PASS Protocol

Active substance AZD1222

Product reference D8111R00007

Version number 3.0

Date 31-January-2022

Real-world effectiveness of the Oxford/AstraZeneca COVID-19 vaccine in England

RAVEN study

(An observational retrospective cohort study using secondary databases to establish effectiveness of the Oxford/AstraZeneca COVID-19 vaccine in England)

Marketing Authorisation Holder(s)

Marketing authorisation holder(s)	[redacted]
MAH contact person	[redacted]

Approved by:

EU Qualified Person Responsible for Pharmacovigilance	Date

AstraZeneca

PASS INFORMATION

Title	*Real-world effectiveness of the Oxford/AstraZeneca COVID-19 vaccine in England (RAVEN)*
Protocol version identifier	*3.0*
Date of last version of protocol	*18/Nov/2021*
EU PAS register number	*EUPAS43571*
Active substance	*AZD1222*
Medicinal product	*Vaxzevria (ChAdOx1-S [recombinant])*
Product reference	*005675*
Procedure number	
Marketing authorisation holder(s)	*AstraZeneca AB*
Joint PASS	*No*
Research question and objectives	*To establish the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine in England.*
Country of study	*England*
Author(s)	[redacted]

	[redacted]

1. TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF APPENDICES

2. LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
A&E	Accident and Emergency
AI/ML	Artificial intelligence/machine learning
APC	Admitted patient care
BMI	Body mass index
CDS	Commissioning Data Set
CHESS	COVID-19 Hospitalization in England Surveillance System
CI	Confidence interval
CMMS	Cambridge Multimorbidity Score
CMR	Computerised medical records
COG-UK	COVID Genomics UK Consortium
COVID-19	Coronavirus disease 2019
DARS	Data access request service
eFI	Electronic frailty index
ECDS	Emergency Care Data Set
GDPPR	GPES data for pandemic planning and research
GEE	Generalised estimating equations
GP	General practice
GPES	General practice extraction service
HES	Hospital episode statistics
HCRU	Healthcare resource use
HRG	Healthcare resource grouper
ICU	Intensive Care Unit
IRAS	Integrated Research Application System
ISRCTN	Registry of research studies originally International Standard Randomised Controlled Trial Number, its scope subsequently broadened
JCVI	Joint Committee of Vaccination and Immunisation
MERS	Middle East Respiratory Syndrome
NHS	National health services
NIMS	National Immunization Management System
NPEx	National pathology exchange
ONS	Office for National Statistics (ONS provide date of issues of death certificates and cause of death)
ORCHID	Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub

Abbreviation or special term	Explanation
PHE	Public Health England
RCGP RSC	Oxford-Royal College of General Practitioners Research and Surveillance Centre
RCT	Randomized controlled trials
RECORD	Reporting of studies Conducted using Observational Routinely-collected Data
RR	Rate ratio
RT-PCR	Reverse transcriptase polymerase chain reaction (sometimes shortened to PCR)
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SGSS	Second Generation Surveillance System
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SUS	Secondary uses service
TRE	Trusted research environment
UK	United Kingdom
UKHSA	UK Health Security Agency
VE	Vaccine effectiveness, defined as 1-Rate ratio (RR)

3. RESPONSIBLE PARTIES

[redacted]

Role	Name, title, qualifications
Scientific lead	[redacted]
Scientific lead	[redacted]
Scientific lead	[redacted]
Study contributor	[redacted]
Study contributor	[redacted]
Statistician	[redacted]

[redacted]

Name	Professional title	Role in study	Affiliation	Email address
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				
[redacted]				

Name	Professional title	Role in study	Affiliation	Email address
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

4. PROTOCOL SYNOPSIS

Background/Rationale: The United Kingdom (UK) is one of the first countries that introduced a mass vaccination campaign for COVID-19 and vaccination of the adult population first focused on the oldest age groups, their carers and health care workers (JCVI, 2020). Three COVID-19 vaccines were licensed and are being used including the Moderna, the BioNTech/Pfizer, and the Oxford/AstraZeneca vaccines. Vaccination with the BioNTech/Pfizer vaccine started in December 2020 and the Oxford/AstraZeneca vaccine started in early January 2021. This study is to primarily assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine. Given the known high efficacy of the mRNA vaccines in randomized controlled trials (RCTs) and real-world evidence studies, the study aims also to evaluate the vaccine effectiveness (VE) of other COVID-19 vaccines as a validation of the study's methods. On 16 September 2021 it was announced that Booster doses would be introduced in the UK to address vaccine waning and for groups with a suboptimal response. Little is known about health care resource utilisation (HCRU) and health care costs for those who have had COVID-19 or by individual clinical risk group. The RAVEN study is a retrospective cohort study to assess the real-world effectiveness of the Oxford/AstraZeneca COVID-19 vaccine in England. The study is using linkage of the English national databases on COVID-19 vaccination, testing, medical records, hospitalization, and death. Analyses will examine the effectiveness of one and two doses, and booster or other additional doses if applicable. This study's extension (October 2021) adds a more detailed exploration of VE in risk groups and an evaluation of the HCRU by people with COVID-19 compared with those who are vaccinated.

Research question and objectives:

Primary: To assess the real-world effectiveness of the Oxford/AstraZeneca COVID-19 vaccine among people who receive at least one dose of the vaccine, overall, and by number of doses, by age group, by comorbidity status, by time period after each dose, by interval between doses.

Secondary: To assess the real-world effectiveness of other COVID-19 vaccines in people vaccinated who received at least one dose; overall, and by number of doses, by age group, by comorbidity status, by time period after each dose, and by interval between the doses.

RAVEN extension objectives:

1. To describe the risk profile of the population at risk due to being ineligible for COVID-19 vaccines due to having suboptimal vaccine response (and thus more likely to have a breakthrough case), predict breakthrough cases and their time to vaccine waning, and describe the predictors of accepting AZ boosters. People who are immunocompromised are the primary population of interest, in our exploration of suboptimal vaccine response.
2. To conduct an analysis of COVID-19 and long COVID HCRU and health care cost for the whole population and for risk groups identified as having a suboptimal vaccine response.

3. To explore VE against the different SARS-CoV-2 variants.
4. To assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine against any positive SARS-CoV-2 test, medically attended COVID-19 and rate of emergency department visit associated with COVID-19.
5. To assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine and other COVID-19 vaccines in people vaccinated with booster/other additional doses.
6. To assess the effectiveness of heterologous dosing (including at least one Oxford/AstraZeneca COVID-19 vaccine)
7. To estimate the impact of vaccination on household incidence/transmission
8. To estimate the impact of vaccination on long COVID

Methods:

Study design: Secondary data analysis using a retrospective cohort design

Population: The study population is people in England who have received at least one dose of the Oxford/AstraZeneca COVID-19 vaccine (the AstraZeneca vaccine arm) or at least one dose of other COVID-19 vaccines (the other COVID-19 vaccine arm) and people in England who were not vaccinated with any COVID-19 vaccine during the same time period (concurrent control arm). All study participants are eligible for COVID-19 vaccination based on age.

Variables

Exposure(s): The main exposure is whether a person was vaccinated with the Oxford/AstraZeneca COVID-19 vaccine or with any other COVID-19 vaccine versus not vaccinated with any COVID-19 vaccine.

Outcome(s): The primary outcome is COVID-19 related hospitalization, Intensive Care Unit (ICU) admission, and death. Events due to any causes of these measures will be secondary. Additionally, the following outcomes will be secondary: any positive SARS-CoV-2 positive test, medically attended COVID-19, COVID-19 related emergency department visit, HCRU related to COVID-19 and associated cost, breakthrough case, time to vaccine waning, and long COVID.

Data sources: Linked secondary databases in England accessed through the Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID) and National Health Services (NHS) Digital's Trusted Research Environments (TRE) will be used to assess VE. The primary care data will be linked with vaccination, hospitalization, COVID-19 test and mortality data at the national level for capture of key study variables. Data assets will be requested through the NHS Digital Data Access Request Service (DARS).

Study size: ORCHID contains records from 18 million people, over 30% of the English population, whilst NHS Digital database comprises complete national data from the national

registries of vaccination and hospitalization with linkage to the computerised medical record (CMR), lab, and mortality data. The NHS Digital dataset is the largest dataset possible for England and covers all regions. As of early March 2021, near 20 million people in England have been vaccinated with at least 1 dose of any COVID-19 vaccines and more than 7 million people have received the Oxford/AstraZeneca COVID-19 vaccine. Given this large number and the use of the national database, the study should have sufficient power to address the study objectives.

Data Analysis: To carry out the VE analysis, vaccinated persons will be matched each week (if feasible) to the unvaccinated individuals by age, gender, general practitioner (GP) practice (or NHS region), and comorbidity.

For each outcome event and for each study cohort, the number of first events, total person-years for the event, number of first events per person-years (rate), the rate ratio (RR) and the VE, calculated as (1 – RR)) will be presented. This will also be provided per age group and per frailty score. Finally, VE will also be provided in shorter periods after dose 1, and between the doses, and by presence of comorbidities. Poisson regression will be used to estimate rates using the matched dataset, adjusting for the matching variables and body mass index (BMI), smoking, prescribed medications, and frailty score.

RAVEN extension analyses (other than VE):

a) Populations at risk: Exploratory analysis of populations at risk of breakthrough COVID-19 and populations with a contraindication to vaccination will be carried out with the aim to better describe these populations at risk of COVID in a highly vaccinated population (e.g. to identify populations, and estimate their size). Artificial intelligence/machine learning (AI/ML) methods will be employed to predict breakthrough cases, time to vaccine waning and predictors of accepting booster doses.
b) HCRU and health care cost related to COVID-19: The non-linear trends over time in healthcare service use and costs will be explored using inverse probability weighted generalized estimating equations (GEE). Estimates of the incremental health service use and incremental health service costs associated with receipt of COVID-19 vaccine will be over 1, 3 and 6 months of follow-up.
c) Household incidence/transmission: Vaccinated and non-vaccinated cases with a positive SARS-CoV-2 test will be identified (index cases), along with their household contacts. The proportion of unvaccinated and vaccinated contacts who test positive for SARS-CoV-2 after the index case (secondary cases) will be compared among the vaccinated and non-vaccinated index cases. Logistic regression will be used to estimate the odds (and 95% confidence intervals (CI)) of being a secondary case, adjusting for confounders.
d) Long COVID: Cases of long COVID will be described, including by initial COVID-19 disease severity. The impact of vaccination on the occurrence of long COVID (in vaccinated individuals, in vaccinated individuals with COVID-19) will be explored.

Milestones

Expected data collection: [redacted]

Interim reports: [redacted]

Final report: [redacted]

5. AMENDMENT HISTORY

Table 1. Amendments and updates

Date	Section of study protocol	Amendment or update	Reason
18-Nov-21 (v1.0 to v2.0)	Title	Addition of study name (RAVEN) to title.	Clarification
	8 (objectives)	Changed number of doses in primary and secondary objectives to 'at least one dose', and added analysis by number of doses.	To reflect the evolution of the vaccination campaign, where most people have now received more than one dose.
	9.1 (study design), 9.2.1 (study population), 9.7.4 (exploratory analysis),	Moved historical control arm from secondary analysis to exploratory analysis and renamed to historical population.	
	9.2.2 (inclusion criteria)	Update of inclusion criteria.	To reflect the evolution of the vaccination campaign, where persons aged <16y are now eligible for COVID-19 vaccine
	8 (objectives), 9.3.1 (exposures), 9.3.2 (outcomes), 9.4 (data sources), 9.7.2 (stratification/subgroup analyses), 9.7.6 (RAVEN extension analyses), **Error! Reference source not found.** (interim analyses)	Addition of RAVEN extension objectives to describe COVID-19 in remaining at risk populations, to conduct health economic analyses, to account for vaccine booster/other additional doses and heterologous dosing, to study the impact of vaccination on household incidence/transmission and long COVID. Addition of outcomes. Addition of data sources. Description of RAVEN extension analyses. Update of interim analyses. Review of stratifications/ subgroups.	Addition of objectives to fill additional knowledge gaps, and corresponding updates to methods.
	3.3 (study population), 9.2.4 (participant follow-up), 9.7.1 (statistical methods – general aspects)	Altered the matching process for vaccinated cases to concurrent controls to change from frequency matching to weekly 1:1 matching. Clarified that concurrent controls that are censored at date of vaccination can re-enter study as newly vaccinated objectives.	Improvement to the matching process and clarification.
	9.7.3 (sensitivity analyses)	Added sensitivity analysis excluding individuals in care homes.	These settings differ from the general population.

Date	Section of study protocol	Amendment or update	Reason
31-Jan-22 (v2.0 to v3.0)	All	Transferred protocol to PASS template	
	4 (protocol synopsis), 9.1 (study design), 9.3.2 (outcomes), 9.7.1. (statistical methods)	Removed unnecessary repetitions	Improving conciseness
	0 (responsible parties)	[redacted]	Update
	0 (milestones)	Revision of milestones to align with PASS template milestones, update timelines	Update
	0 (rationale and background)	Modified to reflect the current stage of the vaccination programme in the UK	Update
	8 (objectives)	Added outcome measure to Table 3 for RAVEN extension objectives	
	8 (objectives), 9.3.2 (outcomes), 9.7.6.1 (populations at risk)	Modifying analysis of the first RAVEN extension objective, i.e. population at risk due to suboptimal response (who may have breakthrough cases) from simple descriptive analysis to descriptive and predictive analysis using AI/ML methods. Including addition of outcomes 'breakthrough case' and 'time to vaccine waning'.	Improved description of methods
	8 (objectives), 9.7.6.3 (household incidence/transmission)	Added household 'incidence' in addition to 'transmission' throughout. Added references to prior household incidence studies.	Incidence to account for the possibility of external common sources of infection
	9.1 (study design)	Removed figure 1 as not accurate	Correction
	9.2.4 (participant follow-up)	Clarified follow-up criteria	Clarification
	9.6 (data management)	Added to comply with PASS template	Addition

Date	Section of study protocol	Amendment or update	Reason
	9.7	Introduced the SAP, and updated to align with SAP	Update
	9.7.1.	Removed analysis restricted to subgroup of people who received an influenza vaccine during the 2020-21 influenza season and before the index date	Simplification
	9.9 (limitations of the research methods)	Limitation added describing change in immunity status of unvaccinated group	Clarification
	9.11.1 (methods to minimize bias)	Rephrased as historical population analysis is no longer the primary analysis	Update
	Appendix B	Addition of EnCEPP checklist	Part of PASS template

6. MILESTONES

Table 2. Study milestones

Milestones	**RAVEN planned dates**
Study design concept	[redacted]
Final protocol	[redacted]
Start of data collection	[redacted]
End of data collection*	[redacted]
Study progress report 1	NA
Interim report 1	[redacted]
Interim report 2	[redacted]
Registration in the ISRCTN registry and EU PAS register (actual dates)	[redacted]
Final report of study results	[redacted]

*Interpreted as date secondary data analysis is complete

7. RATIONALE AND BACKGROUND

The current COVID-19 pandemic constitutes a public health emergency unprecedented in the last century. From a small cluster of cases initially identified in 2019 in Wuhan, China (Zhou, 2020), the disease has spread around the globe. As of 14 March 2021, there have been more than 119 million confirmed cases and more than 2.6 million deaths globally (World Health Organization, 2021). SARS-CoV-2 (the viral agent that causes COVID-19) shares more than 79% of its sequence with SARS-CoV and 50% with the coronavirus responsible for Middle East respiratory syndrome (MERS-CoV) (Lu, 2020). It is believed that evolution of the pandemic will further vary across countries, affected in part by different containment strategies ranging from extreme lockdown to relative inaction. As a result, there may be regional waves of the disease and pockets of deeply affected populations. Globally, governments have acknowledged that an effective vaccine against COVID-19 constitutes a major public health need and may be the only way to guarantee a safe and sustained exit strategy from human movement restrictions while avoiding escalating mortality rates across populations. Accelerated development of safe and effective vaccines and treatments are currently underway and focus on adults, the population in greatest need. Various vaccines against COVID-19 based on different technologies are now in clinical development around the world or have been recently approved, with some being rolled in middle or high-income countries. Many of the vaccines display optimal immune response and protection following 2 vaccinations, a prime and a boost vaccination separated by a number of weeks.

The adenovirus vector ChAdOx1 vaccine technology has been used in the past for the development of candidate vaccines against several infections, including influenza, Zika, and MERS (Antrobus et al, 2014, Folegatti et al, 2020, López-Camacho et al, 2018). Oxford University has adapted and transferred the platform technology used against MERS-CoV to develop a vaccine against SARS-CoV-2, resulting in the rapid development of a vaccine against this infection. The Oxford/AstraZeneca COVID-19 vaccine (AZD1222) is a recombinant replication-defective chimpanzee adenovirus expressing the SARS-CoV-2 S surface glycoprotein (Folegatti et al, 2020). AZD1222 has been shown to be highly efficacious in pivotal randomized clinical trials (RCT), demonstrating 82% overall efficacy against symptomatic infection and 100% efficacy against severe infection. Recent unpublished data from the phase 3 pivotal trial in the United States showed high efficacy (85%) in people 65 years of age or older, and again confirmed 100% efficacy in preventing severe cases. Although a high level of single dose efficacy has been demonstrated in clinical trials between 3 and 12 weeks and confirmed in early effectiveness studies conducted in the United Kingdom (UK), clinical trials and effectiveness study so far have not included detailed vaccine-specific analysis by age group, co-morbidities, nor have they assessed vaccine impact on critical care admission, mortality, and overall outcomes. Similarly, there are few studies taking into account the COVID-19 variant, these data have been aggregated on a national basis by the UK COVID-19 Genomics consortium (UK COG) since September 2020 (Mishra et al, 2021). Thus,

it remains important to better understand VE by ages, time intervals between doses, and to determine if single dose effectiveness may extend beyond 12 weeks.

Given mortality due to COVID-19 is highest in long-term care and assisted living facilities and since the Oxford/AstraZeneca vaccine is prioritized for these facilities due to its ease of reach, it is useful to understand the impact of Oxford/AstraZeneca vaccination on hospitalization and mortality for elderly people in these facilities. It's also important to assess VE by age groups and comorbidities for best guidance for COVID-19 immunization programs.

The UK was one of the first countries that introduced a mass vaccination campaign for COVID-19 and first started vaccinating the elderly population (JCVI, 2020). Three COVID-19 vaccines were licensed and are being used including the Moderna, the BioNTech/Pfizer, and the Oxford/AstraZeneca vaccines. Vaccination with the BioNTech/Pfizer vaccine started in December 2020 and with the Oxford/AstraZeneca vaccine started in early January 2021. The need for booster COVID-19 vaccination was anticipated (Mahase, 2021) and it was introduced in the UK in September 2021. This study is to primarily assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine. Given the known high efficacy of the mRNA vaccine in clinical trials and real-world studies, the study is also to evaluate the VE of other COVID-19 vaccines as a validation of the study's methods.

Whist some early studies have been conducted little is known about the health economic impact of COVID-19 vaccination. There have been exploratory studies of the impact of vaccination on disease transmission and the health-economic impact of COVID-19 (HM Government, 2020; Sandmann et al, 2021).

A further challenge in the management of people with COVID-19 has been the emergence of long COVID. This has been quite disabling for a large number of people. However, clinical terms to record long COVID in clinical records were created late, so data quality of cases in clinical records is suboptimal (Crook et al, 2021; Mayor, 2021).

8. OBJECTIVES

The study's primary objective is to assess the overall effectiveness of at least one dose of the Oxford/AstraZeneca COVID-19 vaccine by number of doses, by age groups, by comorbidity status, by time periods after each dose and by interval between the doses. The secondary objective is to assess the overall effectiveness of at least one dose of other COVID-19 vaccines, by number of doses, by age group, and by time periods after each dose, intervals between doses, and comorbidities that are known to associated with more severe COVID-19 infection and/or suboptimal response to the vaccine.

The RAVEN study extension adds a description of proportion and the risk profile of the people at risk of COVID due to either a suboptimal vaccine response or from being ineligible for vaccination. People who are immunocompromised are the primary population of interest. In addition, we will conduct an analysis of HCRU and health care cost for the whole population and for risk groups identified as having a suboptimal vaccine response. Furthermore, we will assess VE against a wider range of outcomes, and assess VE of booster/other additional vaccine doses. Finally, we will explore the impact of vaccination on household incidence/transmission, using an established method (de Lusignan et al, 2020a; de Lusignan et al, 2020b) and on long COVID.

Table 3. Primary, secondary, exploratory and RAVEN extension objectives

Objectives	Outcome measure
Primary -To assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine among people who receive at least one dose of the vaccine; overall and by a)number of doses b) age group c) comorbidity status d) time period after each dose e) interval between doses	Primary -Hospitalization associated with COVID-19 -Admission to ICU associated with COVID-19 -Mortality associated with COVID-19 Secondary -All-cause hospitalizations -All cause admission to ICU -All-cause mortality
Secondary -To assess the effectiveness of other (e.g. mRNA) COVID-19 vaccines among people who received at least one dose, overall and by a) number of doses b) age group c) comorbidity status d) time periods after each dose e) interval between doses	Same as above
Exploratory -To assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine and other COVID-19 vaccines among people who receive at least one dose of the vaccine, using a historical unvaccinated population as comparator.	Same as above

RAVEN extension objectives: -To describe the population at risk of COVID due to being ineligible for a COVID-19 vaccines. -To explore the population at risk of COVID due to a suboptimal response (breakthrough case) by predicting breakthrough cases, identifying susceptible sub-populations, predicting their time to vaccine waning and identifying predictors of receiving AZ or other booster. People who are immunocompromised are the primary population of interest, in our exploration of suboptimal vaccine response.	-Breakthrough case -Primary outcomes in breakthrough cases -Time to vaccine waning - AstraZeneca COVID-19 booster vaccination
- To conduct an analysis of COVID-19 and long COVID HCRU and health care cost for the whole population and for risk groups identified as having a suboptimal vaccine response.	- HCRU for COVID-19 - HCRU for long COVID - healthcare cost for COVID-19 - healthcare costs for long COVID
- To explore vaccine effectiveness against COVID-19 due to different SARS-CoV-2 variants.	Same as primary objective
To assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine against any positive SARS-CoV-2 test, medically attended COVID-19 and emergency department visit associated with COVID-19.	-Any positive SARS-CoV-2 test -Any medically attended COVID-19 -Emergency department visit associated with COVID-19
-To assess the effectiveness of the Oxford/AstraZeneca COVID-19 vaccine and other COVID-19 vaccines in people vaccinated with booster/other additional doses.	Same as primary objective
-To assess the effectiveness of heterologous dosing (including at least one Oxford/AstraZeneca COVID-19 vaccine)	Same as primary objective
-To estimate the impact of vaccination on household incidence/transmission	Household incidence/transmission of COVID-19
-To estimate the impact of vaccination on long COVID	Long COVID-19

9. RESEARCH METHODS

9.1 Study Design – General Aspects

The effectiveness objectives will be addressed through a retrospective cohort study using linked secondary databases in England accessed through ORCHID and NHS Digital's TRE. Persons vaccinated with Oxford/AstraZeneca COVID-19 vaccine or another COVID-19 vaccine will be compared to matched persons not vaccinated with any COVID-19 vaccine for occurrence of study outcomes.

Due to England's vaccination campaign starting with the older age group, as of March 22 2021, over 90% of people 70 years of age or older were vaccinated with at least one dose of any COVID-19 vaccine (Public Health England, 2021). Therefore, it is expected that the concurrent control arm will have considerable challenges. A greater proportion of younger people might be included who may have lower hospitalization and death (particularly because the Joint Committee of Vaccination and Immunisation (JCVI) recommended early vaccination of people 16 to 64 years with comorbidities). Additionally, people receiving terminal care and those who do not engage much with the health system may also be included. For these reasons, the concurrent control arm is matched with the vaccine arm by age and other characteristics. Events' rates will also be stratified by age groups (e.g. 12-15,16-49, 50-64, 65-69, 70-79, 80+).

We will also explore how COVID-19 vaccination in England coincided with the 3rd wave of COVID-19 and any subsequent wave that may occur. For this analysis, we will include vaccination data periods beyond February 2021.

There have been changes in the age groups vaccinated and the introduction of booster doses. The Moderna and Pfizer BioNTech vaccines have been licenced for children (Medicines and Healthcare products Regulatory Agency (a & b), 2021) and booster doses were introduced in the autumn of 2021. For these reasons the analyses will be extended to include the 12–16-year-old age-group and any booster vaccination (9)(Department of Health and Social Care, 2021). Finally, the UK COG has centralised data on COVID-19 variants allowing VE to reported by variant.

9.2 Setting

9.2.1 Study Population

The study population of this study is people in England who have received at least one dose of the Oxford/AstraZeneca COVID-19 vaccine from January 2021 onwards (the AstraZeneca vaccine arm) or at least one dose of another (e.g. mRNA) COVID-19 vaccine from December 2020 onwards (the Other vaccine arm) and people in England who were not vaccinated with any COVID-19 vaccine during the same time period (concurrent control arm) who will be matched to the vaccinated individuals by age, gender, GP practice (NHS region), and comorbidity (see Section

9.7.1 for a description of matching). Study participants of any age will be considered. For the vaccinated individuals, the study's index date is the date of vaccination of the 1st dose.

Concurrent controls who become vaccinated will be censored at date of vaccination and, then may re-enter the study as newly vaccinated individuals. An appropriate control group will be identified for the booster dose evaluation.

To contextualize, we will describe the impact of SARS-CoV2 on the unvaccinated English population, prior to the availability of COVID-19 vaccines. We will conduct a descriptive analysis of COVID-19 of this historical population of unvaccinated persons in England.

9.2.2 Inclusion Criteria

The study population is required to meet the following inclusion criteria:

1. **For the vaccinated arms:**
   - Any COVID-19 vaccination at the index date
   - Have continuous data coverage for the COVID-19 infection datasets, i.e. Second Generation Surveillance System (SGSS) and National Pathology Exchange (NPEX) from their initiation for history of prior COVID-19 infection
   - Have continuous data coverage in other linked databases for a minimum of 12 months prior to the index date for assessment of baseline variables including socio-economic status, comorbidities, and follow-up of outcome events.

2. **For the control arms:**
   - Eligible for any COVID-19 vaccination based on age at the index date for the concurrent control individuals.
   - Have continuous data coverage for the COVID-19 infection datasets, i.e. SGSS and NPEX from their initiation for history of prior COVID-19 infection
   - Have continuous data coverage in other linked databases for a minimum of 12 months prior to the index date for assessment of baseline variables including socio-economic status, comorbidities, and follow-up of outcome events.
   - People who have not (yet) received any COVID-19 vaccine (Oxford/AstraZeneca, Pfizer, or Moderna COVID-19 vaccine) recorded in their GP record or in NIMS. They will be used as concurrent controls. However, they will be censored at date of vaccination and, then may re-enter the study as newly vaccinated individuals.

9.2.3 Exclusion Criteria

People who meet the following criteria will be excluded:

- Primary analysis: People with a history of COVID-19 infection (confirmed by reverse transcriptase polymerase chain reaction (RT-PCR) or not) prior to vaccination. This group of people is not excluded in the sensitivity analysis.

A sensitivity analysis excluding persons living in an elderly care home will be considered.

9.2.4 Participant Follow-up

For the vaccinated and concurrent control arms, study participants will be followed from the index date to the end of the study, loss to follow-up (deregistration), vaccination (for unvaccinated controls), vaccination of the matched control (for vaccinated people), or death, whichever is earlier.

9.3 Variables and Epidemiological Measurements

9.3.1 Exposures

The main exposure in this study is whether an individual was vaccinated with one, two or more doses of the Oxford/AstraZeneca COVID-19 or other COVID-19 vaccine. This includes next generation vaccines, if and when available. People without record of any COVID-19 vaccination are considered unexposed.

COVID-19 vaccination will be retrieved from the National Immunization Management System (NIMS), which contains detailed information on name of the COVID-19 vaccines, specific date of vaccination, and dose sequence (1^{st}, 2^{nd} or booster dose) of the injection given to each vaccinated individual.

Persons will be considered vaccinated with:

- one dose: from at least 22 days after dose 1 until the day before dose 2
- two doses: from at least 15 days after dose 2 until the day before a third dose
- additional doses: from at least 15 days after the previous dose, until the day before the subsequent dose

A subject is considered to have heterologous dosing after receipt of at least one dose of Oxford/AstraZeneca COVID-19 vaccine and at least one dose of another COVID-19 vaccine.

Table 4. Exposure variables and data sources

Variables	Categories	Data source	Comments
COVID-19 vaccine	-Oxford/AstraZeneca -Pfizer -Moderna	COVID-19 vaccination event from NIMS	Collected daily, weekly batch from TRE; 1 week lag time
Date of vaccination	Specific date		
Dose number	1st, 2nd or booster dose		
Batch number			

NIMS: National Immunization Management System; TRE: trusted research environment

Vaccinated persons will be further stratified by time period after each dose and interval between doses.

9.3.2 Outcomes

The primary outcomes of this study are

- COVID-19 related hospitalization
- COVID-19 related ICU admission
- COVID-19 related death

The secondary outcomes of this study are

- All-cause hospitalization
- All-cause ICU admission
- All-cause mortality

The RAVEN extension outcomes are:

- Any positive SARS-CoV-2 test
- Medically attended COVID-19
- Emergency department visit associated with COVID-19
- HCRU (including but not limited, primary care consultations, prescriptions, medical tests and investigations; hospital outpatient attendances and procedures; hospital emergency department attendances without admission; hospital admissions) and associated cost related to COVID-19 and long COVID
- Breakthrough case
- Time to vaccine waning
- Long COVID (e.g. based on code for long COVID, on combination of code for long COVID and a positive SARS-CoV-2 test, on referral to long COVID clinic)
- A COVID-19 related event (GP attendance, accident department, hospitalization, ICU admission, death) is defined as:

- A medically attended event "Acute symptomatic COVID" – defined as a positive test plus a clinical diagnosis or symptom (e.g. anosmia) which makes COVID plausible
- An emergency department attendance
- An event with an ICD-10 diagnosis of COVID-19
- An event within 28 days of a positive RT-PCR for SARS-CoV-2 infection or
- A visit with a patient presenting with respiratory symptoms or other symptoms credibly associated with COVID-19 within 28 days of a positive RT-PCR for SARS-CoV-2 infection.
    - Events within 14 days will feature in a sensitivity analysis.

Events of interest are those occurring between the index date and the end of follow-up. Only the first outcome event will be considered in the primary and secondary VE analysis.

The Hospital Episode Statistics (HES), Secondary Uses Service (SUS), General Practice Extraction Service data for pandemic planning and research (GDPPR), and Office of National Statistics (ONS) data may include information on COVID-19 infection through specific COVID-19 diagnosis codes. Using a combination of both COVID-19 SGSS/NPEX and diagnosis codes from HES, SUS, GDPPR, and ONS allow capturing the COVID-19 infection status for all study participants. Since testing for COVID-19 and its capture was not complete at the beginning of the pandemic but significantly improved later, restriction of the study period from July 2020 or the 2nd wave of infection by regions in England, whichever is later, for determining COVID-19 related events in the historical population is needed. Medically attended COVID-19 will come from ORCHID, GDPPR, Emergency Care Data Set (ECDS) and HES. Emergency department visits will come from ECDS. SARS-CoV-2 variants will come from COG UK. Data on HCRU and associated cost will come from HES using the Health Resource Grouper (HRG) and primary care attendance and medication costs. Long COVID will come from ORCHID and HES.

Table 5. Outcome variables and data sources

Variables		Data source	Comments
Hospitalization	Date	HES for curated data SUS for more recent raw data HES APC for combined hospital and ICU data; HES CC for ICU specific data	6-8 weeks lag for HES SUS is updated weekly with a maximum of 2 weeks lag
	Frequency		
	Duration		
ICU admission	Date		
	Frequency		
	Duration		
Death	Date	ORCHID/GDPPR for date	2 weeks lag
	Cause of death	ONS for cause of death	6 weeks lag
Medically attended COVID-19		ORCHID, GDPPR, ECDS, HES	
COVID-19 infection	Positive RT-PCR test; date	SGSS for first positive test in hospital and community NPEX for all positive/negative test in the community	Daily update
	ICD-10/SNOMED codes for COVID	HES, ONS, GDPPR	
SARS-CoV-2 variants of interest		COG UK	
Long COVID		ORCHID, HES	
HCRU and cost		HES & HRG ICU datasets Primary care attendance & costs	

APC: Admitted Patient Care; CC: Critical Care COG UK: COVID Genomics UK Consortium; ECDS: Emergency Care Data Set; GDPPR: General Practice Extraction Service data for pandemic planning and research; HES: Hospital Episode Statistics; HRG: Health Resource Grouper; ICU: Intensive Care Unit; NPEX: National Pathology Exchange; ONS: Office for National Statistics; ORCHID: Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub; SGSS: COVID-19 Second Generation Surveillance System; SUS: Secondary Uses Service.

9.3.3 Other Variables and Covariates

Table 6. Variable categories and data sources

Variables	Categories	Data source	Comment
Age	Different categorizations may be used -12 <50; 50-<65; 65-<70; 70-<80; 80+ -18-<40; 40-<50; 50+; 65+; 70+; 80+	GDPPR	BMI is binary as obese or not using HES, SUS, GDPPR or BMI can be created from weight/height if available;
Gender	Male/Female		
Ethnicity	White, Black, Asian, Other, Mixed		
Region	NHS Region		
Sociodemographic	Socioeconomic status – Index of Multiple Deprivation (IMD)		
BMI	-Obese (BMI 30+) -Not-obese (BMI <30)		
Smoking	- Current - Past - Non smoker		
Comorbidity	Specific comorbidities known to relate with increased risk of severe COVID-19 and/or suboptimal vaccine response and/or vaccine contraindication including chronic respiratory disease, chronic kidney disease, chronic heart disease and vascular disease, chronic liver disease, chronic neurological disease, diabetes mellitus, severe mental illness, morbid obesity, asplenia or dysfunction of the spleen, and immunosuppression due to disease or treatment.	ORCIHD, HES and GDPPR	Follow JCVI/Green book list of comorbidities (COVID-19 Greenbook chapter 14a) Using ICD or SNOMED codes to identify comorbidities during the baseline period. For immunosuppressant medications, see Prescribed medication section below.
CMMS	0, and by quartile		Created from specific comorbidities;
Non-COVID-19 vaccination	- Flu vaccination (Y/N) - Receive adequate vaccination (Y/N)	GDPPR	An algorithm will be developed to determine if an individual is adequately vaccinated according to recommended vaccination schedule by age and comorbidity status
Prescription medication	Immunosuppressant vs. not	GDPPR	Prescribed medications at baseline will be grouped into immunosuppressant or not.
LTC /assisted living care facility	Yes/No	Not in any database	To be created for people aged 70+ by comparing their zip codes and LTC

Variables	Categories	Data source	Comment
			address; or vaccination at LTC facility. (ORCHID)
Household composition	Size and age combination.	ORCHID UKHSA VE cohort	Available for 7.2 million subset of ORCHID
Frailty score	eFI grouped into - Mild - Moderate - Severe	Derived from GDPPR	Use an algorithm already developed to derive eFI, only available for population 65years and older. For this sub-group analysis.

BMI: body mass index; CMMS: Cambridge Multimorbidity Score; eFI: electronic frailty index; GDPPR: General Practice Extraction Service data for pandemic planning and research; HES: Hospital Episode Statistics; JCVI: Joint Committee on Vaccination and Immunisation; LTC: long term care; ORCHID: Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub; SUS: Secondary Uses Service; UKHSA: UK Health Security Agency; VE: vaccine effectiveness.

9.4 Data Source(s)

This study will employ linked England's national databases for data on background characteristics, medications, comorbidities, COVID-19 infection status, vaccination status, hospitalization, ICU admission, and mortality. The datasets use unique identifiers allowing linkage across the datasets and avoidance of duplicates.

Through the NHS DARS it is possible to request access to a number of secondary data assets collected as part of routine care and commissioning activities in the NHS. With appropriate approval, these data assets are linkable in their TRE. Data assets available through DARS which may be used for this study include:

- General Practice Extraction Service (GPES) data for pandemic planning and research (GDPPR): central collection of GP patient data for COVID-19 purposes (fortnightly collection – All GP practices in England). Limitations of these data included: (1) Some practices may have poor data quality; (2) The clinical data are restricted to 56,318 concepts from the much larger SNOMED clinical terminology (SNOMED CT international version has 352,567 concepts, the UK version has additional concepts to meet national needs)(SNOMED, 2020). Careful testing will be required for the equivalence of variables between ORCHID and GDPPR. We will carefully compare descriptive statics for key variables, some measures such as Cambridge Multimorbidity Score (CMMS) will require careful validation in GDPPR as some of their component codes do not exist within GDPPR.
- NIMS – includes 2 associated datasets including COVID-19 Vaccination Status and COVID-19 Adverse Reaction.
- HES: this is the transformed data, initially part of the Commissioning Data Set (CDS), covering patients attending accident and emergency (A&E) units, admitted for treatment,

or attending outpatient clinics at NHS hospitals in England, including details about length of stay and HRG to enable health economic analysis. An HRG is a grouping consisting of patient events that have been judged to consume a similar level of resource, which can be linked to an appropriate NHS tariff. Statistical controls have been applied to the HES products. Data on cost will be extracted from HES using HRG. (Time lag 8-12 weeks)

- ECDS: this collects data from A&E Departments (time lag – 8-12 weeks). Only available for linkage to ORCHID.
- Civil Registrations (Deaths): Information including the date, place and certificated cause of death from the Office for National Statistics (ONS) (Time lag – 8-12 weeks, released by NHS Digital with HES)
- COVID-19 SGSS - Demographic and diagnostic information from laboratory test reports for patients tested for COVID-19 in England only. It currently includes the first positive results from pillar 1 (swab testing in Public Health England (PHE) and NHS hospital labs and pillar 2 (swab testing for the wider population) (Time lag – less than 1 week)
- COVID-19 UK Non-hospital Antigen Testing Results (Pillar 2) data includes a range of COVID19 test results, including NPEX. This is broadly similar to SGSS, but only covers Pillar 2 data, however, contains the full result set – i.e. all positive, negative and null results.
- COVID-19 Hospitalization in England Surveillance System (COVID-19 SARI-Watch/CHESS): Epidemiological data on COVID-19 infection in persons requiring hospitalization in ICU or High Dependency Units. It records all demographic, risk factor, treatment and outcome information for patients admitted to hospital with a confirmed COVID-19 diagnosis. This only has data from a limited number of surveillance hospitals. (Time lag – less than 1 week)
- SUS Episodes (including the following concepts: Admitted Patient Care (APC), Outpatient, Critical Care & A&E). This covers all secondary care provided in England and paid for by the NHS. This is essentially the CDS. These data are transformed to create HES product set that is subject to official publication timelines. Therefore, SUS should be considered tactical but carries a shorter latency (Time lag –1-2 weeks). Only available for linkage to ORCIHD.
- ORCHID a TRE for research and surveillance (de Lusignan, 2021). This TRE contains the Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) one of Europe's oldest sentinel systems. The network has been expanded considerably during the SARS-CoV-2 pandemic. It pseudonymised data are linked to the range of data sets described above by NHS Digital, though usage must be approved on a study-by-study basis. It is one of the freshest (either daily or twice weekly data extracts) with data quality as good as it gets in primary care. The University of Oxford team run the national primary care surveillance system – The Oxford-RCGP RSC. This surveillance system is sponsored by PHE and collaborates in reporting vaccine uptake and effectiveness, including of COVID-19 vaccine.

Additionally, the data on SARS-CoV-2 variants will be requested from the COG UK (Time lag – unknown).

Most individual data sources have frequent refresh (daily or weekly) and most have only 1–2-week lag from the time of data entry to when they are ready for analysis. One exception is the HES data source with the longest lag time of approximately 8-10 weeks. The raw hospitalization data source, SUS, which is the source of HES has much shorter time lag of 1-2 weeks. For this reason, our main analysis will include HES data until 8 weeks before the database lock. The sensitivity analysis will include HES data until 8 weeks before database lock.

Validation exercises will be conducted in the NHS Digital TRE, these are to ensure there is no denominator inflation and also to validate the GDPPR variables, as there are a limited number of concepts available. We will ensure the total population denominator is correct. The denominator varies across data sources therefore we will explore any areas of apparent list inflation. In addition, we will validate variables with a recorded prevalence in NHS Digital TRE within 15% of that in ORCHID, and validate or provide a surrogate marker for variables with >15% variation. Finally, we will choose and validate a multimorbidity measure ideally CMMS, but will consider using an alternative comorbidity measure if needed. The method for this exercise will be referenced in the Statistical Analysis Plan (SAP).

9.5 Study Size

The NHS Digital database comprises data from the national registries of vaccination and hospitalization with linkage to the CMR, lab, and mortality data. It is the largest dataset possible for England and covers all regions. As of early March 2021, near 20 million people in England have been vaccinated with at least 1 dose of any COVID-19 vaccines. According to a recent study in Scotland (Vasileiou et al 2021), the Oxford/AstraZeneca vaccine have been used by the majority of the 75+ years old age group and by a significant proportion of the 60-65- and 70-75-years old age groups.

Based on the preliminary count of the cohort of VE from ORCHID in collaboration with PHE (N=5.4 million), as of 7/4/2021, 42.1% of the population have received a COVID-19 vaccine. Overall, 91% of people 80 years and older, 91% of people 65-79 years old, and 79% of people aged 16-64 years old in risk groups were vaccinated. Of those vaccinated, 38% has received the Oxford/AstraZeneca COVID-19 vaccine, i.e. more than 7 million people. It is expected that almost all of England's COVID-19 vaccination data are included in the NHS Digital databases representing the best option for this type of study. Given a large number of people vaccinated with the Oxford/AstraZeneca COVID-19 vaccine and the use of the national database, the study should have sufficient power to address the study objectives.

9.6 Data Management

As a first step, the quality of datasets will be evaluated, and only datasets recognised of sufficient quality will be used. All statistical analyses will use the R statistical software (Rproject, 2020). Full details will be provided in the SAP.

9.7 Data Analysis

9.7.1 Statistical Methods – General Aspects

A detailed SAP will be developed for this analysis. 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. For all study designs, p-values will be quoted to two decimal places, unless they are less than 0.001 (whereby the p-value will be given as < 0.001) or between < 0.005 and > 0.001, in which case they will be stated to three decimal places. Parameter estimates will be reported with 95% CIs. All statistical tests are 2 sided and with a 5% significance level.

A CONSORT 2010 Flow Diagram of the study population detailing each step of the inclusion and exclusion criteria applied will be generated. As indicated in section 9.4, to obtain the most follow-up data, the primary analysis will include all HES data until 8 weeks prior and all raw hospitalization data (SUS) from 8 weeks to 2 weeks prior database lock. A sensitivity analysis including only HES data, therefore a shorter follow-up for hospitalization, will also be done.

For the concurrent control arm, each week starting from 8th December 2020, we will match all newly vaccinated persons in a 1:1 ratio to unvaccinated controls (Dagan et al, 2021). Newly vaccinated persons will be eligible for inclusion in the study, even if they will have previously been selected as a control. We will match vaccine recipients and controls on variables associated with the probability of both vaccination and infection or severity of COVID-19: age (study defined age bands), gender, general practice, (or NHS Region), and comorbidity as defined by CMMS quartile. We will exact match each week (if feasible). To assess the quality of matching, love plots will be created showing the mean standardized difference between each vaccinated cohort (AstraZeneca or other vaccine cohort) and the comparator cohorts in participants' characteristics. If important differences are observed, the matching criteria will be refined.

The demographic and characteristics of study participants in each vaccine arm and its respective control arm at baseline will be determined during the 12-month period before index date using the measurement closest to index date.

Summary tables of all patient characteristics and end points indicated in section 9.3.2 will be created for both the before and after matching situation to provide insight in population characteristics as well as in errors and missingness. Categorical variables will be checked on meaningful categories and frequencies. It is not anticipated that categories will have too low frequency. Pooling will be considered when the frequency for a category becomes below 1000.

Continuous variables will be summarized in terms of min, max, interquartile range, median and mean to get insight in these variables, but also to provide a first indication of use of same unit across the database and to obtain a first impression about possible transformations needed. Missing values will be considered in detail. If the percentage of missingness is below 10%, complete case analyses will be performed. When exceeding 10%, multiple imputation will be run additionally based on the total dataset.

For each outcome event, the summary tables will present for each study cohort the number of first events, total person-years for the event and number of first events per person-years (rate), the RR and the VE which is calculated as 1 - RR. This will also be provided per age group and per frailty score where further stratification/subgroups are indicated below. Finally, VE will also be provided in shorter periods after dose 1, and between the doses, and presence of comorbidities.

Poisson regression with offset for time at risk will be used to estimate rates using the matched dataset, adjusting for the matching variables and BMI, smoking, prescribed medications, and frailty score, which are known to increase risk of hospitalization and death and are likely associated with vaccination.

Checks on linearity for continuous variables will be performed. In order to evaluate interactions of independent variables with the exposure (i.e. vaccinated/comparator), continuous variables will be categorized and the VE will be tabulated per category. For variables for which the maximum and minimum VE across categories is more than 10% different, interaction effects with the exposure will be taken into account in the Poisson model. For interactions among independent variables that are not the exposure, interaction variables will be created. Forward selection is performed, and interaction variables will be included when the computed VE changes with more than 10%.

9.7.2 Stratification/Subgroup Analyses

Stratifications/subgroup analyses will be fully detailed in the SAP, and may include:

- Age: <12, 12-16, 16-49, 50-64, 65-69, 70-79, 80+; or 16-39, 40-<49, 50+, 65+, 80+
- Comorbidity
    - categories of the CMMS
    - specific comorbidities known to indicate high risk for COVID-19 infections and/or suboptimal response to vaccine, in accordance with National guidance from UK Health Security Agency (UKHSA, 2021) or other, including chronic respiratory disease, chronic kidney disease, chronic heart disease and vascular disease, chronic liver disease, chronic neurological disease, diabetes mellitus, severe mental illness, morbid obesity, asplenia or dysfunction of the spleen, and immunosuppression due to disease or treatment.
- Immunosuppressant prescribed medication: current vs. past use, regimen
- Long-term care or assisted living care residents (if possible)

Stratifications/subgroups in the vaccinated population may include:

- Heterologous dosing
- Vaccine dose: people receiving 1 dose only, ≥1 dose, 2 doses, booster/other additional doses
- Intervals after the 1st dose among people receiving 1 dose only: 0-7, 8-14, 15-21, 22-28, 29-35, 36-42, 43-49, 50-70, 71-77, 78-84 days post dose 1, and/or other combinations may be used
- Interval between doses among people receiving 2 (or more) doses: 21 to 41, 42 to 69, 70 to 77, 78 to 84, over 84 days and/or other combinations may be used.
- Vaccine batch number
- Aged over 65 by electronic frailty index (eFI) categories (mild, moderately, and severely frail)

Stratification/subgroups in the population with COVID-19 may include:

- COVID-19 severity
- COVID-19 variant

Subgroup analyses will only be conducted if sufficient events for an outcome within the subgroup allows.

9.7.3 Sensitivity Analyses

Several sensitivity analyses will be considered

- Only HES data are included for follow-up data on hospitalization and ICU-admission
- People with history of COVID-19 infection prior to vaccination not excluded from the study population
- A COVID-related event within 14 days of a positive RT-PCR for SARS-CoV-2 infection.
- Addition of the following exclusion criterion: People whose household had more than 10 people with an average age of 65 years and over. This is a proxy for people living in an elderly care home (and only available for a subset of ORCHID data). They will generate a strong confounding effect of (due to the targeted vaccination of these residents and their high risks of COVID-19-related events) and the lack of available controls.

9.7.4 Exploratory Analysis

In addition to the concurrent control arm, a historical population dating from after the start of the pandemic but before the availability of COVID-19 vaccines will be identified to contextualize the results and assess the potential impact of different time periods for the analysis. The historical population will be defined as people during the periods from July to September (when SARS-CoV-2 testing was more complete than at the beginning of the pandemic, but there was relatively little

SARS-CoV-2 circulation) and September to December 2020 (when there was increased SARS-CoV-2 circulation).

The event's incidence rates in the historical population will be measured in two ways:

1) National rates of events from July to September and September 1st to December 7th, 2020, to loss to follow-up, or death whichever is earlier and

2) Regional rates of events associated with COVID-19 during periods considered to be the 1st and 2nd wave of infection (for all cause outcome events). This will be conducted in each region in England but not before July 1st 2020 (as limited COVID-19 testing was available before July 2020). The exact timing of each wave of infection by regions in England will be determined during the analysis phase.

Contrary to the concurrent control arm, events' rates in the historical population are not impacted by different age distribution but may be impacted by isolation measures (e.g. lockdown) and waves of infections. Therefore, regions and relevant time periods will be taken into account in terms of waves of infection /lockdown. The exact date of start and end of each wave of infection by regions will be determined during the analysis phase.

When taking into account waves of infection, for COVID-19 related outcomes, the historical population will be followed from the designated date of start of 2nd wave of infection by regions in England (but not before September 2020), when COVID-19 test was more complete, to the end of the 2nd wave, loss to follow-up, or death whichever is earlier. For all-cause outcomes, the historical population will be followed from the designated date of start of 1st wave or 2nd of infection by regions in England to the end of the corresponding wave, loss to follow-up, or death whichever is earlier.

9.7.5 Feasibility Assessment for Further Exploratory Analyses

To understand the VE on different variants of the coronavirus, we will describe the VE by a combination of region and time period that aligns with best knowledge of when and where new variants of the virus were present according to available data sources and explore mechanisms for future individual level studies.

To study the VE among people in long term care (LTC) or assisted living facilities in the subset of subjects with data available on household key, median household age (>60), and number of residents (9 or more) and can be linked to the Care Quality Commission register, which lists residences where a degree of nursing care is provided. These variables will be used as a proxy to identify individuals receiving LTC at home or at specialized facilities.

9.7.6 RAVEN Extension Analyses (Other than VE)

9.7.6.1 Populations at Risk and Predictors of Having Boosters

The risk profile of the population at risk of COVID-19 due to either being ineligible for COVID-19 vaccines or due to suboptimal vaccine response (i.e. immunocompromised individuals) will be described. Specific subgroups (including their size) may be identified. AI/ML methods will be applied to explore the risk profile of the population at risk of COVID due to suboptimal vaccine response, and thus who may be more likely to experience breakthrough cases. A breakthrough case is defined as a person who has a positive virology test for the COVID-19 virus 14 or more days after their second vaccination (CDC definition (CDC, 2021)). We will consider predictions of breakthrough cases of SARS-CoV2 infections and clinically significant breakthrough infections (i.e. the primary outcomes), the identification of susceptible sub-populations and predictions of the time to vaccine waning with respect to these outcomes. People who are immunocompromised are the primary population of interest. AI/ML methods will also be used to identify predictors of receiving AZ boosters.

9.7.6.2 Health Care Resource Use and Health Care Cost Related to COVID-19

The objective of the economic analysis will be to estimate incremental health service use and incremental health service costs associated with receipt of at least one dose of the Oxford/AstraZeneca COVID-19 vaccine (the AstraZeneca vaccine arm) or at least one dose of other COVID-19 vaccines (the other COVID-19 vaccine arm) over the short (one and three months of follow-up) and medium term (six months of follow-up) and to disaggregate incremental costs by cost category. Our comparator groups and their construction will mirror those proposed for the statistical analyses, namely: (i) the AstraZeneca vaccine arm; (ii) the other COVID-19 vaccine arm; (iii) the concurrent control arm. The approaches to matching participants based on age group, gender, region, comorbidity status and practice, and to defining the date of study entry, will mirror those proposed for the statistical analyses.

Cumulative health care service use and health care costs covering primary and secondary health care services over 1 month, 3 months and 6 months of follow-up will be estimated using data extracted from the data assets listed in Section 9.4, and grouped by the following clinical settings and resource categories:

- Primary care consultations;
- Primary care prescriptions;
- Primary care medical tests and investigations;
- Hospital outpatient attendances and procedures;
- Hospital A&E (emergency department) attendances without admission
- Hospital admissions, inpatient, and day cases.

Primary and secondary care services will be valued by attaching unit costs derived from national compendia to resource inputs. All costs will be expressed in pounds sterling and valued at 2021 prices with unit costs estimated at earlier price dates inflated to 2021 prices using the NHS Hospital and Community Pay and Prices Index.

Characteristics of each comparator group will be compared using t-tests for continuous variables and chi-square tests for categorical variables. Initial exploration of the data will be conducted to guide selection of the appropriate analytic strategy. Inverse probability weighted GEE (Salazar et al., 2016) will be used to model non-linear trends in healthcare service use and costs over time, accounting for missing data, and adjusting for the matching variables and BMI status, smoking status, and frailty score. In the models, the total follow-up period will be divided into time intervals and interval-specific weights calculated as the inverse cumulative product of the probability of being observed up to a given interval using the method of van der Wal and Geskus (2011). Inverse probability weighting ensures that observations with a high probability of being missing (for example costs incurred in the later months of follow-up) are given more weight than observations with a low probability of being missing (for example costs incurred in the earlier months of follow-up) (Glick et al., 2015).

Adjusted estimates of cumulative healthcare utilisation counts, lengths of stay and costs, and between-group difference in counts, length of stay and costs, will be obtained from the regressions by summing interval-specific estimates over the time period of interest, a strategy originally reported by Lin (2000). Standard errors and confidence intervals around the mean estimates will be obtained via the robust variance/sandwich estimator, which provides a degree of robustness to departures from normality of errors and misspecification of the GEE working correlation matrix. An alternative approach for estimating standard errors and 95% confidence intervals is through the use of non-parametric bootstrapping. This would involve randomly sampling individual patients with replacement from the data to generate, for example, 1000 replicate datasets, and then fitting regression models to each replicate dataset and calculating adjusted bootstrap estimates of cumulative counts, lengths of stay and costs and the respective between-group differences. However, we recognise that the non-parametric bootstrap method can be computationally expensive for large sample applications such as ours. As a sensitivity check of the results generated by our primary modelling approach, we will therefore also apply the nonparametric bootstrap method to a smaller sample of data to obtain bias-corrected standard errors and confidence intervals. For completeness, the analyses will be repeated by restricting to patients with complete data over 1, 3 and 6 months of follow-up. The approaches to handling potential confounding will mirror those adopted by the statistical analyses. The analyses will also be repeated by the subgroups pre-specified by the statistical analyses.

We will also report the health care resource use and health care costs by COVID-19 risk group, using the JCVI risk group classification.

9.7.6.3 Household Incidence/Transmission

The impact of vaccination on household incidence/transmission will be explored, using an established method (de Lusignan et al, 2020a; de Lusignan et al, 2020b). Vaccinated and non-vaccinated cases with a positive SARS-CoV-2 test will be identified (index cases), along with their household contacts (only possible in part of the ORCHID dataset). The proportion of unvaccinated and vaccinated contacts who test positive for SARS-CoV-2 after the index case (secondary cases) will be compared among the vaccinated and non-vaccinated index cases. Logistic regression will be used to estimate the odds (and 95%CI) of being a secondary case, adjusting for confounders.

9.7.6.4 Long COVID

Cases of long COVID will be described, including by initial COVID-19 disease severity. The impact of vaccination on the occurrence of long COVID (in vaccinated individuals, in vaccinated individuals with COVID-19) will be explored.

9.8 Quality Control

The Principal Investigators (PIs) are responsible for ensuring protocol compliance in accordance with AZ standards of quality. The PIs may implement activities that could include but are not limited to:

- ensure appropriate storage of programming codes, code book, variables' definition etc.
- double programming of the data analysis to ensure high quality of data analyses and avoidance of coding errors
- confirm that the research team is complying with the protocol

The PIs will ensure that appropriate training relevant to the Observational Study is given to investigational staff, and that any new information relevant to the performance of this Observational Study is forwarded to the staff involved.

9.9 Limitations of the Research Methods

Even with an established and universal testing system, given the characteristics of the coronavirus and the RT-PCR test, it is expected that some COVID-19 infections may be missed. However, we expect that this is non-differential regarding the vaccination status.

A high vaccination rate by Oxford/AstraZeneca and other non-Oxford/AstraZeneca COVID-19 vaccines is expected to create herd immunity. The more people develop immunity to the coronavirus (due to infection or vaccination), the less likely unvaccinated people will be infected. Therefore, people in the concurrent control arm will have less infection and less likely to experience hospitalization/death due to COVID-19, diluting the effectiveness of the studied vaccine. Therefore, it is very important to study the real-world effectiveness of the vaccine early after the vaccination campaign started. This study uses all the data available immediately after the

vaccination campaign in England started which will help limit this potential impact of herd immunity. As for any other studies using real world data, there are caveats of administrative data captured for non-research purposes, ranging from misdiagnoses, lack of necessary details, to missing data. NHS Digital data lack information on the virus's variants, people's occupation or long-term care residence status which hinder the ability to study the VE on different variants of the virus, or VE in long-term care residents.

As the pandemic progresses, the proportion of persons who have experienced a SARS-CoV-2 infection, and therefore have a degree of natural immunity, increases. This means that, over time, fewer persons are available for the primary analysis (in which documented prior COVID-19 infection is an exclusion criterion), and that in the sensitivity analysis (in which this exclusion criterion is not applied) the measurement moves towards a measurement of relative effectiveness of two different immune states. This would lead to a decrease in the VE estimate over time.

9.10 Strengths of the Research Methods

The use of the national registries of all vaccination, COVID-19 testing, hospitalization, and mortality in England allow for the largest possible sample size and a near complete assessment of exposure, outcomes, and covariates. The relatively short lag time of the registries enable rapid assessment of the VE on the outcomes.

9.11 Bias

9.11.1 Methods to Minimize Bias

The advantages of having a control group is that it provides a direct comparison with less impact of potential time-varying rate of events. The disadvantages are that there may be potential differences between people who chose to vaccinate vs not. Also, few concurrent control people in the older age group may be available leading to different age composition of the study arms. This means the concurrent control arm may have more younger people which is expected to have lower hospitalization/death. For this reason, adjustment for age is critical.

It is expected that rates of outcome events are time-dependent and they are subject to the impact of waves of infection and isolation methods such as the use of masks (which may be less early in the pandemic), travel restrictions, lockdown (by waves of infection). Therefore, studying the historical population in the pre-vaccination periods, when testing was more common and there was increased circulation of infection enables us to contextualize the results and assess the potential impact of different time periods on the analysis.

9.11.2 Adjustment for Multiple Comparisons

With only a small number of comparisons for the VE analysis, no adjustment for multiple comparisons is needed.

9.12 Interim analyses

To quickly generate important VE and other data, data refreshes will be requested on a monthly basis. Interim analyses will be planned based on data availability. The first interim report is planned for Q1 2022.

10. PROTECTION OF HUMAN SUBJECTS

The investigation will be approved by an appropriate ethics board. For this study we will seek research ethics committee approval through the Integrated Research Application System (IRAS). IRAS is managed by the Health Research Agency, given the importance of a timely review we will apply to their fast-track review.

11. MANAGEMENT AND REPORTING OF ADVERSE EVENTS/ADVERSE REACTIONS

N/A since this is a secondary data study

12. PLANS FOR DISSEMINATING AND COMMUNICATING STUDY RESULTS

We will register the study in the ISRCTN registry (https://www.isrctn.com/), as well as other relevant registries (e.g. ENCePP)and list the study on Oxford websites. Dissemination will take place through peer review publications, provision of information to regulators. We will also involve patients and public from the newly formed ORCHID - Patient and Public Involvement and Engagement group in dissemination.

13. REFERENCES

1. Antrobus RD, Coughlan L, Berthoud TK, Dicks MD, Hill AV, Lambe T, et al. Clinical assessment of a novel recombinant simian adenovirus ChAdOx1 as a vectored vaccine expressing conserved Influenza A antigens. Molecular therapy. 2014;22(3):668-74.
2. CDC. COVID-19 Vaccine Breakthrough Case Investigation and Reporting. 2021 [updated 2021 Sept 10; cited 2021 Sept 30]. Available from: https://www.cdc.gov/vaccines/covid-19/health-departments/breakthrough-cases.html.

3. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid—mechanisms, risk factors, and management. BMJ. 2021;374(n1648).
4. Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz MA, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. New England Journal of Medicine. 2021;384(15):1412-23.
5. de Lusignan S, Sherlock J, Ferreira F, O'Brien S, Joy M. Household presentation of acute gastroenteritis in a primary care sentinel network: retrospective database studies. BMC Public Health. 2020;20(1):1-12.
6. de Lusignan S, Sherlock J, Akinyemi O, Pebody R, Elliot A, Byford R, et al. Household presentation of influenza and acute respiratory illnesses to a primary care sentinel network: retrospective database studies (2013–2018). BMC Public Health. 2020;20(1):1-11.
7. de Lusignan S, Bernal JL, Byford R, Amirthalingam G, Ferreira F, Akinyemi O, et al. Influenza and respiratory virus surveillance, vaccine uptake, and effectiveness at a time of cocirculating COVID-19: Protocol for the English primary care sentinel system for 2020-2021. JMIR public health and surveillance. 2021;7(2):e24341.
8. Department of Health and Social Care. Most vulnerable could be offered booster COVID-19 vaccines from September 2021 [updated 2021 June 30; cited 2021 Aug 31]. Available from: https://www.gov.uk/government/news/most-vulnerable-could-be-offered-booster-covid-19-vaccines-from-september.
9. Folegatti PM, Bittaye M, Flaxman A, Lopez FR, Bellamy D, Kupke A, et al. Safety and immunogenicity of a candidate Middle East respiratory syndrome coronavirus viral-vectored vaccine: a dose-escalation, open-label, non-randomised, uncontrolled, phase 1 trial. The Lancet Infectious Diseases. 2020;20(7):816-26.
10. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials: Oxford University Press; 2015.
11. HM Government. Analysis of the health, economic and social effects of COVID-19 and the approach to tiering. London: Department of Health and Social Care; 2020 Nov 30.
12. JCVI. Joint Committee on Vaccination and Immunisation: advice on priority groups for COVID-19 vaccination, 30 December 2020. 2020 [updated 2021 Jan 6; cited 2022 Jan 17]. Available from: https://www.gov.uk/government/publications/priority-groups-for-coronavirus-covid-19-vaccination-advice-from-the-jcvi-30-december-2020/joint-committee-on-vaccination-and-immunisation-advice-on-priority-groups-for-covid-19-vaccination-30-december-2020.
13. Lin D. Linear regression analysis of censored medical costs. Biostatistics. 2000;1(1):35-47.
14. López-Camacho C, Abbink P, Larocca RA, Dejnirattisai W, Boyd M, Badamchi-Zadeh A, et al. Rational Zika vaccine design via the modulation of antigen membrane anchors in chimpanzee adenoviral vectors. Nature communications. 2018;9(1):1-11.

15. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The lancet. 2020;395(10224):565-74.
16. Mahase E. Covid-19 booster vaccines: What we know and who's doing what. 2021;374(n2082).
17. Mayor N, Tsang R, Joy M, Hobbs FR, de Lusignan S. Long covid: coding is caring. bmj. 2021;373:n1262.
18. Medicines and Healthcare products Regulatory Agency (a). Moderna COVID-19 vaccine approved by MHRA in 12-17 year olds. 2021 [updated 2021 Aug 17; cited 2021 Aug 31]. Available from: https://www.gov.uk/government/news/moderna-covid-19-vaccine-approved-by-mhra-in-12-17-year-olds.
19. Medicines and Healthcare products Regulatory Agency (b). The MHRA concludes positive safety profile for Pfizer/BioNTech vaccine in 12- to 15-year-olds. 2021 [updated 2021 June 4; cited 2021 Aug 31]. Available from: https://www.gov.uk/government/news/the-mhra-concludes-positive-safety-profile-for-pfizerbiontech-vaccine-in-12-to-15-year-olds.
20. Mishra S, Mindermann S, Sharma M, Whittaker C, Mellan TA, Wilton T, et al. Changing composition of SARS-CoV-2 lineages and rise of Delta variant in England. EClinicalMedicine. 2021;39:101064.
21. Public Health England. Weekly national Influenza and COVID 19 surveillance report: Week 12 report (up to week 11 data). London: Public Health England; 2021 March 25.
22. Rproject. R: A language and environment for statistical computing. 2020 [cited 2022 Jan 17]. Available from: https://www.R-project.org/.
23. Salazar A, Ojeda B, Dueñas M, Fernández F, Failde I. Simple generalized estimating equations (GEEs) and weighted generalized estimating equations (WGEEs) in longitudinal studies with dropouts: guidelines and implementation in R. Statistics in medicine. 2016;35(19):3424-48.
24. Sandmann FG, Davies NG, Vassall A, Edmunds WJ, Jit M, Sun FY, et al. The potential health and economic value of SARS-CoV-2 vaccination alongside physical distancing in the UK: a transmission model-based future scenario analysis and economic evaluation. The Lancet Infectious Diseases. 2021;21(7):962-74.
25. SNOMED International. 5-Step briefing. 2020. [cited 2022 Jan 17]. Available from: https://www.snomed.org/snomed-ct/five-step-briefing.
26. UK Health Security Agency. COVID-19 Greenbook Chapter 14a. London: UK Health Security Agency; 2021 Sept 21. Report No.: UKHSA gateway number 2020300.
27. van der Wal WM, Geskus RB. ipw: an R package for inverse probability weighting. Journal of Statistical Software. 2011;43:1-23.

28. Vasileiou E, Simpson CR, Robertson C, Shi T, Kerr S, Agrawal U, et al. Effectiveness of first dose of COVID-19 vaccines against hospital admissions in Scotland: national prospective cohort study of 5.4 million people. Under review. 2021.
29. World Health Organization. Weekly epidemiological update on COVID-19 – 16 March 2021. 2021 [updated 2021 March 16; cited 2021 March 16]. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update---16-march-2021.
30. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.

14. APPENDICES

APPENDIX A List of stand-alone documents

None

APPENDIX B ENCePP checklist for study protocols

Study title:
Real-world effectiveness of the Oxford/AstraZeneca COVID-19 vaccine in England (RAVEN)

EU PAS Register® number: EUPAS43571

Study reference number (if applicable): D8111R00007

Section 1: Milestones	**Yes**	**No**	**N/A**	**Section Number**
1.1 Does the protocol specify timelines for				
1.1.1 Start of data collection[1]	☒	☐	☐	0
1.1.2 End of data collection[2]	☒	☐	☐	0
1.1.3 Progress report(s)	☐	☒	☐	
1.1.4 Interim report(s)	☒	☐	☐	9.12
1.1.5 Registration in the EU PAS Register®	☒	☐	☐	PASS information
1.1.6 Final report of study results.	☒		☐	0

Comments:

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1 Does the formulation of the research question and objectives clearly explain:				
2.1.1 Why the study is conducted? (e.g. to address an important public health concern, a risk identified in the risk management plan, an emerging safety issue)	☒	☐	☐	0
2.1.2 The objective(s) of the study?	☒	☐	☐	8

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

Section 2: Research question	**Yes**	**No**	**N/A**	**Section Number**
2.1.3 The target population? (i.e. population or subgroup to whom the study results are intended to be generalised)	☒	☐	☐	9.2.1
2.1.4 Which hypothesis(-es) is (are) to be tested?	☐	☒	☐	
2.1.5 If applicable, that there is no *a priori* hypothesis?	☐	☐	☒	

Comments:

Section 3: Study design	**Yes**	**No**	**N/A**	**Section Number**
3.1 Is the study design described? (e.g. cohort, case-control, cross-sectional, other design)	☒	☐	☐	9.1
3.2 Does the protocol specify whether the study is based on primary, secondary or combined data collection?	☒	☐	☐	9.1
3.3 Does the protocol specify measures of occurrence? (e.g., rate, risk, prevalence)	☒	☐	☐	9.7.1
3.4 Does the protocol specify measure(s) of association? (e.g. risk, odds ratio, excess risk, rate ratio, hazard ratio, risk/rate difference, number needed to harm (NNH))	☒	☐	☐	8 and 9.7.1
3.5 Does the protocol describe the approach for the collection and reporting of adverse events/adverse reactions? (e.g. adverse events that will not be collected in case of primary data collection)	☐	☐	☒	

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:	☐			9.2
4.2.1 Study time period	☒	☐	☐	
4.2.2 Age and sex	☒	☐	☐	
4.2.3 Country of origin	☒	☐	☐	
4.2.4 Disease/indication	☒	☐	☐	
4.2.5 Duration of follow-up	☒	☐	☐	

Section 4: Source and study populations	**Yes**	**No**	**N/A**	**Section Number**
4.3 Does the protocol define how the study population will be sampled from the source population? (e.g. event or inclusion/exclusion criteria)	☒	☐	☐	9.2.2and 9.2.3

Comments:

Section 5: Exposure definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
5.1 Does the protocol describe how the study exposure is defined and measured? (e.g. operational details for defining and categorising exposure, measurement of dose and duration of drug exposure)	☒	☐	☐	9.3.1
5.2 Does the protocol address the validity of the exposure measurement? (e.g. precision, accuracy, use of validation sub-study)	☐	☒	☐	
5.3 Is exposure categorised according to time windows?	☒	☐	☐	9.3.1
5.4 Is intensity of exposure addressed? (e.g. dose, duration)	☒	☐	☐	9.3.1
5.5 Is exposure categorised based on biological mechanism of action and taking into account the pharmacokinetics and pharmacodynamics of the drug?	☒	☐	☐	9.3.1
5.6 Is (are) (an) appropriate comparator(s) identified?	☒	☐	☐	9.2.1

Comments:

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?	☒	☐	☐	8 and 9.3.2
6.2 Does the protocol describe how the outcomes are defined and measured?	☐	☒	☐	
6.3 Does the protocol address the validity of outcome measurement? (e.g. precision, accuracy, sensitivity, specificity, positive predictive value, use of validation sub-study)	☐	☒	☐	

Section 6: Outcome definition and measurement	**Yes**	**No**	**N/A**	**Section Number**
6.4 Does the protocol describe specific outcomes relevant for Health Technology Assessment? (e.g. HRQoL, QALYs, DALYS, health care services utilisation, burden of disease or treatment, compliance, disease management)	☒	☐	☐	8 and 9.3.2

Comments:

How the outcomes are defined and measures will be described in the Statistical Analysis Plan.

Section 7: Bias	**Yes**	**No**	**N/A**	**Section Number**
7.1 Does the protocol address ways to measure confounding? (e.g. confounding by indication)	☒	☐	☐	9.2.1 and 9.7
7.2 Does the protocol address selection bias? (e.g. healthy user/adherer bias)	☒	☐	☐	9.11.1
7.3 Does the protocol address information bias? (e.g. misclassification of exposure and outcomes, time-related bias)	☒	☐	☐	9.7.4

Comments:

Section 8: Effect measure modification	**Yes**	**No**	**N/A**	**Section Number**
8.1 Does the protocol address effect modifiers? (e.g. collection of data on known effect modifiers, sub-group analyses, anticipated direction of effect)	☒	☐	☐	9.7.2

Comments:

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1 Does the protocol describe the data source(s) used in the study for the ascertainment of:				
9.1.1 Exposure? (e.g. pharmacy dispensing, general practice prescribing, claims data, self-report, face-to-face interview)	☒	☐	☐	9.3.1 and 9.4
9.1.2 Outcomes? (e.g. clinical records, laboratory markers or values, claims data, self-report, patient interview including scales and questionnaires, vital statistics)	☒	☐	☐	9.3.2 and 9.4

Section 9: Data sources	**Yes**	**No**	**N/A**	**Section Number**
9.1.3 Covariates and other characteristics?	☒	☐	☐	9.3.3 and 9.4
9.2 Does the protocol describe the information available from the data source(s) on:				
9.2.1 Exposure? (e.g. date of dispensing, drug quantity, dose, number of days of supply prescription, daily dosage, prescriber)	☒	☐	☐	9.3.1
9.2.2 Outcomes? (e.g. date of occurrence, multiple event, severity measures related to event)	☒	☐	☐	9.3.2
9.2.3 Covariates and other characteristics? (e.g. age, sex, clinical and drug use history, co-morbidity, co-medications, lifestyle)	☒	☐	☐	9.3.3
9.3 Is a coding system described for:				
9.3.1 Exposure? (e.g. WHO Drug Dictionary, Anatomical Therapeutic Chemical (ATC) Classification System)	☐	☒	☐	
9.3.2 Outcomes? (e.g. International Classification of Diseases (ICD), Medical Dictionary for Regulatory Activities (MedDRA))	☐	☒	☐	
9.3.3 Covariates and other characteristics?	☐	☒	☐	
9.4 Is a linkage method between data sources described? (e.g. based on a unique identifier or other)	☒	☐	☐	9.4

Comments:

Code lists and linkage method will be provided in the Statistical Analysis Plan.

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

Comments:

Section 11: Data management and quality control	**Yes**	**No**	**N/A**	**Section Number**
11.1 Does the protocol provide information on data storage? (e.g. software and IT environment, database maintenance and anti-fraud protection, archiving)	☐	☒	☐	
11.2 Are methods of quality assurance described?	☒	☐	☐	9.6 and 9.8
11.3 Is there a system in place for independent review of study results?	☐	☒	☐	

Comments:

Section 12: Limitations	**Yes**	**No**	**N/A**	**Section Number**
12.1 Does the protocol discuss the impact on the study results of:				
12.1.1 Selection bias?	☒	☐	☐	9.11.1
12.1.2 Information bias?	☒	☐	☐	9.9
12.1.3 Residual/unmeasured confounding? (e.g. anticipated direction and magnitude of such biases, validation sub-study, use of validation and external data, analytical methods).	☐	☒	☐	
12.2 Does the protocol discuss study feasibility? (e.g. study size, anticipated exposure uptake, duration of follow-up in a cohort study, patient recruitment, precision of the estimates)	☒	☐	☐	9.5

Comments:

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

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:

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

Comments:

Name of the study scientific lead of the protocol: [redacted]

[redacted]

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: d8111r00007-csp-v3		
Document Title:	D8111R00007 Clinical Study Protocol Version 3	
[redacted]	[redacted]	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
[redacted]	[redacted]	[redacted]

[redacted]