PASS Protocol	
Active substance	AZD1222
Product reference	005675
Version number	V 3.0
Date	07 July 2021

Post-Authorisation Safety Study of AZD1222

A post-authorisation/post-marketing observational study to evaluate the association between exposure to AZD1222 and safety concerns using existing secondary health data sources

Marketing Authorisation Holder(s)

Marketing authorisation holder(s)	***AstraZeneca AB*** *151 36 Södertälje, Sweden*
MAH contact person	[redacted] [redacted] AstraZeneca Academy House 136 Hill Road Cambridge CB2 8PA United Kingdom

Approved by:

[redacted]

Cristina Rebordosa, MD, PhD Date
Director, Epidemiology
RTI Health Solutions

Approved by:

PASS INFORMATION

Title	A post-authorisation/post-marketing observational study using existing secondary health data sources to evaluate the association between exposure to AZD1222 and safety concerns
Protocol version identifier	2.1
Date of last version of protocol	07 July 2021
EU PAS Register number	Study will be registered prior to start of data collection
Active substance	ChAdOx1-S [recombinant] (AZD1222) (formerly ChAdOx1 nCoV-19)
Medicinal product	COVID-19 Vaccine AstraZeneca
Product reference	005675
Procedure number	Not applicable
Marketing authorisation holder(s)	AstraZeneca AB
Joint PASS	No
Research question and objectives	This study will seek to evaluate the incidence and relative risk of safety concerns and adverse events of special interest following immunisation with AZD1222 in the real-world setting
Country (-ies) of study	Italy, The Netherlands, Spain, and UK

Author(s)	Cristina Rebordosa, MD, PhD RTI Health Solutions Av. Diagonal 605, 9-1 08028 Barcelona SPAIN Telephone: +34 93 362 2807 Email: crebordosa@rti.org [redacted] Epidemiology AstraZeneca AB [redacted] Telephone: [redacted] Email: [redacted] On behalf of the AZD1222 PASS research team

1. TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

2. LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ABUCASIS	Ambulatory Medical Record in VID
ACCESS	vACCine covid-19 monitoring readinESS
AE	adverse event
AED	Accident and Emergency Department
AESI	adverse event of special interest
ARD	absolute risk difference
ARDS	acute respiratory distress syndrome
ARS	Agenzia regionale di sanità della Toscana (regional health agency of Tuscany)
ATC	Anatomical Therapeutic Chemical
CDM	common data model
CHESS	COVID-19 Hospitalisation in England Surveillance System
CI	confidence interval
COVID-19	coronavirus disease 2019
CPRD	Clinical Practice Research Datalink
DAP	database access provider
DRE	Digital Research Environment
DSRU	Drug Safety Research Unit
EMA	European Medicines Agency
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ETL	extraction, transformation, and loading
EU	European Union
FISABIO	Foundation for the Promotion of Health and Biomedical Research of Valencia Region
GAIA	Pharmaceutical module in VID
GDPR	General Data Protection Regulation
GP	general practitioner
GPP	Good Pharmacoepidemiology Practices
HES	Hospital Episodes Statistics
HR	hazard ratio
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification
ICD-10	International Classification of Diseases, 10th Revision
ICD-10-ES	International Classification of Diseases, 10th Revision, Spanish Edition
ICNARC	Intensive Care National Audit and Research Centre
IDIAP Jordi Gol	Institut Universitari D'Investigació en Atenció Primària Jordi Gol

Abbreviation or special term	Explanation
IR	incidence rate
IRR	incidence rate ratio
ISPE	International Society for Pharmacoepidemiology
MAH	Marketing Authorisation Holder
MBDS	Minimum Basic Data Set at Hospital Discharge
MHRA	Medicines and Healthcare products Regulatory Agency
NHS	National Health Service
NNH	number needed to harm
ONS	Office for National Statistics
ORION	Hospital Medical Record in VID
PASS	post-authorisation safety study
PHE	Public Health England
PPRN	PHARMO Perinatal Research Network
PS	propensity score
PY	person-years
QC	quality control
RMP	risk management plans
RR	risk ratio
RTI-HS	RTI Health Solutions
SAP	statistical analysis plan
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SAS	Statistical Analysis System
SCRI	self-controlled risk interval
SGSS	Second Generation Surveillance System
SIA	Ambulatory Information System
SIDIAP	Information System for Research in Primary Care
SIP	Population Information System
TBD	to be determined
TTS	thrombosis with thrombocytopenia syndrome
UMCU	University Medical Center Utrecht
UK	United Kingdom
VAC4EU	Vaccine monitoring Collaboration for Europe
VID	Valencia Health System Integrated Database
VIS	Vaccine Information System
VTE	venous thromboembolism

3. RESPONSIBLE PARTIES

AstraZeneca Responsible Parties

Role	Name, title, qualifications
Scientific lead, epidemiology	[redacted]
Operational lead	[redacted]
Study physician	[redacted]
Statistician	[redacted]
Pharmacovigilance	[redacted]
Pharmacovigilance	[redacted]
Pharmacovigilance	[redacted]

RTI Responsible Parties

Role	Name, title, qualifications
Principal investigator	Cristina Rebordosa, MD, PhD Director, Epidemiology
Co-principal investigator	[redacted]
Senior adviser	[redacted]
Statistician	[redacted]

Collaborating Institutions (proposed)[a]	Role	Name, title, qualifications
Drug Safety Research Unit (DSRU)	Principal Investigator	[redacted]
	Epidemiologist	[redacted]
	Epidemiologist	[redacted]

Collaborating Institutions (proposed)[a]	**Role**	**Name, title, qualifications**
The Foundation for the Promotion of Health and Biomedical Research of Valencia Region, (FISABIO)	Project Lead	[redacted]
	Project Statistician	[redacted]
	Project Statistician	[redacted]
	Pharmacovigilance	[redacted]
	Project Manager	[redacted]
ARS Toscana	Data scientist, Pharmacoepidemiologist	[redacted]
	Pharmacoepidemiologist	[redacted]
	Data scientist, programmer	[redacted]
Research Institute in Primary Care (IDIAP), Jordi Gol (IDIAP)	Pharmacoepidemiology researcher. AZ PASS leader at IDIAP Jordi Gol	[redacted]
	Pharmacoepidemiology researcher	[redacted]
	Senior statistician	[redacted]
	Pharmacoepidemiology researcher. Coordinator of the Medicines Studies Unit	[redacted]
	SIDIAP Senior Data Scientist	[redacted]
	Primary health care researcher	[redacted]
PHARMO Institute	Principal investigator	[redacted]
	Co-principal investigator	[redacted]

Collaborating Institutions (proposed)[a]	Role	Name, title, qualifications
	Senior adviser	[redacted]
University Medical Center Utrecht (UMCU)	Principal investigator	[redacted]
Vaccine Monitoring Collaboration for Europe (Vac4EU)	VAC4EU oversight	[redacted]

[a] Members from each proposed collaborating institution have had the opportunity for a high-level review of this protocol.

4. ABSTRACT

Title

A Post-authorisation/Post-marketing Observational Study to Evaluate the Association Between Exposure to AZD1222 and Safety Concerns Using Existing Secondary Health Data Sources

Rationale and background

The novel coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), the cause of coronavirus disease 2019 (COVID-19), has led to a global pandemic. AZD1222 is a vaccine developed to prevent COVID-19. Now known as COVID-19 Vaccine AstraZeneca, the vaccine has received emergency use authorisation in the United Kingdom (UK) and conditional approval by the European Commission. Several important potential risks have been identified based on the experience of other non–COVID-19 vaccines. The clinical development programme also had limited enrolment of certain patient populations, including pregnant or breastfeeding women, individuals who are immunocompromised, frail persons with comorbidities, those with autoimmune or inflammatory disorders, and use with other vaccines such as an influenza vaccine.

Research question and objectives

What are the incidence rates (IRs) of safety events of interest (based on adverse events of special interest [AESIs]) among individuals vaccinated with AZD1222 and in individuals who have not received any vaccination for COVID-19, overall and in subpopulations of interest, within selected European data sources? How do the IRs compare with one another? What are the baseline characteristics of individuals who received at least one dose of AZD1222? How many of them received a second dose of a COVID-19 vaccine, which vaccine did they receive, and when did they receive it?

Study design

A multi-country, retrospective cohort design will be used to estimate the incidence of AESIs after receiving AZD1222 and will compare this incidence with that occurring in an unvaccinated comparator group. Where appropriate, the study will also use a self-controlled risk interval (SCRI) design. The study period will start on 04 January 2021, when the vaccine was first used in the UK, and will end approximately 24 months after it is introduced in the last country among participating data sources.

Population

The source population will comprise all individuals registered in each of the health care data sources. An exposed cohort will be identified based on first vaccination with AZD1222 (index date). A concurrent comparator population will be identified among subjects who have not received any vaccination for COVID-19 matched on the vaccinee's index date, age, and gender.

Variables

Receipt of AZD1222, other SARS-CoV-2 vaccines, and dates of vaccination will be obtained from all possible sources that capture COVID-19 vaccination, such as pharmacy dispensing records, general practice records, and immunisation registers.

Safety outcomes include safety concerns and other AESIs. These outcomes will be identified using algorithms based on codes for diagnoses, procedures, and treatments in electronic data, and they will be defined uniformly across the data sources to the fullest extent possible. Operational case definitions from the ACCESS (vACcine Covid-19 monitoring readinESS) project will be implemented for the AESIs for which they have been developed.

Covariates will be defined at the index date and will be used to define and describe the study cohorts and populations of special interest, their baseline characteristics, and/or to control for confounding. Covariates will include sociodemographic and lifestyle characteristics, comorbidities, comedications, and health care resource utilisation.

Data sources

The proposed data sources for the study are the Clinical Practice Research Datalink (CPRD) in the UK, the Valencia Integrated Database (VID) and the SIDIAP database in Spain, the Agenzia Regionale di Sanità Toscana (ARS Toscana) in Italy, and PHARMO Database Network in the Netherlands.

Study size

Across the proposed databases, the source population for the study includes approximately 33.9 million subjects. All available data in each set will be used to identify vaccinated subjects. The size of the exposed population in this study will depend on the use of AZD1222,

and the size of the comparator population will depend on the pace of mass immunisation during the study period. The precision of risk estimates will depend on the background rate and the duration of vaccine-associated risk assumed for each AESI. For example, for Bell's palsy (background rate of 38 per 100,000 person-years and 42 days at risk per dose), with 2 million vaccinated subjects, we anticipate a 97% probability that the upper bound of the observed incidence rate ratio (IRR) would be below 1.5, assuming a 1-to-1 ratio between vaccinated and comparator person-time and that the true IRR is 1.0.

Data analysis

Baseline characteristics of the subjects who received AZD1222 will be described overall and in sequential periods over time. Patterns of use of a second SARS-CoV-2 vaccine dose (AZD1222 and others) will be described similarly.

For the cohort study, exposure propensity scores (PS) will be used to exclude noncomparable subjects and refine the balance between study cohorts, which were initially matched on calendar date of vaccination, age, and gender. Propensity scores will be used to control for confounding either by PS matching or by analytic methods involving stratification or weighting. For AESIs for which the risk interval is characterised, crude IRs and 95% confidence intervals (CIs) for the vaccinated population and for the comparator cohort will be estimated. Poisson regression models are proposed to estimate crude and adjusted IRRs and IR differences with 95% CIs comparing vaccinated and comparator cohorts. These analyses will combine all person-time at risk after dose 1 and dose 2 (if it was received). Exploratory analyses will report the IRs by specific dose. For AESIs with unknown risk windows, cumulative incidence will be estimated using Kaplan-Meier methods starting after dose 1. Cox regression models are proposed to estimate crude and adjusted hazard ratios and 95% CIs.

For comparative analysis using the SCRI approach, conditional Poisson regression will be used to estimate IRRs and 95% CIs of specific AESIs, where appropriate.

Milestones

Milestone	Planned date [a]
Progress report	*3 months after protocol endorsement (expected October 2021)*
Interim report 1	*9 months after protocol endorsement (expected April 2022)*
Interim report 2	*15 months after protocol endorsement (expected Oct 2022)*
Interim report 3	*21 months after protocol endorsement (expected April 2023)*
Final report of study results	*27 months after protocol endorsement (expected Oct 2023)*

[a] Schedule is dependent on protocol endorsement date, market uptake of AZD1222 vaccine, approvals for data extraction in each data source, and contracts with the research team.

5. AMENDMENTS AND UPDATES

None.

6. MILESTONES

Table 1 Study Milestones

Milestone	Actual/Planned date[a]
Protocol submission	01 April 2021
Protocol endorsement by EMA/MHRA	TBD (expected July 2021)
Registration in the EU PAS Register	No later than 6 months after EMA protocol endorsement and before start of data collection
Statistical analysis plan submission	4 months after protocol endorsement (expected November 2021)
Start of data collection[b]	4 months after protocol endorsement (expected November 2021-first interim analysis)
End of data collection[c]	20-21 months after protocol endorsement (expected March/April 2023 - final report)
Progress report	3 months after protocol endorsement (expected October 2021)
Interim report 1	9 months after protocol endorsement (expected April 2022)
Interim report 2	15 months after protocol endorsement (expected October 2022)
Interim report 3	21 months after protocol endorsement (expected April 2023)
Final report of study results	27 months after protocol endorsement (expected October 2023)

[a] Contract with the research team is pending, as is a fully detailed protocol review by new research partners with access to data sources added in version 2.0 of the protocol. The schedule is dependent on protocol endorsement date, market uptake of AZD1222 vaccine, approvals for data extraction in each data source, and contracts with the research team.

[b] Start of data collection: the date from which information on the first study subject is first recorded in the study data set or, in the case of secondary use of data, the date from which data extraction starts [IR Art 37(1)]. Simple counts in a database to support the development of the study protocol, eg, to inform the sample size and statistical precision of the study, are not part of this definition.

[c] End of data collection: the date from which the analytical data set is completely available [IR Art 37(2)]. Analytical data set: the minimum set of data required to perform the statistical analyses leading to the results for the primary objective(s) of the study.

EMA, European Medicines Agency; EU PAS, European Union electronic Register of Post-Authorisation Studies; MHRA, The Medicines and Healthcare products Regulatory Agency; TBD, to be determined.

Timing of the interim reports will be driven by dates of protocol endorsement by the European Medicines Agency (EMA), contracting between research institutions, a minimum time of data accrual and the needed documentation of analytical instructions, data extraction, analysis, and reporting. Content of the progress report and interim reports is described in Section 9.7.11.

7. RATIONALE AND BACKGROUND

The novel coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), the cause of coronavirus disease 2019 (COVID-19), has led to a global pandemic. AZD1222 (called Vaxzevria® in Europe) is a vaccine that was developed to prevent COVID-19.

Coronaviruses are enveloped viruses with positive-sense single-stranded RNA genomes. The spike glycoprotein is a coronavirus surface protein involved in receptor binding and mediating virus entry into host cells during infection (Li 2016). AZD1222 is a recombinant replication-defective chimpanzee adenovirus expressing the SARS-CoV-2 spike glycoprotein. Development of AZD1222 was initiated by the University of Oxford with subsequent transfer of development activities to the Sponsor.

The University of Oxford is investigating the safety, immunogenicity, and efficacy of AZD1222 in four ongoing controlled (meningococcal vaccine or placebo) clinical studies. A pooled interim analysis of the four ongoing studies found AZD1222 to have an acceptable safety profile in adults following vaccination (Voysey et al 2021). The incidence of serious adverse events (AEs) and AEs of special interest (AESIs) was similar between the analysis groups (participants receiving AZD1222 vs controls), and all four of the non-COVID-19 deaths (one in an AZD1222 recipient and three in control recipients) were considered unrelated to the study product (Voysey et al 2021). Local and systemic reactogenicity of AZD1222 was tolerable and decreased in incidence and severity in older adults and after the second dose (Voysey et al 2021).

In addition to the Oxford studies, the Sponsor has ongoing clinical studies, including a phase III double-blind placebo-controlled study.

Results of the blinded, randomised, controlled trials conducted in the United Kingdom (UK), Brazil, and South Africa reported an efficacy of 70.4% after two doses of AZD1222 against symptomatic disease with no safety concerns. These studies involved approximately 24,000 people in both active treatment and placebo arms.

On 30 December 2020, the UK Medicines and Healthcare products Regulatory Agency (MHRA) provided authorisation for emergency supply of AZD1222, and on 29 January 2021, the European Commission granted conditional marketing authorisation for the vaccine. Now known as COVID-19 Vaccine AstraZeneca, the vaccine in both jurisdictions is indicated for active immunisation to prevent COVID-19 caused by SARS-CoV-2, in individuals aged 18 years and older. The vaccination course consists of two separate intramuscular injections. The second dose should be administered between 4 and 12 weeks (28 to 84 days) after the first dose.

The protocol is prepared based on the most current version of the European Union (EU) risk management plans (RMPs) under review by EMA/PRAC, which, at the time of protocol approval, is EU RMP Version 4.0. Based on this version of the EU RMP, the safety concerns for AZD1222 include nervous system disorders (including immune-mediated neurological conditions), vaccine-associated enhanced disease (including vaccine-associated enhanced respiratory disease), thrombocytopenia with associated bleeding, anaphylaxis, thrombosis, and 'thrombosis with thrombocytopenia syndrome'(TTS). Areas of missing information include use of AZD1222 in pregnant or breastfeeding women, use in immunocompromised individuals, use in frail individuals with comorbidities, use in those with autoimmune or inflammatory disorders, and interactions with other vaccines and long-term safety.

This post-authorisation safety study (PASS), which is required in AstraZeneca's approved EU RMP, will evaluate the incidence and relative risk of safety concerns and AESIs, as defined in the approved EU RMP, following immunisation in the real-world setting. A retrospective, longitudinal cohort study will be conducted using existing secondary automated electronic health data sources to address these safety concerns in the general population as well as within special populations for which there is missing information. Self-controlled risk interval (SCRI) analyses will also be implemented for specific AESIs for which the risk interval for vaccine-associated effects is known.

8. RESEARCH QUESTION AND OBJECTIVES

The primary study objectives are as follows:

1 To describe the baseline characteristics (eg, demographics, medical history) of all individuals who receive at least one dose of AZD1222 over the study period

2 To describe, among subjects who receive a first dose of AZD1222, the timing and type of second dose of any COVID-19 vaccine (AZD1222 or other) over the study period

3 To describe the incidence of prespecified AESIs in subjects who have received at least one dose of AZD1222 and in matched subjects who have not received any vaccination against COVID-19 (unvaccinated subjects)

4 To estimate any increased risk of prespecified AESIs following vaccination with AZD1222 using study retrospective cohort and SCRI designs

Secondary objectives are as follows:

1 To describe the baseline characteristics (eg, demographics, medical history) of all individuals who receive at least one dose of AZD1222 over the study period among the specific populations considered as missing information (see Section 9.2.1.2)

2 To describe, among subjects who receive a first dose of AZD1222, the timing and type of second dose of any COVID-19 vaccine (AZD1222 or other) over the study period among the specific populations considered as missing information

3 To describe the incidence of prespecified AESIs in subjects who have received at least one dose of AZD1222 and matched subjects who have not received any vaccination against COVID-19 (unvaccinated subjects) among the specific populations considered as missing information

4 To estimate any increased risk of prespecified AESIs following vaccination with AZD1222 using study retrospective cohort and SCRI designs, which allow for control of confounding among the specific populations considered as missing information

No specific hypotheses will be tested during the study described in this protocol.

9. RESEARCH METHODS

9.1 Study design

Using information from several health care databases, this observational population-based multi-country study will primarily employ a retrospective cohort design to compare subjects who have received at least one dose of AZD1222 with subjects who have not yet received any COVID-19 vaccine for the occurrence of AESIs. Secondarily, it will employ an SCRI design for a subset of AESIs that are appropriate for this approach, as described in Section 9.1.2.

This study, which will be conducted in the VAC4EU (Vaccine monitoring Collaboration for Europe, https://vac4eu.org/) research environment, will use a common protocol across all study sites, a common data model (CDM) and common analytics. It will use case definitions developed by the vACCine covid-19 monitoring readinESS (ACCESS) project, which was funded by the EMA to prepare a European infrastructure to monitor COVID-19 vaccines. This protocol is adapted from a template developed by the ACCESS project entitled “Safety evaluation of COVID-19 vaccines in electronic healthcare databases” (ACCESS Project 2020).

The study will be conducted using information collected in electronic health care data sources in Europe. The proposed data sources are the Clinical Practice Research Datalink (CPRD) in the UK, the Valencia Health System Integrated Database (VID) and the SIDIAP database in Spain, the Agenzia Regionale di Sanità Toscana (ARS Toscana) in Italy, and PHARMO Database Network in the Netherlands. These data sources were selected because they are able to capture the data elements needed to conduct this study, cover areas where AZD1222 use was expected, have been used in prior studies of vaccine safety, and appear to have reasonably short data lags, which is necessary to conduct interim analyses (Willame et al 2021) (see Section 9.4). The study period will run from the introduction of AZD1222 in each country in early 2021 for 24 months for the evaluation of all AESIs.

9.1.1 Retrospective cohort design

To address Objectives 1, 2, and 3 (both primary and secondary), a cohort of exposed subjects will be selected. Subjects will enter the exposed cohort upon their first vaccination with AZD1222.

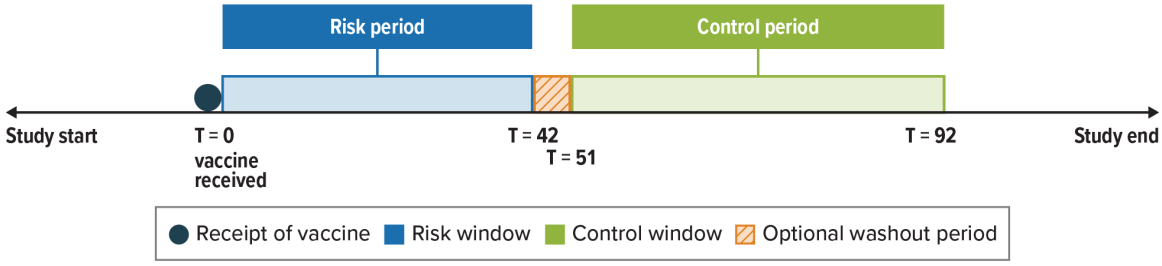
To address Primary Objectives 3 and 4, a cohort of concurrent matched subjects will be selected. The comparison cohort will comprise subjects who have not yet received any COVID-19 vaccine. The comparison cohort will be matched to the exposed population based on calendar date of vaccination, age, and gender. The rationale for matching on calendar date is to control for the temporal variation in the SARS-CoV-2 circulation (COVID-19 infection may affect the risk of some AESIs) and expected changes in vaccinee characteristics over time. Propensity scores (PS) will be used to address confounding, balancing multiple covariates such as prior COVID-19 infection, comorbidities, and markers of health care utilisation between vaccinated and unvaccinated cohorts (Austin 2014, Webster-Clark et al 2021). Relative risk will be the main measure of effect, which will be estimated using incidence rate ratios (IRRs) or hazard ratios (HRs), depending on the analysis and outcome. Similar analyses will be conducted in the subpopulations of interest to address Secondary Objectives 3 and 4.

In the cohort study, the initial plan is to use concurrent, rather than historical, comparator (unvaccinated) subjects because they are at risk for COVID-19, and COVID-19 may function as an important covariable in the analyses of several of the safety endpoints. Moreover, both health care-seeking behaviours and health care delivery during the pandemic differ from those during the prepandemic era. The potential challenge with this approach is that during the pandemic, unvaccinated subjects in the comparator group may themselves become vaccinated over time. It is conceivable that these patterns may vary by geography and clinical characteristics. Once unexposed subjects are vaccinated, it is no longer reasonable to use their subsequent person-time for comparisons. The extent to which unvaccinated subjects are vaccinated will be monitored during the first year of the study. If it is extensive and substantially limits the duration of longer-term follow-up, historical rather than concurrent subjects will be used for the reference cohort for events not suitable for self-controlled designs (see Table 2 in Section 9.3.2). The rationale for not selecting historical prepandemic subjects for primary comparison pertains to concerns over noncomparability. The background rate of AESIs may have changed during the pandemic due to the effects of COVID-19 or because of less intensive ascertainment due to access issues or health-seeking behaviours. Using **historical controls** from the pandemic era before the availability of COVID-19 vaccines could theoretically overcome this challenge. One could potentially select comparators from this era matched on the first vaccination date of the exposed subjects minus 1 year—an approach that could control for seasonality.

9.1.2 Self-controlled risk interval design

A complementary approach for Primary Objectives 3 and 4, an SCRI analysis, will be used to assess relative risks for AESIs meeting the criteria for this specific design (ie, acute onset, short latency, risk intervals that are relatively well known, that the occurrence of the event must not affect the probability of vaccination, and that the occurrence of the event of interest must not censor or affect the observation period) (Weldeselassie et al 2011). The SCRI is a case-only study that includes only individuals who were vaccinated with AZD1222 who experienced an outcome during the study period (Tokars et al 2012). The IRR comparing the IR of the AESI(s) in a period hypothesised to be at increased risk due to exposure ("risk period" or "exposed person-time") will be compared with a prespecified postvaccination period within the same individual that falls after the risk period ("control period" or "unexposed person-time"). Limiting the selection of control periods to after vaccination will minimise the potential biases of other self-controlled designs that include comparator time before vaccination and require the assumption that prevaccination events do not affect the probability of subsequent vaccination (Weldeselassie et al 2011). Figure 1 illustrates the design of the SCRI analysis.

Figure 1 Self-controlled Risk Interval Design



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

ACCESS, vACCine covid-19 monitoring readinESS; COVID-19, coronavirus disease 2019; T, time measured in days.

Source: "Safety evaluation of COVID-19 vaccines in electronic health care databases: a protocol template from the ACCESS project".

The cohort study is proposed as the primary approach because, unlike the SCRI design, this approach can be used to study outcomes with gradual onset (eg, multiple sclerosis), long latency, and/or risk intervals that cannot be well defined (eg, chronic fatigue syndrome), and it also allows estimation of risk difference. On the other hand, the main strengths of the SCRI design are that it implicitly adjusts for time-invariant confounders and that it can still be used to estimate relative risk even if everyone is vaccinated during the study period. Moreover, in contrast to the cohort design, including only exposed individuals minimises exposure misclassification due to imperfect capture of vaccinations in secondary data sources.

9.2 Setting

9.2.1 Population

The source population for each of the study designs will comprise all individuals registered in each health care data source (see Section 9.4) during the study period.

9.2.1.1 Cohort design: study population

Conceptually, the study population will include subjects who are exposed to the AZD1222 vaccine and concurrent subjects who have not received any SARS-CoV-2 vaccine, who will serve as comparators.

Subjects will be required to meet the following inclusion criteria:

- Have at least 12 months of data available before cohort entry
- Have no record of vaccination with any SARS-CoV-2 vaccine on or before the date of cohort entry

Subjects will be eligible to enter the AZD1222 cohort if they receive at least one dose of AZD1222 within the study period. The date when a patient receives the first AZD1222 dose within the study period will be considered the "index date".

Identification of the comparator cohort will involve several steps, as follows, that will take place within each data source:

1 For each exposed subject, all potential comparator subjects will be identified as matched on index date, sex, and age.

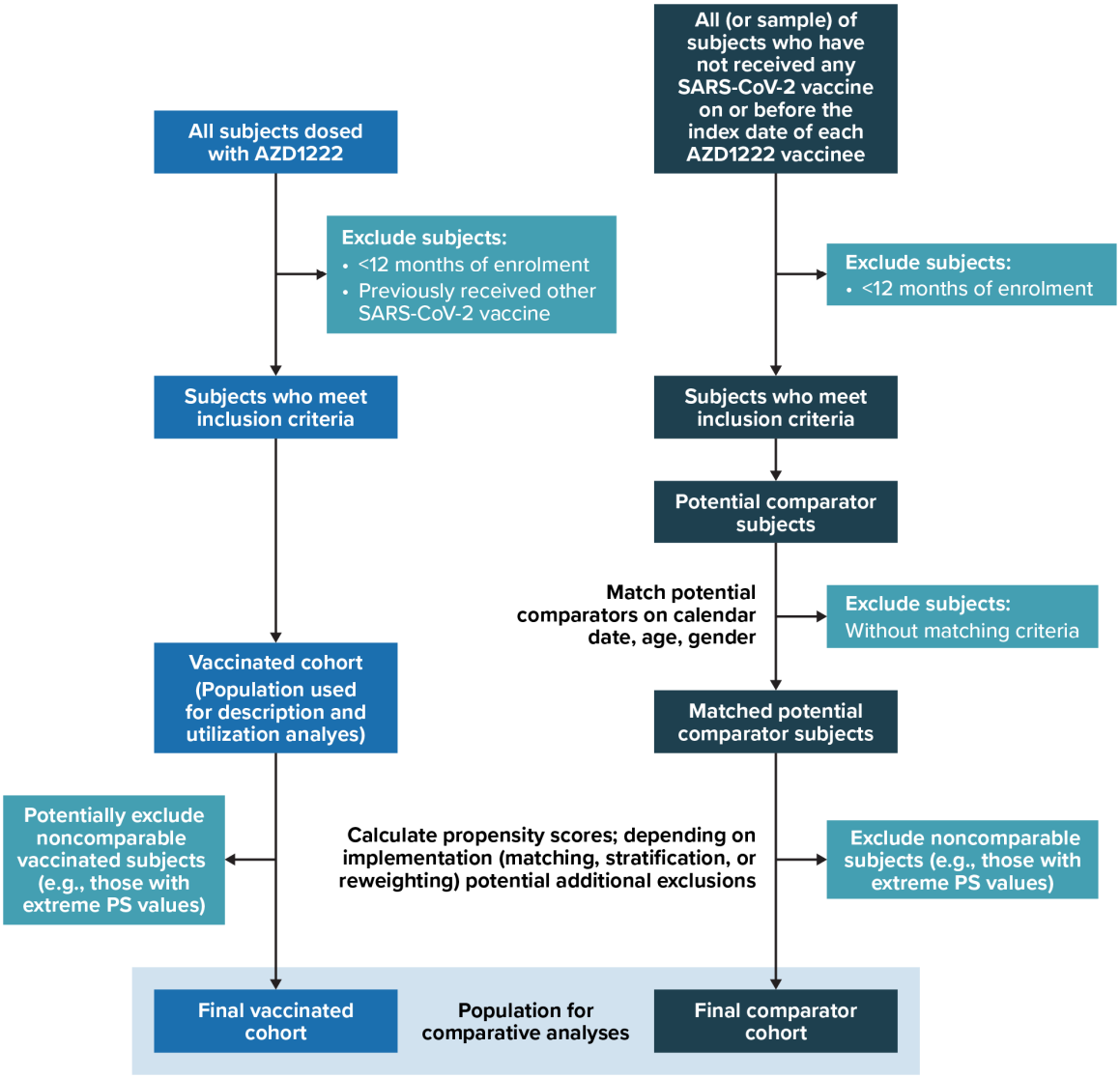
2 All candidates who do not meet all the study inclusion criteria will be excluded.

3 To balance baseline characteristics that may differ between cohorts, we will use PS methods. The PS will be estimated based on characteristics of the comparator and exposed cohorts ascertained at the time of dose 1 (see Section 9.7.1).

4 Final selection of the comparator cohort will then occur using the PS to exclude noncomparable subjects, according to a method that will be specified in the statistical analysis plan (SAP).

Whether all potential comparators will be included in the selection of the comparator cohort, or only a sample, will depend on the data governance rules and data management capabilities of each research partner, which is still to be determined. The SAP will document this decision, along with other details about matching criteria. Figure 2 illustrates the assembly of the study cohorts.

Figure 2 Assembly of Study Cohorts



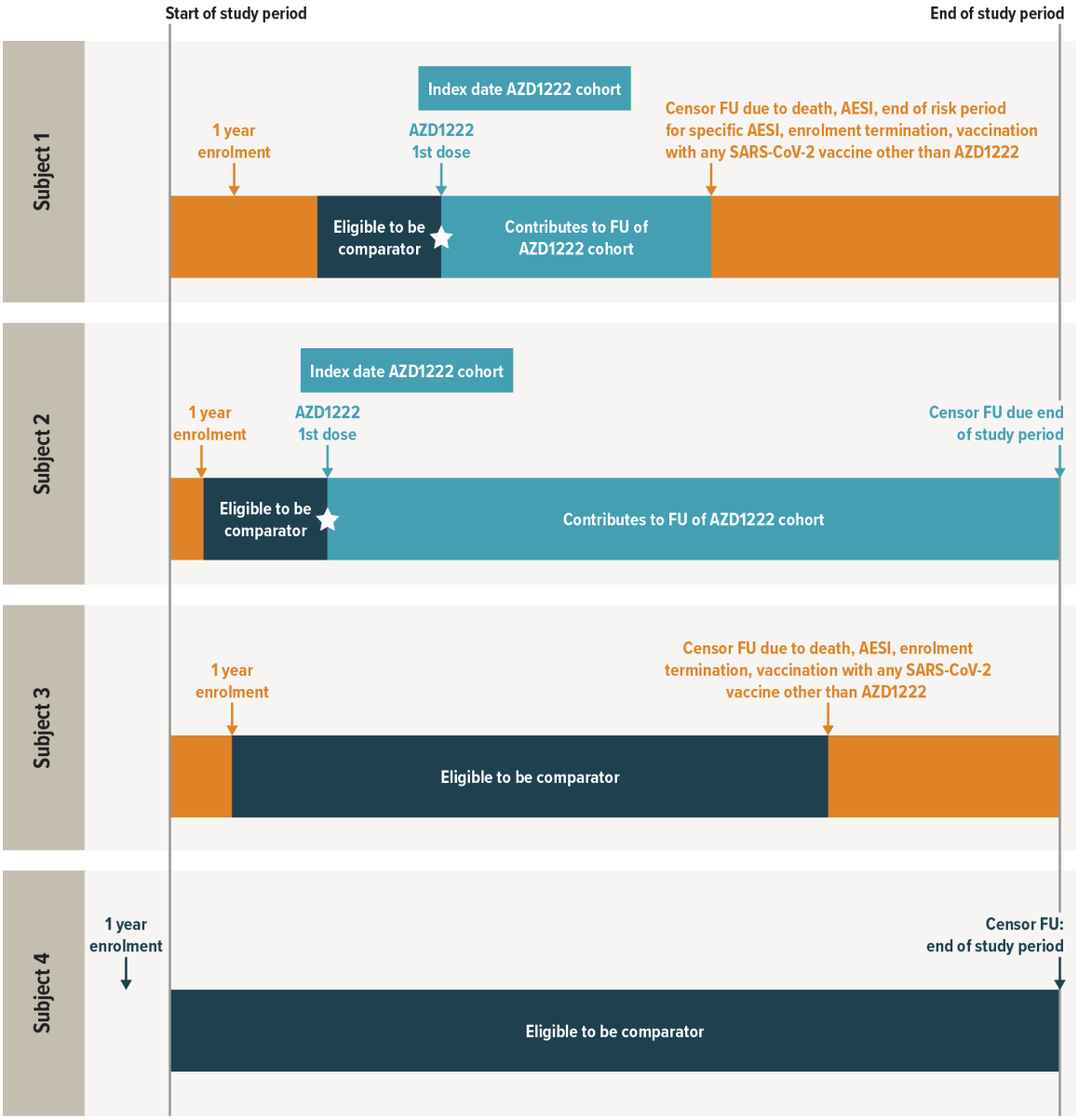
Note: The process illustrated here will be conducted separately in each data source. Index date refers to the date of the first dose of AZD1222 for each vaccinee.

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; PS, propensity score.

Subjects in the comparator cohort will be allowed to enter the AZD1222 cohort if they later receive at least one dose of AZD1222. Subjects in the comparator cohort will be censored from that cohort if and when they receive any SARS-CoV-2 vaccine.

Figure 3 illustrates cohort design concepts of eligibility, index dates, and the potential to move from a comparator (unexposed) to the exposed cohort.

Figure 3 Eligibility Criteria and Index Date Definition



AESI, adverse event of special interest; FU, follow-up; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

9.2.1.2 Special populations

In addition to the general study population, safety concerns will also be evaluated in prespecified subgroups for which information was deemed missing in the RMP.

- Women who are pregnant or breastfeeding
- Immunocompromised patients
- Frail patients with comorbidities (eg, chronic obstructive pulmonary disease, diabetes, chronic neurological disease, cardiovascular disorders)
- Patients with autoimmune or inflammatory disorders
- Patients who, at cohort entry, had recently received vaccines to prevent diseases other than COVID-19, such as pneumococcal infection, influenza, and shingles (herpes zoster).

Concepts to identify the special populations are listed in Appendix C, Annex Table 3. Details on algorithms to identify each of these subgroups will be documented in a SAP.

9.2.1.3 SCRI design: study population

For each outcome of interest to be evaluated using the SCRI design, the eligible population will include subjects from the AZD1222 cohort who experienced the outcome of interest during the study period. Only selected outcomes with well-characterised risk intervals will be evaluated using this approach (see Section 9.3.2). Consistent with other vaccine studies that have used this approach (Lee et al 2011), events of interest must not have occurred in the recent past in order to distinguish between ongoing care for a historical event and an incident or recurrent event. Accordingly, each outcome will require a "clean" (ie, event-free) interval before the start of observation, which will vary by outcome; proposed specifications for this "clean look-back interval" appear in Table 2. Specific criteria for selection of individuals for the SCRI design will require that:

- The subject received at least one dose of AZD1222
- The subject experienced the event during the risk or control interval
- There is 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 Study period

The study period will start on the date AZD1222 vaccination began in each country. The first vaccinations started approximately 1 week after approval date, which was 30 December 2020 in the UK and 29 January 2021 in the EU. The study period duration will be 24 months in each data source or until latest data available at the time of start of data collection (data extraction), ie, this duration may include a few more months in the data sources where the lag time to obtain the data is shorter.

9.2.3 Follow-up and time at risk

Follow-up time will start on the date that all inclusion criteria are met and will end at the earliest of the occurrence of censoring conditions or the last data extraction/data availability as defined for each study design.

9.2.3.1 Cohort design: follow-up

Follow-up will start on the index date and end on the earliest of the following possible follow-up termination dates for both cohorts:

- The end date of the study period.
- 365 days after the second vaccine dose. This rationale is two-fold. First, because we expect that most comparators will become vaccinated during the study period, this choice will reduce the anticipated mismatch of follow-up time duration between the exposed and comparator cohorts. Second, although the risk window for some AESIs is not defined, vaccine-associated AEs are unlikely to manifest beyond 1 year. To include person-time in the risk window when a vaccine effect could not reasonably occur may attenuate risk estimates, if a true risk exists.
- The subject's enrolment termination date in the health plan or system.
- The subject's date of death.
- The date on which the subject receives a dose of any SARS-CoV-2 vaccine other than AZD1222.
- For subjects in the comparator cohorts, the date on which the subject receives a dose of AZD1222, at which time the subject switches from the comparator to the AZD1222 cohort.
- For the specific evaluation of each AESI:
  - The date of the first diagnosis of a specific AESI recorded during the study period. For composite outcomes, follow-up will be censored at the date of the first occurrence of any of the components.
  - For AESIs for which the risk interval is well characterised, the date on which the defined risk interval ends; this risk interval applies to both the vaccinee and matched comparator subjects.
  - For subjects who receive two doses of AZD1222, a second risk interval will follow the second dose; and for comparator subjects, a corresponding second risk interval will also apply, mirroring the timing of the second risk interval for the matched vaccinee. As previously noted, the start and duration of the risk windows will be specific to each outcome, corresponding to the period when the biologic effect of a vaccine is hypothesised to increase the risk for the particular AE (see Section 9.1.2).
- For outcomes for which a specific risk window is not defined, follow-up will comprise all person-time after the index date and extend until the occurrence of any censoring criteria, regardless of administration of a second dose.

9.2.3.2 SCRI design: follow-up and time at risk

The SCRI evaluates occurrence of a particular AESI during a "risk interval" (exposed person-time) and a subsequent "control interval" (unexposed person-time) following AZD1222 vaccination (see Figure 1). Proposed outcome-specific risk intervals for the AESIs appear in Section 9.3.2. Control intervals, which by design occur following the risk intervals, will have the same duration as the risk interval, except under the special circumstances explained below. Control intervals will start on the day following the end of the risk interval, except in situations for which the risk interval is ambiguous. In these situations, a washout period will be introduced, after which the control interval will begin.

Figure 1 illustrates this arrangement for a risk window that starts 1 day after vaccination and continues for 42 days. After a 7-day washout period, the control interval begins and ends after 42 days as well. It is important to note that in contrast to the cohort design, the occurrence of the outcome is not a censoring event for SCRI evaluation; the full duration of the risk interval and control intervals factors into the analysis. For subjects who receive a second dose, corresponding risk and control intervals follow the second dose as well.

In general, if the AESI occurs after dose 1 but before dose 2, the SCRI analysis includes only the risk and control intervals following dose 1. Similarly, if the AESI occurs after dose 2, the SCRI analysis includes only the risk and control intervals following dose 2.

The scenario depicted in Figure 1 may be complicated if a subject receives a second vaccine dose before the control interval for the first dose has fully elapsed. In this situation, the first control period will be censored at the time of the second dose. Death of a subject during the control interval will also serve as a censoring event.

If a subject receives a second dose before the risk interval for dose 1 has completed, the risk periods will be concatenated and extended for the duration of the risk interval following the second dose. In this scenario, the SCRI analysis will include the person-time in the concatenated risk intervals and the control period following dose 2.

Details about AESI-specific risk intervals and washout periods will be specified in the SAP.

9.3 Variables

9.3.1 Exposures

Receipt of AZD1222, other SARS-CoV-2 vaccines, and dates of vaccination will be obtained from all possible sources that capture COVID-19 vaccination, such as pharmacy dispensing records, general practice records, and immunisation registers. Depending on the data source, vaccines will be identified via nationally used product codes, including batch numbers, where possible. The extent of capture of COVID-19 vaccination in the target populations and of

specific product types and batch numbers is not currently known in candidate data sources. The duration of the risk periods will be specific for each of the outcomes and defined to establish an accurate relationship and patterns in that relationship (see Section 9.3.2). Algorithms and codes to ascertain relevant vaccinations will be included in the SAP.

9.3.2 Outcomes

Outcomes include the safety concerns and other AESIs listed in the current approved AZD1222 EU RMP and also new safety events of interest raised by the EMA after evaluation of cases involving thrombocytopenia with thrombosis or bleeding. The AESIs differ in terms of latency, acuity of onset, availability of empirical estimates for appropriate risk periods, and the effect of the event on subsequent likelihood of vaccination. Table 2 lists these AESIs, indicates which of the events are deemed suitable for analysis using SCRI analysis, and indicates the proposed risk interval for those that are deemed suitable. In addition, risk intervals are proposed for some AESIs that are deemed not suitable for SCRI. Although available evidence suggests that any biological effect of a vaccine is expected to occur during these intervals, uncertainty about the end of this period raises concerns about misclassification bias that could occur if the control interval (which follows the risk interval in the proposed SCRI analysis) includes person-time that is actually at risk. The Sponsor also judges that it is reasonable to use these intervals in the cohort analysis. This allows for more certainty that comparator person-time does not include time at risk (barring unrecorded vaccinations). Person-time remote from the vaccination date is deemed very unlikely to be at risk from any adverse vaccine effect, and to include it could substantially attenuate estimates of a vaccine-associated risk that operated only closer to the index date. There is, of course, the potential cost of this approach limiting the number of outcomes, but we judge that the trade-off is reasonable to reduce misclassification bias that could mask a true signal.

Table 2 Adverse Events of Special Interest and Other Safety Concerns

Adverse event	Suitable for SCRI analysis according to ACCESS protocol template[a]	Risk interval after index date (days)	Clean look-back interval (days)*
Vaccine-associated enhanced disease, including vaccine-associated enhanced respiratory disease, including ARDS	No	14-365[b]	365
Multisystem inflammatory syndrome in adults/children	Yes	1-42[c]	365[c]
Sudden death	No	0-6[b]	Not applicable
Autoimmune thyroiditis	No[d]	1-180[e]	365
Anosmia, ageusia	No[f]	1-365[e]	365
Anaphylaxis	Yes	0-2[c]	180[c]
Type III hypersensitivity reactions	No[d]	1-365[e]	365

Table 2 Adverse Events of Special Interest and Other Safety Concerns

Adverse event	Suitable for SCRI analysis according to ACCESS protocol template[a]	Risk interval after index date (days)	Clean look-back interval (days)*
ARDS	Yes	1-28[j]	365
Guillain-Barré syndrome	Yes	1-42[c]	365[c]
Other peripheral and polyneuropathies	No[d]	1-42[c]	365[c]
Multiple sclerosis[d], transverse myelitis, and other demyelinating disorders[d]	Yes (for transverse myelitis only)	1-90[c] for transverse myelitis only; all others 1-365[e]	365[c]
Optic neuritis/neuromyelitis optica spectrum disorder	Yes	1-42[c]	365[c]
Encephalitis (including acute disseminated encephalomyelitis)	Yes	1-42[b]	183[c]
Myasthenia gravis	No[d]	1-365[e]	365
Bell's palsy	Yes	1-42[c]	365[c]
Generalised convulsions	Yes	0-14[c]	365
Narcolepsy	No	1-42[c]	365[c]
Myocarditis/Pericarditis	Yes	1-42[c]	365[c]
Postural orthostatic tachycardia syndrome	No[d]	1-42[b]	365[c]
Myocardial infarction	Yes	1-28[c]	365[c]
Acute cardiac injury including microangiopathy, cardiogenic shock, heart failure, stress cardiomyopathy	No[f]	1-90[e] 1-42 for stress cardiomyopathy	365
Thrombocytopenia	Yes	1-42[b]	365[c]
Thrombocytopenia with associated bleeding	No[d]	1-42[g]	365[c]
Thrombosis (embolic and thrombotic events) without thrombocytopenia	Yes	1-42[c]	365[c]
Thrombosis with thrombocytopenia syndrome	No[d]	1-42[g]	365
Capillary leak syndrome	No[d]	1-365[e]	365
Acute kidney injury	Yes	1-14[b]	365
Acute liver injury	Yes	1-14[b]	365
Acute pancreatitis	No[d]	1-365[e]	365
Acute aseptic arthritis	Yes	1-42[b]	365
Fibromyalgia	No[d]	91-365[h]	365
Rhabdomyolysis	No[d]	1-42[b]	365

Table 2 Adverse Events of Special Interest and Other Safety Concerns

Adverse event	Suitable for SCRI analysis according to ACCESS protocol template[a]	Risk interval after index date (days)	Clean look-back interval (days)*
Chronic fatigue syndrome/ME/PVFS	No[d]	183-365[i]	365
Erythema multiforme	No[f]	1-365[e]	365
Chilblain-like skin lesions	No[f]	1-365[e]	365

[a] Suitability according to ACCESS protocol template entitled "Safety evaluation of COVID-19 vaccines in electronic healthcare databases." See Section 9.1.2 and Section 9.2.1.3 for additional information on SCRI design.

[b] Risk interval based on consensus definition from the AESI Working Group of Vaccines Europe, March 2021.

[c] Based on Lee et al (2011) or COVID-19 Vaccine Safety Surveillance: Active Monitoring Master Protocol (FDA 2020).

[d] Not addressed in ACCESS protocol template but is deemed not suitable for one or more of the following reasons: onset may be gradual, the diagnosis date may substantially lag the onset of symptoms, or the risk period is not clearly established.

[e] No consensus evidence for a risk period was identified; defaults to one year following index date.

[f] ACCESS protocol template indicates "yes," or SCRI may be suitable if a purported risk period can be identified, which is not yet known for this AESI. Absent such information, deemed not suitable.

[g] Sensitivity analyses will evaluate alternative risk windows. For TTS, they will be 1-14, 1-21, and 1-28 days after the index date. For thrombocytopenia with bleeding, the alternative risk window will be 1-28 days after the index date, following Pottegård et al (2021).

[h] No consensus evidence for a risk period was identified. However, because symptoms need to be present for at least 3 months before patients qualify for the diagnosis (Arnold et al 2019), the start of the risk window was set to 91 days after the index date.

[i] No consensus evidence for a risk period was identified. However, because chronic fatigue needs to be present for at least 6 months before patients qualify for the diagnosis (IOM 2015), the start of the risk window was set to 183 days after the index date.

[j] Based on Serazin et al (2021).

* Clean interval refers to the look-back period before cohort entry during which no AESI was observed. Unless indicated otherwise, the interval duration is based on the clinical judgement of the research team.

Note: Risk window definitions may evolve during the course of the study as more knowledge is gathered about individual AESIs.

AESI, adverse event of special interest; ARDS, acute respiratory distress syndrome; PVFS, Postviral Fatigue Syndrome; ME, myalgic encephalitis; SAP, statistical analysis plan; SCRI, self-controlled risk interval; TTS, thrombosis with thrombocytopenia syndrome.

Outcomes will be identified using algorithms based on codes for diagnoses, procedures, and treatments in electronic data, and they will be defined uniformly across the data sources to the fullest extent possible. Operational case definitions from the ACCESS project (https://drive.google.com/drive/folders/1Y_3cuGRN1g-jBv2ec1fC0aYcpxEjtrY9) will be implemented for the AESIs for which they have been developed. The SAP will include detailed information on codes and algorithms.

If feasible, algorithms to identify selected AESIs will be validated in a sample of cases based on manual review of electronic records or chart abstraction conducted by clinicians blinded to COVID-19 vaccine exposure. If validation is performed, certainty of an event diagnosis will

be classified against existing and as-yet developed standardised definitions such as those created by the Brighton Collaboration.

9.3.2.1 Thrombosis with thrombocytopenia syndrome

As previously noted, case definitions will be specified in the SAP, however, given its clinical importance, additional details for defining TTS are provided here. As of May 2021, no TTS standard case definition has been accepted for use by all countries (Brighton Collaboration 2021b). The Brighton Collaboration interim case definition for TTS (version 10.16.2, 18 May 2021) defines TTS as "Any patient presenting with both acute venous or arterial thrombosis AND new-onset thrombocytopenia" (as confirmed by both the Brighton Case Definitions for thrombocytopenia and thrombosis (Brighton Collaboration 2021a, Wise et al 2007)). The Brighton Collaboration case definition for thrombocytopenia (Wise et al 2007) requires evidence of a platelet count of less than 150,000/μL^3.

Because laboratory measurements of platelet counts are not expected to be available for most proposed data sources, the case definition of TTS for the current study has been adapted to qualify based on a diagnosis code. In addition, diagnoses that qualify for thrombosis do not include acute myocardial infarction or stroke, which is consistent with the May 2021 CDC working case definition for TTS following COVID-19 vaccination (CDC 2021). It should be emphasised that the proposed case definition for this study is provisional and subject to change as the Brighton case definition evolves and in response to findings from validation exercises. Because the concept of "thrombosis" also includes embolic events, for purposes of the case definition, the concept has been relabeled "thromboembolic event" (TE).

The working TTS case definition for this study requires:

- A new qualifying TE (defined below) AND
- A new diagnosis of thrombocytopenia (or laboratory evidence of the same, where available) that is made from 7 days before and up to 7 days after that event

Figure 4 shows the look-back ("washout") periods proposed to define "new" diagnoses.

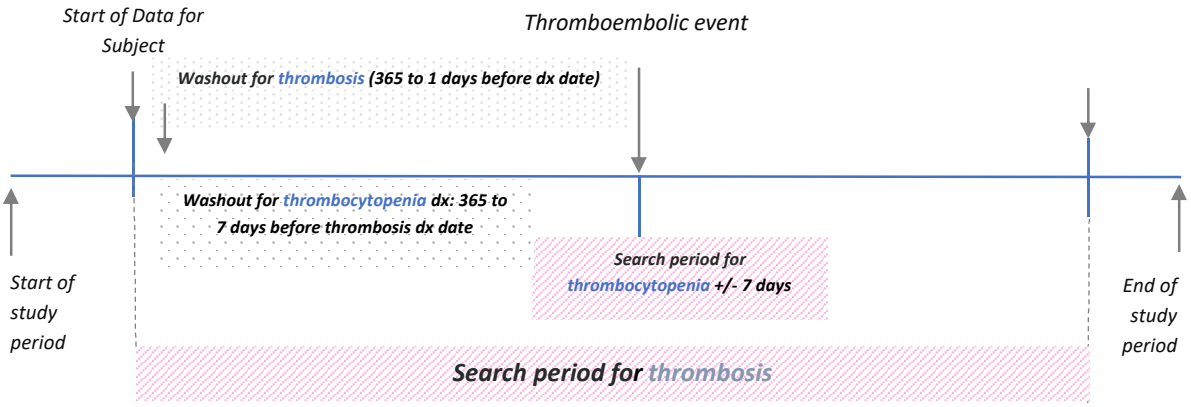
Thromboembolic events will include new diagnoses of embolism or deep vein thrombosis (DVT) (including splanchnic, intracranial, limb locations) or thromboembolism. According to the working case definition, infarctions (including myocardial and cerebral) and stroke will not qualify as TE. In addition to evaluating TE overall, it is proposed to evaluate DVT outcomes separately by location of the event into central venous sinus thrombosis (CVST), limb thrombosis, and splanchnic area thrombosis.

In the primary analysis, the risk window for TTS extends from 1 to 42 days after vaccination (or the corresponding index date for the unexposed comparator subject). Sensitivity analyses

will also evaluate alternative risk windows of 1 to 14 days, 1 to 21 days, and 1 to 28 days after the index date.

Both TE and thrombocytopenia will be ascertained using codes for diagnoses, procedures, and treatments in electronic data that are also under development within the studies conducted under the VAC4EU framework.

Figure 4 Time Windows to Define TTS



dx, diagnosis; TTS, thrombosis with thrombocytopenia syndrome.

Identification of the clinical concerns "Thromboses without thrombocytopenia" and "Thrombocytopenia without thrombosis" will follow a similar logic.

9.3.2.1.1 Expert review of TTS cases identified in automated data

For cases of TTS identified in automated data (as well as other AESIs selected for further review [ie, validation]), the electronic patient profile, a cumulative chronologically ordered record of all available electronic linkable information, will be reviewed by the members of each local research team who are blinded to vaccination status, wherever possible. Additional review of clinical records in hospital charts by research team clinicians will be requested as per access and approval of hospitals in PHARMO, VID, and ARS Toscana. In SIDIAP and CPRD, additional review can be requested from the general practitioners (GPs), who have access to the discharge letters.

It is anticipated that reviewers will be physicians and other life science professionals who will apply a standardised case definition of TTS (to be finalised). Data sought will include platelet counts, evidence documenting the thrombus and location, evidence of recent heparin exposure, and any evidence that a specific alternative diagnosis to TTS (eg, thrombotic thrombocytopenic purpura) accounted for the clinical presentation.

9.3.2.2 Thrombocytopenia with bleeding

As of June 2021, there appears to be no standard case definition to study thrombocytopenia with bleeding following vaccination. The approach proposed to identify this outcome in automated data will follow the same approach used for TTS (see Section 9.3.2.1), using bleeding instead of TE to anchor the search. It will use the same look-back period to define "new" diagnoses (see Figure 4). Briefly, the working case definition for thrombocytopenia with bleeding for this study requires 2 components:

- A new qualifying bleeding event (defined below) AND
- A new diagnosis of thrombocytopenia (or laboratory evidence of the same, where available) that is made from 7 days before and up to 7 days after that event

Bleeding events will include new diagnoses of clinically significant bleeds in the central nervous system, respiratory tract, gastrointestinal tract, or genitourinary tract, where "clinically significant" means requiring hospitalisation or emergency department care. Both bleeding and TCP will be ascertained using codes for diagnoses, procedures, and treatments in electronic data that are also under development within the studies conducted under the VAC4EU framework. However, as for TTS, definitions and risk window definitions may evolve during the course of the study as more knowledge is gathered about individual AESIs and while definitions are under development within the studies conducted under the VAC4EU framework.

In the primary analysis, the risk window for thrombocytopenia with bleeding extends from 1 to 42 days after vaccination (or the corresponding index date for the unexposed comparator subject). A sensitivity analyses will also evaluate the alternative risk windows of 1 to 28 days, following Pottegård et al (2021). As feasible, cases of thrombocytopenia with bleeding identified from automated data will be reviewed using the approach outlined for TTS (Section 9.3.2.1.1). Information will be sought about the magnitude of thrombocytopenia (ie, platelet count) and any specific cause of the thrombocytopenia.

9.3.3 Covariates

Covariates will be defined at the index date. Covariates will be used to define and describe the study cohorts and population of special interest, their baseline characteristics, and/or to control for confounding. Covariates proposed to define the study population and cohorts are listed in Appendix C, Annex Table 1. Covariates proposed to define subgroups of interest are listed in Appendix C, Annex Table 2. Covariates proposed to define baseline characteristics and to be considered in PS estimation appear in Appendix C, Annex Table 3. A final list of covariates, including time-varying covariates, and their operational definitions will be specified in the SAP.

The look-back observation period to define covariates of interest will be all available information before the index date in each data source, except when an alternative look-back window is specified.

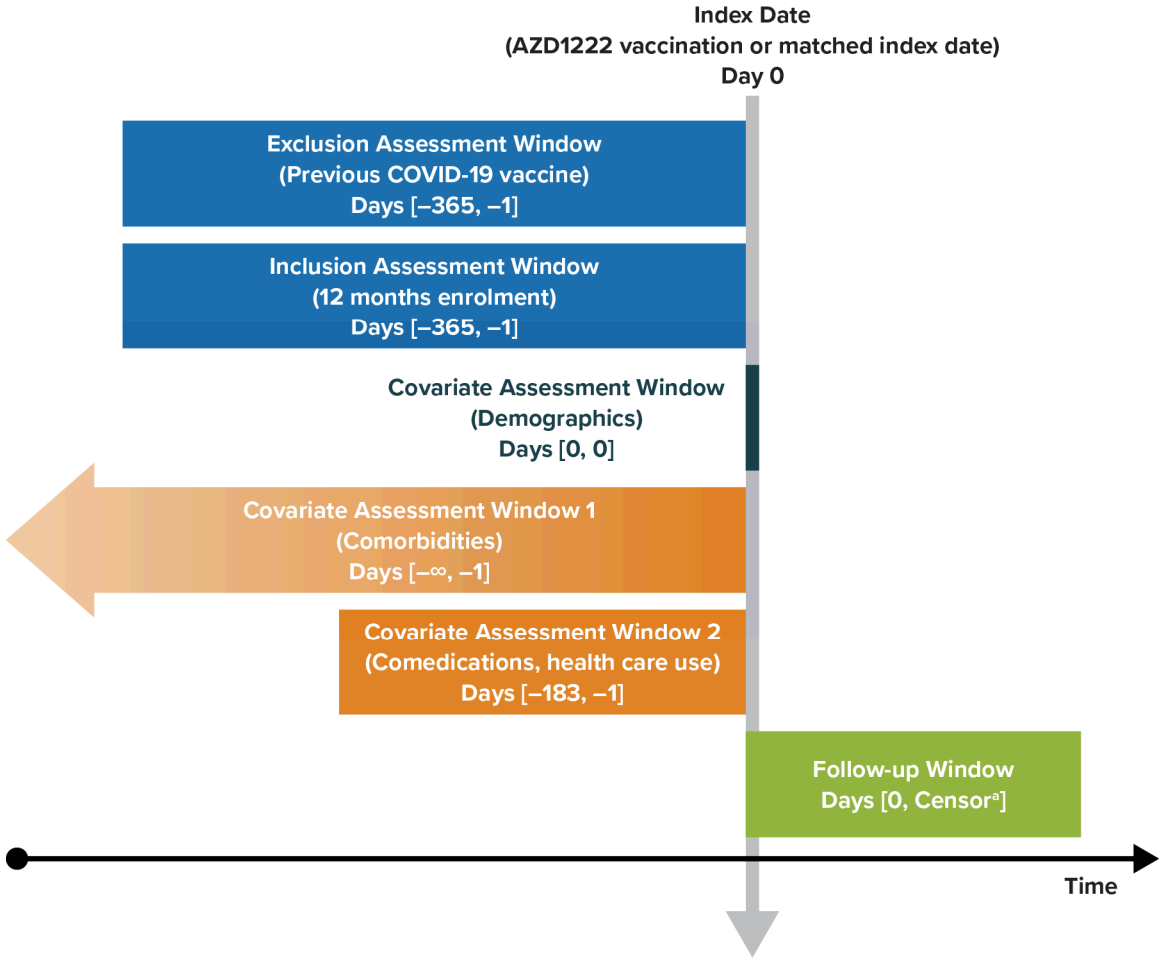
Sociodemographic characteristics and lifestyle factors will be described based on the information closest to the index date as available in each data source. Among the sociodemographic characteristics, age, sex, month/year at the index date, and geographic region are available in all data sources. Race/ethnicity is only available in the CPRD for the majority of patients. In VID, only nationality is captured, while in ARS Toscana no information on race/ethnicity is available. Socioeconomic status information is available in all data sources. Health care or essential worker status is partially available in the CPRD and is fully available in ARS Toscana. However, in VID, only employment status is available. Residence in a long-term facility or nursing home is available in the CPRD, VID, and SIDIAP. Smoking status is available in a high percentage of patients in the CPRD and in SIDIAP and is defined as partially complete in VID and PHARMO. Use of smoking cessation drugs is available in the five data sources. Body mass index (BMI) is only available in the CPRD and is partially available in SIDIAP and PHARMO. Obesity diagnosis and obesity surgery are available in all data sources. History of alcohol abuse is available in the CPRD. Alcohol-related disorders diagnosis codes are available in all data sources. Missing information on these variables will be described.

Medical conditions will be defined based on diagnoses (and medications and/or procedures when applicable) as recorded in the respective data sources included in the study (see Section 9.4) at any time before or on the index date, unless otherwise specified. Absence of a diagnosis code or proxies for a condition will be regarded as absence of the condition.

Use of medications at baseline will be defined based on prescriptions or dispensing as recorded in each data source (see Section 9.4) in the 183 days before or on the index date, unless otherwise specified. Absence of any prescription or dispensing for a medication will be regarded as absence of the medication use at baseline. Note that some variables may not be available in all data sources, may be underrecorded, or may be available only for a subset of the study population. Figure 5 shows the ascertainment windows for covariates in this study, according to their intended use.

The data source-specific definitions of the extraction algorithms of the covariates will be guided and informed through available algorithms from VAC4EU COVID-19 vaccines studies (VAC4EU 2021).

Figure 5 Summary of Covariate Ascertainment in Cohort Analyses



[a] Earliest end date of study period, enrolment termination date, death, vaccination with any SARS-CoV-2 vaccine other than AZD1222, also for each specific AESI: date of the AESI, end of risk window for that specific AESI. For subjects in the comparator cohorts, also the date on which they receive a dose of AZD1222.

Note that information on inclusion, exclusion, and covariate assessment windows will all be available before the index date in each data source, except when an alternative look-back-specific risk window is specified. For comedications and health care utilisation, this time window will be 183 days before the index date, except if a different time window is specified for specific medications or health care utilisation measures.

AESI, adverse event of special interest; COVID-19, coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Source: Original design diagram template can be found at www.repeatinitiative.org/projects.html.

9.4 Data sources

This study will use data from existing secondary automated electronic health data sources that are population based and can provide high quality data on COVID-19 vaccines (ie, product type and dates), outcomes (ie, diagnoses, procedures, and treatments), and important covariates (eg, clinical characteristics, demographics). Data sources have also been selected based on the ability to provide periodic updates during the study period.

The five research partners at the proposed collaborating institutions with access to the data in each of the proposed countries have indicated their interest in participating in the study and have reviewed the study protocol. It also may be possible to increase the study size in the UK by extending the study beyond practices included in the CPRD. Additional linked electronic health data may be available through the OpenSAFELY programme (Grint et al 2021, NHS Digital 2020, Wong et al 2021). This data source may be proposed as exploratory.

9.4.1 CPRD, United Kingdom

The CPRD collates the computerised medical records of a network of GPs in the UK who act as the gatekeepers of health care and maintain patients' life-long electronic health records. The data are sourced from over 2,000 primary care practices and include 50 million patients, of whom 16 million are currently registered and active (https://cprd.com/Data, accessed 1 March 2021). 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 feedback information to GPs about their patients, including key diagnoses. The data in the CPRD are updated monthly and include demographic information, prescription details, clinical events, preventive care, specialist referrals, hospital admissions, and major outcomes, including death (Herrett et al 2015, Wolf et al 2019). Most of the data are coded using Read or SNOMED codes. Data validation with original records (specialist letters) is available.

Currently, approximately 59 million individuals (acceptable for research purposes) are recorded in the CPRD, 16 million of whom are active (ie, still alive and registered with the GP practice)—in over 2,000 primary care practices (https://cprd.com/Data). Depending on the type of electronic medical software used by the general practice, data are collected into either the CPRD GOLD (General Practitioner Online Database) or the CPRD Aurum versions. The data set is generalisable to the UK population based on age, sex, socioeconomic class, and national geographic coverage when the CPRD GOLD and the CPRD Aurum versions are used. Data include demographics, all GP/health care professional consultations, diagnosis and symptoms, results from laboratory tests, information about treatments (including prescriptions), 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). Lag time for the CPRD GOLD and Aurum is 1 month. Information about vaccinations from mass vaccination campaigns during the pandemic is expected to transfer to GPs and into the patient's medical records (via National Health Service [NHS] systems rather than patients informing the GP), however, the time lag is not yet clear.

Linkage of the CPRD primary care data with other patient-level data sets is available for English practices that have consented to participate in the linkage scheme. In more than 80% of the CPRD panel practices, the GPs have agreed to permit the CPRD to link at the patient level to these patient-level data sets. The Hospital Episodes Statistics (HES) database contains

details of all admissions to NHS hospitals in England (Accident and Emergency, Admitted Patient Care, Outpatients); approximately 44.7 million individuals in the CPRD are linked to the HES database. Not all patients in the CPRD have linked data (eg, if they live outside England, if their GP has not agreed that their data should be used in this way). As with standard CPRD patients, HES data are limited to patients who are research standard. The CPRD records are linked to the HES using a combination of the patient's NHS number, sex, and date of birth (Williams et al 2012). Additional CPRD-linked data sets include Death Registration data from the Office for National Statistics (ONS), which includes information on the official date and causes of death (using International Classification of Diseases [ICD] codes), and the Mother-Baby Link, and an algorithm-based Pregnancy Register, which are only available with the CPRD GOLD.

In addition, other CPRD-linked COVID-19 data sets, which may provide further follow-up information on AESIs, include the Public Health England (PHE) Second Generation Surveillance System (SGSS) COVID-19 positive virology test data, PHE COVID-19 Hospitalisation in England Surveillance System (CHESS), and the Intensive Care National Audit and Research Centre (ICNARC) data on COVID-19-related intensive care admissions.

Linked data sets are usually updated every 6 months, and the lag time between data recording and data availability varies by data set. The latest linkage set (set 21) contains an update of priority linkages to support COVID-19 research along with the CPRD-linked SGSS COVID-19 positive virology test data and CHESS hospitalisation and ICU/HDU data up to the end of September 2020. ONS deaths data (up to 16 November 2020), HES APC (up to the end of October 2020), and small area data are also available.

The present study will include only active CPRD practices (GOLD and Aurum) that have agreed to be linked at the patient level with the HES data, as hospital data are critical for ascertainment of most AESIs. These practices include an estimated 13.4 million current patients. The CPRD is listed under the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database, and access will be provided by the Drug Safety Research Unit (DSRU).

9.4.2 VID, Spain

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

- The Population Information System (SIP) is a database that provides basic information on health system coverage (eg, dates and causes of Valencia health system entitlement or disentitlement, insurance modality, pharmaceutical copayment status, assigned health care department) as well as some sociodemographic data (eg, sex, date of birth, nationality,

employment status, geographic location). Importantly, the SIP database includes the date of death captured from the Mortality Registry. The SIP database is paramount to the VID, as it is the source of the individual, exclusive, and permanent identifier number associated with each individual (the SIP number), which is then used throughout the rest of the databases, thereby allowing data linkage across the multiple databases in the network.

- The Ambulatory Medical Record (ABUCASIS) is the electronic medical record for primary and specialised outpatient activity, with 96% population coverage since 2009. ABUCASIS is integrated by two main modules: the Ambulatory Information System (SIA) and the Pharmaceutical Module (GAIA), including paediatric and adult primary care, mental health care, prenatal care, and specialist outpatient services, as well as providing information about dates, visits, procedures, laboratory test results, diagnoses, and clinical and lifestyle information. It also includes information on several health programmes (eg, healthy children, vaccines, pregnancy, notifiable diseases), the primary care nurse clinical record, and the health-related social assistance record. The SIA module uses the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) for coding diagnoses (and, partially, the International Classification of Diseases, 10th Revision, Spanish Edition [ICD-10-ES] from 2019). The SIA also uses the Clinical Risk Groups system to stratify the morbidity of the entire population.
- The GAIA Pharmaceutical module stores data on all outpatient pharmaceutical prescriptions and dispensings, including both primary care and outpatient hospital departments, using the Anatomical Therapeutic Chemical (ATC) classification system and the National Pharmaceutical Catalogue, which allow the identification of the exact content of each dispensing. GAIA does not include in-hospital medication or medication administered in the Accident and Emergency Department (AED). GAIA provides detailed information on prescriptions issued by physicians, such as the duration of treatment and dosage.
- The Hospital Medical Record (ORION) provides comprehensive information covering all areas of specialised care, from admission, outpatient consultations, hospitalisation, emergencies, diagnostic services (eg, laboratory tests, imaging, microbiology, pathology), pharmacy, surgical block including day surgery, critical care, prevention and safety, social work, at-home hospitalisation, and day hospitalisation. ORION is currently in the process of being integrated for the whole region, with several databases already fully integrated and available for all hospitals, including the Minimum Basic Data Set at Hospital Discharge (MBDS) and the AED clinical record.
- The MBDS is a synopsis of clinical and administrative information on all hospital admissions and major ambulatory surgery in the Valencia health system hospitals, including public-private partnership hospitals (approximately 450,000 admissions per year in the region). The MBDS includes admission and discharge dates, age, sex, geographic area and zone of residence, main diagnosis at discharge, up to 30 secondary diagnoses (comorbidities or complications), clinical procedures performed during the hospital episode, and the diagnosis-related group(s) assigned at discharge. The MBDS used the ICD-9-CM system for coding through December 2015 and ICD-10-ES afterward. The MBDS was extended in 2015 to include the “present on admission” diagnosis marker and information on tumour morphology.

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

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

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

Information on pregnancy and pregnancy outcomes is available in the VID database, although some end dates of pregnancy may be missing. However, the mother-baby linkage is not available.

9.4.3 SIDIAP, Spain

The Information System for Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària [SIDIAP]) in Catalonia, Spain, is a primary care database set up by the Institute of Research in Primary Care (Institut Universitari D'Investigació en Atenció Primària Jordi Gol [IDIAP Jordi Gol]) and Catalan Institute of Health (Institut Català de la Salut). The database collects information from 278 primary health care centres and includes more than 5.8 million patients (approximately 78% of the Catalan population) covered by the Catalan Institute of Health and is highly representative of the Catalan population (Willame et al 2021).

The SIDIAP data comprise the clinical and referral events registered by primary care health professionals (ie, 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 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 International Classification of Diseases, 10th Revision (ICD-10) codes, ATC codes, and structured forms designed for the collection of variables relevant to primary care clinical management, such as country of origin, sex, age, height, weight, body mass index, tobacco and alcohol use, blood pressure measurements, and blood urine test results. In relation to vaccines, information on all routine childhood and adult immunisations is included in addition to the antigen and the number of administered doses.

Currently, because of the COVID-19 pandemic, having shorter-term updates to monitor the evolution of the pandemic is a possibility. 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 (http://www.encepp.eu/encepp/viewResource.htm?id=4646).

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

9.4.4 ARS Toscana, Italy

The ARS Toscana is a research institute of the Tuscan regional government. Tuscany is an Italian region with approximately 3.6 million inhabitants. The ARS database comprises all the data that are collected in Tuscany related to health care delivered to those who are official residents of the region. Additionally, ARS collects data tables from regional initiatives.

The ARS database routinely collects primary care and secondary care prescriptions of drugs for outpatient use and is able to link them at the individual level with hospital admissions, admissions to emergency care, records of exemptions from copayment, dispensings of diagnostic tests and procedures, causes of death, and a pathology registry, which has been available for the last few years and includes complete information only for morphology and topography. Occasionally, ARS may request retrieval of information from medical records or laboratory results regarding specific subpopulations and link this information to its core data.

Patients in ARS can be characterised in terms of age, sex, comorbidities (via algorithms), socioeconomic indicators, medication taken regularly on an outpatient basis, date of death, and health care utilisation (including visits to specialists, visits to ambulatory cancer care units, and visits to an emergency department or urgent care centre). Cause of death is available with a lag time of 3 years.

The lag time from a health care encounter to incorporation of data for research purposes is approximately 3 to 4 months. All patients in the ARS database can be linked to mortality data through deterministic linkage. There is no restriction on reporting small numbers.

Participation of the ARS database in a research study includes mandatory compliance with the ENCePP Code of Conduct (ENCePP 2018b). The ENCePP Code of Conduct “provides a set of rules and principles for pharmacoepidemiology and pharmacovigilance studies to promote scientific independence and transparency throughout the research process”.

Information on pregnancy, pregnancy outcomes, and mother-baby linkage is available in the ARS Toscana database.

9.4.5 PHARMO, Netherlands

The PHARMO Database Network, which is maintained by the PHARMO Institute for Drug Outcomes Research, is a population-based network of electronic health record databases that combines anonymous data from different primary and secondary health care settings in the Netherlands. These different data banks—including data from general practices, in- and outpatient pharmacies, clinical laboratories, hospitals, the cancer register, the pathology register, and the perinatal register—are linked on the 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 PHARMO Institute is always seeking new opportunities to link with additional databanks and is currently exploring linkage with the COVID-19 immunisation register, which is collected by the Dutch National Institute for Public Health and the Environment (RIVM).

Currently, the PHARMO Database Network covers over 6 million active persons out of 17 million inhabitants of the Netherlands. 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. The lag time of all databases is 1 year, except for the General Practitioner Database, which is updated every 3 months or less. A detailed description of the different data sources is given below (Willame et al 2021).

- The Hospitalisation Database includes discharge dates, discharge diagnoses, and procedures for hospitalisations longer than 24 hours (or shorter if the patient required a bed). Hospital discharge diagnoses are available from the Dutch National Basic Hospital Care Registration (Landelijke Basisregistratie Ziekenhuiszorg [LBZ]) and are recorded using ICD-10 codes. Procedures are coded according to the Dutch Hospital Data Foundation registration system.
- 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. Primary care data are available for a portion of the population of approximately 3.2 million inhabitants (approximately 20% of the Dutch population). Information on lifestyle variables (eg, body mass index, smoking, alcohol consumption) is available in the General Practitioner Database if recorded by GPs in the electronic medical records.
- 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 Netherlands Perinatal Registry is maintained by Perined and comprises data on pregnancies, births, and neonatal outcomes of births in the Netherlands, which are voluntarily collected by perinatal caregivers, mainly for benchmarking. For the current study, permission has been obtained from PHARMO as well as Perined to link the data with the PHARMO Database Network via the TTP and use the PHARMO Perinatal Research Network (PPRN). Records include information on mothers (eg, maternal age, obstetric history, parity), pregnancy (eg, mode of conception, mode of delivery), and children (eg, birth weight, gestational age, Apgar score). Diagnoses and symptoms are coded according to the Perinatal Registry code lists. Perinatal data are available with a lag time of 1 year.

9.5 Study size

Across the proposed databases, the source population for the study includes approximately 33.8 million subjects (Table 3). All available data in each set will be used to identify subjects who received any SARS-CoV-2 vaccine.

Table 3 Estimates of Study Size and Exposed Population, by Data Source

Data source	Approximate current enrolled population (million)	Cumulative exposure estimate for the data source based on ECDC vaccine country distribution (doses)[a]	Actual exposure data (date)	Other
CPRD Aurum, UK[b]	13.4	4.9 million first doses and 1.8 million second doses	3.7 million patients who have had at least 1 dose (May 2021 database release)	
VID (FISABIO), Valencia, Spain[c]	5	605,662	484,145 first doses and 171 second doses (communication 30 May 2021)	
SIDIAP, Catalonia, Spain[d]	5.8	714,225		Number of exposures up to 02 June 2021 in Catalonia: 916,384 AZ doses
ARS Toscana, Italy[e]	3.6	411,719	197,267 first doses (communication 11 April 2021)	

Table 3 Estimates of Study Size and Exposed Population, by Data Source

Data source	Approximate current enrolled population (million)	Cumulative exposure estimate for the data source based on ECDC vaccine country distribution (doses)[a]	Actual exposure data (date)	Other
PHARMO, Netherlands[f]	3.5	426.725	34,000 first doses (20 May 2021)	Based on GP data only; data from CIMS (Dutch COVID-19 vaccination registration) will potentially become available

[a] Number of doses for EU data sources were estimated using ECDC distribution data per country as of 24 May 2021, multiplied by the proportion of the national population covered in each database. This number also assumes that vaccine distribution in the country is uniform and that all distributed vaccines will be administered (https://qap.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#distribution-tab). For the CPRD Aurum, the estimates are based on vaccine distribution in the UK. See footnote b.

[b] The CPRD Aurum contains information on 13.4 million inhabitants in England and Northern Ireland, representing 20.4 % of the 65.8 million UK population. As of 12 May 2021, an estimated 23.9 million first doses and 9.0 million second doses of the AZ COVID-19 vaccine had been administered (https://www.gov.uk/government/publications/coronavirus-covid-19-vaccine-adverse-reactions/coronavirus-vaccine-summary-of-yellow-card-reporting).

[c] VID: 10.6% coverage among the 5,713,800 total AZ doses administered in Spain as of 21 May 2021, resulting in an estimated 605,662 doses captured in VID.

[d] SIDIAP: 12.5% coverage among the 5,713,800 total AZ doses administered in Spain as of 21 May 2021, resulting in an estimated 714,225 doses captured in SIDIAP.

[e] ARS Toscana: 6% coverage among the 6,861,980 total AZ doses administered in Italy as of 21 May 2021, resulting in an estimated 411,719 doses captured in ARS Toscana.

[f] PHARMO: 20.2% coverage among the 2,112,500 total AZ doses administered in the Netherlands as of 21 May 2021, resulting in an estimated 426,725 doses captured in PHARMO.

ARS, Agenzia regionale di sanità della Toscana; AZ, AstraZeneca; CIMS, COVID-19 vaccination Information and Monitoring System; CPRD, Clinical Practice Research Datalink; COVID-19, coronavirus disease 2019; ECDC, European Centre for Disease Prevention and Control; EU, European Union; FISABIO, Foundation for the Promotion of Health and Biomedical Research of Valencia Region; GP, general practitioner; PHARMO, PHARMO Database Network of the PHARMO Institute for Drug Outcomes Research; SIDIAP, Information System for Research in Primary Care (Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària); UK, United Kingdom; VID, Valencia Health System Integrated Database.

The size of the exposed population in this study will depend on the use of AZD1222 during the study period. Estimates of vaccine uptake are shown in Table 3; these estimates are subject to several assumptions and are difficult to translate into the number of fully vaccinated individuals. Given the uncertainty about the size of the exposed population, we have estimated the precision of potential study results for various study sizes ranging from 1,000,000 to 5,000,000 vaccinated subjects shows the probability that the upper bound of the 95% CI around the observed relative risk will be below 1.5, 2.0, 2.5, and 3.0 for various study sizes and AE rates. These calculations assume a 1-to-4 ratio between vaccinated and comparator

person-time and that the true IRR between those vaccinated and those unvaccinated is 1.0. For the rarest anticipated event, anaphylaxis, assuming a background rate of 6 per 100,000 person-years (PY), 6 days at risk (3 days per dose), and a study size of 5 million vaccinated subjects, we anticipate a 59% probability that the upper bound of the observed IRR would be below 3.0. For the most common outcome, stroke, with a rate of 212 per 100,000 PY, a study size of 500,000 vaccinated subjects would ensure that the upper bound of the 95% CI would be below 1.5, assuming there is no increased risk associated with the vaccine. Table 5 shows the same predicted probabilities as in Table 4, that it assumes a 1-to-1 ratio of vaccinated-to-comparator person-time. For both sets of scenarios, as expected, estimates become more precise with increasing sample size, background rates, risk interval duration, and ratio of vaccinated-to-comparator person-time.

Table 4 Probability that the Upper 95% Confidence Limit of the Observed IRR Will Be Below 1.5, 2, 2.5, and 3 for Various Study Sizes of Vaccinated Subjects, Assuming that the True RR is 1 and the Ratio of Vaccinated-to-Comparator Person-time is 1 to 4

AE (duration of risk interval after each dose)	Estimated background rate per 100,000 PY	People vaccinated	Upper confidence limit of IRR			
			1.5	2.0	2.5	3.0
Anaphylaxis (VAC4EU 2021, Willame et al 2021)	6	1,000,000	0.05	0.09	0.13	0.16
(3 days)	6	2,000,000	0.07	0.14	0.21	0.28
	6	3,000,000	0.09	0.19	0.29	0.39
	6	4,000,000	0.11	0.23	0.37	0.50
	6	5,000,000	0.12	0.28	0.44	0.59
GBS (VAC4EU 2021, Willame et al 2021)	2	1,000,000	0.12	0.26	0.42	0.56
(42 days)	2	2,000,000	0.19	0.47	0.70	0.85
	2	3,000,000	0.27	0.63	0.86	0.95
	2	4,000,000	0.34	0.76	0.94	0.99
	2	5,000,000	0.41	0.84	0.98	> 0.99
Bell's palsy (Morales et al 2013) (42 days)	38	1,000,000	0.92	> 0.99	> 0.99	> 0.99
VTE[a] (Heit 2015) (42 days)	100	1,000,000	> 0.99	> 0.99	> 0.99	> 0.99

[a] This AE is a constituent of the composite AESI "thrombosis."

Note: Calculations assume a true IRR of 1.0 and a ratio of 1-to-4 exposed to unexposed person-years. Calculations also assume that each person contributes 6 days of follow-up for anaphylaxis (3-day risk interval × 2 doses), 56 days for stroke (28-day risk interval × 2 doses), and 84 days (42-day risk interval × 2 doses) for all other AEs (Rothman and Greenland 2018). Calculations follow the method of Rothman and Greenland (2018).

AE, adverse event; AESI, adverse event of special interest; GBS, Guillain-Barré syndrome; IRR, incidence rate ratio; PY, person-years; RR, risk ratio; VTE, venous thromboembolism.

Table 5 Probability that the Upper 95% Confidence Limit of the Observed IRR Will Be Below 1.5, 2, 2.5, and 3 for Various Study Sizes of Vaccinated Subjects, Assuming that the True RR is 1 and the Ratio of Vaccinated-to-Comparator Person-time is 1 to 1

AE (duration of risk interval after each dose)	Estimated background rate per 100,000 PY	People vaccinated	Upper confidence limit of IRR			
			1.5	2.0	2.5	3.0
Anaphylaxis (VAC4EU 2021, Willame et al 2021)	6	1,000,000	0.05	0.07	0.09	0.12
(3 days)	6	2,000,000	0.06	0.10	0.15	0.19
	6	3,000,000	0.07	0.13	0.20	0.27
	6	4,000,000	0.08	0.16	0.25	0.34
	6	5,000,000	0.09	0.19	0.30	0.41
GBS (VAC4EU 2021, Willame et al 2021)	2	1,000,000	0.09	0.18	0.28	0.38
(42 days)	2	2,000,000	0.14	0.32	0.50	0.65
	2	3,000,000	0.19	0.44	0.67	0.82
	2	4,000,000	0.23	0.56	0.79	0.92
	2	5,000,000	0.28	0.65	0.87	0.96
Bell's palsy (Morales et al 2013)	38	1,000,000	0.76	> 0.99	> 0.99	> 0.99
(42 days)	38	2,000,000	0.97	> 0.99	> 0.99	> 0.99
VTE[a] (Heit 2015) (42 days)	100	1,000,000	0.99	> 0.99	> 0.99	> 0.99

[a] This AE is a constituent of the composite AESI "thrombosis."

Note: Calculations assume a true IRR of 1.0 and a ratio of 1-to-1 exposed to unexposed PY. Calculations also assume that each person contributes 6 days of follow-up for anaphylaxis (3-day risk interval × 2 doses), 56 days for stroke (28-day risk interval × 2 doses), and 84 days (42-day risk interval × 2 doses) for all other AEs. Calculations follow the method of Rothman and Greenland (2018).

AE, adverse event; AESI, adverse event of special interest; GBS, Guillain-Barré syndrome; IRR, incidence rate ratio; PY, person-years; RR, risk ratio; VTE, venous thromboembolism.

For the outcome of TTS, study size implications were evaluated in terms of precision of the absolute risk difference (ARD), rather than IRR, because the background risk for this newly described syndrome is unknown and conceivably could be zero (Table 6). Relative risks cannot be quantified when the risk in the reference group is zero. Moreover, effect measures in terms of absolute risk are preferred to relative risk to assess the public health impact of a safety concern and how it may factor into benefit-risk balance. Among vaccinated individuals, the risk has been estimated from reporting rates, which have ranged between 1/100,000 and 1/250,000 per vaccinee (WHO 2021). The duration of the risk period for TTS is not known, but for this exercise, it has been assumed to be 4 weeks; based on case reports published by 11 May 2021, the onset of this outcome appears to occur within 4 weeks of vaccination

(Greinacher et al 2021, Scully et al 2021). Table 6 shows the probability that the upper 95% confidence limit of the ARD for TTS will fall below various thresholds, across a range of different study sizes, assuming a 1-to-1 ratio of vaccinated to unvaccinated subjects. For example, assuming that the true 4-week risk of TTS is 1 per 100,000 vaccinees and that the background risk is zero, with 500,000 people vaccinated there is a 99% probability that the upper confidence limit of the ARD will fall below 3.0 per 100,000. The scenarios displayed in Table 6 were also evaluated assuming that the 4-week incidence of TTS among those who are unvaccinated is 1 per 100,000,000 or 1 per 250,000,000. Results are identical to those shown in Table 5.

Table 6 Probability that the Upper 95% Confidence Limit of the Observed 4-week ARD for TTS Will Be Below Various Thresholds for Different Study Sizes, Assuming that a True 4-week Risk Among Vaccinees is 1/100,000 or 1/250,000

Estimated 4-week incidence among vaccinated individuals	Estimated 4-week incidence among unvaccinated individuals	People vaccinated	Upper confidence limit of ARD (per 100,000)			
			1.5	2.0	2.5	3
1 per 100,000	0	1,000,000	0.35	0.89	> 0.99	> 0.99
		2,000,000	0.61	0.99	> 0.99	> 0.99
		3,000,000	0.78	> 0.99	> 0.99	> 0.99
		4,000,000	0.89	> 0.99	> 0.99	> 0.99
		5,000,000	0.94	1.00	1.00	1.00
			Upper confidence limit of ARD (per 250,000)			
			1.5	2.0	2.5	3
1 per 250,000	0	1,000,000	0.17	0.52	0.85	0.98
		2,000,000	0.29	0.81	0.99	> 0.99
		3,000,000	0.41	0.93	> 0.99	> 0.99
		4,000,000	0.52	0.98	> 0.99	> 0.99
		5,000,000	0.61	0.99	> 0.99	> 0.99

Note: The actual incidence of TTS is not known; estimates here are based on reporting rates among vaccinated subjects (WHO 2021). The actual background incidence of TTS is also unknown. Calculations presented here assume a 1-to-1 ratio of vaccinated to unvaccinated subjects and follow the methods described in Rothman and Greenland (2018).

ARD, absolute risk difference; TTS, thrombosis with thrombocytopenia syndrome.

9.6 Data management

This study will be conducted in a distributed manner using a common protocol, CDM, and common analytics programmes based on routinely collected 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 (findable, accessible, interoperable, and reusable) 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. 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 (eg, 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. Specifications for both ETL and semantic harmonisation will be shared in the catalogue.

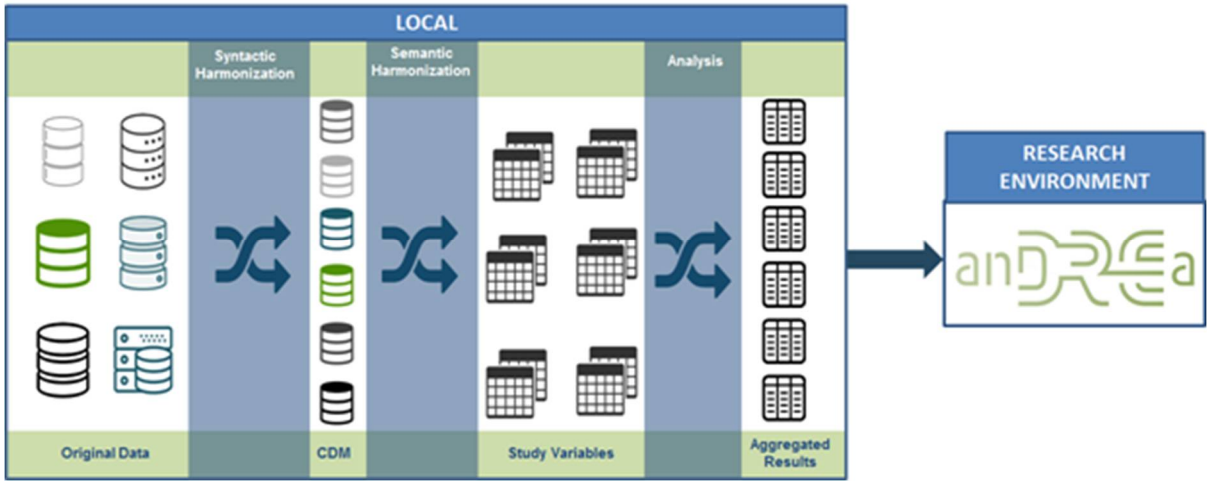
3 Third, following conversion to harmonised study variable sets, additional R and/or SAS (SAS Institute; Cary, North Carolina) scripts for calculation of analytical data sets will be distributed to data access providers for local deployment. The aggregated results produced by these scripts will then be uploaded to the Digital Research Environment (DRE) for pooled analysis and visualisation (see Figure 6). The DRE is made available through 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/).

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

9.6.1 Data extraction

Each database access provider (DAP) will create ETL specifications using the standard ConcePTION ETL design template (accessible via this link: https://docs.google.com/document/d/1SWi31tnNJL7u5jJLbBHmoZa7AvfcVaqX7jiXgL9uAWg/edit). Following completion of this template and review by study statisticians, each DAP will extract the relevant study data locally using their software (eg, Stata, SAS, R, Oracle). These data will be loaded into the CDM structure in csv format. These data remain local (Figure 6).

Figure 6 Data Management Plan



ACCESS, vACCine covid-19 monitoring readinESS; CDM, common data model; COVID-19, coronavirus disease 2019.
Source: Figure from ACCESS Project Protocol: Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines (Version 1.1, 21 September 2020). EUPAS37273. Available at: http://www.encepp.eu/encepp_studies/indexRegister.shtml. Accessed 16 March 2021.

9.6.2 Data processing and transformation

Data processing and transformation will be conducted using R and/or 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 6). Following creation of study variables, the data will be characterised. This characterisation will include calculation of code counts and IRs, 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. 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.

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

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

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

9.6.4 Record retention

The investigators must obtain AstraZeneca'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, final statistical report, statistical programmes, and output files will be archived on a specific and secured central drive.

Documents that individually and collectively permit evaluation of the conduct of a study and the quality of the data produced will be retained for a period of 5 years in accordance with the Guidelines for Good Pharmacoepidemiology Practices (GPP) (ISPE 2015). However, these documents could be retained for a longer period if required by the applicable regulatory requirements or by an agreement between study partners. It is the responsibility of the principal investigator to inform the other investigators and/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.

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 finalised before any comparative analyses begin. The SAP 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.

Analyses will be conducted separately within each contributing database, and pooling will occur using meta-analytic methods, if appropriate, at the end of the study (Section 9.7.10). Data extraction and analysis of AESIs are planned to occur periodically during the study period.

9.7.1 Descriptive analysis of subjects who received AZD1222

The analyses described below will address Primary Objective 1.

Baseline characteristics (eg, demographics, medical history, comedications) of subjects who received at least one dose of AZD1222 will be described for the overall study population and for the specific populations of interest.

To describe potential differences in the baseline characteristics of AZD1222 vaccinees across the study period, descriptions will also be stratified by sequential 3-month periods since the start of the study.

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.

9.7.2 Descriptive analysis of second SARS-CoV-2 vaccine use

The analyses described below will address Primary Objective 2.

Among subjects who received dose 1 of AZD1222, descriptive measures will be used to characterise patterns of second COVID-19 vaccine use over the study period. These distributions will be described overall and in sequential 3-month periods since the start of the study period.

Description of the patterns of second SARS-CoV-2 vaccine will include the following:

- Counts and proportions of subjects who received only the first dose of AZD1222
- Counts and proportions of subjects who received a second dose of SARS-CoV-2 vaccine, specifying whether it was the AZD122 vaccine or another SARS-CoV-2 vaccine and type (Brand name)
- Among those who received two doses: the time (days) between the first and second doses will be described as a continuous variable (mean, standard deviation, median, quartiles, min, and max) as well as categorically (< 2, 2-4, 5-8, 9-12, 13-18, > 18 weeks)

Patterns of second COVID-19 vaccine use will be described for the general study population and subgroups of special interest.

9.7.3 Selection of the population to describe and compare incidence of AESIs

For the cohort study, exposure PS will be used to refine the initial matched comparator populations for use in analyses to describe and compare incidence of AESIs. This approach is proposed to improve exchangeability, reduce non-positivity, and control for confounding (Webster-Clark et al 2021). Propensity scores will be estimated separately within each database. The PS will serve as a within-database variable that summarises the confounding from a large set of variables. This advantage may be important in this study, given the low number of events for many of the study endpoints expected in the study population (Austin 2014, Cepeda et al 2003, Webster-Clark et al 2021).

9.7.3.1 Selection of covariates to estimate propensity scores

The strategy for selecting candidate covariates to estimate PS focuses on variables related to the probability of receiving the vaccine as well as those associated with the risk of one or more of the study outcomes. Candidate variables are listed in Appendix C (Annex Table 3). It is proposed to estimate only one PS for the analysis of all outcomes. Using a single PS model to control for confounding in settings with more than one outcome, a generic outcome PS model, is preferred in terms of precision and bias. These generic outcome models include all covariates affecting any of the outcomes, while excluding covariates that only affect treatment (Brookhart et al 2006, Wyss et al 2013).

The PS estimated in the full cohort has also been shown to be a valid approach for inference within subgroups given a minimum number of patients and events. This one PS will serve to adjust for confounding in the analysis of all AESIs, including those AESIs that have specific exclusion criteria and to adjust for confounding in the subgroup analyses (Girman et al 2014, Rassen et al 2012).

9.7.3.2 Estimation of propensity scores

Logistic regression models will be used to estimate PS. From the fitted PS model, a PS will be estimated for each subject for each potential index date the subject was included in the model. Propensity score distributions for each cohort will be produced.

9.7.3.3 Selection of cohorts for comparative analyses and assessment of balance

After PS have been estimated, they will be used to exclude noncomparable subjects and refine the balance between study cohorts, which were initially matched on calendar date, age, and gender, using an approach that will be specified in the SAP. Whether all potential comparators will be included, or only a sample, will depend on the approach selected, as well as data governance rules and data management capabilities at each research partner, which are to be determined. Covariate balance will be assessed using the standardised differences and will be displayed graphically. If balance cannot be achieved for all important covariates, an option would be to refit the model to include interactions or higher-order terms to improve balance.

Any attrition of vaccinated individuals and comparators excluded during final cohort assembly will be described. For example, to satisfy assumptions about exchangeability, it may be necessary to exclude subjects with extremely high or low PS. Baseline characteristics of the final study cohorts will be presented as well as information about the potentially excluded vaccinated subjects.

9.7.4 Exclusions for analysis of specific AESIs

In general, the entire population identified in the preceding section (Section 9.7.3) will be included in the analysis of IRs and estimates of relative risk. However, in the analysis of specific AESIs, certain subjects who experienced the outcome in the recent past will be excluded from that analysis. The rationale for this approach is to be able to differentiate ongoing care for the historical event from a truly recurrent or incident event. Definition of the "recent past" (or "event-free interval") will vary by outcome, and proposed specifications appear in Table 2. Some AESIs are chronic conditions, characterised by daily symptoms, that may not have discretely identifiable clinical recurrences, such as chronic fatigue syndrome or fibromyalgia. Subjects with a history of these conditions at baseline will be excluded from analysis of their incidence and relative risk. The SAP will also specify such conditions. Due to these exclusions, the exact composition of the analysis population may vary in the evaluation of different AESIs. The number of patients excluded from each cohort for each analysis will be reported.

9.7.5 Missing data

Because the underlying data represent attended medical care, we generally assume that the absence of information of clinical events or prescriptions/dispensings indicates an absence of the condition or the treatment. Therefore, no missing data on diagnosis and medications are expected.

Missingness is only expected for some lifestyle or biometric data, such as smoking status or body mass index in primary care electronic medical records. In the other data sources, these variables are not captured and will be defined based on proxies that do not have missing data (ie, use of smoking cessation drugs or diagnosis of obesity). In the main analysis, when describing a variable, patients with missing values will be reported as a separate category.

The presence of missingness will be analysed. In the case of < 10% of missingness, a complete-case analysis is proposed for the main analysis. A complete-case analysis consists of performing the adjusted analyses described in the next subsections only among patients with no missing values, under the assumption that data are missing completely at random. However, in the case of > 10% of missingness for key covariates, multiple imputation will be explored.

9.7.6 Analysis of incidence rates

The analyses described below will address Primary Objective 3.

For AESIs for which the risk interval is characterised, crude incidence rates (IRs), and 95% CIs of each of the AESIs will be estimated for the vaccinated cohort and comparator cohort. These analyses will combine person-time at risk after dose 1 and dose 2 (if it was received), using definitions of time at risk presented in Section 9.2.3.1. Exploratory analyses will report the IRs by specific dose.

For AESIs with unknown risk windows, cumulative incidence will be estimated using Kaplan-Meier survival curves starting at the index date. For the vaccinated cohort, we will potentially also repeat these analyses stratified by vaccine dose number. In addition, for AESIs with unknown risk windows, we will also describe the IR over sequential time periods (eg, fortnights) to explore potential changes in risk over the follow-up period.

In the primary analysis, IRs and cumulative incidence of AESIs will be described for the vaccinated cohort selected for comparative analyses (Section 9.7.3.3); secondary analysis will present these results for the entire vaccinated population described in Section 9.7.1.

9.7.7 Crude and adjusted measures of association: cohort design

The analyses described below will address Primary Objective 4.

The main analysis will compare patients vaccinated with AZD1222 with unvaccinated subjects to estimate the effect of receiving any dose (a first or a second), under the assumption that the effect of either dose is homogeneous. The effect of each individual dose on AESIs, where appropriate, will be estimated as secondary analyses, as prior vaccine exposure may "prime" the immune system and because patients who receive the second dose may have different characteristics than those who only received the first dose.

For outcomes that have well-defined risk intervals, we propose to estimate crude and adjusted IRRs and IR differences with 95% CIs using Poisson regression models. Specifically, the IRRs would be estimated using Poisson regression with robust estimation of the variance to account for individuals who may contribute person-time at different points to the exposed and unexposed cohorts (Zou 2004).

For endpoints that do not have well-characterised risk intervals, we propose estimating crude and adjusted relative risks and absolute risk differences. The relative risk (HRs) and 95% CI will be estimated using Cox regression models with robust estimation of the variance time-to-event analyses (Allison 2010). Using Cox regression models allows the flexibility to potentially adjust for time-dependant covariates, such as intercurrent COVID-19 infection.

Adjustment for confounding will occur using PS, either through PS matching or by analytic methods involving stratification or weighting, which will be specified in the SAP. In the event that important covariates remain imbalanced after final selection of the cohorts for comparative analyses, additional approaches for confounding adjustment will be considered, including double-robust methods (ie, including the covariate in the outcome model as a separate term) or through analyses that restrict the population to subjects with more similar baseline characteristics (eg, subjects without substantial comorbidity).

9.7.8 Crude and adjusted measures of association: SCRI design

The analyses described below will address Primary Objective 4.

For comparative analysis using the SCRI approach, conditional Poisson regression will be used to estimate IRRs and 95% CIs. The primary SCRI analysis will include all qualifying events following dose 1 or dose 2 of AZD1222 (see Sections 9.2.1.3 and 9.2.3.2). Exploratory analyses may stratify by dose. Time-invariant confounders will be inherently adjusted for using the SCRI design, however, time-varying confounders may be included as covariates in regression models. Definition of any time-varying confounders and analytic approach will be specified in the SAP.

9.7.9 Subgroup analyses

Subgroup analyses of special populations of interest (Section 9.2.1.2) will be conducted using the analogous approaches used for the general study population described in Sections 9.7.1 (*Secondary Objective 1*), 9.7.2 (*Secondary Objective 2*), 9.7.6 (*Secondary Objective 3*); and 9.7.7 and 9.7.8 (*Secondary Objective 4*). No dose-specific analyses are planned for subgroups, given the expected limited precision of estimates.

Subgroup analyses may also be conducted based on age groupings.

No quantitative evaluations of effect modification are planned, however, for each AESI, forest plots will be generated for each subgroup and for the general population to facilitate visual comparison.

9.7.10 Meta-analysis

Main estimates of association from the participating data sources will be pooled using meta-analytic techniques. Crude and/or adjusted IRs, IRRs, and HRs along with 95% CIs will be estimated.

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

Some of the analysis may be limited due to a small number of events and/or data privacy driven cell count restrictions at each research partner.

9.7.11 Progress and interim analysis

A first progress report will be submitted to the regulatory agencies to update the status of the data source participation being explored at the time of the protocol endorsement, to confirm that data extractions and analysis as well as other study activities are being conducted as planned, and to discuss the impact of the changes in the uptake of the vaccine on the study and whether the inclusion of other databases should be explored.

Interim reports will include information from the study period available at the time of each data extraction, a description of the number of subjects vaccinated with one or two doses of AZD1222, the number of subjects vaccinated with one dose of AZD122 and a second dose of another SARS-CoV-2 vaccine, IRs of the AESI, and comparative analyses (when the data have been accrued in the database and are available to support those analyses with sufficient precision).

The study periods to be included in each interim report will depend on the database, linkages, lag times, and the time required to obtain the data by the research institutions. In addition, for the first interim report, study period will also depend on time needed to allow for protocol endorsement by the EMA, contracting between research institutions, a minimum set of analytical instructions (or the whole SAP), data extraction, analysis, and reporting. For AESIs that require linkage with other databases such as hospital database or cause of death registers, time lags may be longer and may not be included in the first interim report.

Descriptions of the cohorts and estimations of IRs and any other measures will be performed as described in Section 9.7.

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.

CPRD (UK)

The DSRU has information security policies in place to preserve the confidentiality, integrity, and availability of the organisation's systems and data by ensuring the following: 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 Guideline on Good Pharmacovigilance Practices (GVP) (EMA 2017) and according to the ENCePP Code of Conduct (ENCePP 2018b). 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, eg, the legal basis for data collection, data flows, data access rights, data retention periods
- Routine data cleaning to screen for errors, missing values, and extreme values and diagnose their cause
- System process logs to document staff access and other items

FISABIO (Spain)

After Ethical Review Board approval, raw data will be extracted in text file format and will undergo a data quality check. Data will be stored on secure servers at FISABIO (Foundation for the Promotion of Health and Biomedical Research of Valencia Region) in accordance with Spanish and data protection requirements and ensuring that no identifiable data will be stored longer than required.

All the procedures that will be implemented for data collection, storage, protection, retention, and destruction will comply with national and EU legislation. The research team will stay up to date with the detailed provisions of the EU GDPR, which came into force in May 2018, and which will supersede national legislation within the 28 EU Member States.

SIDIAP (Spain)

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

ARS Toscana (Italy)

One or two researchers will review the study documents. ARS Toscana receives data on a bimonthly basis from the Tuscany region, and these data then undergo a first QC. The ARS Toscana statistical office appends the data to an Oracle database and checks them using a dashboard to identify any inconsistencies with historical data.

The Pharmacoepidemiology 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 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.

PHARMO (Netherlands)

PHARMO is ISO 9001:2015 certified for its quality management system. At the study level, all aspects of the study from protocol development to the reporting of the results are conducted within the work frame of the PHARMO quality management system.

The PHARMO Database Network combines data from different sources (eg, pharmacy, hospital, laboratory). These different sources are probabilistically linked through validated algorithms to ensure that patient privacy is maintained. Before databases are linked, those patients for whom linkage-critical information (eg, date of birth, gender, GP) is missing are removed. All data are handled in a way that meets the full requirements for managing and storing sensitive patient data. Involved researchers have signed a confidentiality agreement. The anonymised data are stored on an internal network drive. Relevant extractions will be stored in a project folder. Specific checks on the linked data are performed, depending on which data sources are used. The study data folder, including all extracted and derived data tables, will be archived after study closure.

All programming is developed in accordance with standard operating procedures, prepared by the lead analyst, and reviewed/quality controlled by an experienced analyst at PHARMO. Additionally, all results and reports are audited by the QC department, using a standardised check list.

The use of the PHARMO data is controlled by the independent Compliance Committee STIZON/PHARMO Institute. The Compliance Committee STIZON/PHARMO Institute consists of representatives of the participating data suppliers and a privacy expert (chairman of this Committee). Each study requires permission from this Committee, according to the applicable legislation in the Netherlands, eg, the Personal Data Protection Act and the Medical Treatment Contract Act. Within this legal framework, the Code of Conduct "Use of Data in Health Research" is an important document for the interpretation of the use of this kind of data for scientific research in the Netherlands and is approved by the Dutch Data Protection Authority (www.dutchdpa.nl).

RTI Health Solutions

At RTI Health Solutions, as the coordination 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.

All key study documents, such as the analysis plan and study reports, will undergo QC review, senior scientific review, and editorial review.

Procedures will be consistent with the International Society for Pharmacoepidemiology (ISPE) GPP (ISPE 2015).

A quality assurance audit of this study may be conducted by the Sponsor or the Sponsor's designees.

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.

9.9.1 Bias and confounding

Misclassification bias can occur when exposure or outcome status are ascertained with error. Exposure identification may be based on pharmacy dispensing records, general practice records, immunisation registers, medical records, or other secondary data sources. Currently, the ability to identify specific COVID-19 vaccine products and dates of vaccination in these data sources is only incompletely understood. It is likely that some subjects vaccinated outside the health care system will not be recorded in secondary electronic health records databases; this could occur, for example, if health care workers receive the vaccination in the workplace rather than through usual clinical care. The general direction of misclassification would be that some subjects who appear unexposed actually are exposed. Furthermore, the extent to which vaccine brands and batch numbers/lot numbers will be available in the secondary data sources is unknown. To the degree that vaccination registries are in place that capture immunisation history regardless of care setting, the potential for such misclassification will be ameliorated. Concerns about undetected vaccination are addressed in part through use of the SCRI design, which includes only individuals who are known to be exposed.

Outcome definition relies on the accuracy of codes and algorithms to identify outcomes and the data available in each data source. To reduce the probability of **outcome misclassification**, this study plans to use the codes and algorithms proposed by ACCESS. These codes and algorithms are based on literature reviews, have been reviewed by epidemiology and medical experts, and have been used in the "Background rates of AESI protocol" (EUPAS 37273). If feasible, identification algorithms will be validated in a sample of cases based on manual review of electronic records or chart abstraction conducted by clinicians blinded to COVID-19 vaccine exposure. Bias analyses may be conducted to evaluate the potential effect of misclassification.

This study is potentially subject to **detection bias**, which can occur when knowledge of exposure leads to differential intensity of outcome ascertainment. A vaccinated individual may be more attuned to the possibility of an adverse reaction following vaccination and have a different threshold for seeking medical evaluation of symptoms than someone who is unvaccinated. Physicians who are aware of a patient's vaccination status may also have a different threshold for evaluating nonspecific findings in the postvaccination period. We suspect that certain outcomes may be more subject to detection bias, particularly those with a more gradual onset or early symptoms that can be nonspecific, such as autoimmune thyroiditis. In contrast, someone with a new-onset seizure nearly always seeks prompt medical attention. Concerns that publicity about potential adverse effects may lead to differential

ascertainment over time can be addressed through inspection of IRs in sequential periods during the study.

Uncertainty about risk periods may lead to misclassification and potential attenuation of risk estimates in the cohort design, particularly in the SCRI design. For AESIs with gradual onset, such as multiple sclerosis or peripheral neuropathies, the date of disease onset will likely precede the date of first recorded diagnosis; for events diagnosed shortly after vaccination, there is the possibility that the order of events (exposure/outcome) may actually be reversed. These concerns can be partially addressed through sensitivity analyses that introduce lag intervals. Approaches to address this limitation will be discussed in the SAP.

For AESIs with known risk intervals, the main cohort analysis will pool risk estimates from risk windows after the first and second doses of AZD1222. Under the assumption that the effect of a first or second dose in both populations is homogeneous, this pooling will increase study size and statistical precision. If this assumption is inaccurate, eg, because receiving a first dose sensitises the immune system to react against a second dose, the estimates of the main analysis will be biased. This factor may be especially important when evaluating immune-mediated AESIs, such as anaphylaxis. To assess whether this occurs, we will evaluate the IRs after each dose and if there are substantial differences, we will report the relative effect measures separately.

For AESIs evaluated using SCRI, it is assumed that the occurrence of the event must not censor or affect the observation period, however, if such events are frequently fatal, this may relevantly impact the evaluation. To assess whether this outcome occurs, we will evaluate and report the number of fatal AESI that led to a censoring of the risk interval.

This study plans to estimate the risk of AESIs among subjects who have been vaccinated compared with unvaccinated with similar health conditions. Subjects who do not receive vaccination (or receive it later) may have different health status and patterns of health care compared with those who are vaccinated. These differences may include a higher prevalence of frailty or underlying chronic conditions (Webster-Clark et al 2021) among vaccinees. In studies involving vaccines, special attention should be given to bias related to **healthy vaccinee bias** and **confounding by indication** (Remschmidt et al 2015). While healthy vaccinee bias refers to the situation in which subjects who have better health conditions are more likely to be vaccinated; thus, confounding by indication may be present if subjects with underlying chronic diseases are more likely to be vaccinated than healthy study subjects. Although both situations could occur in this study, it is more likely that vaccination, at least at the beginning of the study period, will be preferentially directed to subjects with more underlying conditions. To address these potential bias/confounding issues, the current study is designed to first match subjects in the vaccinated cohort with comparators of the same age,

sex, and at the same calendar date, and then use PS to adjust for potential confounders, including patterns of prior health care use.

It is expected that the patient groups targeted for vaccination will change over the course of the study in all countries. Matching on calendar date and age will address this concern and may also reduce potential confounding due to changing patterns of SARS-CoV-2 infection (including the potential for herd immunity to develop over time) and for changes in health care-seeking behaviours that have been observed during the pandemic. Being unvaccinated may also be related to lifestyle choices that are difficult to measure. It is unlikely that the data sources will have information on all potential confounders. Failure to account for variables that are unmeasured (or imperfectly measured) may lead to **residual confounding** (Webster-Clark et al 2021). To address potential confounding, the SCRI, which automatically adjusts for time-invariant confounders, will be used as a secondary approach, and using only postvaccination control periods will avoid the potential bias that use of prevaccination control periods may introduce, as prevaccination health events may affect the probability of exposure. However, the SCRI design is not well suited to study outcomes with gradual onset, long latency, or risk periods that are not well known.

It is conceivable that over the study period, as more of the population has received vaccines to prevent COVID-19, the population of potential **comparators will dwindle** or that people who remain unvaccinated may be systematically different from those choose to be vaccinated. Additionally, the age restrictions in the use of AZ1222 imposed in the UK, Spain, Italy, or the Netherlands may make it difficult to find comparable controls for the cohort study. If such a scenario develops, it is proposed to use the SCRI analysis, which only includes exposed individuals, as the primary analysis to evaluate vaccine/AESI associations for outcomes that are suitable for this design (see Table 2). Using recent historical control subjects could also be considered. Such an approach, however, introduces other potential biases given that health care-seeking behaviours have been shown to differ substantially between the pre- and post-COVID-19 eras. In the first year of the current study, the availability of comparators and the duration of their follow-up time will be monitored; if it is judged to be insufficient for reasonable inferences in comparative analyses, the protocol may be amended to include a secondary historical comparator cohort for the study outcomes for which self-controlled designs are not suitable.

Finally, data for some of the covariates proposed to define the subgroups of interest included as missing information in the RMP may not be complete. Patients who are immunocompromised will be identified using several diagnoses of diseases that involve immunosuppression and use of immunosuppressive medications. It is acknowledged that availability of information on immunosuppressive drugs that are given in the hospital or recorded in separate databases will be limited, but whenever possible, this information will be used and ascertainment of the covariates to identify patients who are immunosuppressed will be complemented using information on diagnosis and procedures. Identification of haematological malignancies will be complemented whenever possible with data from all data available in each data source, including cancer registries. For the CPRD, for example, it is known that primary care data on neoplasias are of limited completeness, however, when neoplasias are also ascertained using data from the hospital (HES) and from the linked cancer registry, completeness of the diagnosis is high (Margulis et al 2018). Similarly, human immunodeficiency virus (HIV) infection is known to be incompletely recorded in some databases such as the CPRD, where only around 55% to 67% of patients who are HIV positive have a diagnosis in the primary care database (Gompels et al 2019). To overcome this limitation, all data available in each data source will be used, including, eg, hospital data (HES) in the CPRD, and data prescriptions or dispensing of anti-retroviral medications that are specific to the disease and expected to be available. Additionally, the identification of women who are pregnant or breastfeeding might be challenging in the selected data sources. To identify pregnancy onset, information on delivery or birth dates, proxies for them (eg, delivery hospital admission date), or other valid methods will be used for estimation (Margulis et al 2015). To identify women who are breastfeeding, the mother-baby linkage and other approaches will be used.

9.9.2 Challenges to interpretation

This study involves multiple endpoints, which will be evaluated using several analytic approaches, including evaluation in several subgroups. The probability of **false positive results** increases as the number of comparisons increases. Any positive association will be interpreted in the context of the totality of evidence.

The study size will be limited by the number of subjects who receive AZD1222 and who do not receive any vaccination. The purpose of using multiple databases is to maximise the accrual of vaccinated subjects before the entire population has been vaccinated. Precision of the estimates for some AESIs may be low, especially in subgroup analyses or for very rare endpoints, such as anaphylaxis.

9.9.3 Generalisability of study findings

Generalisability from the study findings will depend on the category of the finding (Brunelli et al 2013, Rothman 2014). Findings related to vaccine utilisation and patient characterisation will apply to the patient populations included in the study. The AZD1222 vaccine is approved

by the EMA and MHRA to be used among adults of all ages, however, some countries may have additional recommendations or vaccination strategies that would limit the population who will receive the vaccine.

The study will provide insight into the situation during the pandemic period, where contact patterns depend on governmental rules and where these rules vary both from country to country and from period to period. Accordingly, it may be challenging to generalise some results of the study to other countries and time periods.

9.10 Other aspects

None.

10. PROTECTION OF HUMAN SUBJECTS

The proposed study is a noninterventional study reusing health care data. All data collected in the study will be de-identified with no breach of confidentiality with regard to personal identifiers or health information. Each data access provider will apply for an independent ethics committee review according to local regulations, and the local data protection impact assessment should be informed. Data protection and privacy regulations (GDPR) should be respected in collecting, forwarding, processing, and storing data from study participants.

10.1 Other good research practice

This study will adhere to the Guidelines for GPP and has been designed in line with the ENCePP Guide on Methodological Standards in Pharmacoepidemiology (ENCePP 2018c). The ENCePP Checklist for Study Protocols (ENCePP 2018a) will be completed (see Appendix B).

The study is a post-authorisation study of vaccine safety and will comply with the definition of the noninterventional (observational) study referred to in the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use tripartite guideline Pharmacovigilance Planning E2E (ICH 2004) and provided in the EMA Guidelines on Good Pharmacovigilance Practices (GVP) Module VIII: Post-Authorisation Safety Studies (EMA 2017), and with the 2012 EU pharmacovigilance legislation, adopted on 19 June 2012 (European Commission 2012). The study will comply with the study reporting requirements specified in Module VIII Section VIII.B.6.3.1, “Progress reports” and VIII.B.6.3.2. “Final Study Report” of the Guideline of Good Pharmacovigilance Practices (EMA 2017).

The study will be registered in the EU PAS Register (ENCePP 2021) before the study implementation commences.

The study will be conducted according the ENCePP Code of Conduct (ENCePP 2018b). Adherence to the ENCePP Code of Conduct supports the broadest participation requirements for European research partners with access to data instrumental for this project and is in line with VAC4EU requirements.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

Based on current guidelines from ISPE (ISPE 2015) and the EMA GVP (EMA 2017), noninterventional studies using secondary data such as those described in this protocol, conducted using medical chart reviews or electronic claims and health care records, do not require expedited reporting of AEs/reactions (EMA 2017, ISPE 2015).

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

The study protocol, study progress reports, interim reports, and final study report will be included in regulatory communications in line with the RMP, periodic benefit-risk evaluation report, and other regulatory milestones and requirements. Study reports will be prepared using a template following the GVP Module VIII Section B.6.3 (EMA 2017).

The progress report will include status updates (ie, progress against milestones, number of vaccinees) and will report and address any challenges in the progress of the project. For information on the content of the progress, interims, and final study reports, refer to Progress and Interim Analysis Section 9.7.11.

In its GPP, the ISPE contends that “there is an ethical obligation to disseminate findings of potential scientific or public health importance” (ISPE 2015), eg, results pertaining to the safety of a marketed medication. Study results will be published following guidelines, including those for authorship, established by the International Committee of Medical Journal Editors (ICMJE 2018). When reporting results of this study, the appropriate Strengthening the Reporting of Observational Studies in Epidemiology Checklist will be followed (von Elm et al 2008). The Consolidated Standards of Reporting Trials statement (Schulz et al 2010) refers to randomised studies but also provides useful guidance applicable to non-randomised studies.

Communication via appropriate scientific venues (eg, ISPE) will be considered. The Marketing Authorisation Holder (MAH) and the investigators will agree upon a publication policy: the principal and coinvestigators will coauthor scientific manuscript(s) of the results to be published, irrespective of data ownership. In line with EMA GVP Module VIII, the research team will have independent publication rights. The MAH will be entitled to view the results and interpretations included in the manuscript(s) and provide comments before

submission of the manuscript(s) for publication (EMA 2017), however, final decisions rest with the research team.

13. REFERENCES

ACCESS Project. Safety evaluation of COVID-19 vaccines in electronic healthcare databases: a protocol template from the ACCESS project. 2020. Available at: https://vac4eu.org/wp-content/uploads/2021/02/3c.Protocol_ACCESS_Safety-Evaluation-EHR.pdf. Accessed 1 March 2021.

Allison PD. Survival analysis using SAS®: a practical guide, second edition (chapter 8, page 266). Cary, NC: SAS Institute Inc.; 2010.

Arnold LM, Bennett RM, Crofford LJ, Dean LE, Clauw DJ, Goldenberg DL, et al. AAPT diagnostic criteria for fibromyalgia. J Pain. 2019 Jun;20(6):611-28.

Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat Med. 2014 Mar 30;33(7):1242-58.

Brighton Collaboration. Case definition of thrombosis and thromboembolism (draft v0.1). 15 March 2021a. Available at: https://brightoncollaboration.us/wp-content/uploads/2021/03/CASE-DEFINITION-OF-THROMBOSIS-AND-THROMBOEMBOLISM_DRAFT.15MAR.2021.pdf. Accessed 21 May 2021.

Brighton Collaboration. Updated proposed Brighton Collaboration process for developing a standard case definition for study of new clinical syndrome X, as applied to thrombosis with thrombocytopenia syndrome (TTS). 18 May 2021b. Available at: https://brightoncollaboration.us/wp-content/uploads/2021/05/TTS-Interim-Case-Definition-v10.16.2-May-18-2021.pdf. Accessed 21 May 2021.

Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol. 2006 Jun 15;163(12):1149-56.

Brunelli SM, Gagne JJ, Huybrechts KF, Wang SV, Patrick AR, Rothman KJ, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. Pharmacoepidemiol Drug Saf. 2013 May;22(5):542-50.

CDC. Update: Thrombosis with thrombocytopenia syndrome (TTS) following COVID-19 vaccination. Advisory Committee on Immunization Practices (ACIP); 2021. Available at: https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2021-05-12/07-COVID-Shimabukuro-508.pdf. Accessed 22 May 2021.

Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol. 2003 Aug 1;158(3):280-7.

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/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-viii-post-authorisation-safety-studies-rev-3_en.pdf. Accessed 15 December 2020.

ENCePP. ENCePP checklist for study protocols (revision 4). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; 15 October 2018a. Available at: http://www.encepp.eu/standards_and_guidances/checkListProtocols.shtml. Accessed 13 February 2021.

ENCePP. The ENCePP code of conduct for scientific independence and transparency in the conduct of pharmacoepidemiological and pharmacovigilance studies (Revision 4). 15 March 2018b. Available at: http://www.encepp.eu/code_of_conduct/. Accessed 13 February 2021.

ENCePP. EU PAS Register. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; 28 March 2021. Available at: http://www.encepp.eu/encepp_studies/indexRegister.shtml. Accessed 19 March 2021.

ENCePP. Guide on methodological standards in pharmacoepidemiology (EMA/95098/2010 Rev. 7). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; July 2018c. Available at: http://www.encepp.eu/standards_and_guidances/. Accessed 13 February 2021.

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. 2012. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:159:0005:0025:EN:PDF. Accessed 25 February 2014.

FDA. Biologics Effectiveness and Safety (BEST) initiative. COVID-19 vaccine safety surveillance: active monitoring master protocol. Center for Biologics Evaluation and Research (CBER); 16 December 2020 2020. Available at: https://www.bestinitiative.org/wp-content/uploads/2020/12/C19-Vaccine-Safety-Protocol-2020.pdf Accessed 13 February 2021.

Garcia-Sempere A, Orrico-Sanchez A, Munoz-Quiles C, Hurtado I, Peiro S, Sanfelix-Gimeno G, et al. Data resource profile: the Valencia Health System Integrated Database (VID). Int J Epidemiol. 2020 Jun 1;49(3):740-1.

Girman CJ, Gokhale M, Kou TD, Brodovicz KG, Wyss R, Stürmer T. Assessing the impact of propensity score estimation and implementation on covariate balance and confounding control within and across important subgroups in comparative effectiveness research. Med Care. 2014 Mar;52(3):280-7.

Gompels M, Michael S, Davies C, Jones T, Macleod J, May M. Trends in HIV testing in the UK primary care setting: a 15-year retrospective cohort study from 2000 to 2015. BMJ Open. 2019 Nov 24;9(11):e027744.

Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. N Engl J Med. 2021 Apr 9.

Grint DJ, Wing K, Williamson E, McDonald HI, Bhaskaran K, Evans D, et al. Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England, 16 November to 5 February. Euro Surveill. 2021 Mar;26(11):2100256.

Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015 Aug;12(8):464-74.

Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun;44(3):827-36.

ICH. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Pharmacovigilance planning. E2E. 2004. 2004. Available at: http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/pharmacovigilanceplanning.html. Accessed 13 February 2021.

ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors; December 2018. Available at: http://www.icmje.org/recommendations/. Accessed 17 January 2019.

IOM. Beyond myalgic encephalomyelitis/chronic fatigue syndrome: redefining an illness. Washington, DC: Institute of Medicine of the National Academy of Sciences; 2015. Available at: https://www.nap.edu/resource/19012/MECFScliniciansguide.pdf Accessed 22 May 2021.

ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Revision 3. International Society for Pharmacoepidemiology. International Society for Pharmacoepidemiology; June 2015. Available at: https://www.pharmacoepi.org/resources/policies/guidelines-08027/. Accessed 20 January 2021.

Lee GM, Greene SK, Weintraub ES, Baggs J, Kulldorff M, Fireman BH, et al. H1N1 and seasonal influenza vaccine safety in the vaccine safety datalink project. Am J Prev Med. 2011 Aug;41(2):121-8.

Li F. Structure, function, and evolution of coronavirus spike proteins. Annu Rev Virol. 2016 Sep 29;3(1):237-61.

Margulis AV, Fortuny J, Kaye JA, Calingaert B, Reynolds M, Plana E, et al. Validation of cancer cases using primary care, cancer registry, and hospitalization data in the United Kingdom. Epidemiology. 2018 Mar;29(2):308-13.

Margulis AV, Palmsten K, Andrade SE, Charlton RA, Hardy JR, Cooper WO, et al. Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. Pharmacoepidemiol Drug Saf. 2015 Apr;24(4):335-42.

Morales DR, Donnan PT, Daly F, Staa TV, Sullivan FM. Impact of clinical trial findings on Bell's palsy management in general practice in the UK 2001-2012: interrupted time series regression analysis. BMJ Open. 2013;3(7):e003121.

NHS Digital. COVID-19 public health directions 2020. Available at: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notices/secretary-of-state-directions/covid-19-public-health-directions-2020#learn-more-about-this-direction. Accessed 28 March 2021

Pottegård A, Lund LC, Karlstad Ø, Dahl J, Andersen M, Hallas J, et al. Arterial events, venous thromboembolism, thrombocytopenia, and bleeding after vaccination with Oxford-AstraZeneca ChAdOx1-S in Denmark and Norway: population based cohort study. Bmj. 2021 May 5;373:n1114.

Rassen JA, Glynn RJ, Rothman KJ, Setoguchi S, Schneeweiss S. Applying propensity scores estimated in a full cohort to adjust for confounding in subgroup analyses. Pharmacoepidemiol Drug Saf. 2012 Jul;21(7):697-709.

Remschmidt C, Wichmann O, Harder T. Frequency and impact of confounding by indication and healthy vaccinee bias in observational studies assessing influenza vaccine effectiveness: a systematic review. BMC Infect Dis. 2015 Oct 17;15:429.

Rothman KJ. Six persistent research misconceptions. J Gen Intern Med. 2014 Jul;29(7):1060-4.

Rothman KJ, Greenland S. Planning study size based on precision rather than power. Epidemiology. 2018 Sep;29(5):599-603.

Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med. 2010;8:18.

Scully M, Singh D, Lown R, Poles A, Solomon T, Levi M, et al. Pathologic antibodies to platelet factor 4 after ChAdOx1 nCoV-19 vaccination. N Engl J Med. 2021 Apr 16.

Serazin NA, Edem B, Williams SR, Ortiz JR, Kawade A, Das MK, et al. Acute respiratory distress syndrome (ARDS) as an adverse event following immunization: case definition & guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine. 2021 May 21;39(22):3028-36.

Tokars JI, Lewis P, DeStefano F, Wise M, Viray M, Morgan O, et al. The risk of Guillain-Barré syndrome associated with influenza A (H1N1) 2009 monovalent vaccine and 2009-2010 seasonal influenza vaccines: results from self-controlled analyses. Pharmacoepidemiol Drug Saf. 2012 May;21(5):546-52.

VAC4EU. VAccine monitoring Collaboration for Europe. 2021. Available at: https://vac4eu.org/. Accessed 29 March 2021.

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344-49.

Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim

analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet. 2021 Jan 9;397(10269):99-111.

Webster-Clark M, Stürmer T, Wang T, Man K, Marinac-Dabic D, Rothman KJ, et al. Using propensity scores to estimate effects of treatment initiation decisions: state of the science. Stat Med. 2021 Mar 30;40(7):1718-35.

Weldeselassie YG, Whitaker HJ, Farrington CP. Use of the self-controlled case-series method in vaccine safety studies: review and recommendations for best practice. Epidemiol Infect. 2011 Dec;139(12):1805-17.

WHO. Global Advisory Committee on Vaccine Safety (GACVS) review of latest evidence of rare adverse blood coagulation events with AstraZeneca COVID-19 vaccine (Vaxzevria and Covishield). 2021. Available at: https://www.who.int/news/item/16-04-2021-global-advisory-committee-on-vaccine-safety-(gacvs)-review-of-latest-evidence-of-rare-adverse-blood-coagulation-events-with-astrazeneca-covid-19-vaccine-(vaxzevria-and-covishield)#. Accessed 12 May 2021.

Willame C, Dodd C, Gini R, Durán C, Ehrenstein V, Thomsen R, et al. Background rates of adverse events of special interest for monitoring COVID-19 vaccines, an ACCESS study. D3-Draft final report. 30 April 2021. Available at: http://www.encepp.eu/documents/ACCESS_Report_BGR_20210430_v.1.2.pdf. Accessed 12 April 2021.

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.

Wise RP, Bonhoeffer J, Beeler J, Donato H, Downie P, Matthews D, et al. Thrombocytopenia: case definition and guidelines for collection, analysis, and presentation of immunization safety data. Vaccine. 2007 Aug 1;25(31):5717-24.

Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. Int J Epidemiol. 2019 Dec 1;48(6):1740-g.

Wong AY, MacKenna B, Morton CE, Schultze A, Walker AJ, Bhaskaran K, et al. Use of non-steroidal anti-inflammatory drugs and risk of death from COVID-19: an OpenSAFELY cohort analysis based on two cohorts. Ann Rheum Dis. 2021 Jan 21.

Wyss R, Girman CJ, LoCasale RJ, Brookhart AM, Stürmer T. Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study. Pharmacoepidemiol Drug Saf. 2013 Jan;22(1):77-85.

Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004 Apr 1;159(7):702-6.

14. ANNEXES

Appendix A List of stand-alone documents

None

Appendix B ENCePP Checklist for Study Protocols

Doc.Ref. EMA/540136/2009

ENCePP Checklist for Study Protocols (Revision 4)

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

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

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

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

Study title: A post-authorisation/post-marketing observational study to evaluate the association between exposure to AZD1222 and safety concerns using existing secondary health data sources

EU PAS Register® number:
Study reference number (if applicable): not yet registered

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

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:	☒	☐	☐	7 and 8
2.1.1 Why the study is conducted? (eg, to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	7
2.1.2 The objective(s) of the study?	☒	☐	☐	8
2.1.3 The target population? (ie, population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	9.1, 9.2.1
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☐	☒	8
2.1.5 If applicable, that there is no *a priori* hypothesis?	☒	☐	☐	8

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (eg, cohort, case-control, cross-sectional, other design)	☒	☐	☐	9.1
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	9.1, 9.4

[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.

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.3 Does the protocol specify measures of occurrence? (eg, rate, risk, prevalence)	☒	☐	☐	9.7.6
3.4 Does the protocol specify measure(s) of association? (eg, relative risk, odds ratio, excess risk, incidence rate ratio, hazard ratio, number needed to harm [NNH])	☒	☐	☐	9.7.7, 9.7.8
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (eg, adverse events that will not be collected in case of primary data collection)	☒	☐	☐	11

Comments:

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.1 Is the source population described?	☒	☐	☐	9.2.1
4.2 Is the planned study population defined in terms of:				
4.2.1 Study time period	☒	☐	☐	9.2.2
4.2.2 Age and sex	☐	☐	☒	9.2.1
4.2.3 Country of origin	☒	☐	☐	9.1, 9.4
4.2.4 Disease/indication	☐	☐	☒	9.2.1
4.2.5 Duration of follow-up	☒	☐	☐	9.2.3
4.3 Does the protocol define how the study population will be sampled from the source population? (eg, event or inclusion/exclusion criteria)	☒	☐	☐	9.2.1.1, 9.2.1.3

Comments:

4.2.2 The study is population based without any restriction on age or sex. 4.2.4 During a period of mass vaccination for COVID-19, it is assumed that eventually all subjects will be eligible for vaccination.

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (eg, operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	9.3.1, 9.4.1, 9.4.2, 9.4.3
5.2 Does the protocol address the validity of the exposure measurement? (eg, precision, accuracy, use of validation sub-study)	☒	☐	☐	9.3.1, 9.9.1

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.3 Is exposure categorised according to time windows?	☒	☐	☐	9.2.3.1,9.2.3.2, 9.3.1, 9.3.2
5.4 Is intensity of exposure addressed? (eg, dose, duration)	☒	☐	☐	9.7.2, 9.7.6
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☒	☐	☐	9.3.2
5.6 Is (are) an appropriate comparator(s) identified?	☒	☐	☐	9.2.1.1, 9.2.1.3

Comments:

5.2 Validity of COVID-19 vaccinations has not been characterised yet.

5.5 Risk intervals shown in Table 2, Adverse Events of Special Interest and Other Safety Concerns, are informed in part based on known timing of immunologic response to vaccines.

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.1 Does the protocol specify the primary and secondary (if applicable) outcome(s) to be investigated?	☒	☐	☐	9.3.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☒	☐	☐	9.3.2
6.3 Does the protocol address the validity of outcome measurement? (eg, precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☒	☐	☐	9.3.2
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (eg, HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☐	☐	☒	

Comments:

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (eg, confounding by indication)	☒	☐	☐	9.7.3
7.2 Does the protocol address selection bias? (eg, healthy user/adherer bias)	☒	☐	☐	9.1.1, 9.9.2

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.3 Does the protocol address information bias? (eg, misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	9.3.2, 9.9.1

Comments:

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (eg, collection of data on known effect modifiers, subgroup analyses, anticipated direction of effect)	☒	☐	☐	9.2.1.2, 9.7.9

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? (eg, pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☒	☐	☐	9.3.1, 9.4.1, 9.4.2, 9.4.3
9.1.2 Outcomes? (eg, clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	9.3.2, 9.4.1, 9.4.2, 9.4.3
9.1.3 Covariates and other characteristics?	☒	☐	☐	9.3.3, 9.4.1, 9.4.2, 9.4.3
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (eg, date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	9.4.1, 9.4.2, 9.4.3
9.2.2 Outcomes? (eg, date of occurrence, multiple events, severity measures related to event)	☒	☐	☐	9.4.1, 9.4.2, 9.4.3
9.2.3 Covariates and other characteristics? (eg, age, sex, clinical and drug use history, comorbidity, comedications, lifestyle)	☒	☐	☐	9.4.1, 9.4.2, 9.4.3
9.3 Is a coding system described for:				

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.3.1 Exposure? (eg, WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☒	☐	☐	9.4.1, 9.4.2, 9.4.3
9.3.2 Outcomes? (eg, International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☒	☐	☐	9.4.1, 9.4.2, 9.4.3
9.3.3 Covariates and other characteristics?	☒	☐	☐	9.4.1, 9.4.2, 9.4.3
9.4 Is a linkage method between data sources described? (eg, based on a unique identifier or other)	☒	☐	☐	9.4.1, 9.4.2, 9.4.3

Comments:

Section 10: Analysis plan	**Yes**	**No**	**N/A**	**Section Number**
10.1 Are the statistical methods and the reason for their choice described?	☒	☐	☐	9.7.6, 9.7.7, 9.7.8
10.2 Is study size and/or statistical precision estimated?	☒	☐	☐	9.5
10.3 Are descriptive analyses included?	☒	☐	☐	9.7.1, 9.7.2, 9.7.9
10.4 Are stratified analyses included?	☒	☐	☐	9.7, 9.7.1, 9.7.2, 9.7.5, 9.7.9
10.5 Does the plan describe methods for analytic control of confounding?	☒	☐	☐	9.7.3 and subsections
10.6 Does the plan describe methods for analytic control of outcome misclassification?	☒			9.9.1
10.7 Does the plan describe methods for handling missing data?	☒	☐	☐	9.7.5
10.8 Are relevant sensitivity analyses described?	☒	☐	☐	9.9.1

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (eg, software and IT environment, database maintenance and anti-fraud protection, archiving)	☒	☐	☐	9.6.3, 9.6.4
11.2 Are methods of quality assurance described?	☒	☐	☐	9.8
11.3 Is there a system in place for independent review of study results?	☐	☒	☐	

Comments:

To date, there are no plans for an independent advisory board

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?	☒	☐	☐	9.9.1
12.1.2 Information bias?	☒	☐	☐	9.9.1
12.1.3 Residual/unmeasured confounding? (eg, anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods)	☒	☐	☐	9.9.1
12.2 Does the protocol discuss study feasibility? (eg, study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	9.1.1, 9.9.2

Comments:

Section 13: Ethical issues	**Yes**	**No**	**N/A**	**Section Number**
13.1 Have requirements of Ethics Committee/ Institutional Review Board been described?	☒	☐	☐	9.8, 10
13.2 Has any outcome of an ethical review procedure been addressed?	☐	☐	☒	
13.3 Have data protection requirements been described?	☒	☐	☐	9.4.1, 9.6.3, 10

Comments:

13.2 Protocol has not yet been submitted for Ethics Committee or IRB review

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

Comments:

This is version 1 of protocol. No amendments or deviations yet, but they would go in Section 5.

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

Comments:

Name of the main author of the protocol: Cristina Rebordosa

Date: 07/July/2021

Signature: [redacted]

Appendix C Additional information

Annex Table 1 Covariates Proposed to Define the Study Population and Cohorts

Related to AZD1222 vaccination
Date of dose 1
Date of dose 2
Number of doses
Timing between doses
Batch of the vaccine
Related to non-AZD1222 vaccines against COVID-19
Receipt of any SARS-CoV-2 vaccine other than AZD1222
Date of vaccination
Duration of available look-back period
Earliest date of data availability
Exclusion and censoring criteria
Latest date of data availability
For each AESI: dates of occurrence
Date of death

AESI, adverse event of special interest; COVID-19, coronavirus disease 2019; SAP, statistical analysis plan; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Note: A final list of covariates and their operational definitions will be specified in the SAP.

Annex Table 2 Covariates Proposed to Define Subgroups of Interest Included as Missing Information in the EU RMP[a]

To identify women who are pregnant or breastfeeding
Pregnancy status at cohort entry (algorithms including pregnancy outcomes such as miscarriage or live birth will be used to determine duration of pregnancy)
Breastfeeding status at cohort entry (as available, or measured by proxy, such as first 6 months postpartum)
To identify immunocompromised patients
Receipt of organ transplantation
Specific white cell haematological neoplasias (eg, chronic lymphocytic leukaemia, chronic myeloid leukaemia, multiple myeloma)
Human immunodeficiency virus (HIV) infection
Primary immunodeficiencies (eg, chronic granulomatous disease [CGD])
Recent or current use of immunosuppressive medications, eg, cytotoxic chemotherapy, biologic immunomodulatory therapies, methotrexate, chronic corticosteroid use
To identify autoimmune and inflammatory disorders
Rheumatoid arthritis
Polymyalgia rheumatica
Systemic lupus erythematosus
Crohn's disease and ulcerative colitis
Other autoimmune and connective tissue diseases (eg, primary biliary cholangitis, systemic sclerosis, psoriasis, vasculitis)
Neuroimmune conditions (eg, multiple sclerosis, myelitis, previous GBS)
To identify frail patients with comorbidities
Indicators of frailty
Dispensing of or reimbursement for durable medical equipment (eg, wheelchairs, home oxygen)
Residence in long-term facility or nursing home
Hip fracture
Palliative care
Metastatic cancer
Cachexia
Dementia
Pressure ulcers
Bladder incontinence
Indicators of other relevant comorbidities (diagnosis and procedure codes as applicable)
Chronic obstructive pulmonary disease
Diabetes mellitus
Chronic neurological disease (eg, Parkinson's disease, motor neuron disease, cerebrovascular disease)

Metastatic cancer
Heart failure
Peripheral vascular disease
Chronic liver disease
Substance use disorders
Chronic pancreatic diseases
Chronic kidney disease
Peritoneal dialysis or haemodialysis
To identify subjects who received vaccines to prevent diseases other than COVID-19 recently before cohort entry
Pneumococcal vaccine
Influenza vaccine
Shingles (H. Zoster) vaccine
Other routine scheduled vaccinations
To explore potential dose effect
AZD1222 dose number
To explore effects by age group
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, and greater than 80 years) although this may be collapsed if needed

[a] Missing information as per the most current approved EU RMP at the time of protocol approval.

Note: A final list of covariates and their operational definitions will be specified in the SAP.

ACCESS, vACCine covid-19 monitoring readinESS; COVID-19, coronavirus disease 2019; EU, European Union; GBS, Guillain-Barré syndrome; RMP, risk management plans; SAP, statistical analysis plan.

Annex Table 3 Covariates Proposed to Define Baseline Characteristics and to Be Considered in Estimation of the Propensity Score

Sociodemographic and lifestyle
Age (in years)
Sex
Month/year at index date
Ethnicity and/or race
Geographic region
Socioeconomic status (employment/occupational status, income, education, index of multiple deprivation)
Health care or essential worker status
Residence in long-term facility or nursing home
Smoking or use of smoking cessation drugs
Body mass index (BMI) > 30, obesity diagnosis or obesity surgery
History of alcohol abuse/dependence or alcohol-related diseases
Duration of available health history before cohort entry
Calculated as index date minus earliest date of data availability
Health care resource utilisation (look-back period for each variable to be specified in the SAP)
Number of hospitalisations
Number of hospitalisations requiring intensive care
Number of emergency department visits
Any stay in a skilled nursing facility or nursing home
Number of outpatient specialist visits
Number of outpatient primary care visits
Number of prescriptions/dispensing for any drug
Receipt of influenza vaccine in prior 12 months
Cancer screening tests (eg, mammography, colonoscopy) in prior 12 months, as indicated
Diagnostic testing for COVID-19
COVID-19 history
Previous COVID-19 diagnosis
Positive test result for COVID-19
Other vaccination history
Recent pneumococcal vaccine
Recent influenza vaccine
Recent shingles vaccine
Recent other scheduled vaccines

Indicators of pregnancy or breastfeeding
Pregnancy status at cohort entry
Breastfeeding status at cohort entry (as available)
Indicators of immunosuppression
Organ transplant recipient
Specific haematological neoplasias
Human immunodeficiency virus (HIV) infection
Other congenital or acquired immunodeficiencies or immunosuppressive diseases
Recent or current use of immunosuppressive medications (eg, cytotoxic chemotherapy, biologic immunomodulatory therapies, azathioprine, methotrexate, chronic corticosteroid use)
Indicators of frailty
Dispensing of or reimbursement for durable medical equipment (eg, wheelchairs, home oxygen)
Residence in long-term facility or nursing home
History of hip fracture
Palliative care
Metastatic cancer
Cachexia
Dementia
Pressure ulcers
Bladder incontinence
Indicators of autoimmune or inflammatory disorders
Rheumatoid arthritis
Polymyalgia rheumatica
Psoriasis
Psoriatic arthritis
Vasculitis (any)
Spondylarthritis (any)
Systemic lupus erythematosus
Inflammatory bowel disease (ulcerative colitis and Crohn’s disease)
Potential indicators of anaphylaxis risk
History of anaphylaxis
Drug allergies
Food/latex/insect bite allergies
Atopic dermatitis
Epinephrine auto-injector prescription or redemption

Other comorbidities
History of allergies not elsewhere mentioned
Diabetes mellitus (types 1 and 2)
Diabetic complications (renal, ophthalmic, neurologic, amputations and other)
Hypertension
Heart failure
Peripheral vascular disease
Hyperlipidaemia
Ischaemic heart disease
Supraventricular tachyarrhythmias including atrial fibrillation
History of ventricular tachycardia or fibrillation
Receipt of internal cardiac defibrillator
Cardiac conduction system disorder/pacemaker insertion
Valvular heart disease
Peripheral vascular disease
Cerebrovascular disease
Chronic neurological diseases (eg, Parkinson's disease, motor neuron disease, epilepsy)
Chronic respiratory diseases including asthma, emphysema, COPD, interstitial lung disease
Recent respiratory insufficiency
Cancer diagnoses by organ
Chronic liver disease
Recent acute liver injury
Infectious hepatitis
Substance use disorders
Chronic and acute pancreatic diseases
Biliary tract diseases
Chronic kidney disease
Peritoneal dialysis or haemodialysis
Arthrosis and arthritis not elsewhere mentioned
Recent respiratory infections
Recent influenza-like illness
Peptic ulcer disease
Anaemia
Charlson Comorbidity Index

Indicators of other relevant comorbidities (medications)
Anticoagulants (warfarin and novel oral anticoagulants)
Diuretics
Beta blockers
Calcium channel blockers
Angiotensin-converting enzyme inhibitors
Angiotensin-receptor blockers
Lipid-lowering drugs
Antibiotics
Antihistamines
Non-steroidal anti-inflammatory drugs
Oral glucose lowering drugs
Insulin
Antidepressants
Psychotropics
Bronchodilators
HIV anti-retroviral therapy
Other antiviral medications
Drugs with hepatoxic potential (including paracetamol and others to be specified)

Note: A final list of covariates and their operational definitions will be specified in the SAP.

COPD, chronic obstructive pulmonary disease; COVID-19, coronavirus disease 2019; HIV, human immunodeficiency virus, SAP, statistical analysis plan.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d8111r00006-pass-clinical-study-protocol		
Document Title:	D8111R00006 PASS Clinical Study Protocol version 3	
Document ID:	Doc ID-004619144	
Version Label:	2.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
09-Jul-2021 16:58 UTC	Christina Guiton	Qualified Person Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.