 95% CIs) will be estimated at different time intervals, which can be adapted to the specific nature of each outcome.

Crude risks, cumulative incidence, and measures of association for each AESI after vaccination will be estimated in the entire population. Subgroup analyses will be conducted by subgroups defined by demographic and clinical characteristics, as well as other covariates of interest. In addition, risks or cumulative incidence of AESI in persons vaccinated with Pfizer-BioNTech COVID-19 vaccine will be compared with the expected rates in the absence of vaccination, which will be calculated using unexposed and historical data (i.e., from the ACCESS project).

9.7.1.5. Adjustment for baseline imbalances

Individuals following each vaccination category under study may have different characteristics that may determine their risk of AESI. To account for such potential confounding, propensity score methods will be used to estimate the adjusted risk ratios and 95% CIs. Propensity scores represent the probability of being vaccinated at any calendar time given a set of baseline covariates. More details will be provided in the SAP.

9.7.1.6. Adjustment for adherence to recommended vaccination schedule

The censoring described in Section 9.7.1.1 may introduce selection bias if individuals receiving a non-Pfizer-BioNTech COVID-19 vaccine during the follow-up (the main reason for being censored) are different from those who do not; these differences may be present at baseline (e.g., age, sex) or may become apparent during follow-up (e.g., a reaction after a first dose). To account for such potential selection bias, we will use weighting by inverse probability of censoring. More details will be specified in the SAP. This population with censoring will be described and compared with the population in the main analysis.

9.7.1.7. Meta-analysis

Using the main estimates from each data source, appropriate random-effects meta-analytic methods will be used to obtain a combined effect estimate. The heterogeneity across data sources will be checked, and a forest plot will be produced with the data sources and the pooled estimate.

9.7.2. Self-controlled risk interval

9.7.2.1. Descriptive statistics

The number of cases and incidence rates of each AESI will be reported, overall and by important covariates.

9.7.2.2. Measures of association

Conditional Poisson regression will be used to estimate incidence rate ratios and 95% CIs, and risk differences and 95% CIs will be estimated using an appropriate method. The primary analysis will be pooled across both doses, and the secondary analysis will stratify by dose number to assess potential effect modification. AESI for which the SCRI design will be a complementary design and risk windows for such AESI are described in [Table 1](#). The control period will follow the risk window, will be of the same lengths, and will be detailed in the SAP.

The SCRI inherently adjusts for both measured and unmeasured time constant factors such as sex and chronic health conditions with onset before the start of follow-up. Time-varying confounders may be included as covariates in regression models.

Subgroup analyses will be conducted by subgroups defined by demographic and clinical characteristics, dose number, and other covariates of interest.

9.8. Quality control

Rigorous QC will be applied to all deliverables. Data transformation into the CDM will be conducted by each subcontracted research partner in its associated database, with processes as described in the following corresponding sections. Standard operating procedures or internal process guidance at each research centre will be used to guide the conduct of the study. These procedures include rules for secure and confidential data storage, backup, and recovery; methods to maintain and archive project documents; QC procedures for programming; standards for writing analysis plans; and requirements for scientific review by senior staff.

At UMCU, as the scientific coordinating centre responsible for central data management and analysis, all documents undergo QC review and senior scientific review. Data management and statistical analysis follow standard operating procedures. All statistical analysis programmes will be double coded.

At RTI Health Solutions (RTI-HS), as the project coordinating centre and scientific coleader centre, all key study documents will undergo QC review, senior scientific review, and editorial review. Senior reviewers with expertise in the appropriate subject matter area will provide advice on the design of research study approaches and the conduct of the study and will review results, reports, and other key study documents.

9.8.1. PHARMO (NL)

PHARMO adheres to high standards throughout the research process based on robust methodologies, transparency, and scientific independence. PHARMO conducts studies in accordance with the ENCePP *Guide on Methodological Standards in Pharmacoepidemiology* [20] and the ENCePP *Code of Conduct* [21]. PHARMO is ISO 9001:2015 certified. Standard operating procedures, work instructions, and checklists are used to guide the conduct of a study. These procedures and documents include internal quality audits, rules for secure and confidential data storage, methods to maintain and archive project documents, rules and procedures for execution and QC of SAS programming, standards for writing protocols and reports, and requirements for senior scientific review of key study documents.

9.8.2. ARS Toscana (IT)

One or two researchers will review study documents. ARS Toscana receives data on a bimonthly basis from the Tuscany region (where it undergoes a first QC); the ARS Toscana statistical office appends it to an Oracle database and checks it using a dashboard to identify any inconsistencies with historical data.

The Pharmacoepi Unit has standardised parametric procedures in Structured Query Language (SQL) and Stata to extract data from the Oracle database. Parametric procedures are also available to convert the data into various CDMs. Study-specific procedures are developed, based on the study protocol and/or SAP, as well as by composing standard parametric procedures in Stata. Standard procedures in R are currently under development in the context

of the ConcePTION project. The Unit also regularly generates simulated data sets and double programming in R programmes that are originally developed in SAS or Stata.

9.8.3. Pedianet/HSD (IT)

Pedianet data processing includes, in addition to standardised procedures in SQL and Microsoft Access to extract data from database, QC steps aimed at verifying the correspondence between a diagnostic code and its open-text descriptor that are conducted through manual validation of clinical histories. Quality control checks of patient general data are conducted through the detection of outlier values and validation rules; grouping of diseases; and regular monitoring of aggregate clinical and drug data. All transformations in the data are logged in R scripts. To ensure code reliability, double programming in R and Stata or Python is in place for all scripts.

HSD data processing includes, in addition to standardised procedures in SQL and Access to extract data from a database, QC steps aimed at verifying the correspondence between a diagnostic code and its open-text descriptor, which are conducted through manual validation of clinical histories. Quality control of patient general data is conducted through the detection of outlier values and validation rules, grouping of diseases, and regular monitoring of aggregate clinical and drug data. All transformations in the data are logged in SQL scripts through version control. Furthermore, to ensure code reliability, double programming in Stata is in place for all scripts.

9.8.4. EpiChron - Aragon data sources (ES)

The data QC process in Aragon is conducted in three steps (i.e., data collection, data request and extraction, and data processing). Common data collection software and procedures guarantee standardised data input by all health care professionals. In the case of the hospital CMBD register and the drug dispensation database, their completion is systematic, uniform, and normative according to legal orders. Online specific training and chart documentation on the use of EHR software is regularly provided to physicians and nurses in Aragon. The data contained in each of the registers is routed to a specific service of the Department of Health, which performs a pseudonymisation of the data to encrypt individual-level identification codes, protecting individuals' privacy and complying with data protection laws. This new encrypted code is applied in all registers, enabling the linkage of data at the patient level. The resulting databases are stored on a central computer server, and access to the files is restricted to members of the research group by a double-entry password. The research group is a multidisciplinary qualified team including public health specialists, epidemiologists, clinicians, pharmacists, statisticians, and data managers; they are all trained in data management and patient data protection. Given that original databases are in different formats (e.g., Microsoft Access, Microsoft Excel, plain text), the SQL programming language is employed to extract the data. Stata statistical software (Release 12) is used for data processing, which includes a number of systematic steps aimed at improving the quality, accuracy, and reliability of the data for research purposes (e.g., QC of diagnoses to verify the correspondence between a diagnostic code and its open-text descriptor through manual validation of clinical histories and use of specific algorithms to search for specific key words

or roots of words in open-text fields, QC of patient general data through the detection of outlier values and validation rules, grouping of diseases, and regular monitoring of aggregate clinical and drug data). The original databases also have their own QC processes. All changes conducted in the data are logged in Stata scripts, which are continuously revised and updated given the dynamic nature of the data processing.

9.8.5. CPRD (UK)

The DSRU has information security policies in place to preserve the confidentiality, integrity and availability of the organisation's systems and data. These include ensuring that the premises provide suitable physical and environmental security, all equipment is secure and protected against malicious software, the network can be accessed only by authorised staff, telecommunication lines to the premises are protected from interception by being routed overhead or underground, and personnel receive training regarding security awareness. The study will be conducted according to the *Guidelines for Good Pharmacoepidemiology Practices (GPP)* [22] and according to the *ENCePP Code of Conduct* [21]. Data quality is a high priority at the DSRU and is assured through a number of methods based on staff training, validated systems, error prevention, data monitoring, data cleaning, and documentation, including the following:

- Staff training on data processing standard operating procedures
- Data management plan for every research study outlining the legal basis for data collection, data flows, data access rights, data retention periods, etc.
- Routine data cleaning to screen for errors, missing values, and extreme values and diagnose their cause
- System process logs to document staff access, etc.

9.8.6. SIDIAP (ES)

Data quality processes are implemented at each phase of the data flow cycle. Quality control checks are performed at the extraction and uploading steps. To assess data completeness the elements presence are described by geographical areas, registering physician, time and the distribution function of values. Correctness is assessed by validity checks on outliers, out of range values, formatting errors and logical dates incompatibilities. Completeness and correctness measures are used to inform decisions on the required transformations to improve data quality (e.g., harmonisation, normalisation, and clean-up) and the data fitness for the purpose of specific research projects.

9.9. Limitations of the research methods

This study is subject to limitations related to both the study design and use of secondary health care data.

A data-related limitation of this study is the reliance on the accuracy of codes and algorithms to identify outcomes. Outcomes and their dates of occurrence will be validated, but the extent of validation may be limited because of the use of medical records. Exposure identification may be based on pharmacy dispensing records, general practice records, immunisation registers, medical records, or other secondary data sources. The ability to identify specific COVID-19 vaccine products and dates of vaccination in these data sources is reflected in Section 9.3.1. The vaccine roll-out is just underway in some countries and has yet to begin in others. It is possible that individuals vaccinated outside the health care system will not be recorded in secondary EHR databases, thereby leading to potential bias because of exposure misclassification with the cohort design. Furthermore, it is unknown the extent to which vaccine brands and batch numbers/lot numbers will be available in the secondary data sources. It is also possible that some AESI are the result of immunisation errors occurring during the administration of the Pfizer-BioNTech COVID-19 vaccine. This information is not collected regularly and will not be able to be taken into account with the current protocol.

In some databases, the mother-baby link may not be available. This is the case for the CPRD Aurum database, while there are Pregnancy Register linked data sets available for the CPRD GOLD database. It is expected that, in the near future, the CPRD GOLD database will be able to contribute to the overall CPRD study.

A study design-related limitation of both the cohort and SCRI designs is that any uncertainty regarding risk periods will lead to misclassification and attenuation of risk estimates. A limitation of the cohort design is the potential for residual or unmeasured confounding, as it is unlikely that the data sources will have information on all potential confounders. To address potential confounding, the SCRI, which automatically adjusts for time-invariant confounders, will be used as a secondary approach. However, the SCRI is not well suited to study outcomes with gradual onset, long latency, or risk periods that are not well known. It also may be subject to bias for outcomes that affect the probability of exposure. The SCRI design will be complementary to the cohort design for prespecified AESI with defined risk intervals. In Italy, the COVID-19 vaccination campaign started in December 2020. Each region ($n = 20$) has adopted different vaccination strategies involving hubs and/or general practices.

The primary care setting was actively involved in the vaccination campaign only at the beginning of April 2021, and only certain age categories and/or type of vaccines were available for direct administration by GPs. Thus, for the period between January and March 2021, Italian GPs have likely recorded vaccine injections according to three main pathways: a) some regions automatically informed GPs regarding their patients' COVID-19 vaccination status; b) GPs referred patients to a specific hub to register their vaccination status there; and c) patients autonomously reported their vaccination to their GPs. For the first semester of 2021, HSD expects to find complete data for certain age categories, while in the first trimester and for some other age categories, we could only find incomplete data for some regions. In HSD, after preliminary evaluation of data completeness, the study design (e.g., self-controlled or cohort design) will be chosen for the specific objectives.

The main analysis for both the cohort and SCRI analysis will pool together the population used to estimate the effect of a first dose of the Pfizer-BioNTech COVID-19 vaccine and the population used to estimate the effect of a second dose of such vaccine. This pooling is done to gain statistical precision, under the assumption that the effect of a first or second dose in both populations is homogeneous. If this assumption is inaccurate, e.g., because receiving a first dose sensitises the immune system to react against a second dose, the estimates of the main analysis will be biased.

9.10. Other aspects

Not applicable

10. PROTECTION OF HUMAN SUBJECTS

This is a non-interventional study using secondary data collection and does not pose any risks for individuals. Each data source research partner will apply for an independent ethics committee review according to local regulations.

Data protection and privacy regulations will be observed in collecting, forwarding, processing, and storing data from study participants.

10.1. Patient information

This study mainly involves data that exist in anonymised structured format and contain no patient personal information.

All parties will comply with all applicable laws, including laws regarding the implementation of organisational and technical measures to ensure protection of patient personal data. Such measures will include omitting patient names or other directly identifiable data in any reports, publications, or other disclosures, except where required by applicable laws.

Patient personal data will be stored at DAPs in encrypted electronic form and will be password protected to ensure that only authorised study staff have access.

DAPs will implement appropriate technical and organisational measures to ensure that personal data can be recovered in the event of disaster. In the event of a potential personal data breach, DAPs shall be responsible for determining whether a personal data breach has in fact occurred and, if so, providing breach notifications as required by law.

To protect the rights and freedoms of natural persons with regard to the processing of personal data, when study data are compiled for transfer to Pfizer and other authorised parties, any patient names will be removed and will be replaced by a single, specific, numerical code. All other identifiable data transferred to Pfizer or other authorised parties will be identified by this single, patient-specific code. In the case of data transfer, Pfizer will maintain high standards of confidentiality and protection of individuals' personal data consistent with the vendor contract and applicable privacy laws.

10.2. Patient consent

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

10.3. Institutional review board (IRB)/Independent ethics committee (IEC)

Each DAP will be following the local country and data custodian requirements to apply for access to the data. At the coordinating centre, RTI-HS will ask approval for exemption from review by the RTI International institutional review board. All correspondence with the institutional review board or independent ethics committee and applicable documentation will be retained as part of the study materials.

10.4. Ethical conduct of the study

This study will adhere to the *Guidelines for Good Pharmacoepidemiology Practices (GPP)* [22] and has been designed in line with the *ENCePP Guide on Methodological Standards in Pharmacoepidemiology* [20]. The *ENCePP Checklist for Study Protocols* [23] will be completed (see ANNEX 2).

The study is a post-authorisation study of vaccine safety and will comply with the definition of the non-interventional (observational) study referred to in the International Conference on Harmonisation tripartite guideline *Pharmacovigilance Planning E2E* [24] and provided in the *EMA Guideline on Good Pharmacovigilance Practices (GVP) Module VIII: Post-Authorisation Safety Studies* [25], and with the 2012 EU pharmacovigilance legislation, adopted 19 June 2012 [26].

The study will be registered in the EU PAS Register [27] before data collection commences.

The research team and study sponsor should adhere to the general principles of transparency and independence in the *ENCePP Code of Conduct* [21] and the *ADVANCE Code of Conduct* [25].

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

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

This study involves a combination of existing structured data and unstructured data, which will be converted to structured form during the implementation of the protocol solely by a computer using automated/algorithmic methods, such as natural language processing.

In these data sources, it is not possible to link (i.e., identify a potential association between) a particular product and medical event for any individual. Thus, the minimum criteria for reporting an adverse event (AE) (i.e., identifiable patient, identifiable reporter, a suspect product, and event) cannot be met.

For non-interventional study designs that are based on secondary use of data, such as studies based on medical chart reviews or EHRs, systematic reviews, or meta-analyses, reporting of adverse events/adverse drug reactions is not required. Reports of AEs/adverse drug reactions should only be summarised in the study report, where applicable.

According to the EMA *Guideline on Good Pharmacovigilance Practices (GVP), Module VI – Management and Reporting of Adverse Reactions to Medicinal Products* ^[29],

“All adverse events/reactions collected as part of [non-interventional post-authorisation studies with a design based on secondary use of data], the submission of suspected adverse reactions in the form of [individual case safety reports] is not required. All adverse events/reactions collected for the study should be recorded and summarised in the interim safety analysis and in the final study report”.

Module VIII – Post-Authorisation Safety Studies ^[25] echoes this approach. Legislation in the EU further states that for certain study designs such as retrospective cohort studies, particularly those involving EHRs, it may not be feasible to make a causality assessment at the individual case level.

This study protocol requires human review of patient-level unstructured data; unstructured data refer to verbatim medical data, including text-based descriptions and visual depictions of medical information, such as medical records, images of physician notes, neurological scans, x-rays, or narrative fields in a database. The reviewer is obligated to report AEs with explicit attribution to any Pfizer drug that appears in the reviewed information (defined per the patient population and study period specified in the protocol). Explicit attribution is not inferred by a temporal relationship between drug administration and an AE but must be based on a definite statement of causality by a health care provider linking drug administration to the AE.

The requirements for reporting safety events on the non-interventional study (NIS) adverse event monitoring (AEM) Report Form to Pfizer Safety are as follows:

- All serious and non-serious AEs with explicit attribution to **any Pfizer drug** that appear in the reviewed information must be recorded on the data collection tool (e.g., chart abstraction form) and reported, within 24 hours of awareness, to Pfizer Safety using the NIS AEM Report Form.
- Scenarios involving drug exposure, including exposure during pregnancy, exposure during breast feeding, medication error, overdose, misuse, extravasation, lack of efficacy,

and occupational exposure associated with the use of a Pfizer product, must be reported, within 24 hours of awareness, to Pfizer Safety using the NIS AEM Report Form.

For AEs with an explicit attribution or scenarios involving exposure to a Pfizer product, the safety information identified in the unstructured data reviewed is captured in the Event Narrative section of the report form, and constitutes all clinical information known regarding these AEs. No follow-up on related AEs will be conducted.

All the demographic fields on the NIS AEM Report Form may not necessarily be completed, as the form designates, as not all elements will be available due to privacy concerns with the use of secondary data sources. While not all demographic fields will be completed, at the very least, at least one patient identifier (e.g., gender, age as captured in the narrative field of the form) will be reported on the NIS AEM Report Form, thus allowing the report to be considered valid in accordance with pharmacovigilance legislation. All identifiers will be limited to generalities, such as the statement, “A 35-year-old female...” or “An elderly male...” Other identifiers will have been removed.

Additionally, the onset/start dates and stop dates for “Illness”, “Study Drug”, and “Drug Name” may be documented in month/year (mmm/yyyy) format rather than identifying the actual date of occurrence within the month /year of occurrence in the day/month/year (DD/MMM/YYYY) format.

All research staff members must complete the following Pfizer training requirements:

- “YRR Training for Vendors Working on Pfizer Studies (excluding interventional clinical studies and non-interventional primary data collection studies with sites/investigators)”.

These trainings must be completed by research staff members before the start of data collection. All trainings include a “Confirmation of Training Certificate” (for signature by the trainee) as a record of completion of the training, which must be kept in a retrievable format. Copies of all signed training certificates must be provided to Pfizer.

Re-training must be completed on an annual basis using the most current Your Reporting Responsibilities training materials.

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

As per EMA GVP Module VIII, the study and its protocol will be registered in the EU PAS Register prior to the start of data collection.

Results of analyses and interpretation will be delivered in report form.

- The first report will be a progress report describing the status of the study at each study site, including ethical or related approvals, and results for the population covered by ARS Toscana, Italy.

- Each of the following six interim analysis will be presented as interim reports, to be delivered every 6 months. Pfizer-BioNTech proposes to include in the interim reports descriptive results, incidence rates, and comparative analyses when the data exist and are accessible to support those analyses with sufficient precision.
- At the end of the third year of follow-up, the final report will be produced, including the analysis and interpretation of each outcome, including pregnancy outcomes.

Study results will be published following guidelines, including those for authorship, established by the ICMJE ^[30]. When reporting results of this study, the appropriate Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist will be followed ^[31]. Independent publication rights will be granted to the research team in line with Section VIII.B.5., Publication of study results, of the EMA *Guideline on Good Pharmacovigilance Practices (GVP) Module VIII: Post-Authorisation Safety Studies* ^[25].

Upon study completion and finalisation of the study report, the results of this PASS will be submitted for publication, preferably in a relevant peer-reviewed journal and posted in the EU PAS Register.

Communication via other appropriate scientific venues will be considered.

In the event of any prohibition or restriction imposed (e.g., clinical hold) by an applicable competent authority in any area of the world, or if the investigator party responsible for collecting data from the participant is aware of any new information that might influence the evaluation of the benefits and risks of a Pfizer product, Pfizer should be informed immediately.

13. REFERENCES

1. Pfizer-BioNTech. Fact sheet for healthcare providers administering vaccine (vaccination providers); emergency use authorisation (EUA) of the Pfizer-BioNTech COVID-19 vaccine to prevent coronavirus disease 2019 (COVID-19). 2020. Available at: <https://www.fda.gov/media/144413/download>. Accessed 17 December 2020.
2. ECDC. Overview of COVID-19 vaccination strategies and vaccine deployment plans in the EU/EEA and the UK. Stockholm: European Centre for Disease Prevention and Control. 2 December 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/Overview-of-EU_EEA-UK-vaccination-deployment-plans.pdf. Accessed 17 December 2020.
3. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016 Nov;79:70-5.
4. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016 Apr 15;183(8):758-64.

5. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol*. 2017 Jun;32(6):495-500.
6. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004 Sep;15(5):615-25.
7. Suissa S. Immortal time bias in observational studies of drug effects. *Pharmacoepidemiol Drug Saf*. 2007 Mar;16(3):241-9.
8. Gini R, Dodd CN, Bollaerts K, Bartolini C, Roberto G, Huerta-Alvarez C, et al. Quantifying outcome misclassification in multi-database studies: the case study of pertussis in the ADVANCE project. *Vaccine*. 2020 Dec 22;38 Suppl 2:B56-b64.
9. Kuiper JG, Bakker M, Penning-van Beest FJA, Herings RMC. Existing data sources for clinical epidemiology: the PHARMO Database Network. *Clin Epidemiol*. 2020;12:415-22.
10. Sturkenboom M, Braeye T, van der Aa L, Danieli G, Dodd C, Duarte-Salles T, 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 Dec 22;38:B8-B21.
11. Prados-Torres A, Poblador-Plou B, Gimeno-Miguel A, Calderón-Larrañaga A, Poncel-Falcó A, Gimeno-Feliú LA, et al. Cohort profile: the epidemiology of chronic diseases and multimorbidity. The EpiChron Cohort study. *Int J Epidemiol*. 2018;47(2):382-4f.
12. Poblador-Plou B, Carmona-Pérez J, Ioakeim-Skoufa I, Poncel-Falcó A, Bliiek-Bueno K, Cano-Del Pozo M, et al. Baseline chronic comorbidity and mortality in laboratory-confirmed COVID-19 cases: results from the PRECOVID study in Spain. *Int J Environ Res Public Health*. 2020 Jul 17;17(14).
13. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK primary care data resource. *Ther Adv Drug Saf*. 2012 Apr;3(2):89-99.
14. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Revision 3. International Society for Pharmacoepidemiology. June 2015. Available at: <https://www.pharmacoepi.org/resources/policies/guidelines-08027/>. Accessed 20 January 2021.
15. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009;28(25):3083-107.

16. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. Cary, NC: SAS Institute Inc. 2012. Available at: <https://support.sas.com/resources/papers/proceedings12/335-2012.pdf>. Accessed 22 October 2020.
17. Hernan MA, Robins JM. Causal survival analysis. In: Causal inference. Boca Raton: Chapman & Hall/CRC; 2020. Available at: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>. Accessed 22 October 2020.
18. ENCePP. Guide on methodological standards in pharmacoepidemiology (EMA/95098/2010 Rev. 7). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. July 2018. Available at: http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml. Accessed 22 October 2020.
19. ENCePP. The ENCePP code of conduct for scientific independence and transparency in the conduct of pharmacoepidemiological and pharmacovigilance studies (Revision 4). 15 March 2018. Available at: http://www.encepp.eu/code_of_conduct/. Accessed 17 January 2020.
20. ENCePP. ENCePP checklist for study protocols (revision 4). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. 15 October 2018. Available at: http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml. Accessed 22 October 2020.
21. ICH. Pharmacovigilance planning. E2E. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 2004. Available at: https://database.ich.org/sites/default/files/E2E_Guideline.pdf. Accessed 22 October 2020.
22. EMA. Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies (EMA/813938/2011 Rev 3). European Medicines Agency. 13 October 2017. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf. Accessed 17 January 2020.
23. European Commission. Commission implementing Regulation (EU) No 520/2012 of 19 June 2012 on the performance of pharmacovigilance activities provided for in Regulation (EC) No 726/2004 of the European Parliament and of the Council and Directive 2001/83/EC of the European Parliament and of the Council. 20 June 2012. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:159:0005:0025:EN:PDF>. Accessed 22 October 2020.
24. ENCePP. The European Union electronic register of post-authorisation studies (EU PAS Register). European Network of Centres for Pharmacoepidemiology and

- Pharmacovigilance. 20 December 2018. Available at:
http://www.encepp.eu/encepp_studies/indexRegister.shtml. Accessed 22 October 2020.
25. Kurz X, Bauchau V, Mahy P, Glismann S, van der Aa LM, Simondon F, et al. The ADVANCE Code of Conduct for collaborative vaccine studies. *Vaccine*. 2017 Apr 4;35(15):1844-55.
26. ENCePP. The ENCePP seal. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. 08 September 2020. Available at:
http://www.encepp.eu/encepp_studies/index.shtml. Accessed 22 October 2020.
27. EMA. Guideline on good pharmacovigilance practices (GVP). Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). European Medicines Agency. 22 November 2017. Available at:
https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf. Accessed 16 June 2020.
28. ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors. December 2019. Available at: <http://www.icmje.org/recommendations/>. Accessed 22 October 2020.
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008 Apr;61(4):344-9.
30. Yih WK, Lee GM, Lieu TA, et al. Surveillance for adverse events following receipt of pandemic 2009 H1N1 vaccine in the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) System, 2009-2010. *Am J Epidemiol*. 2012;175(11):1120-1128.
31. Liu CH, Yeh YC, Huang WT, Chie WC, Chan KA. Assessment of pre-specified adverse events following varicella vaccine: A population-based self-controlled risk interval study. *Vaccine*. 2020;38(11):2495-2502.

14. LIST OF TABLES

Table 1. List of Selected Adverse Events of Special Interest.....28
 Table 2. Number of individuals needed to detect different risk ratios for select AESI^a with a range of background rates.....47

15. LIST OF FIGURES

Figure 1. Self-controlled risk interval design.....23
 Figure 2. Study period and follow-up periods25
 Figure 3. Data management plan49

ANNEX 1. LIST OF STAND ALONE DOCUMENTS

None

ANNEX 2. ENCEPP CHECKLIST FOR STUDY PROTOCOLS

ENCePP Checklist for Study Protocols (Revision 4)

Study title: Post Conditional Approval Active Surveillance Study Among Individuals in Europe Receiving the Pfizer-BioNTech Coronavirus Disease 2019 (COVID-19) Vaccine

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	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
1.1.2 End of data collection ²	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
1.1.3 Progress report(s)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
1.1.4 Interim report(s)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
1.1.5 Registration in the EU PAS Register	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6
1.1.6 Final report of study results	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6

Comments:

¹ Date from which information on the first study is first recorded in the study data set or, in the case of secondary use of data, the date from which data extraction starts.

² Date from which the analytical data set is completely available.

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Section 2: Research question		Yes	No	N/A	Section Number
2.1	Does the formulation of the research question and objectives clearly explain:	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8
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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	7
2.1.2	The objective(s) of the study?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8.1
2.1.3	The target population? (i.e., population or subgroup to whom the study results are intended to be generalised)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8.1
2.1.4	Which hypothesis(-es) is (are) to be tested?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	8.1
2.1.5	If applicable, that there is no <i>a priori</i> hypothesis?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Comments:

--

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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.1
3.2	Does the protocol specify whether the study is based on primary, secondary or combined data collection?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.1, 9.6, 9.6.4
3.3	Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.1
3.4	Does the protocol specify measure(s) of association? (e.g., relative risk, odds ratio, excess risk, incidence rate ratio, hazard ratio, number needed to harm [NNH])	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.2.2
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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	11

Comments:

--

Section 4: Source and study populations		Yes	No	N/A	Section Number
4.1	Is the source population described?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.1
4.2	Is the planned study population defined in terms of:				
4.2.1	Study time period	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.2
4.2.2	Age and sex	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.3.1
4.2.3	Country of origin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.1
4.2.4	Disease/indication	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.3.1
4.2.5	Duration of follow-up	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.3.1
4.3	Does the protocol define how the study population will be sampled from the source population? (e.g., event or inclusion/exclusion criteria)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.2.3, 9.2.4

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 *Non-Interventional Study Protocol Template For Secondary Data Collection Study*

01-Jun-2020

Page 69

Comments:

--

<u>Section 5: Exposure definition and measurement</u>		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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.1
5.2	Does the protocol address the validity of the exposure measurement? (e.g., precision, accuracy, use of validation sub-study)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
5.3	Is exposure categorised according to time windows?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2.1
5.4	Is intensity of exposure addressed? (e.g., dose, duration)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.1, 9.7.1.3
5.5	Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.4
5.6	Is (are) an appropriate comparator(s) identified?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.1.1, 9.3.1.2

Comments:

--

<u>Section 6: Outcome definition and measurement</u>		Yes	No	N/A	Section Number
6.1	Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2
6.2	Does the protocol describe how the outcomes are defined and measured?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2.1
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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2.1.1
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)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

Comments:

--

<u>Section 7: Bias</u>		Yes	No	N/A	Section Number
7.1	Does the protocol address ways to measure confounding? (e.g., confounding by indication)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.2
7.2	Does the protocol address selection bias? (e.g., healthy user/adherer bias)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.6
7.3	Does the protocol address information bias? (e.g., misclassification of exposure and outcomes, time-related bias)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2.1.1

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 Non-Interventional Study Protocol Template For Secondary Data Collection Study

01-Jun-2020

Page 70

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Comments:

--

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, subgroup analyses, anticipated direction of effect)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.3, 9.7.1.4, 9.7.2.2

Comments:

--

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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.1
9.1.2	Outcomes? (e.g., clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2
9.1.3	Covariates and other characteristics?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.3
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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.4
9.2.2	Outcomes? (e.g., date of occurrence, multiple event, severity measures related to event)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.4
9.2.3	Covariates and other characteristics? (e.g., age, sex, clinical and drug use history, comorbidity, co-medications, lifestyle)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.1
9.3.2	Outcomes? (e.g., International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2
9.3.3	Covariates and other characteristics?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.4
9.4	Is a linkage method between data sources described? (e.g., based on a unique identifier or other)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	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?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7
10.2	Is study size and/or statistical precision estimated?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.5

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

PFIZER CONFIDENTIAL

CT24-WI-GL02-RF02 2.0 Non-Interventional Study Protocol Template For Secondary Data Collection Study

01-Jun-2020

Page 71

Section 10: Analysis plan	Yes	No	N/A	Section Number
10.3 Are descriptive analyses included?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.2, 9.7.2.1
10.4 Are stratified analyses included?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.2.2
10.5 Does the plan describe methods for analytic control of confounding?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.5, 9.7.1.6, 9.7.2.2
10.6 Does the plan describe methods for analytic control of outcome misclassification?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.2.1.1
10.7 Does the plan describe methods for handling missing data?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.7.1.2
10.8 Are relevant sensitivity analyses described?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.3.1.1, 9.3.1.2

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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.6.3, 9.6.5
11.2 Are methods of quality assurance described?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.8
11.3 Is there a system in place for independent review of study results?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10.4

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?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.9
12.1.2 Information bias?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.9
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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9.9
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)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Comments:

--

090177e19714f608Approved On: 20-May-2021 15:15 (GMT)

Section 13: Ethical issues	Yes	No	N/A	Section Number
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10.3
13.2 Has any outcome of an ethical review procedure been addressed?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
13.3 Have data protection requirements been described?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10.1

Comments:

Section 14: Amendments and deviations	Yes	No	N/A	Section Number
14.1 Does the protocol include a section to document amendments and deviations?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5

Comments:

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)?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12
15.2 Are plans described for disseminating study results externally, including publication?	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	12

Comments:

Name of the main author of the protocol: Alejandro Arana

Date: 24/02/2021

Signature: To be signed upon PRAC endorsement

ANNEX 3. ADDITIONAL INFORMATION

Not applicable

090177e19714f608\Approved\Approved On: 20-May-2021 15:15 (GMT)

Document Approval Record

Document Name: C4591021_PROTOCOL_20MAY2021

Document Title: C4591021_PROTOCOL_20MAY2021

Signed By:	Date(GMT)	Signing Capacity
Campbell, Ulka	20-May-2021 15:12:02	Final Approval
De Bernardi, Barbara	20-May-2021 15:15:15	EUQPPV Approval