| **PASS Protocol** | |
|---|---|
| Active substance | AZD1222 |
| Product reference | D8111R00006 |
| Version number | V4.0 |
| Date | 10 October 2022 |

# 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***<br>*151 85 Södertälje, Sweden* |
| **MAH contact person** | PPD<br>AstraZeneca AB<br>151 85 Södertälje<br>Sweden |

**Approved by:**

Cristina Rebordosa

*Electronically signed by: Cristina Rebordosa
Reason: I am an authorized signatory and I approve this document.
Date: Oct 27, 2022 12:30 GMT+2*

Cristina Rebordosa, PPD Date
PPD
RTI Health Solutions

# 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** | 4.0 |
| **Date of last version of protocol** | 10 October 2022 |
| **EU PAS Register number** | EUPAS43556 |
| **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, PPD<br>RTI Health Solutions<br>PPD<br>08028 Barcelona SPAIN<br>PPD<br>PPD<br>AstraZeneca AB<br>PPD<br><br>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 |
| AED | Accident and Emergency Department |
| AESI | adverse event of special interest |
| AMI | acute myocardial infarction |
| AR | assessment report |
| ARD | absolute risk difference |
| ARDS | acute respiratory distress syndrome |
| ARS Toscana | Regional Health Agency of Tuscany [Agenzia regionale di sanità della Toscana] |
| ATC | Anatomical Therapeutic Chemical |
| BMI | body mass index |
| CDC | Centers for Disease Control and Prevention (United States) |
| CDM | common data model |
| CHESS | COVID-19 Hospitalisation in England Surveillance System |
| CI | confidence interval |
| CIMS | COVID-19 vaccination Information and Monitoring System (Netherlands) |
| COVID-19 | coronavirus disease 2019 |
| CPRD | Clinical Practice Research Datalink |
| DAP | data access partner |
| DRE | Digital Research Environment |
| DSRU | Drug Safety Research Unit |
| DVT | deep vein thrombosis |
| ECDC | European Centre for Disease Prevention and Control |
| EMA | European Medicines Agency |
| ENCePP | European Network of Centres for Pharmacoepidemiology and Pharmacovigilance |
| ETL | extract, transform, and load |
| EU | European Union |
| FAIR | findable, accessible, interoperable, and reusable |
| FDA | Food and Drug Administration (United States) |
| FISABIO | Foundation for the Promotion of Health and Biomedical Research of Valencia Region |
| FU | follow-up |
| GAIA | Pharmaceutical module in VID |

| Abbreviation or special term | Explanation |
|---|---|
| GBS | Guillain-Barré syndrome |
| GDPR | General Data Protection Regulation |
| GP | general practitioner |
| GPP | Good Pharmacoepidemiology Practices |
| GVP | Good Pharmacovigilance Practices |
| HES | Hospital Episodes Statistics |
| HIV | human immunodeficiency virus |
| 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 |
| ICH | International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use |
| ICMJE | International Committee of Medical Journal Editors |
| ICNARC | Intensive Care National Audit and Research Centre |
| IDIAP Jordi Gol | Institute of Research in Primary Care [Institut Universitari D'Investigació en Atenció Primària Jordi Gol] |
| IOM | Institute of Medicine |
| IR | incidence rate |
| IRR | incidence rate ratio |
| ISPE | International Society for Pharmacoepidemiology |
| KM | Kaplan-Meier estimate |
| MAH | marketing authorisation holder |
| MBDS | Minimum Basic Data Set at Hospital Discharge |
| ME | myalgic encephalitis |
| MHRA | Medicines and Healthcare products Regulatory Agency |
| NHS | National Health Service |
| ORION | Hospital Medical Record in VID |
| PASS | post-authorisation safety study |
| PHARMO | PHARMO Database Network or PHARMO Institute for Drug Outcomes Research (Netherlands) |
| PHE | Public Health England |
| PP | prevalence proportion |
| PPD | prevalence proportion difference |
| PPR | prevalence proportion ratio |
| PPRN | PHARMO Perinatal Research Network |

| Abbreviation or special term | Explanation |
|---|---|
| PPV | positive predictive value |
| PRAC | Pharmacovigilance Risk Assessment Committee (of the EMA) |
| PS | propensity score |
| PVFS | postviral fatigue syndrome |
| PY | person-years |
| QC | quality control |
| RIVM | Dutch National Institute for Public Health and the Environment |
| RMP | risk management plans |
| RR | risk ratio |
| SAP | statistical analysis plan |
| SARS-CoV-2 | severe acute respiratory syndrome coronavirus 2 |
| SCRI | self-controlled risk interval |
| SGSS | Second Generation Surveillance System (UK) |
| SIA | Ambulatory Information System in VID |
| SIDIAP | Information System for Research in Primary Care [Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària] |
| SIP | Population Information System in VID |
| TBD | to be determined |
| TCP | thrombocytopaenia |
| TE | thromboembolic event |
| TTS | thrombosis with thrombocytopaenia syndrome |
| UMCU | University Medical Center Utrecht |
| UK | United Kingdom |
| VAC4EU | Vaccine monitoring Collaboration for Europe |
| VID | Valencia Health System Integrated Database (Spain) |
| VIS | Vaccine Information System in VID |
| VTE | venous thromboembolism |
| WHO | World Health Organization |

# 3. RESPONSIBLE PARTIES

## AstraZeneca Responsible Parties

| Role | Name, title, qualifications |
|---|---|
| Project lead | PPD |
| Scientific adviser | PPD |
| Epidemiologist | PPD |
| Epidemiologist | PPD |
| Operational lead | PPD |
| Study physician | PPD |
| Statistician | PPD |
| Statistician | PPD |
| Statistician | PPD |
| Data Management | PPD |
| Pharmacovigilance | PPD |
| Pharmacovigilance | PPD |

## RTI Health Solutions Responsible Parties

| Role | Name, title, qualifications |
|---|---|
| Principal investigator | Cristina Rebordosa, PPD PPD |
| Coprincipal investigator | PPD |
| Statistician | PPD |
| Statistician | PPD |
| Senior adviser | PPD |
| Senior adviser | PPD |

## Collaborating Institutions

| Collaborating Institution | Role | Name, title, qualifications |
|---|---|---|
| Drug Safety Research Unit (DSRU) | Principal Investigator | PPD |
| | Epidemiologist | PPD |
| | Epidemiologist | PPD |
| The Foundation for the Promotion of Health and Biomedical Research of Valencia Region, (FISABIO) | Project Lead | PPD |
| | Project Statistician | PPD |
| | Project Statistician | PPD |
| | Project Statistician | PPD |
| | Pharmacovigilance | PPD |
| | Project Manager | PPD |

| Collaborating Institution | Role | Name, title, qualifications |
|---|---|---|
| ARS Toscana | Data Scientist, Pharmacoepidemiologist | PPD |
| | Pharmacoepidemiologist | PPD |
| | Data Scientist, Programmer | PPD |
| | Pharmacoepidemiologist | PPD |
| | Pharmacoepidemiologist | PPD |
| Research Institute in Primary Care (IDIAP), Jordi Gol (IDIAP) | Pharmacoepidemiology Researcher. AZ PASS leader at IDIAP Jordi Gol | PPD |
| | Pharmacoepidemiology Researcher | PPD |
| | Senior Statistician | PPD |
| | Senior Adviser, Pharmacoepidemiology. Coordinator of the Medicines Studies Unit | PPD |
| | SIDIAP Senior Data Scientist | PPD |
| | Primary Healthcare Researcher | PPD |
| PHARMO Institute | Principal Investigator | PPD |
| | Coprincipal Investigator | PPD |
| | Investigator | PPD |
| | Senior Adviser | PPD |
| University Medical Center Utrecht (UMCU) | Principal Investigator | PPD |
| | Senior Statistician | PPD |
| | Data Engineer/ Epidemiologist | PPD |

| Collaborating Institution | Role | Name, title, qualifications |
|---|---|---|
| | Data Engineer | PPD |
| | Data Engineer | PPD |
| | Statistician | PPD |
| Vaccine Monitoring Collaboration for Europe (Vac4EU) | VAC4EU oversight | PPD |

## VAC4EU Scientific Advisory Board

| Role | Name, title, affiliation |
|---|---|
| Scientific Advisory Board Member | PPD Taipei, Taiwan |
| Scientific Advisory Board Member | PPD Montreal, Canada |

# 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 questions and objectives**

What are the incidence rates (IRs) of adverse events of special interest [AESIs] in individuals vaccinated with AZD1222 and in comparators, 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 multinational, retrospective cohort design will be used to estimate the incidence of AESIs after receiving AZD1222 and will compare this incidence with that occurring in 3 different comparator cohorts: concurrent unvaccinated comparators, active comparators, and historical comparators. The AZD1222 cohort will be matched, as applicable, independently to the 3 different comparator cohorts on calendar date of vaccination, age, sex, region, prior COVID-19, and status according to each of the 5 special populations[1]. Matching will be done with replacement in a ratio of 1 vaccinated to 1 comparator subject. Where appropriate, the study will also use a self-controlled risk interval (SCRI) design.

[1] See section 9.2.1.2 for the definition of special populations

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 healthcare data sources.
- The AZD1222 cohort will be identified based on the first vaccination with AZD1222 (index date).
- A concurrent unvaccinated comparator cohort will be identified among subjects who have not received any vaccination for COVID-19 matched (to the extent possible) on the vaccinee's index date, age, sex, prior diagnosis of COVID-19, and status according to each of the 5 special populations.
- The active comparator cohort will be initially identified based on the first vaccination with an mRNA vaccine (Comirnaty or Spikevax) matched (to the extent possible) on the vaccinee's index date (first dose; a second matching will be done using second dose for the comparative analysis), age, sex, prior diagnosis of COVID-19, and status according to each of the 5 special populations.
- A historical comparator cohort will be identified among subjects who were enrolled in the study data sources at any time during 2017 and 2018 matched on age, sex, and status according to each of the 5 special populations.

**Variables**

Receipt of AZD1222, other COVID-19 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 healthcare resource utilisation.

**Data sources**

The proposed data sources for the study are the Clinical Practice Research Datalink (CPRD) in the UK, the Valencia Health System Integrated Database (VID) and the Information System for Research in Primary Care (SIDIAP) database in Spain, the Regional Health Agency of

Tuscany (ARS Toscana) database in Italy, and the PHARMO Database Network (PHARMO) 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 populations 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 1 million vaccinated subjects, we anticipate a 90% probability that the upper bound of the observed incidence rate ratio (IRR) would be below 2 assuming a 3-to-1 ratio between person-time for vaccinated subjects and concurrent unvaccinated comparators, respectively, and that the true IRR is 1.0 and a 100% probability that the upper bound of the observed IRR would be below 2, assuming a 1-to-1 ratio between person-time for vaccinated and active comparators or historical comparators, respectively, 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 COVID-19 vaccine dose (AZD1222 and others) will be described.

For the cohort study, exposure propensity scores (PS) will be used to refine the balance between study cohorts, which will be matched initially on index date, age, sex, region, prior diagnosis of COVID-19, and status according to each of the 5 special populations, as applicable for each comparator. Propensity scores will be used to control for confounding either by analytic methods involving stratification or weighting. For AESIs with known risk windows, ie, for which the risk interval is characterised, crude IRs and 95% confidence intervals (CIs) for the vaccinated population and for the comparator cohorts 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. Comparative analysis between AZD1222 and active comparators will be limited to those with 2 doses of AZD1222 or an mRNA COVID-19 vaccine, and to the risk window after dose 2. 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 with known risk windows, where appropriate.

### Milestones

| Milestone | Planned date [a] | Actual/revised date |
|---|---|---|
| Progress Report 1 | 3 months after protocol endorsement (expected Oct 2021) | 22 Oct 2021 |
| Interim Report 1 | 9 months after protocol endorsement (expected Apr 2022) | 22 Apr 2022 |
| Interim Report 2 | 15 months after protocol endorsement (expected Oct 2022) | Jan 2023 (planned) |
| Progress Report 2 | April 2023 | |
| Feasibility assessment report for comparative analysis | June 2023 | |
| Final report of study results | 27 months after protocol endorsement (expected Oct 2023) | Jun 2024 (planned revised) |

[a] Schedule is dependent on protocol endorsement date, data extraction in each data source, and contracts with the research team.

# 5. AMENDMENTS AND UPDATES

Version 3.0 (22 July 2021) of the protocol was endorsed by the EMA Pharmacovigilance Risk Assessment Committee (PRAC). Adjustments were made in the course of study implementation and statistical analysis plan (SAP) development and review of the SAP and Interim Report 1 by the PRAC. Changes in the milestones are driven by the 10th of August 2022 EMA communication (email from Lopez Fauqued Marta marta.lopezfauqued@ema.europa.eu) whereby:

- PRAC Rapporteur team proposed replacing Interim Report 3 (April 2023) with a progress report for the same period.
- PRAC Rapporteur accepted delaying the final report of study results from October 2023 to March 2024.
- Considering that no data are collected from the US, PRAC Rapporteur accepted removing Study D811R00002 (US).

All changes occurred before the start of data collection for the final report (Table 1). The third column of Table 3 of the PRAC communication describes which changes were implemented in IR1, IR2, and which will be implemented in the final analysis, as needed.

## Table 1 Amendments, Updates, and Adjustments

| Version number | Date | Section of study protocol | Amendment or update | Reason |
|---|---|---|---|---|
| 4.0 | 10 Oct 2022 | 9.2.1.1.3 | Added matching variables: prior diagnosis of COVID-19 and geographic region (applied in analyses for Interim Report 1) | Based on PRAC-endorsed SAP (V1.0, 10 Nov 2021). Prior diagnosis of COVID-19 was added to obtain more comparable populations and use study design to address confounding. Geographic region was added to facilitate data extraction in data sources in which converting the whole database into the common data model is an issue due to size. |
| 4.0 | 10 Oct 2022 | 9.4.1 | CPRD data include only Aurum and not GOLD (applied in analyses for Interim Report 1) | Aurum was chosen because it has a larger number of subjects and because of its high percentage of practices with data linkable to hospital and death data, as opposed to CPRD GOLD. This was erroneously stated in the protocol version 2.1 |
| 4.0 | 10 Oct 2022 | 9.2.1.1.3 | Matching with replacement in a ratio of 1:1 between AZD1222 and comparators (applied in analyses for Interim Report 2) | Based on PRAC recommendation in the AR for Interim Report 1 (7 Jul 2022). To reduce the number of times that a concurrent unvaccinated subject is included for matching (prior matching ratio was 1:5) |
| 4.0 | 10 Oct 2022 | 9.3.2.1 | Time window between “thrombosis and thrombocytopaenia for thrombotic thrombocytopaenia syndrome” and “thrombocytopaenia with bleeding” adjusted from 7 to 10 days (applied in analyses for Interim Report 2) | To align with VAC4EU COVID-19 vaccine PASS |

## Table 1 Amendments, Updates, and Adjustments

| Version number | Date | Section of study protocol | Amendment or update | Reason |
|---|---|---|---|---|
| 4.0 | 10 Oct 2022 | 9.3.2 Table 2 | Split AESI group categories when events in the category have different risk windows, ie, split stress myocardiopathy from "acute cardiac injury" and transverse myelitis from "multiple sclerosis and other demyelinating disorders" (applied in analyses for Interim Report 2) | Alignment with the analysis conducted for Interim Reports 1 and 2. To avoid analysing together the components within a composite AESI that have different risk windows |
| 4.0 | 10 Oct 2022 | 9.7.7 | For 2 AESIs with known risk windows (anaphylaxis and sudden death) prevalence proportion, prevalence proportion rate, and prevalence proportion difference will be estimated instead of IR, IRR, and IRD (applied in analyses for Interim Report 2) | Based on PRAC-endorsed SAP (V1.0, 10 Nov 2021). For anaphylaxis and sudden death, due to extremely short risk windows, the time to the event is not considered relevant in the analysis |
| 4.0 | 10 Oct 2022 | 9.7.10 | New sensitivity analysis: inclusion of a sensitivity analysis where TTS also includes AMI or stroke (applied in analyses for Interim Report 2) | To assess the potential effect of AZD1222 on arterial thrombosis and thromboembolisms |
| 4.0 | 10 Oct 2022 | 9.7.10.1 | New sensitivity analysis: negative control outcome analysis. The 2 negative control outcomes selected are fractures and urinary tract infections (applied in analyses for Interim Report 2) | Based on PRAC recommendation in the AR for the SAP (10 Jun 2022); to evaluate the presence of residual confounding |
| 4.0 | 10 Oct 2022 | 9.7.3.4 | Feasibility assessment of cohort comparability for the active comparator population (to be applied in the Feasibility Assessment Report) | To assess the comparability between the AZD1222 and the active comparator cohorts |

## Table 1 Amendments, Updates, and Adjustments

| Version number | Date | Section of study protocol | Amendment or update | Reason |
|---|---|---|---|---|
| 4.0 | 10 Oct 2022 | 9.2.1.3 | Changed SCRI inclusion criteria from "full accrual" to "at least 1 day of data accrual" to define the event in each interval (to be applied in analyses for the final report) | Based on PRAC-endorsed SAP (V1.0, 10 Nov 2021). Duration of follow-up is taken into account in the SCRI analysis; will allow inclusion of subjects with shorter risk or control intervals |
| 4.0 | 10 Oct 2022 | 8, 9.1.1, 9.2.1.1.3, 9.2.3.1 | Added 2 additional comparator cohorts: an active comparator cohort (subjects vaccinated with mRNA vaccine) and historical comparator cohort (subjects 2017-2018) (to be applied in analyses for the final report) | Based on PRAC recommendation in the AR for the progress report (13 Jan 2022) and in the AR for the SAP (10 Feb 2022). By including different comparator cohorts, conclusions of the present study will provide a deeper insight into the safety of AZD1222 vaccine |
| 4.0 | 10 Oct 2022 | 9.7.1 | New descriptive analysis: To describe the follow-up time by matching covariates for AZD1222 and each comparator cohort (to be applied in analyses for the final report) | To understand the impact of censoring follow-up |
| 4.0 | 10 Oct 2022 | 9.7.10 | New sensitivity analysis to estimate the effect of the vaccine on the AESIs among those subjects without prior COVID-19 at baseline and censoring at the first positive test or diagnosis of COVID-19 during follow-up (direct effect) (to be applied in analyses for the final report) | Based on PRAC-endorsed SAP (V1.0 10 Nov 2021). To disentangle the direct effect of the vaccine on the risk of AESI from the indirect effect of the vaccine averted by prevention of symptomatic SARS-CoV-2 infection |
| 4.0 | 10 Oct 2022 | 9.7.10 | New sensitivity analysis: Analysis in CPRD will be repeated using hospital data (HES) for the final analysis (to be applied in analyses for the final report) | To assess the magnitude of potential outcome misclassification |

## Table 1 Amendments, Updates, and Adjustments

| Version number | Date | Section of study protocol | Amendment or update | Reason |
|---|---|---|---|---|
| 4.0 | 10 Oct 2022 | 9.7.10 | New sensitivity analysis: Analysis to estimate the effect of partial censoring within matched pairs on the AESIs will be estimated by censoring both members of a matched pair when one member is censored (to be applied in analyses for final report) | To assess potential informative censoring |
| 4.0 | 10 Oct 2022 | 6 | Final report to be submitted in June 2024 | Based on PRAC communication 10 Aug 2022. To be able to conduct all the additional activities requested with additional comparator cohorts |
| 4.0 | 10 Oct 2022 | 5 | Analyses for Interim Report 3 will not be conducted, so Interim Report 3 will not be produced | Based on PRAC communication 10 Aug 2022. The last period of administration of AZD1222 vaccines was covered in Interim Report 2. The MAH proposes to conduct only the final analyses including the maximum length of follow-up which (see row above) require considerable additional resources and time |
| 4.0 | 10 Oct 2022 | 5 | Progress Report 2 | Based on PRAC communication 10 Aug 2022. To provide status update in the absence of Interim Report 3 |
| 4.0 | 10 Oct 2022 | 5 | Feasibility assessment report for comparative analysis | To assess the comparability between the AZD1222 and the active comparator cohorts |
| 4.0 | 10 Oct 2022 | Throughout document, especially 9.4.1 and 9.6 | Edits to align language with SAP V1.0 and Interim Reports 1 & 2 | To achieve alignment of terminology used across the study documents |
| 3.0 | 07 Jul 2021 | | Submitted to EMA. Endorsed by EMA | Modified in response to PRAC review |

**Table 1 Amendments, Updates, and Adjustments**

| Version number | Date | Section of study protocol | Amendment or update | Reason |
|---|---|---|---|---|
| 2.0 | 14 Jun 2021 | | Submitted to EMA. Not endorsed<br>Submitted to MHRA. Endorsed | Modified in response to PRAC review |
| 1.0 | 29 Mar 2021 | | Submitted to EMA. Not endorsed | |

AESI, adverse event of special interest; AMI, acute myocardial infarction; AR, assessment report, COVID-19, coronavirus disease 2019; CPRD GOLD, General Practitioner Online Database of CPRD; CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics; MAH, marketing authorisation holder; mRNA, messenger ribonucleic acid; PASS, post-authorisation safety study; PHARMO, PHARMO Database Network or PHARMO Institute for Drug Outcomes Research; PRAC, Pharmacovigilance Risk Assessment Committee; SAP, statistical analysis plan; SCRI, self-controlled risk interval; TTS, thrombosis with thrombocytopaenia syndrome; VAC4EU, Vaccine monitoring Collaboration for Europe.

# 6. MILESTONES

The study milestones are summarised in Table 2.

**Table 2 Study Milestones**

| Milestone | Planned date[a] | Actual/revised date |
|---|---|---|
| Protocol submission | 01 Apr 2021 | 01 Apr 2021 |
| Protocol endorsement by EMA/MHRA | TBD (expected Jul 2021) | Version 3.0: 9 Jul 2021 (MHRA); 22 Jul 2021 (EMA) |
| Registration in the EU PAS Register | No later than 6 months after EMA protocol endorsement and before start of data collection | Registered 7 Oct 2021 (EUPAS43556) |
| Statistical analysis plan submission | 4 months after protocol endorsement (expected Nov 2021) | Version 1.0, 22 Nov 2021 |
| Start of data collection[b] | 4 months after protocol endorsement (expected Nov 2021-first interim analysis) | 18 Feb 2022 |
| End of data collection[c] | 20-21 months after protocol endorsement (expected Mar/Apr 2023 - final report) | 30-31 months after endorsement of protocol revised version 3.0 (expected Oct-Dec 2023 - final report) |
| Progress Report 1 | 3 months after protocol endorsement (expected Oct 2021) | 22 Oct 2021 |
| Interim Report 1 | 9 months after protocol endorsement (expected Apr 2022) | 22 Apr 2022 |

**Table 2 Study Milestones**

| Milestone | Planned date[a] | Actual/revised date |
|---|---|---|
| Interim Report 2 | 15 months after protocol endorsement (expected Oct 2022) | Jan 2023 (planned revised date) |
| Interim Report 3 | 21 months after protocol endorsement (expected Apr 2023) | Will be replaced with a progress report for the same period |
| Amended protocol endorsement by EMA/MHRA | TBD (expected Jan 2023) | |
| Progress Report 2[d] | April 2023 | This will replace Interim Report 3 |
| Feasibility assessment report for comparative analysis | June 2023 | |
| Final report of study results | 27 months after protocol endorsement (expected Oct 2023) | June 2024 (planned, revised) |

[a] The schedule is dependent on protocol endorsement date, 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.

[d] After communication with EMA (10 Aug 2022), the PRAC Rapporteur team proposed to replace Interim Report 3 (April 2023) with a progress report covering the same period.

EMA, European Medicines Agency; EU PAS, European Union electronic Register of Post-Authorisation Studies; MHRA, Medicines and Healthcare products Regulatory Agency; PRAC, Pharmacovigilance Risk Assessment Committee (of the EMA); 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 9.7.12.

# 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.

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 2 separate intramuscular injections. The second dose should be administered between 4 and 12 weeks (28 to 84 days) after the first dose. Since May 2022, a booster dose may be given at least 3 months after the second dose. A booster dose of AZD1222 can also be given to adults who have had 2 doses of an authorised mRNA COVID-19 vaccine but is beyond the scope of this study. At the national level, public health bodies may issue official recommendations, taking into account emerging effectiveness data and the limited safety data.

The protocol was 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 5.0. 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), thrombocytopaenia with associated bleeding, anaphylaxis, thrombosis, and “thrombosis with thrombocytopaenia 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 subjects in the *matched population*

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

3 To describe the incidence rates (IRs) of prespecified AESIs in subjects who received at least 1 dose of AZD1222 in the *matched population* and subjects who did not receive any vaccination against COVID-19 (concurrent unvaccinated comparators) in the *matched population*

4 To estimate the relative and absolute risk of prespecified AESIs in subjects who received at least 1 dose of AZD1222 compared with concurrent unvaccinated comparators in the matched populations, using a retrospective cohort design and an SCRI design

Secondary objectives are as follows:

1 To describe the baseline characteristics (eg, demographics, medical history) of all subjects in the matched population among the specific populations considered to have missing information (Section 9.2.1.2)

2 To describe, among subjects who receive a first dose of AZD1222, (ie, *in the all AZD1222 vaccinated first dose population*), the timing and type of second dose of any COVID-19 vaccine (AZD1222 or other) over the study period among the specific populations considered to have missing information

3 To describe the IRs of prespecified AESIs in subjects who received at least 1 dose of AZD1222 in the *matched population* and subjects who did not receive any vaccination against COVID-19 (concurrent unvaccinated comparators) in the *matched population*, among the specific populations considered to have missing information

4 To estimate the relative and absolute risk of prespecified AESIs in subjects who received at least 1 dose of AZD1222 compared with concurrent unvaccinated comparators in the matched populations, among the specific populations considered to have missing information, using a retrospective cohort design and an SCRI design

Exploratory objectives are as follows:

1 To describe the IRs of prespecified AESIs in subjects who received an mRNA vaccine against COVID-19 (either Comirnaty or Spikevax) (active comparators) and in subjects

from the pre-pandemic period (2017-2018) (historical comparators) in the matched population

2 To estimate the relative and absolute risk of prespecified AESIs in subjects who received at least 1 dose of AZD1222 in the matched population compared with historical comparators in the matched population.

3 To estimate the relative and absolute risk of prespecified AESIs in subjects who received 2 doses of AZD1222 in the matched population compared with subjects who received 2 doses of active comparator (Comirnaty or Spikevax as per homologous vaccination regimen) in the matched population.

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

*Additional considerations (exploratory objective 3):*
In order to more robustly handle the known temporal variability of AESIs across the various vaccines (AZD1222, Comirnaty or Spikevax) and within vaccine by dose, the relative and absolute risk will be estimated only in a restricted study population considered homologous in terms of vaccine exposure and who received 2 doses as per recommended vaccine schedule.

# 9. RESEARCH METHODS

## 9.1 Study Design

Using information from several healthcare databases, this observational population-based multinational 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 (concurrent unvaccinated comparators), subjects who received an mRNA vaccine (either Comirnaty or Spikevax) (active comparators – 2 doses), and subjects enrolled in the study data sources any time during the pre-pandemic period (ie, 2017-2018) (historical comparators) 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 healthcare 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 Information System for Research in Primary Care (SIDIAP) database in Spain, the Regional Health Agency of Tuscany (ARS Toscana) in Italy, and PHARMO Database Network (PHARMO) 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) (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

A retrospective cohort of exposed subjects will be selected. Subjects will enter the exposed cohort upon their first vaccination with AZD1222. Subjects in the AZD1222 cohort will be matched, as applicable, to 3 different comparator cohorts: concurrent unvaccinated subjects, active comparators (ie, subjects vaccinated with an mRNA vaccine against COVID-19), and historical comparators enrolled in the study data sources during the pre-pandemic period (ie, from 2017 to 2018). The AZD1222 cohort will be matched, as applicable, independently to the 3 different comparator cohorts on calendar date of vaccination, age, sex, region, prior COVID-19, and status according to each of the 5 special populations (Section 9.2.1.2). Propensity scores (PSs) will be used to address confounding, balancing multiple covariates such as comorbidities, markers of healthcare utilisation and comedications between the vaccinated and the different comparator 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), prevalence proportion ratios (PPRs), or hazard ratios (HRs), depending on the analysis and outcome. Absolute risk difference measures will include incidence rate differences, prevalence proportion differences (PPD), and 1 – KM[2] differences.

In the cohort study, the initial plan was 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 healthcare-seeking behaviours and healthcare delivery during the pandemic differ from those during the pre-pandemic era. The potential challenge with this approach is that during the pandemic, concurrent unvaccinated subjects in the comparator group may themselves become vaccinated over time. Results from Interim Reports 1 and 2 confirmed that the duration of follow-up among concurrent unvaccinated subjects was short (median, < 2 months in all data sources) and probably not long enough to evaluate AESIs with unknown risk window—(those with a risk window that ends at 180 or 365 days). Thus, additional comparators for whom the follow-up is expected to be longer were requested by the PRAC and have been added to the

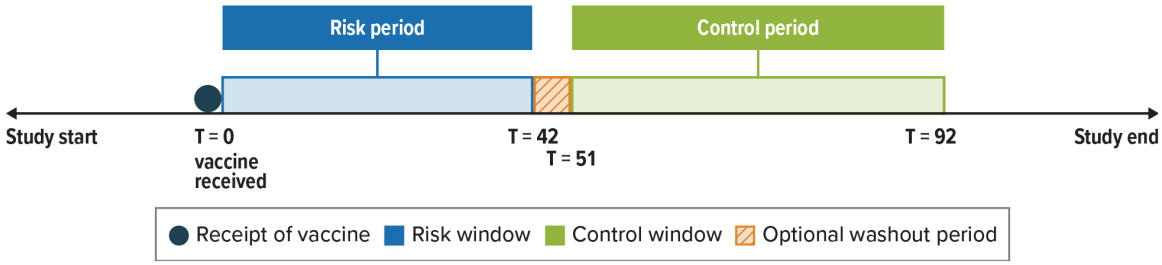
[2] KM, Kaplan-Meier estimate.

protocol. These include historical comparators from the pre-pandemic era and active comparators. This study will select historical comparators from the pre-pandemic era (2017-2018) matched by age, sex, region, and the first vaccination date of the exposed subjects minus 3 or 4 years—an approach that could control for seasonality. Concerns about non-comparability remain as 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 healthcare-seeking behaviours. Additionally, this study will select active comparators (subjects vaccinated with mRNA vaccines, either Comirnaty or Spikevax) matched by age, sex, prior COVID-19, region, and the first vaccination date. The inclusion of active comparators is also challenging due to lack of overlapping age distribution between AZD1222 subjects and subjects vaccinated with mRNA vaccines observed in Interim Report 2. A pre-matching feasibility assessment of cohort comparability will be performed to examine the overlap of matching covariates before matching in the AZD1222 cohort and the active comparator cohort. This will henceforth be referred to as the feasibility assessment of cohort comparability. Section 9.7.3.4 details the content of the feasibility assessment of cohort comparability, the results of which will guide and inform the suitability of the active comparator cohort analysis.

### 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 healthcare 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 healthcare data source (Section 9.4) during the study period.

#### 9.2.1.1 Cohort Design: Study Population

##### 9.2.1.1.1 ALL AZD1222 VACCINATED POPULATION

The *all AZD1222 vaccinated population* will include all subjects vaccinated with at least 1 dose of AZD1222 during the study period. No other inclusion or exclusion criteria will be applied. For each AESI, subjects who had an event of a specific AESI (Table 3) during the clean look-back interval were excluded from the cohort included in the analysis for the specific AESI with which they had history, but not from the analysis cohorts for other AESIs.

The **index date** will be the date when a subject received the first AZD1222 dose within the study period. Subjects who received AZD1222 after another COVID-19 vaccine as a second or third dose only will also be included in this population.

This population will be included in the analysis to address primary objective 4 and secondary objective 4 of the study (SCRI design).

#### 9.2.1.1.2 ALL AZD1222 VACCINATED FIRST DOSE POPULATION

Among all subjects in the *all AZD1222 vaccinated population*, only those subjects that received AZD1222 as their first COVID-19 vaccine dose (subjects who received AZD1222 as their second or third, but not as first, COVID-19 vaccine dose will be excluded) will be included in the *all AZD1222 vaccinated first dose population*. To ensure a minimum quality of the data, subjects with missing values for age and sex will be excluded. This population will be included in the analysis to address primary objective 2 and secondary objective 2 (patterns of use). The **index date** will be the date when a subject received the first AZD1222 dose as their first COVID-19 vaccine dose ever.

An *all mRNA vaccinated first dose population* will also be generated in order to describe the patterns of use of the vaccines included in the active comparator cohort. This will include subjects who received at least 1 dose of an mRNA vaccine (Comirnaty or Spikevax) during the study period and for whom that dose was their first COVID-19 vaccine dose (subjects who received an mRNA vaccine as their second or third COVID-19 vaccine dose after a first dose with a vaccine other than mRNA will be excluded). Subjects with missing values for age and sex will be excluded. The **index date** will be the date when a subject received the first mRNA dose as their first COVID-19 vaccine dose ever.

#### 9.2.1.1.3 MATCHED POPULATIONS

Subjects in the *all AZD1222 vaccinated first dose population* fulfilling eligibility criteria will be matched to 3 different comparator cohorts. Three independent sets of matched populations will be constructed: AZD1222 concurrent unvaccinated comparators; AZD1222 active comparators in the *all mRNA vaccinated first dose population*; and AZD1222 historical comparators. Each of these 3 sets of matched populations will be used to address primary and secondary objectives 1, 3 and 4. Figure 2, located at the end of this section (after the descriptions of the cohorts), illustrates the assembly of the study cohorts.

**Inclusion Criteria**

AZD1222 Matched Cohort

Subjects in the **AZD1222 matched cohort** will include those subjects in the *all vaccinated first dose population* (ie, received at least 1 dose of AZD1222 within the study period, and this was their first COVID-19 vaccine dose) that fulfil *all* the following inclusion criteria:

- Have at least 12 months of data available before the index date
- Have no record of vaccination with any other COVID-19 vaccine on or before the index date

The **index date** will be the date when a subject receives the first AZD1222 dose within the study period.

For the comparison with the active comparator cohort, only AZD1222 vaccinees with a first and second dose of AZD1222 will be selected. The **index date** for this comparison will be the date of the second dose.

Concurrent Unvaccinated Comparators Cohort

A potential comparator subject (**concurrent unvaccinated comparators**) will be eligible to be matched from the date the subject meets *all* the following inclusion criteria:

- Has at least 12 months of prior data available.
- Has not received any COVID-19 vaccine, including AZD1222.

The **index date** for the unvaccinated subjects will be assigned as the index date of the matched vaccinated subject.

Unexposed subjects that are eligible at the index date of an exposed subject with the same age (± 2 years), sex, prior diagnosis of COVID-19, and status according the 5 special populations will be **matched** to the exposed subject. Matching will be done with replacement in a ratio of 1 vaccinated to 1 unvaccinated subject.

Active Comparators Cohort

Subjects in the active comparator cohort will include those subjects in the *all mRNA vaccinated first dose population* (ie, received at least 1 dose of an mRNA vaccine within the study period, and this was their first COVID-19 vaccine dose) that fulfil *all* the following inclusion criteria:

- Have at least 12 months of data available before the index date
- Have no record of vaccination with any other COVID-19 vaccine on or before the index date

The **index date** is the date when a subject received the first identified Comirnaty dose or the first identified Spikevax dose within the study period.

All subjects in the active comparator cohort will be **matched** to subjects in the AZD1222 vaccinated cohort on index date (by ± 2 weeks), age (± 2 years), sex, region, prior diagnosis of COVID-19, and status according to the 5 special populations. Matching will be done with replacement at a ratio of 1:1. The same matching approach will be used for the comparative analysis with the 2 doses population using as **index date** the date when a subject received the second consecutive dose of Comirnaty or Spikevax within the study period.

Note that the distribution for matching covariates was driven by the vaccination policies that steered mRNA and AZD1222 vaccines into specific age and special population categories that varied over time. Where pre-matching populations are deemed significantly different, the suitability of this analysis will be reconsidered as the ability to make comparative inference on the safety profile is limited to a specific sub-population that may not be representative of the general AZD1222 and active comparator populations. To examine this risk, a pre-matching feasibility assessment of cohort comparability will be performed (see section 9.7.3.4).

**Historical Comparators Cohort**

A potential comparator subject (**historical comparators**) identified during the period between 2017 and 2018 will be eligible to be matched from the date the subject met all the following inclusion criteria:

- Had at least 12 months of prior data available.

The **index date** for the historical comparators will be same day and month of the AZD1222 index date minus 3 years (subjects from 2018) or 4 years (subjects from 2017)—an approach that intends to address seasonality.

Subjects will be **matched** to AZD1222-vaccinated subjects on age (± 2 years), sex, region, and status according to the 5 special populations. Matching will be done with replacement at a ratio of 1:1.

Whether all potential comparators, or only a sample, will be included in the selection of the comparator cohorts 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.

**Exclusion Criteria**

The following exclusion criterion will be considered in all the cohorts:

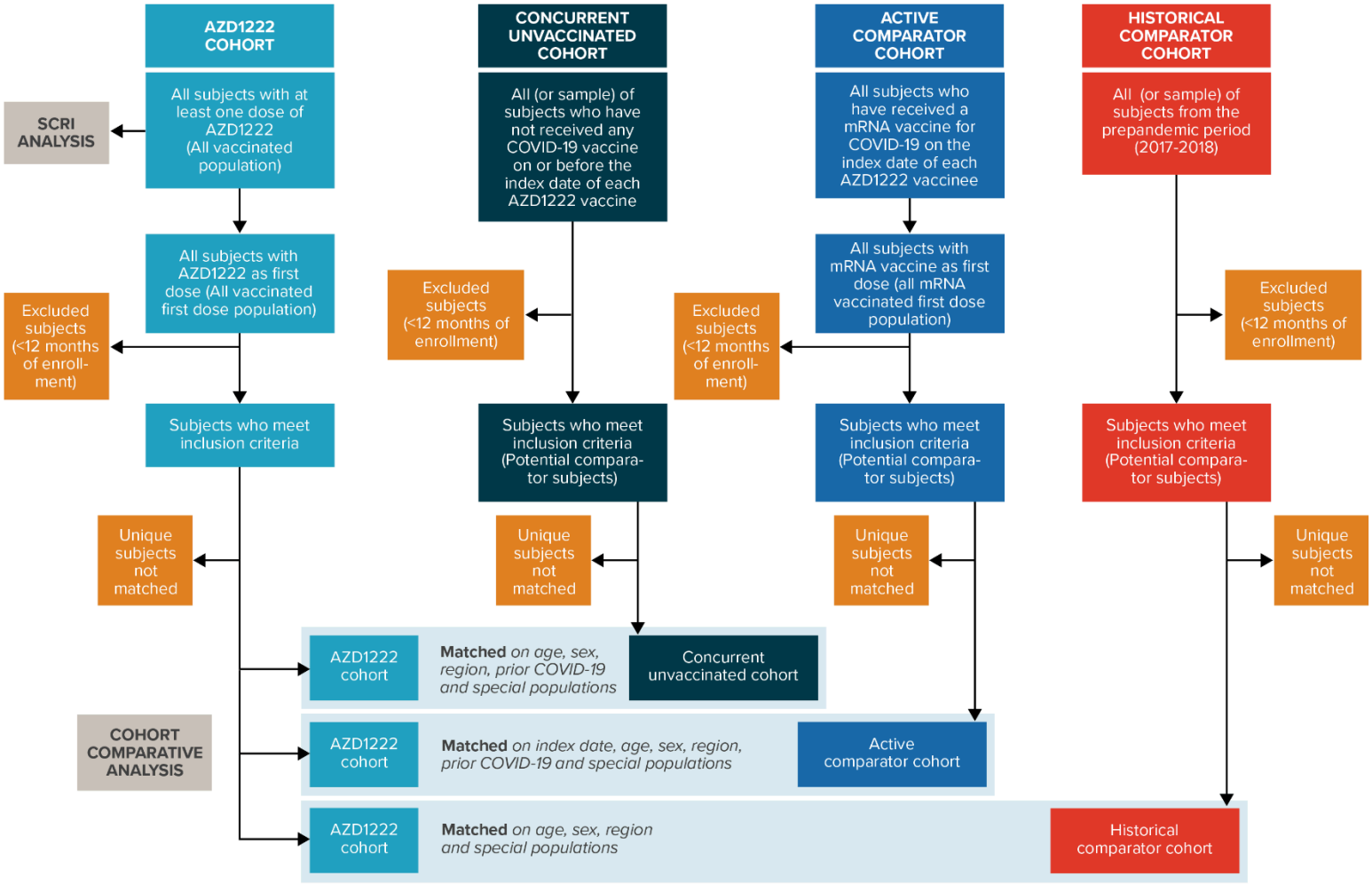
- For each AESI, subjects (in any matched cohort) who had a specific AESI (Table 3) during the look-back period before the index date will be excluded from the cohort for the analysis of that specific AESI, but not from the cohorts for analysis of other AESIs.
  - If a subject had a specific AESI between the index date and the start of the risk window or had any censoring criteria within this time period, the subject will be excluded from the analysis of that specific AESI.

The number of subjects in the *all AZD1222 vaccinated first dose population* and in the *all mRNA vaccinated first dose population* excluded because they had no matches will be described. Similarly, the number of subjects in the initial pool of potential concurrent

unvaccinated comparators and historical comparators that fulfil the inclusion criteria but are excluded because they had no matches will be described.

In the analysis of a specific AESI, if a subject in the *AZD1222 matched cohort* is excluded due to having a prior history of that specific AESI, all its matched subjects will also be excluded from that analysis. Analogously, if the matched subject in the *concurrent unvaccinated comparator* matched cohort or in the *active comparator* matched cohort or in the *historical comparator* matched cohort is excluded, then the corresponding subject in the *AZD1222 matched cohort* will also excluded from the analysis of that specific AESI.

**Figure 2 Assembly of Study Cohorts**



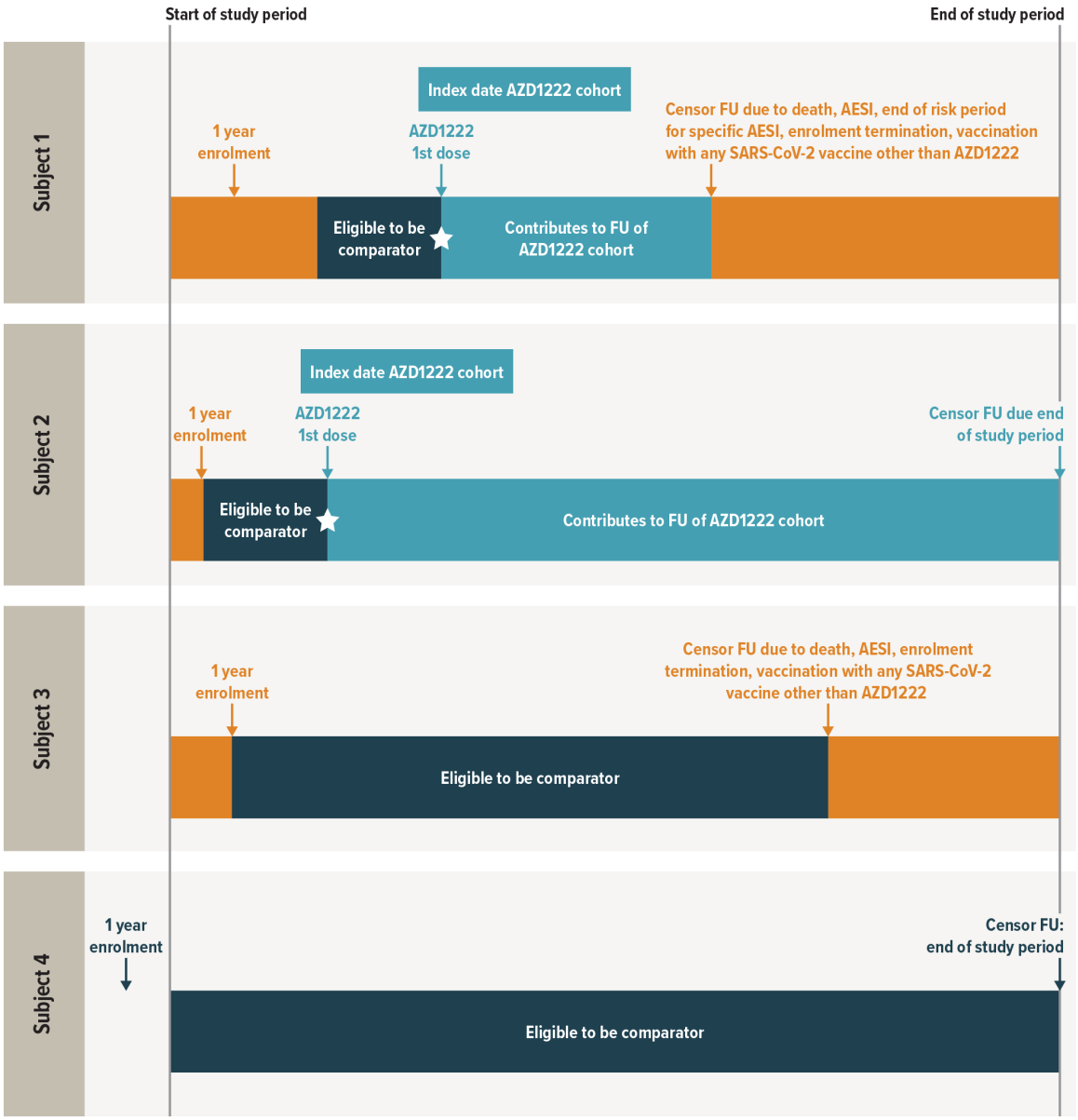
Note: The process illustrated here will be conducted separately in each data source. Index date refers to the date of the first dose of AZD1222 for each vaccinee and for each unvaccinated comparator subject (minus 3 or 4 years for each matched historical control) or the date of the first dose of an mRNA vaccine for each active comparator subject. Subjects in the concurrent unvaccinated cohort will be allowed to enter the AZD1222 or the active comparator cohort if they later fulfil specific cohort entry criteria.

Note that the comparative analysis between AZD1222 and active comparator cohorts will be conducted among vaccinees who have received 2 doses.

COVID-19, coronavirus disease 2019; SCRI, self-controlled risk interval.

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.

Subjects in the concurrent unvaccinated cohort will be allowed to enter any of the AZD1222 or active comparator cohorts if they later fulfil specific cohort entry criteria. Subjects in the active comparator cohort will not be allowed to enter the AZD1222 cohort, and vice versa. Subjects in the historical comparators cohort will be allowed to enter the any of the other cohorts if they later fulfil specific cohort entry criteria.

#### 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 subjects
- Frail subjects with comorbidities (eg, chronic obstructive pulmonary disease, diabetes, chronic neurological disease, cardiovascular disorders)
- Subjects with autoimmune or inflammatory disorders
- Subjects 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 (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 3. 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 at least 1 day of data accrual, which is 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

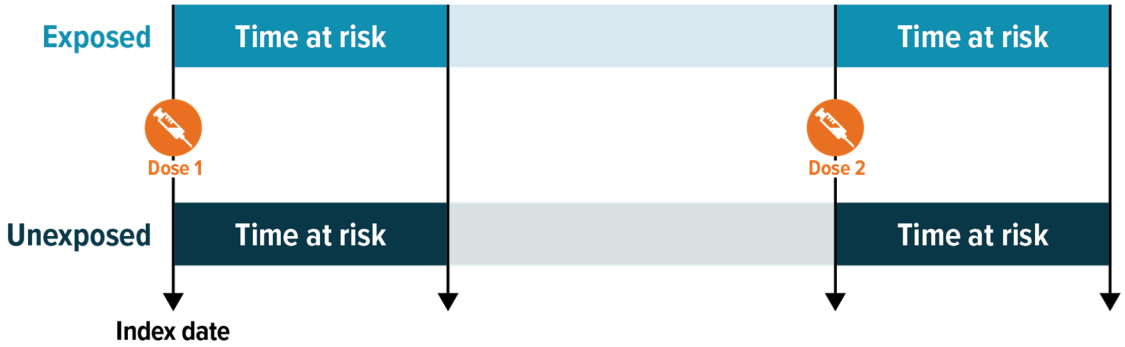
Follow-up time will start on the date that all inclusion criteria are met and will end at the earliest of the occurrence of censoring conditions or the last data extraction/data availability as defined for each study design.

#### 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:

- The end date of the study period.
- The subject's enrolment termination date in the health plan or system.
- The subject's date of death.
- For the specific evaluation of each AESI:
  - The date of the first diagnosis of a specific AESI recorded after the index date, 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 the evaluation of AESIs with **known risk windows** (ie, those for which the end of the risk window is other than 180 or 365 days), the date on which the defined risk window for the last dose ends (the last dose was dose 1 if only 1 dose was received and dose 2 if 2 or more doses were received).
    - For subjects who received 2 doses of AZD1222, or of an mRNA vaccine, and were not censored due to an AESI or other censoring criteria before the second dose, a second risk window will follow the second dose; and for concurrent unvaccinated comparators or historical comparator subjects, a corresponding second risk interval will also apply, mirroring the timing of the second risk interval for the matched vaccinee (Figure 4). 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 AESI (Section 9.1.2). If an AESI occurs during a gap between doses, this AESI will not be included as an event in the study, and as stated above, the follow-up will be censored at the time of the AESI.
    - This risk window applies to both the vaccinee and matched comparator subjects.

**Figure 4. Time at Risk for an AESI With a Known Risk Window in Subjects With 2 Doses of AZD1222**




AESI, adverse event of special interest.

  - For the evaluation of AESIs with an **unknown risk window** (those with a risk window that ends at 180 or 365 days), follow-up will comprise all person-time after the index date and will extend 180 or 365 days (Table 3) after the last AZD1222 or mRNA vaccine dose (dose 1 if only 1 dose was received and dose 2 if 2 or more doses were received) for vaccinated subjects or the corresponding time for the unvaccinated and historical comparators.

- For the AZD1222 cohort:
  - 365 days after the second vaccine dose. This rationale is 2-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 adverse events 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 date on which the subject received a dose of any COVID-19 vaccine other than AZD1222.
  - The date on which a subject receives a 3rd dose of any COVID-19 vaccine
- For the active comparator cohort:
  - 365 days after the second vaccine dose.
  - The date on which the subject received a dose of any COVID-19 vaccine other than Comirnaty or Spikevax.
  - The date on which a subject receives a 3rd dose of any COVID-19 vaccine
- For the concurrent unvaccinated cohort, the date on which the subject received a dose of AZD1222, at which time the subject switches from the comparator to the AZD1222 cohort.

- For the historical comparator cohort, 31 October 2019, an estimate of the date when SARS-CoV-2 started to circulate in Europe.

#### 9.2.3.2 SCRI Design: Follow-up and Time at Risk

The SCRI will be used for most of the AESI with known risk windows as per Table 3. The SCRI evaluates occurrence of a particular AESI during a "risk interval" (exposed -person time) and a subsequent "control interval" (unexposed person-time) following vaccination with AZD1222 (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.

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 COVID-19 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, where possible. The extent of capture of COVID-19 vaccination in the target populations and of specific product types 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 (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 thrombocytopaenia 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 3 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 3 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)[k] |
|---|---|---|---|
| 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] |

## Table 3 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)[k] |
|---|---|---|---|
| Type III hypersensitivity reactions | No[d] | 1-365[e] | 365 |
| 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, and other demyelinating disorders | No[d] | 1-365[e] | 365[c] |
| Transverse myelitis | Yes | 1-90[c] | 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, and heart failure | No[f] | 1-90[e] | 365 |
| Stress cardiomyopathy | No[f] | 1-42 | 365 |
| Thrombocytopaenia | Yes | 1-42[b] | 365[c] |
| Thrombocytopaenia with associated bleeding | No[d] | 1-42[g] | 365[c] |
| Thrombosis (embolic and thrombotic events) without thrombocytopaenia | Yes | 1-42[c] | 365[c] |
| Thrombosis with thrombocytopaenia 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 |

**Table 3 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)[k] |
|---|---|---|---|
| Acute aseptic arthritis | Yes | 1-42[b] | 365 |
| Fibromyalgia | No[d] | 91-365[h] | 365 |
| Rhabdomyolysis | No[d] | 1-42[b] | 365 |
| 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 1 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 thrombocytopaenia 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 subjects 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 subjects 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).

[k] 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; COVID-19, coronavirus disease 2019; FDA, Food and Drug Administration; IOM, Institute of Medicine; ME, myalgic encephalitis; PVFS, postviral fatigue syndrome; SCRI, self-controlled risk interval; TTS, thrombosis with thrombocytopaenia 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 Thrombocytopaenia 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 thrombocytopaenia (TCP)" [as confirmed by both the Brighton Case Definitions for thrombocytopaenia and thrombosis (Brighton Collaboration 2021a, Wise et al 2007)]. The Brighton Collaboration case definition for TCP (Wise et al 2007) requires evidence of a platelet count of less than 150,000/$\mu L^3$.

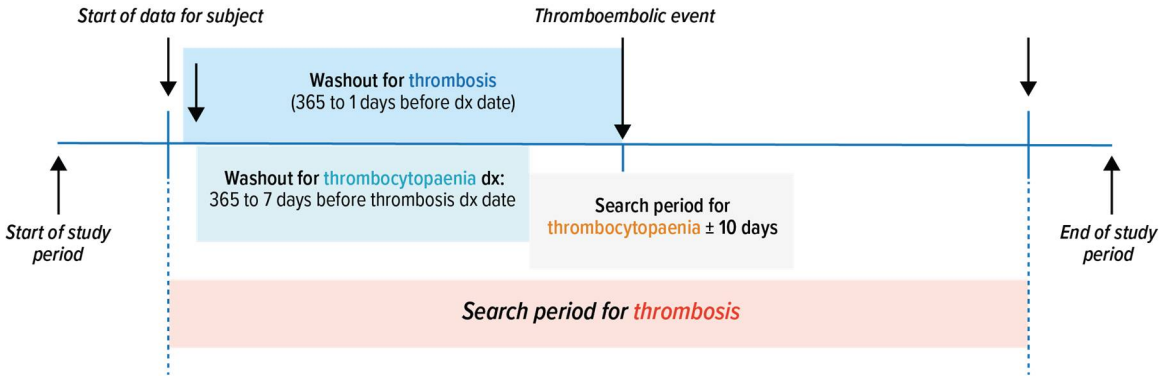
Because laboratory measurements of platelet counts are not expected to be available for most proposed data sources, the case definition of TTS for the current study has been adapted to qualify TCP based on a diagnosis code. In addition, diagnoses that qualify for thrombosis do not include acute myocardial infarction (AMI) or stroke, which is consistent with the May 2021 Centers for Disease Control and Prevention (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 TCP (or laboratory evidence of the same, where available) that is made from 10 days before and up to 10 days after that event (https://zenodo.org/record/5255870#.Yx8KBHZBxPZ) (Burn et al 2022)

Figure 5 shows the look-back ("washout") periods proposed to define "new" diagnoses.

**Figure 5 Time Windows to Define Thrombosis With Thrombocytopaenia Syndrome**




dx, diagnosis.

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 a TE in the main analysis but will be included in a sensitivity analysis in order to evaluate the potential effect of AZD1222 on arterial thrombosis and thromboembolisms. In addition to evaluating TEs overall, it is proposed to evaluate DVT outcomes separately by location of the event into central venous sinus thrombosis, 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 be conducted:

- Including the TTS definition described previously but adding infarctions (including myocardial and cerebral) and stroke (Interim Report 2 and final study report)
- Using alternative risk windows of 1 to 14 days, 1 to 21 days, and 1 to 28 days after the index date (final study report)

Both TE 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.

Identification of the clinical concerns “Thromboses without thrombocytopaenia” and “Thrombocytopaenia without thrombosis” will follow a similar logic.

#### 9.3.2.1.1 Expert Review of TTS Cases Identified in Automated Data

Review of clinical records in hospital charts for potential cases of TTS by research team clinicians will be requested as per access and approval by hospitals in PHARMO, VID, and ARS Toscana. In SIDIAP, additional review can be requested from the general practitioners

(GPs), who have access to the discharge letters. In CPRD, and for any potential cases of TTS in other data sources where hospitals do not grant access for record review, patient profile review will be performed. The electronic patient profile is a cumulative chronologically ordered record of all available electronic linkable information. This record will be reviewed by members of each local research team who are blinded to vaccination status, wherever possible.

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 thrombocytopaenic purpura) accounted for the clinical presentation.

#### 9.3.2.2 Thrombocytopaenia With Bleeding

As of June 2021, there appears to be no standard case definition to study thrombocytopaenia with bleeding following vaccination. The approach proposed to identify this outcome in automated data will follow the same approach used for TTS (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 (Figure 5). Briefly, the working case definition for thrombocytopaenia with bleeding for this study requires 2 components:

- A new qualifying bleeding event (defined below) AND
- A new diagnosis of TCP (or laboratory evidence of the same, where available) that is made from 10 days before and up to 10 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 thrombocytopaenia 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 TCP 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 thrombocytopaenia (ie, platelet count) and any specific cause of the TCP.

### 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 subjects. 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. Healthcare 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 subjects 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 5 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.

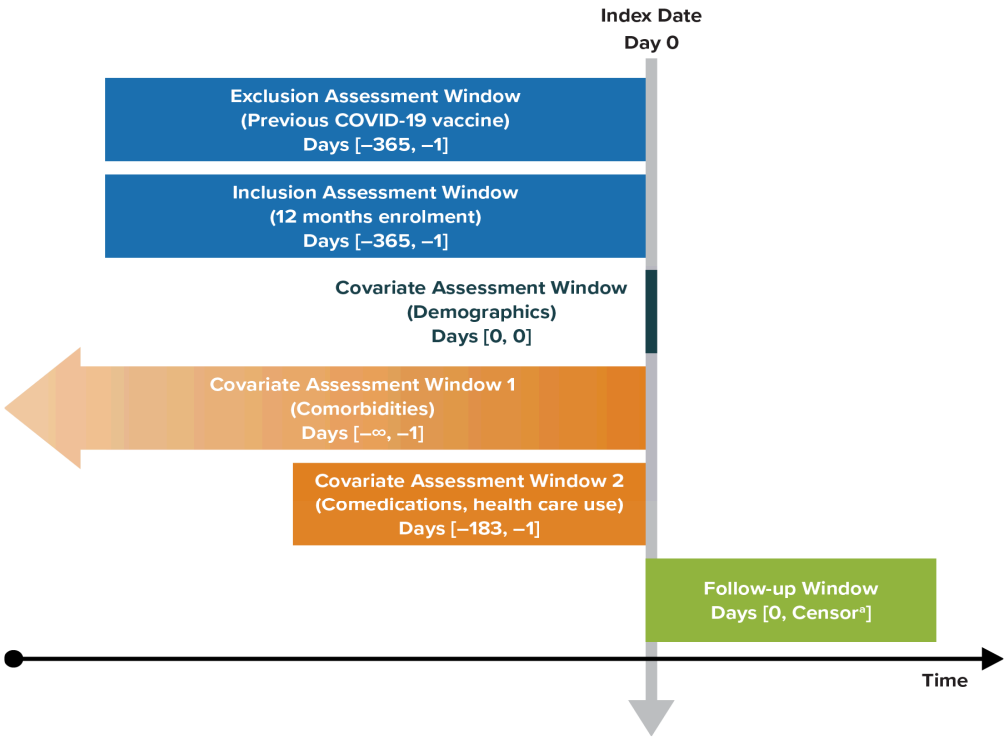
Medical conditions will be defined based on diagnoses (and medications and/or procedures when applicable) as recorded in the respective data sources included in the study (see Section 9.4) at any time before or on the index date, unless otherwise specified. Absence of a diagnosis code or proxies for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

the study population. Figure 6 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 6 Summary of Covariate Ascertainment in Cohort Analyses**




[a] Earliest end date of study period, enrolment termination date, death, vaccination with any COVID-19 vaccine other than the vaccine granting access to the cohort, also for each specific AESI: date of the AESI, end of risk window for that specific AESI.

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 healthcare utilisation, this time window will be 183 days before the index date, except if a different time window is specified for specific medications or healthcare utilisation measures.

AESI, adverse event of special interest; COVID-19, coronavirus disease 2019.

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 5 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 healthcare and maintain subjects' 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.

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 database. Data include demographics, all GP/healthcare professional consultations, diagnoses and symptoms, results from laboratory tests, information about treatments (including prescriptions), data on referrals to other care providers, hospital discharge summaries (date and Read/SNOMED codes), hospital clinic summaries, preventive treatment and immunisations, and death (date and cause). Lag time for CPRD Aurum is usually 1 month, although this may be extended due to CPRD database technical updates. Information about vaccinations from mass vaccination campaigns during the pandemic is expected to transfer to GPs and into the subject's medical records (via National Health Service [NHS] systems rather than subjects informing the GP); however, the lag time for this transfer is not yet clear.

Linkage of the CPRD primary care data with other subject-level datasets 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 subject level to these subject-level datasets. The Hospital Episode Statistics (HES) database contains details of all admissions to NHS hospitals in England (Accident and Emergency, Admitted Subject Care, and Outpatient); approximately 46.8 million individuals in the CPRD are linked to the HES database. Not all subjects in the CPRD have linked data (eg, if they live outside

England, if their GP has not agreed that their data may be used in this way). As with standard CPRD subjects, HES data are limited to subjects with research-standard data. The CPRD records are linked to the HES using a combination of the subject's NHS number, sex, and date of birth (Williams et al 2012). Additional CPRD-linked datasets include Death Registration data from the Office for National Statistics, which includes information on the official date and causes of death (using International Classification of Diseases codes). The Pregnancy Register will be used to identify pregnant women at the index date.

In addition, other CPRD-linked COVID-19 datasets, 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 datasets are usually updated every 6 months, and the lag time between data recording and data availability varies by dataset. The latest linkage set (set 22) 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 intensive care unit/high dependency unit data up to the end of March 2021.

The present study included active CPRD Aurum practices. These practices include an estimated 13.3 million current subjects. CPRD is listed under the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) resources database, and access was 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 healthcare department) as well as some sociodemographic data (eg, sex, date of birth, nationality, employment status, geographic location). Importantly, the SIP database includes the date of death captured from the Mortality Registry. The SIP database is paramount to the VID, as it is the source of the individual, exclusive, and permanent identifier number associated with each individual (the SIP number), which is then used throughout the rest of the databases, thereby allowing data linkage across the multiple databases in the network.
- The Ambulatory Medical Record (ABUCASIS) is the electronic medical record for primary and specialised outpatient activity, with 96% population coverage since 2009. ABUCASIS is integrated by 2 main modules: the Ambulatory Information System (SIA)

and the Pharmaceutical Module (GAIA), including paediatric and adult primary care, mental healthcare, 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, 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 (SIDIAP) in Catalonia, Spain, is a primary care database set up by the Institute of Research in Primary Care (IDIAP Jordi Gol) and Catalan Institute of Health (Institut Català de la Salut). The database collects information from 278 primary healthcare 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, BMI, 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 healthcare 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 healthcare 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 healthcare 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 healthcare 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 subject 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 healthcare 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, BMI, 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 healthcare 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 World Health Organization (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 trusted third party 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 4). All available data in each set will be used to identify subjects who received any COVID-19 vaccine.

**Table 4 Estimates of Study Size and Exposed Population, by Data Source**

| Data source | Approximate current enrolled population (million, protocol V2.1) | Cumulative exposure estimate for the data source based on ECDC vaccine country distribution (doses, protocol V2.1)[a] | Actual exposure data (protocol V2.1) | Actual exposure data (Interim Report 2) |
|---|---|---|---|---|
| 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) | 4,301,194 (of these 3,938,490 were matched) (preliminary results from Interim Report 2) |
| VID (FISABIO), Valencia, Spain[c] | 5 | 605,662 | 484,145 first doses and 171 second doses (communication 30 May 2021) | 530,226 (of these, 517,015 were matched) |
| SIDIAP, Catalonia, Spain[d] | 5.8 | 714,225 | Number of exposures up to 02 Jun 2021 in Catalonia: 916,384 AZ doses | 614,472 (of these, 609,842 were matched) |
| ARS Toscana, Italy[e] | 3.6 | 411,719 | 197,267 first doses (communication 11 Apr 2021) | 343,478 (of these, 336,889 were matched) (preliminary results from Interim Report 2) |

**Table 4 Estimates of Study Size and Exposed Population, by Data Source**

| Data source | Approximate current enrolled population (million, protocol V2.1) | Cumulative exposure estimate for the data source based on ECDC vaccine country distribution (doses, protocol V2.1)[a] | Actual exposure data (protocol V2.1) | Actual exposure data (Interim Report 2) |
|---|---|---|---|---|
| PHARMO, Netherlands[f] | 3.5 | 426.725 | 34,000 first doses (20 May 2021) Based on GP data only; data from COVID-19 vaccination Information and Monitoring System (CIMS) may become available | 175,149 (of these, 154,858 were matched) |

[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 Toscana, Regional Health Agency of Tuscany; AZ, AstraZeneca; CIMS, COVID-19 vaccination Information and Monitoring System (Netherlands); COVID-19, coronavirus disease 2019; CPRD, Clinical Practice Research Datalink; 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; UK, United Kingdom; VID, Valencia Health System Integrated Database.

The precision of study results was estimated for each comparator cohort: concurrent unvaccinated, active comparators, and historical comparators. Additionally, a power estimation for the SCRI design analysis was performed. All calculations assumed that the true IRR between the AZD1222-vaccinated cohort and each comparator cohort was 1.0. All the

calculations estimated the probability that the upper bound of the 95% confidence interval (CI) around the observed IRR was below 1.5, 2.0, 2.5, and 3.0 for various study sizes and AESI rates. The ratio between AZD1222 and comparator person-time varied depending on the type of comparator.

First, the precision of the estimates for the comparison of **AZD1222-vaccinated subjects with concurrent unvaccinated subjects** was estimated assuming a 3:1 ratio for the person-time of AZD1222-vaccinated subjects to concurrent unvaccinated subjects. In Interim Report 1, the follow-up of subjects in the concurrent unvaccinated cohort was short, leading to a 2:1 ratio of person-time despite a 1:5 matching ratio. Results from Interim Report 2 included a longer follow-up and almost the same number of subjects vaccinated with AZD1222, since AZD1222 was no longer administered. The results suggested that the actual ratio of person-time is likely to be 3:1 in the final analysis, with some variations across data sources. The anticipated precision of the study was estimated for one of the least frequent AESIs, Guillain-Barré syndrome (GBS), assuming a background rate of 2 events per 100,000 person-years, 84 days at risk (42 days per dose). With a study size of 5 million vaccinated subjects, it was anticipated that there was a 75% probability that the upper bound of the 95% CI around the observed IRR would be below 3.0 (Table 5).

The PRAC AR for Interim Report 1 (7 July 2022) requested that the MAH discuss the need for additional data sources. As per results from this Interim Report 1, 5.8 million subjects included in the study have received at least one dose of AZD1222. The number of AZD1222 vaccinees and data sources included in the current study at this time is considered acceptable for the purpose of interpreting findings as continuous measures that are estimated with varying degrees of precision (Rothman and Greenland 2018). Therefore, no additional data sources are considered necessary.

**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-Concurrent Unvaccinated Person-time is 3 to 1**

| Adverse event (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 | 0.04 | 0.05 | 0.07 | 0.08 |
| (3 days) | 6 | 2,000,000 | 0.05 | 0.07 | 0.09 | 0.12 |
| | 6 | 3,000,000 | 0.05 | 0.09 | 0.12 | 0.16 |
| | 6 | 4,000,000 | 0.06 | 0.10 | 0.15 | 0.19 |
| | 6 | 5,000,000 | 0.07 | 0.12 | 0.17 | 0.23 |
| GBS (VAC4EU 2021, Willame et al 2021) | 2 | 1,000,000 | 0.06 | 0.11 | 0.16 | 0.22 |
| (42 days) | 2 | 2,000,000 | 0.09 | 0.18 | 0.28 | 0.38 |
| | 2 | 3,000,000 | 0.11 | 0.25 | 0.40 | 0.53 |
| | 2 | 4,000,000 | 0.14 | 0.32 | 0.50 | 0.65 |
| | 2 | 5,000,000 | 0.16 | 0.38 | 0.59 | 0.75 |
| Bell's palsy (Morales et al 2013) (42 days) | 38 | 1,000,000 | 0.47 | 0.90 | 0.99 | 1.00 |
| VTE[a] (Heit 2015) (42 days) | 100 | 1,000,000 | 0.87 | 1.00 | 1.00 | 1.00 |

[a] This adverse event 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 x 2 doses), 56 days for stroke (28-day risk interval x 2 doses), and 84 days (42-day risk interval x 2 doses) for all other adverse events (Rothman and Greenland 2018). Calculations follow the method of Rothman and Greenland (2018).

AESI, adverse event of special interest; GBS, Guillain-Barré syndrome; IRR, incidence rate ratio; PY, person-years; RR, risk ratio; VAC4EU, Vaccine monitoring Collaboration for Europe; VTE, venous thromboembolism.

Second, the precision for the results comparing **AZD1222-vaccinated subjects with active comparators and historical comparators** was estimated. The estimation of the precision of the results assumed a 1:1 ratio for the person-time of AZD1222-vaccinated subjects to active comparators and to historical controls (Kaura et al 2022). For one of the rarest anticipated events, GBS, assuming a background rate of 2 events per 100,000 person-years, 84 days at risk (42 days per dose), and a study size of 5 million vaccinated subjects, it was anticipated that there was a 96% probability that the upper bound of the 95% CI around the observed IRR would be below 3.0 (Table 6).

The number of AZD1222 vaccinees and data sources for the comparison of AZD1222 vaccinees with active and historical comparators is considered acceptable for the purpose of

interpreting findings as continuous measures that are estimated with varying degrees of precision. Assuming 1:1 ratio of matching will be achieved with the 5.8 million AZD1222 vaccinated subjects, no additional data sources are considered necessary.

**Table 6 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 Subjects to Active Comparators Person-time and to Historical Comparators Person-time is 1 to 1**

| Adverse event (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) (42 days) | 38 | 1,000,000 | 0.76 | 1.00 | 1.00 | 1.00 |
| VTE[a] (Heit 2015) (42 days) | 100 | 1,000,000 | 0.99 | 1.00 | 1.00 | 1.00 |

[a] This adverse event 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 x 2 doses), 56 days for stroke (28-day risk interval x 2 doses), and 84 days (42-day risk interval x 2 doses) for all other adverse events. Calculations follow the method of Rothman and Greenland (2018).

AESI, adverse event of special interest; GBS, Guillain-Barré syndrome; IRR, incidence rate ratio; PY, person-years; RR, risk ratio; VAC4EU, Vaccine monitoring Collaboration for Europe; VTE, venous thromboembolism.

Third, for the outcome of TTS, study size implications were evaluated in terms of precision of the absolute risk difference, rather than IRR, because the background risk for this newly described syndrome is unknown and conceivably could be zero (Table 7). 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 per 100,000 and 1 per 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 7 shows the probability that the upper 95% confidence limit of the absolute risk difference 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 1,000,000 people vaccinated there is a 99% probability that the upper confidence limit of the absolute risk difference will fall below 3.0 per 100,000.

**Table 7 Probability That the Upper 95% Confidence Limit of the Observed 4-Week Absolute Risk Difference for TTS Will Be Below Various Thresholds for Different Study Sizes, Assuming that a True 4-Week Risk Among Vaccinees is 1 per 100,000 or 1 per 250,000**

| Estimated 4-week incidence among vaccinated individuals | Estimated 4-week incidence among concurrent 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).

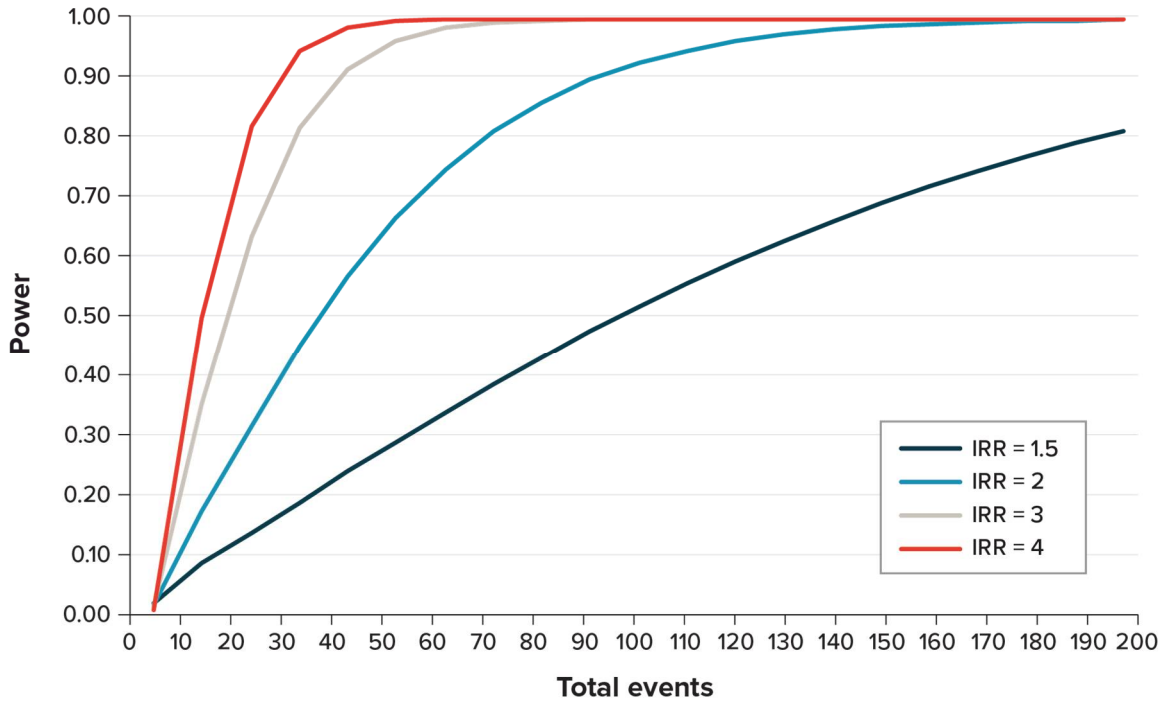
ARD, absolute risk difference; TTS, thrombosis with thrombocytopaenia syndrome; WHO, World Health Organization.

Finally, the study size was estimated for the **SCRI** study component. The methods described by Musonda et al (2006) were used. These methods follow a traditional power and statistical

significance approach and do not currently allow a precision approach. The 2 parameters of most interest and specific to the SCRI design are the proportion of the observation period in the exposed state (ie, 50% in this study) and the relative incidence of events during the observation period in the exposed state compared with the incidence of events during the observation period in the unexposed state (ie, IRR). For this estimation, we assumed several IRR scenarios of 1.5, 2, 3 and 4. The study size estimation is not based on independent subjects but rather on the occurrences of the event of interest, which may happen more than once in the same subject. The SCRI includes only subjects that experienced the event of interest at least once.

Figure 7 displays different scenarios of IRRs for AESIs assuming a 50% of observation period in the exposed state, an alpha error of 5%, and a range of 0 to 200 events. In scenarios with an IRR of 1.5, including 200 events, a maximum power of 80% will be expected. In the scenario considering an IRR of 2 for certain AESIs, approximately 70 events will be required to have a power of 80%. Finally, in scenarios considering an IRR of 3 with 40 events or an IRR of 4 with 20 events, the study will have more than 90% power.

**Figure 7 Different Scenarios for Self-controlled Risk Interval Design**

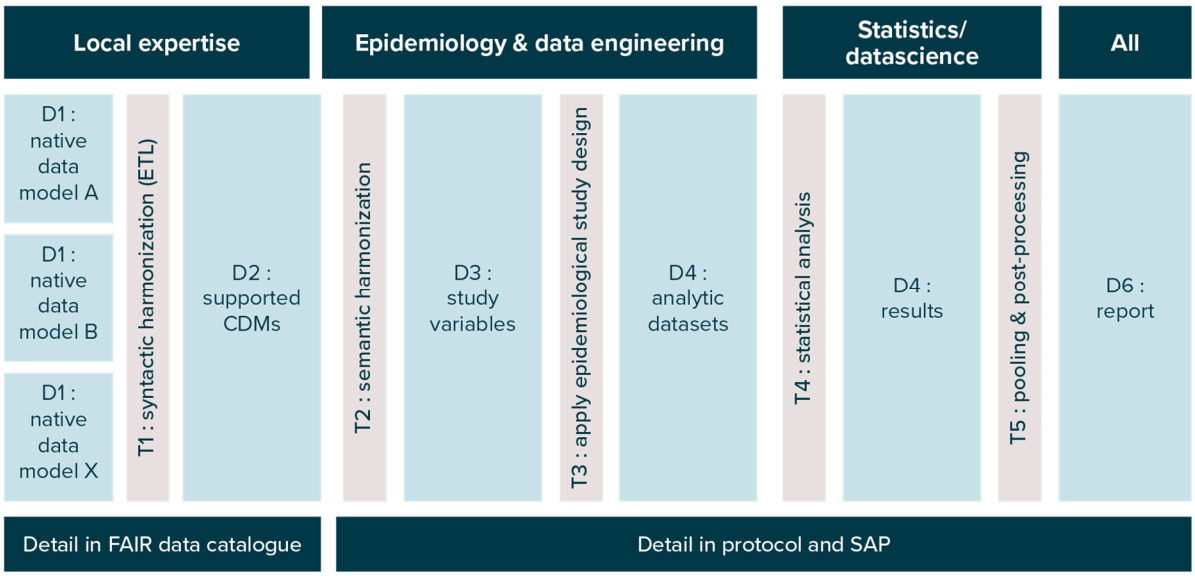



IRR, incidence rate ratio.

## 9.6 Data Management

The following transformation steps (T) will be implemented (Figure 8):

**Figure 8. Data Transformation Process**




CDMs, common data models; D*n*, data type; ETL, extract, transform, and load; FAIR, findable, accessible, interoperable, and reusable; SAP, statistical analysis plan, T*n*, transformational step (for explanation, see below).

### 9.6.1 T1: Syntactic Harmonisation (ETL)

T1: Syntactic harmonisation through an extract, transform, and load (ETL) process of native data into the ConcePTION CDM (ConcePTION Study 2020). In this CDM, data are represented using a common structure, but the content of the data remain in their original format. The CDM version that will be used was v2.2, which is available as an open-source CDM. The CDM was developed as part of the IMI-ConcePTION project (project number IMI-821520). The ETL process has various structured steps as described in Thurin et al (2021):

1 Data access partners (DAPs) will be asked to share the data dictionaries of their data banks (tables and variable names/structure) with the principal investigator of the study.

2 Based on the data dictionaries of the data banks, an interview will be conducted by the Principal Investigator to explore what action(s) prompt the creation of a record, what is missing, and the context of each of the data banks.

3 Metadata (descriptive data about the data sources and databanks), data dictionaries, and interview answer sheets will be uploaded in the VAC4EU FAIR (findable, accessible, interoperable, and reusable) data catalogue, according to the metadata structure for electronic health data that was defined in IMI-ConcePTION and the EMA-funded MINERVA project.

4 An overview will be created of all the required study variables and definitions in order to create the code lists to identify the outcomes and covariates (see T2).

5 Instructions for the ETL design will be provided by the scientific coordinator, these instructions will comprise the required CDM tables, mandatory variables, the calendar period over which data need to be extracted, and the code lists for the data to be extracted.

Once the ETL design will be approved, it will be executed by each DAP using its programming language. The output files will be CSV (comma-separated values) files. Once the ETL is conducted, Level 1 and 2 data quality checks will be conducted to measure the integrity of the ETL, as well as internal consistency within the context of the CDM.

1 Level 1 data checks will be performed to assess the completeness and content for each variable in all CDM tables to ensure that the mandatory variables contain data and conform to the formats specified by the CDM specifications.
2 Level 2 data checks will be performed to assess the logical relationship and integrity of data values within a variable, or between 2 or more variables within and between tables.

### 9.6.2 T2: Semantic Harmonisation

To reconcile differences between terminologies and native data availability, a shared semantic foundation will be applied for the creation of study variables. This is a multistep process:

1 Definition of study variables, which will be done using an event definition form and will be a living document that will be closed upon AZD1222 PASS ending.
2 Initial code lists will be created using the VAC4EU CodeMapper tool (https://vac4eu.org/codemapper/) which was developed during the IMI-ADVANCE project to assist in the creation of code sets from event definitions for several coding systems simultaneously while keeping a record of the complete mapping process.
3 Study variables will be named in a standard hierarchical fashion based on body system.
4 Review of the codes by DAPs.
5 5. Consolidation: Comments from DAPs will be consolidated by the study team.
6 Based on the relevant diagnostic medical codes and keywords, the algorithms will be constructed to operationalise the measurement of each study variable. These algorithms could differ by database, as the components relevant for the study variables may differ.
7 During the T2 step, transformations will occur for a series of steps related to completion of missing features in the data, eg, dose of vaccines, sorting on record level, combination of concepts for algorithm, and rule-based creation of study variables on a personal level for the study population (specific, if needed, per DAP).
8 Once the study variables are created, Level 3 checks will be deployed, which will be targeted to assess the patterns of study variables.

### 9.6.3 T3: Application of the Epidemiological Study Design

Based on the creation of the study variables on a person level or a medicines level, epidemiological designs will be applied. Specifically, matching on specific variables and selection based on inclusion and exclusion criteria.

### 9.6.4 T4: Statistical Analysis

This step in the data transformation pipeline produced statistical estimates such as descriptives (counts, percentages), distributions (mean, percentiles), and rates (prevalence, incidence). This was conducted using R.

#### 9.6.4.1 Scripting and Deployment

The analytical R scripts that produced the T2-T4 steps will be produced on VAC4EU GitHub for version control; links to the latest script will be distributed to DAPs for local deployment. Any issues will be notified, and the data engineers who are responsible for the R code will work with the local DAP to resolve issues. Scripts will be developed independently, based on data engineering programmes and the codebook in R (by University Medical Center Utrecht [UMCU]) will be validated using SAS (by RTI Health Solutions), comparing the output from both analysis programs was against each other for consistency.

### 9.6.5 T5: Results and Pooling After Processing

The aggregated results that will be produced through T4 will be uploaded to the Digital Research Environment (DRE) for pooled analyses and visualisation (T5, Figure 8). The DRE will be made available through VAC4EU and UMCU (The anDREa consortium 2021). The DRE is a Microsoft Azure cloud-based, research environment with double authentication where researchers can collaborate using data that are stored and organised securely. VAC4EU can use this facility through the UMCU tenant. UMCU is responsible for data processing and data security.

## 9.7 Data Analysis

In accordance with the recommendations of the American Statistical Association, the International Committee for Medical Journal Editors (ICMJE 2019), and expert opinion on the misuse of significance testing (Greenland et al 2016, *Nature* editorial 2019, Rothman and Lash 2021), we avoid relying on statistical significance to interpret study results. Instead of a dichotomous interpretation based on p-values and significance testing, we rely on a quantitative interpretation that considers the magnitude, precision, and possible bias in the estimates that we derive and report. We believe that this is a more appropriate approach than one that ascribes to chance any result that does not meet conventional criteria for statistical significance.

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 9.7.11). 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 in the *matched population* (ie, subjects vaccinated with AZD1222 and the 3 different comparator cohorts) 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.

In order to understand the impact of censoring follow-up for individuals rather than for the matched pairs, the follow-up time by matching covariates for AZD1222 and each comparator cohort will be described.

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 COVID-19 Vaccine Use

The analyses described below will address primary objective 2.

- Among subjects who received dose 1 of AZD1222 (ie, *all AZD1222 vaccinated first dose population*) and those who received dose 1 of an mRNA vaccine (ie, *all mRNA vaccinated first dose population)*, 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 COVID-19 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 COVID-19 vaccine, specifying whether it was the AZD1222 vaccine or another COVID-19 vaccine and type (brand name)

  - Among those who received 2 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 1 PS for the analysis of all outcomes for each comparison. 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 subjects 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

Separate PSs will be estimated for each of the 3 comparisons conducted: AZD1222 cohort with concurrent unvaccinated comparator cohort, AZD1222 cohort with active comparator cohort, and AZD1222 cohort with historical comparator cohort.

Logistic regression models will be used to estimate PSs. 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 will be produced for each cohort for each comparison.

#### 9.7.3.3 Selection of Cohorts for Comparative Analyses and Assessment of Balance

After PSs have been estimated, using an approach that will be specified in the SAP, they will be used to refine the balance between study cohorts, which were initially matched, as applicable, on calendar date, age, sex, region, prior diagnosis of COVID-19, and status according the 5 special populations. Whether all potential comparators, or only a sample, will be included 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.

#### 9.7.3.4 Feasibility Assessment of Cohort Comparability for the Active Comparator Population

The feasibility assessment of cohort comparability will be conducted in the pre-matching populations (ie, after applying inclusion criteria and prior to matching), for AZD1222 vaccinees with 2 consecutive doses of AZD1222) and active comparators cohort, also with 2 doses of mRNA vaccines.

The distribution of the baseline covariates used for matching will be described within each data source. Covariate balance will be assessed using the standardised mean differences and will be displayed graphically. Histograms detailing the overlap in age by calendar time between subjects in the AZD1222 cohort and subjects in the active comparator cohort will be provided, both overall and separately stratified by special population group. The denomination of calendar time will be defined in the SAP.. Additionally, a logistic regression model will estimate propensity scores for receiving AZD1222 with the matching variables as covariates.. Following the estimation of the propensity scores, the predictive ability of the model can be assessed using the area under the receiver operating characteristic (ROC) curve. A ROC curve plots the sensitivity against 1 minus the specificity of a model. Where our propensity score model is an excellent predictor of treatment (i.e. AUC > 0.8;(Mandrekar 2010)), this suggests that the covariates are strong predictors of treatment and indicates fundamental differences between the AZD1222 cohort and the active comparator cohort. Large values of AUC of ROC curves indicate the importance of adjusting for all confounders (observed and unobserved) correctly to prevent unfair treatment comparisons. In the situation of large values of AUC, the suitability of the analysis with the active comparator will be reconsidered. Additionally, the examination of the effective sample size will be considered. Should we be able to proceed with the comparative analysis, the matching will be conducted as described in Section 9.2.1.1.3.

### 9.7.4 Exclusions for Analysis of Specific AESIs

In general, the entire population identified in Section 9.7.3 will be included in the analysis of IRs, and with restrictions as appropriate for 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 3. 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 subjects 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 BMI 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, subjects 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 subjects 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 and exploratory objective 1.

- For AESIs with **known risk windows** (those for which the end of the risk window is other than 180 or 365 days), crude IRs, and 95% CIs for each AESI will be estimated for the vaccinated cohort and the comparator cohorts. 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 2 specific AESIs, anaphylaxis and sudden death, since time to the event was not of interest, the prevalence proportions (PPs) and 95% CIs will be estimated.

- For AESIs with **unknown risk windows** (those with a risk window that ends at 180 or 365 days), cumulative incidence will be estimated using 1 − Kaplan-Meier survival curves starting at the index date. In addition, for AESIs with **unknown risk windows**, we will also describe the cumulative incidence over sequential time periods (eg, fortnights) to explore potential changes in risk over the follow-up period.
- In the primary analysis, IRs, PPs, and cumulative incidence of AESIs will be described for the AZD1222 cohort and for the 3 comparator cohorts (concurrent unvaccinated comparators, active comparators, and historical comparators) selected for comparative analyses (Section 9.7.3.3).
- For exploratory objective 1, in the AZD1222 matched cohort to active comparator cohort, the AESIs following a single dose of AZD1222 will be described.

### 9.7.7 Crude and Adjusted Measures of Association: Cohort Design

The analyses described below will address primary objective 4 and exploratory objectives 2.

- The analysis will compare subjects vaccinated with AZD1222 with the concurrent unvaccinated comparators and historical comparators 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, will be estimated as prior vaccine exposure may "prime" the immune system.
- For outcomes that have **known risk windows**, 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 anaphylaxis and sudden death, crude and adjusted PPRs and PPDs with 95% CIs will be estimated using Poisson regression models with robust estimation of the variance to account for individuals who may contribute person-time to the exposed and unexposed cohorts at different timepoints.
- For outcomes **with unknown risk windows**, 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).

The analysis described below will address exploratory objective 3

- The analysis will compare subjects vaccinated with 2 doses of AZD1222 with the active comparator cohort consisting of those vaccinated with 2 doses of an mRNA vaccine.
- For outcomes that have **known risk windows**, we propose to estimate crude and adjusted IRRs and IR differences with 95% CIs using Poisson regression models. For anaphylaxis and sudden death, crude and adjusted PPRs and PPDs with 95% CIs will be estimated using Poisson regression models.
- For outcomes **with unknown risk windows**, 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 (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 Sensitivity Analyses

For each of the matched cohort populations (ie, AZD1222 cohort with concurrent unvaccinated comparator cohort, AZD1222 cohort with active comparator cohort, and AZD1222 cohort with historical comparator cohort), the following sensitivity analyses will be conducted analogously to the main analysis:

- The direct effect of vaccination with AZD1222 on the AESIs will be estimated, excluding subjects with prior COVID-19 diagnosis and thus censoring their follow-up at the occurrence of a COVID-19 diagnosis.
- The effect of partial censoring within matched pairs on the AESIs will be estimated by censoring both members of a matched pair when one member is censored.
- For thrombocytopaenia with bleeding, an alternative risk window of 1 to 28 days at the index date (instead of the 1 to 42 days in the primary analysis) will be assessed.
- For TTS; 3 alternative risk windows of 1 to 14 days, 1 to 21 days, and 1 to 28 days after the index date (instead of 1 to 42 days in the primary analysis) will be assessed.
- For TTS; an alternative definition of the outcome will include not only thrombocytopaenia with venous thromboembolisms but also thrombocytopaenia with AMI or stroke. The risk window between thrombocytopaenia and AMI or stroke will be the same as the one used for venous thromboembolism, ie, 10 days.
- For TTS; AESI cases will be restricted to confirmed cases (ie, excluding cases that could not be validated). The resulting positive predictive values (PPVs) will be used to correct the observed PPV from the main results. This analysis will be performed depending on the total number of outcomes that will be validated (as detailed in Section 9.3.2.1.1).
- In CPRD, a sensitivity analysis using hospital data to assess the AESIs will be performed at the time of the final analysis. However, depending on the duration of the lag time for hospital data at the time of the final analysis, the analysis using hospital data may be considered as the primary analysis.

#### 9.7.10.1 Negative Control Outcome

For the matched cohorts, cumulative incidence (1 − KM estimate) and curves of urinary tract infections and non-pathological fractures within 12 days after the index date will be produced using the inverse probability of treatment weighting from the main analysis. The KM estimator will be estimated as for the main analysis (Section 9.7.6).

### 9.7.11 Meta-analysis

Main estimates of association from the participating data sources will be pooled using meta-analytic techniques. Crude and/or adjusted IRs, PPs, cumulative incidence, IRRs, PPRs, 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.

The analysis will be conducted and reported separately for each data source. The main estimates of association from the participating data sources will be pooled in the final report using meta-analytic techniques only if appropriate, based on a review of the effect measures, population characteristics, type and availability of data in the data sources included, and disease knowledge.

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.

### 9.7.12 Progress and Interim Analysis

A first progress report was 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. The second progress report, which replaced interim report 3, as agreed with EMA, will be delivered in April 2023, followed by the feasibility assessment report for the comparative analysis in June 2023.

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 1 or 2 doses of AZD1222, the number of subjects vaccinated with one dose of AZD1222 and a second dose of another COVID-19 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 quality-control (QC) procedures 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, United Kingdom**

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 2017b) 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 General Data Protection Regulation (GDPR),

which came into effect in May 2018, and which will supersede national legislation within the 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 2 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 subjects 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) *Guidelines for Good Pharmacoepidemiology Practices (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 healthcare 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 healthcare system will not be recorded in secondary electronic health records databases; this could occur, for example, if healthcare 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" (EUPAS37273). 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 subject'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 windows**, 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. As this assumption is likely inaccurate, eg, because receiving a first dose sensitises the immune system to react against a second dose, the estimates of the main analysis could be biased. This factor may be especially important when evaluating immune-mediated AESIs, such as anaphylaxis. We will evaluate the IRs after each dose and 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 healthcare 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 healthcare use.

It is expected that the subject 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 healthcare-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 **concurrent unvaccinated comparators will dwindle** or that people who remain unvaccinated may be systematically different from those who choose to be vaccinated. Additionally, the age restrictions on the use of AZD1222 imposed in the UK, Spain, Italy, or the Netherlands, as in most European countries, 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 includes only exposed individuals, to evaluate vaccine-AESI associations for outcomes that are suitable for this design (Table 3). In the first year of the current study, the availability of comparators and the duration of their follow-up time have been monitored. The results from analyses for the first and second interim reports confirm that the number of concurrent unvaccinated comparators was appropriate; however, the duration of follow-up was too short to evaluate outcomes with long risk windows. For this reason, the protocol was amended to include an active comparator cohort and a historical comparator cohort.

Inclusion of historical comparators raises concerns over non-comparability. 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 healthcare-seeking behaviours. Using historical controls from the pandemic era before the availability of COVID-19 vaccines could theoretically overcome this challenge. The current protocol plans to select comparators from 2017-2018 matched on the day and month of the first vaccination of the exposed subjects—an approach that could control for seasonality.

Results from Interim Report 2 also showed that vaccination against COVID-19 occurred within targeted age groups with different COVID-19 vaccines and at different calendar times, and that there was little overlap for some age groups, and when some overlap of age groups occurred, this occurred in different calendar periods. A pre-matching feasibility assessment of cohort comparability will be performed to examine the overlap of matching covariates in the AZD1222 cohort and the **active comparator cohort** (Section 9.7.3.4). Large values of AUC of ROC curves indicate the importance of adjusting for all confounders (observed and unobserved) correctly to prevent unfair treatment comparisons. In the situation of large values of AUC, the suitability of the analysis with the active comparator will be reconsidered. The active comparator analysis will be focused on the two-dose vaccination strategy, which may lead to selection bias, which may or may not be differential.

Since mRNA doses are much closer in time than AZD1222 vaccination doses, events following dose 1 for mRNA vaccinees could fall in the dose 2 risk window. This will be less likely with the dose 2 risk window in the AZD1222 cohort.

**Multiple use of the same comparator**: In the PRAC assessment report for Interim Report 1 (Assessment Report for the Post-Authorisation Measure 007.6, 21 July 2022), the rapporteurs expressed concerns regarding the multiple use of the same concurrent unvaccinated subject based on the matching with replacement in a ratio of 1 AZD1222-vaccinated subject to 5 concurrent unvaccinated subjects. Based on the PRAC's suggestion, the matching ratio for AZD1222 to concurrent unvaccinated subjects was modified from 1:5 to 1:1. This measure reduced the number of times that a concurrent unvaccinated subject was matched. In the analysis for Interim Report 1, the percentage of comparators used only 1 time for matching ranged from 35% in CPRD to 60% in ARS Toscana, while in the analysis for Interim Report 2, it ranged from 65% in CPRD to 85% in ARS Toscana. Similarly, the percentage of comparators used 3 times or more ranged from 18% in ARS Toscana to 46% in CPRD in Interim Report 1, while in Interim Report 2, it ranged from 2% in ARS Toscana to 13% in CPRD (results for Interim Report 2 are preliminary). Robust estimation of the variance will be used to account for the remaining repeated matches to concurrent unvaccinated subjects, as well as to account for the repeated matches when using active comparators or historical comparators. PRAC Rapporteurs also expressed their concern about the use of general estimating equations (GEEs) taking into account only one of the sources of correlation to estimate the variance of the incidence rate ratio of AESIs and proposed the use of GEE methods with the combination of 3 covariance matrixes (Austin and Cafri 2020, Stuart 2010). The 2 GEE methods together with bootstrapping will be evaluated for their use in the final study report. Bootstrapping may be considered if it is the only way to provide measures of uncertainty given the constraints of study size and issues with convergence.

Finally, data for some of the covariates proposed to define the subgroups of interest included as having 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 subjects 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. 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 subjects 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 subject 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 non-interventional study reusing healthcare 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 subjects.

## 10.1 Other Good Research Practice

This study will adhere to the *Guidelines for Good Pharmacoepidemiology Practices (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 non-interventional (observational) study referred to in the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) 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 2017b), 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 2017b).

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 *Guideline on Good Pharmacovigilance Practices (GVP) Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2)* (EMA 2017a), non-interventional studies using secondary data such as those described in this protocol, conducted using medical chart reviews or electronic claims and healthcare records, do not require expedited reporting of adverse events/adverse reactions (EMA 2017a, 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 *Guideline on Good Pharmacovigilance Practices (GVP)* Module VIII Section B.6.3 (EMA 2017b).

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 Section 9.7.12, Progress and Interim Analysis.

In its *Guidelines for Good Pharmacoepidemiology Practices (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 *Guideline on Good Pharmacovigilance Practices (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 2017b), 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.

Austin PC, Cafri G. Variance estimation when using propensity-score matching with replacement with survival or time-to-event outcomes. Stat Med. 2020 May 20;39(11):1623-40.

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.

Burn E, Li X, Kostka K, Stewart HM, Reich C, Seager S, et al. Background rates of five thrombosis with thrombocytopenia syndromes of special interest for COVID-19 vaccine safety surveillance: incidence between 2017 and 2019 and patient profiles from 38.6 million people in six European countries. Pharmacoepidemiol Drug Saf. 2022 Feb 22.

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.

ConcePTION Study. D7.5 Report on existing common data models and proposals for ConcePTION 2020. Available at: https://www.imi-conception.eu/wp-content/uploads/2020/10/ConcePTION-D7.5-Report-on-existing-common-data-models-and-proposals-for-ConcePTION.pdf. Accessed 2 July 2021.

EMA. Guideline on good pharmacovigilance practices (GVP). Module VI – Collection, management and submission of reports of suspected adverse reactions to medicinal products (Rev 2). European Medicines Agency; 22 November 2017a. Available at: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-good-pharmacovigilance-practices-gvp-module-vi-collection-management-submission-reports_en.pdf. Accessed 8 April 2022.

EMA. Guideline on good pharmacovigilance practices (GVP). Module VIII – Post-authorisation safety studies (EMA/813938/2011 Rev 3). European Medicines Agency; 13 October 2017b. 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 8 April 2022.

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 8 April 2022.

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 8 April 2022.

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 8 April 2022.

ENCePP. Guide on methodological standards in pharmacoepidemiology (EMA/95098/2010 Rev. 7). European Network of Centres for Pharmacoepidemiology and Pharmacovigilance; July 2018c. Available at: https://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml. Accessed 8 April 2022.

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 11 November 2021.

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. Available at: https://www.bestinitiative.org/wp-content/uploads/2020/12/C19-Vaccine-Safety-Protocol-2020.pdf Accessed 4 November 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.

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016 Apr;31(4):337–50.

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

ICMJE. Preparing a manuscript for submission to a medical journal. 2. Reporting guidelines. International Committee of Medical Journal Editors; 2019. Available at: http://icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html#two. Accessed 8 April 2022.

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 4 November 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 8 April 2022.

Kaura A, Trickey A, Shah ASV, Benedetto U, Glampson B, Mulla A, et al. Comparing the longer-term effectiveness of a single dose of the Pfizer-BioNTech and Oxford-AstraZeneca COVID-19 vaccines across the age spectrum. EClinicalMedicine. 2022 Apr;46:101344.

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.

Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010 Sep;5(9):1315-6.

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.

Musonda P, Farrington CP, Whitaker HJ. Sample sizes for self-controlled case series studies. Stat Med. 2006 Aug 15;25(15):2618-31.

*Nature* editorial. It's time to talk about ditching statistical significance. Nature. 2019 Mar;567(7748):283.

NHS Digital. COVID-19 public health directions 2020. 17 March 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.

Rothman KJ, Lash TL. Precision and study size. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. Modern epidemiology, 4th edition. Wolters Kluwer; 2021.

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.

Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci. 2010 Feb 1;25(1):1-21.

The anDREa consortium. anDREa workspace. 2021. Available at: https://mydre.org/. Accessed 2 July 2021.

Thurin NH, Pajouheshnia R, Roberto G, Dodd C, Hyeraci G, Bartolini C, et al. From inception to ConcePTION: genesis of a network to support better monitoring and communication of medication safety during pregnancy and breastfeeding. Clin Pharmacol Ther. 2021 Nov 26;111(1):321-31.

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 4 November 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.

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). World Health Organization; 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 4 November 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 (2.0). Zenodo. 30 April 2021. Available at: https://doi.org/10.5281/zenodo.5255870. Accessed 25 August 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.

1.

# 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 | | | | |
| 1.1.1 Start of data collection[3] | ☒ | ☐ | ☐ | 6 |
| 1.1.2 End of data collection[4] | ☒ | ☐ | ☐ | 6 |
| 1.1.3 Progress report(s) | ☒ | ☐ | ☐ | 6 |
| 1.1.4 Interim report(s) | ☒ | ☐ | ☐ | 6 |
| 1.1.5 Registration in the EU PAS Register® | ☒ | ☐ | ☐ | 6 |
| 1.1.6 Final report of study results | ☒ | ☐ | ☐ | 6 |

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 |
| 3.3 Does the protocol specify measures of occurrence? (eg, rate, risk, prevalence) | ☒ | ☐ | ☐ | 9.7.6 |

[3] 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.

[4] Date from which the analytical data set is completely available.

| **Section 3: Study design** | **Yes** | **No** | **N/A** | **Section Number** |
|---|---|---|---|---|
| 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.<br>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 |
| 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 |

| **Section 5: Exposure definition and measurement** | **Yes** | **No** | **N/A** | **Section Number** |
|---|---|---|---|---|
| 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 3, 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, healthcare 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 |
| 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: | | | | |
| 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 |

| **Section 9: Data sources** | **Yes** | **No** | **N/A** | **Section Number** |
|---|---|---|---|---|
| 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, 9.4.5 |

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? | ☒ | ☐ | ☐ | 9.8, 10 |
| 13.3 Have data protection requirements been described? | ☒ | ☐ | ☐ | 9.4.1, 9.6.3, 10 |

Comments:

| |
|---|
| |

| **Section 14: Amendments and deviations** | **Yes** | **No** | **N/A** | **Section Number** |
|---|---|---|---|---|
| 14.1 Does the protocol include a section to document amendments and deviations? | ☒ | ☐ | ☐ | 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: 10/October/2022

Signature: _____

# 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 |
| **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) |
| Healthcare 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 |
| **Healthcare 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 4 | |
| **Document ID:** | Doc ID-004619144 | |
| **Version Label:** | 3.0 CURRENT LATEST APPROVED | |
| **Server Date** **(dd-MMM-yyyy HH:mm 'UTC'Z)** | **Signed by** | **Meaning of Signature** |
| 27-Oct-2022 12:38 UTC | PPD | Management Approval |
| 27-Oct-2022 13:08 UTC | PPD | Qualified Person Approval |

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