

Study Code	D8111R00007
Version	1.0
Date	22-Mar-21

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

**Lead investigator (Oxford):**

[REDACTED]

[REDACTED]

## TABLE OF CONTENTS

## PAGE

TITLE PAGE.....	1
TABLE OF CONTENTS .....	2
LIST OF ABBREVIATIONS AND DEFINITION OF TERMS.....	4
RESPONSIBLE PARTIES .....	5
PROTOCOL SYNOPSIS .....	6
Feasibility assessment for the following exploratory analyses .....	9
AMENDMENT HISTORY .....	11
MILESTONES .....	12
BACKGROUND AND RATIONALE .....	13
1. OBJECTIVES AND HYPOTHESES.....	15
2. METHODOLOGY .....	16
2.1 Study Design – General Aspects .....	16
2.2 Data Source(s) .....	19
2.3 Study Population.....	21
2.4 Inclusion Criteria .....	21
2.5 Exclusion Criteria .....	22
2.6 Participant Follow-up.....	22
3. VARIABLES AND EPIDEMIOLOGICAL MEASUREMENTS .....	23
3.1 Exposures .....	23
3.2 Outcomes.....	23
3.3 Other Variables and Covariates .....	25
4. STATISTICAL ANALYSIS PLAN .....	27
4.1 Statistical Methods – General Aspects.....	27
4.2 Stratification/ subgroup analyses .....	30
4.3 Sensitivity analyses .....	30
4.4 Feasibility assessment for exploratory analyses .....	31
4.5 Bias.....	31



4.5.1	Methods to Minimize Bias .....	31
4.5.2	Adjustment for Multiple Comparisons.....	32
4.5.3	Strengths and Limitations.....	32
4.6	Interim Analyses .....	33
4.7	Sample Size and Power Calculations.....	34
5.	STUDY CONDUCT AND REGULATORY DETAILS.....	34
5.1	Study Conduct .....	34
5.1.1	Study Flow Chart and Plan.....	34
5.1.2	Procedures .....	35
5.1.3	Quality Control .....	35
5.2	Protection of Human Subjects .....	35
5.2.1	Subject Informed Consent (Primary Data Collection Only) .....	36
5.2.2	Confidentiality of Study/Subject Data (Primary Data Collection Only) .....	36
5.3	Collection and Reporting of Adverse Events/Adverse Drug Reactions .....	36
6.	LIST OF REFERENCES .....	36
8.	Public Health England weekly vaccine effectiveness report, 22 March 2021 .....	37
7.	APPENDICES .....	39
8.	ATTACHMENTS.....	40
9.	SIGNATURES .....	41



## LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

Abbreviation or special term	Explanation
APC	Admitted patient care
BMI	Body mass index
CC	Critical care
CMMS	Cambridge Multimorbidity Score
COVID-19	SARS-CoV-2
DARS	Data access request service
EFI	Electronic frailty index
GDPPR	GPES data for pandemic planning and research
GPES	General practice extraction service
HDU	High dependency unit
HES	Hospital episode statistics
HRA	Health Research Authority (UK)
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
NHS	National health services
NPEx	National pathology exchange
ONS	Office for national statistics
ORCHID	Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub
REC	Research Ethics Committee
SGSS	Second generation surveillance system
SUS	Secondary uses service
TRE	Trusted research environment
VE	Vaccine effectiveness



RESPONSIBLE PARTIES

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]
[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]

## PROTOCOL SYNOPSIS

---

This 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. The interim analysis will focus on effectiveness of the 1<sup>st</sup> dose. The final analysis will look at both doses.

---

### **Background/Rationale:**

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 US 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 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. Thus, it remains important to better understand vaccine effectiveness by ages, time intervals between doses, and also to determine if single dose effectiveness may extend beyond 12 weeks. It's also important to assess vaccine effectiveness by age groups and comorbidities for best guidance for COVID-19 immunization programs.

The UK is one of the first countries that are introducing mass vaccination campaign for COVID-19 and is currently vaccinating the elderly population starting from the oldest age groups (JCVI advice 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. This study is to primarily assess the effectiveness of the Oxford/Astrazeneca COVID-19 vaccine. Given the known high efficacy of the Pfizer vaccine in RCT and RWE studies, the study is also to evaluate the Pfizer vaccine effectiveness as a validation of the study's methods.

### **Objectives and Hypotheses:**

#### *Primary:*

To assess the real world effectiveness of the Oxford/Astrazeneca COVID-19 vaccine among people who receive one dose of the vaccine, overall and by age group and time period after 1 dose



*Secondary:*

1. To assess the real world effectiveness of the Oxford/Astrazeneca COVID-19 vaccine in people vaccinated with two doses; and by timing after the 1<sup>st</sup> and 2<sup>nd</sup> dose, interval between the two doses, and comorbidity status
2. To assess the effectiveness of the Pfizer COVID-19 vaccine among people who received one dose and two doses, overall and by age group, time periods after each dose, and interval between the two doses

**Methods:**

**Study design:**

This is a retrospective cohort study using linked secondary databases in England accessed through the NHS Digital Trusted Research Environment (TRE). The primary care data will be linked with vaccination, hospitalization, COVID-19 test results, mortality data at the national level for capture of key study variables. All individuals aged 16 years or older will be included. People in the vaccinated arm will be compared with two comparator cohorts:

- a) The concurrent control arm (primary): People not vaccinated with any COVID-19 vaccine from January 2021 onward (for the Oxford/Astrazeneca arm) or from December 2020 onward (for the Pfizer arm)
- b) The historical control arm (secondary): People during the period from July-December 2020 for primary outcomes or from March-December 2020 for secondary outcomes, before the COVID-19 vaccine was available in England

Only the first outcome event, i.e. hospitalization, ICU admission will be considered in the analysis.

**Data Source(s):**

NHS Digital Data Access Request Service (DARS): Through this service 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. Data assets available through DARS which may be used for this study include:

- GPES data for pandemic planning and research (GDPPR): central collection of GP patient data for COVID-19 purposes
- NIMS National Immunization Management System – includes COVID-19 Vaccination Status
- Hospital Episode Statistics (HES) covers patients attending accident and emergency units, admitted for treatment or attending outpatient clinics.
- Civil Registrations (Deaths): Information including the date, place and certificated cause of death from the Office for National Statistics (ONS)
- COVID-19 Second Generation Surveillance System (SGSS) - diagnostic information from laboratory test reports for patients tested for COVID-19 in England only.
- Covid-19 UK Non-hospital Antigen Testing Results data includes a range of COVID19 test results, including NPEX.



- COVID-19 Hospitalization in England Surveillance System (COVID-19 SARI-Watch/CHES): Epidemiological data on COVID-19 infection in persons requiring hospitalization in Intensive Care Units (ICU) or High Dependency Units (HDU).
- Secondary User Services (SUS) Episodes (including the following concepts: Admitted Patient Care, Outpatient, Critical Care & Accident and Emergency).
- Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID) (de Lusignan et al, 2021). This TRE hosts the Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) one of Europe's oldest sentinel systems.

### **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 (the Astrazeneca vaccine arm) or at least one dose of the Pfizer COVID-19 vaccine (the Pfizer vaccine arm) and people in England who were not vaccinated with any COVID-19 vaccine during the same time period (concurrent control arm) or during March-Dec 2020 (historical control arm) who were matched to the vaccinated individuals by age, gender, region, and comorbidity. All study participants are 16 years of age or older.

### **Exposure(s):**

The main exposure is whether a person was vaccinated with the Oxford/Astrazeneca COVID-19 vaccine or with the Pfizer COVID-19 vaccine versus not vaccinated with any COVID-19 vaccine.

### **Outcome(s):**

The primary outcome is COVID-19 related hospitalization, ICU admission, and death. Events due to any causes of these measures will be secondary.

### **Sample Size Estimations:**

The NHS Digital database comprises data from the national registries of vaccination and hospitalization with linkage to the EMR, 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 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.

### **Statistical Analysis:**

A flow diagram of the study population detailing each step of the inclusion and exclusion criteria applied will be generated. The index date of the vaccinated subjects is the date of the 1<sup>st</sup> vaccination. For the concurrent control arm, the index date is generated by assigning a random value to their 2021 follow-up. After this index date is assigned, participants in the control arms will be frequency matched to the respective vaccinated arms (Astrazeneca or Pfizer vaccine arm), where matching is based on age group, gender, region, and comorbidity status.

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





the measurement closest to index date. 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 Rate Ratio and the vaccine effectiveness (VE) which is calculated as  $1 - \text{Rate Ratio}$ . 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 2 doses, and presence of comorbidities.

Poisson regression 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. Other patient characteristics will be included in the model unless the number of events per covariates is less than 7. In that case, we will use stepwise regression for the additional variables to select these variables into the model if including or excluding them results in change of at least 10% of the point estimate of the vaccination effect.

In order to address concern regarding potential confounding due to differences in health care seeking behaviour, we will also duplicate the analysis in a subgroup of people who received an influenza vaccine during the 2020-2021 flu season and before the index date.

Stratifications/subgroup analyses may include:

- +Age: 16-49, 50-64, 65-69, 70-79, 80+; or 16-39, 40-<49, 50+, 65+, 80+
- +Vaccine dose: people receiving 1 dose only, 2 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
- +Interval between doses among people receiving 2 doses: 3-<6, 6-<10, and 10-11, 12, >12 weeks and/or other combinations may be used
- +Comorbidity
  - categories of the Cambridge Multimorbidity Score
  - specific comorbidities known to indicate high risk for covid-19 infections in accordance with JCVI's guidance
- +Vaccine type (1st dose with Oxford/Astrazeneca vaccine and 2nd dose with a non-Oxford/Astrazeneca vaccine)
- +Long-term care or assisted living care residents (if possible)

Several sensitivity analyses will be considered

- Only HES data are included for follow-up data
- People with history of covid-19 infection prior to vaccination not excluded
- A covid-19 related event is defined as an event with a ICD-10 diagnosis of COVID-19 or an event within 14 days of a positive RT-PCR for SARS-CoV-2 infection.

Feasibility assessment for the following exploratory analyses

In order 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 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 also can be linked to the Care Quality Commission (CQC) 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.



## AMENDMENT HISTORY

Date	Section of study protocol	Amendment or update	Reason
		N/A	



## MILESTONES


Milestones	Planned dates
Study design concept approved	██████
External service provider/contract research organization selected (if relevant)	██████
Study protocol approved	██████
First subject/patient in (or database start date)	██████
Registration of the study with ISRCTN	██████
Last subject/patient in (or database end date)	██████
Last subject/patient last visit	██████
Database lock for interim analysis, i.e. primary objective	██████
Interim data analysis completed for primary objective	██████
Final database lock	██████
Clinical study report completed	██████
Operational information	Details
Approximate study budget	
Budget holder(s), including cost-sharing	██████
Delivery model (internal – Global Medical Affairs, Marketing Company, Site Management and Monitoring – or external)	██
Approach towards patient centricity (e.g. engagement with patient groups related to study)	No
International coordinating investigator or executive steering committee	TBD

████████████████████  
████████████████████  
████████████████████

## BACKGROUND AND RATIONALE

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 (Zhu et al, 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 et al, 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 US 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 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. Thus, it remains important to better understand vaccine effectiveness by ages, time intervals between doses, and also 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 vaccine effectiveness by age groups and comorbidities for best guidance for COVID-19 immunization programs.

The UK is one of the first countries that are introducing mass vaccination campaign for COVID-19 and is currently vaccinating the elderly population starting from the oldest age groups (JCVI advice 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. This study is to primarily assess the effectiveness of the Oxford/Astrazeneca COVID-19 vaccine. Given the known high efficacy of the Pfizer vaccine in RCT and RWE studies, the study is also to evaluate the Pfizer vaccine effectiveness as a validation of the study's methods.



## 1. OBJECTIVES AND HYPOTHESES

The study's primary objective is to assess the overall effectiveness of the 1st dose of the Oxford/Astrazeneca COVID-19 vaccine among people with one dose of the vaccine and by age groups and by time periods after the 1<sup>st</sup> dose. The secondary objective is to assess the overall effectiveness among people with 2 doses, and by time periods after the 1<sup>st</sup> and 2<sup>nd</sup> dose, intervals between the 2 doses, and comorbidities that are known to associated with more severe COVID-19 infection.

<b>Primary objective(s)</b> -To assess the effectiveness of the Oxford/Astrazeneca COVID-19 vaccine among people who receive one dose of the vaccine +overall +by age group +time period after 1 dose	<b>Outcome measure</b> <i>Primary</i> -Rate of hospitalizations associated with COVID-19 -Rate of admission to ICU associated with COVID-19 -Rate of mortality associated with COVID-19  <i>Secondary</i> -Rate of any hospitalizations -Rate of any admission to ICU -Overall mortality rate	<b>Hypothesis tested (if relevant)</b> The Oxford/Astrazeneca COVID-19 vaccine is effective in preventing severe COVID-19 infections after just one dose
<b>Secondary objective(s)</b> -To assess the effectiveness of the Oxford/Astrazeneca COVID-19 vaccine in people vaccinated with two doses; and by	<b>Outcome measure</b> Same as above	<b>Hypothesis tested (if relevant)</b> The Oxford/Astrazeneca COVID-19 vaccine provides protection after 2 doses; The vaccine is effective with different intervals between



a) timing after the 1 <sup>st</sup> and 2 <sup>nd</sup> dose b) interval between the two doses c) comorbidity status  -To assess the effectiveness of the Pfizer COVID-19 vaccine among people who received one dose and two doses, overall and by a) age group b) time periods after each dose c) interval between the two doses		the two doses and in high risk patients with comorbidities.
--	--	---

## 2. METHODOLOGY

### 2.1 Study Design – General Aspects

This is a retrospective cohort study using linked secondary databases in England accessed through the NHS Digital Trusted Research Environment (TRE). The primary care data will be linked with vaccination, hospitalization, COVID-19 test results, mortality data at the national level for capture of key study variables. The main exposure is whether a person was vaccinated with the Oxford/Astrazeneca COVID-19 vaccine or with the Pfizer COVID-19 vaccine versus not vaccinated with any COVID-19 vaccine. The primary outcome is COVID-19 related hospitalization, ICU admission, and death. The secondary outcome is overall hospitalization, ICU admission, and death due to any causes. All individuals aged 16 years or older will be



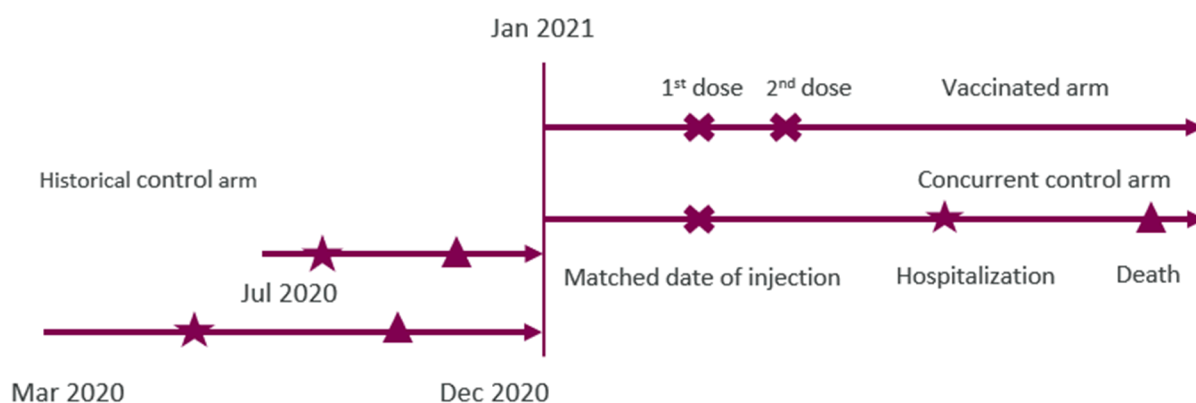


included but people 65 years of age or older will be the focus of this study. People in the vaccinated arm will be compared with two comparator cohorts:

- c) The concurrent control arm (primary): People not vaccinated with any COVID-19 vaccine from January 2021 onward (for the Oxford/Astrazeneca vaccine arm) or from December 2020 onward (for the Pfizer vaccine arm)
- d) The historical control arm (secondary): People during the period from July-December 2020 for primary outcomes or from March-December 2020 for secondary outcomes, before the COVID-19 vaccine was available in England

The primary outcome will be severe COVID-19 related events including admission to hospitals, admission to the ICU, and deaths; events due to any causes of these measures will be secondary. For the vaccinated individuals, the study's index date is the date of vaccination of the 1<sup>st</sup> dose. Primarily, outcomes will be evaluated until the end of follow-up from day 22 after the 1<sup>st</sup> dose and among people with 1 dose, and from day 15 after the 2<sup>nd</sup> dose for people with 2 doses. Since people may not be protected right after vaccination, we will look at the 3-week post dose 1 period and the 2-week post dose 2 period separately. Only the first outcome event, i.e. hospitalization, ICU admission will be considered in the analysis.

Figure 1: The study design diagram



In order to ensure comparability between the study arms, participants in the control arms will be matched using frequency matching methods to the respective vaccinated arms (Astrazeneca or Pfizer vaccine arm) on age, gender, region, and comorbidity status. Participants in the concurrent control arm will also be matched to the vaccinated arm on date of vaccination (for the index date). Matched variables and other confounders e.g. frailty score, BMI, smoking status, prescribed medication etc will be included in multivariable regression to ensure comparability between the study arms.

For the vaccinated arm and the concurrent control arm, the incidence rate of the outcome events (i.e. COVID-19 related and overall hospitalization, admission to ICU, and death) will be measured from the index date to the end of follow-up, loss to follow-up, or death whichever is earlier. These event's rates may be stratified by different time periods of interest from the index date, after each dose, or between the two vaccine doses. For the historical control arm, the event's incidence rates will be measured in two ways: 1) rates of events associated with COVID-19 from July 1<sup>st</sup> to December 31<sup>st</sup>, 2020 (rates of all-cause events from March 1<sup>st</sup> – Dec 31<sup>st</sup> 2020), loss to follow-up, or death whichever is earlier and 2) during periods considered to be the 1<sup>st</sup> and 2<sup>nd</sup> wave of infection (for all cause outcome events) and during periods considered to be the 2<sup>nd</sup> wave of infection (for COVID-19 related events) in each respective region in England but not before July 1<sup>st</sup> 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. The 2<sup>nd</sup> way, i.e. 2) above, is to correspond to the analysis using vaccination data from January-February of 2021 when COVID-19 vaccination started in England which coincided with the 3<sup>rd</sup> wave of COVID-19 infection. For the analysis that includes vaccination data periods beyond February 2021, the historical control arm will focus on the event rates as mentioned as 1) above.

Due to England's vaccination campaign starting with the older age group, as of 22-March, 2021, around 90% of people 70 years of age or older were vaccinated with at least one dose of any COVID-19 vaccine (PHE weekly VE report 2021). Therefore, it is expected that the concurrent control arm will have more younger people and will therefore be subject to artificially lower



hospitalization and death. For this reason, the concurrent control arm is matched with the vaccine arm by age. Events' rates will also be stratified by age groups (e.g. 16-49, 50-64, 65-69, 70-79, 80+). For the historical control arm, events' rates 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.

## **2.2 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.

- NHS Digital Data Access Request Service (DARS). Through this service 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 trusted research environment (TRE). Data assets available through DARS which may be used for this study include:
- 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)
- NIMS National Immunization Management System – includes 2 associated datasets including COVID-19 Vaccination Status and COVID-19 Adverse Reaction.
- Hospital Episode Statistics (HES): this is the transformed data, initially part of the Commissioning Data Set (CDS), covering patients attending accident and emergency units, admitted for treatment or attending outpatient clinics at NHS hospitals in England. Statistical controls have been applied to the HES products. (Time lag 8-12 weeks)



- Civil Registrations (Deaths): Information including the date, place and certificated cause of death from the Office for National Statistics (ONS) (Time lag –1-2 weeks)
- COVID-19 Second Generation Surveillance System (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 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 Intensive Care Units (ICU) or High Dependency Units (HDU). It records all demographic, risk factor, treatment and outcome information for patients admitted to hospital with a confirmed COVID-19 diagnosis (Time lag – less than 1 week)
- SUS Episodes (including the following concepts: Admitted Patient Care, Outpatient, Critical Care & Accident and Emergency). This covers all secondary care provided in England and paid for by the NHS. This is essentially the Commissioning Data Set (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)
- Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID). 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 pseudonymized data are linked to the range of data sets described above by NHS Digital, though usage has to 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-Royal College of General Practitioners Research and Surveillance Centre (RSC). This surveillance system is sponsored by Public Health England and collaborates in reporting vaccine uptake and effectiveness, including of COVID-19 vaccine.

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 and SUS data from 8 weeks to 2 weeks before the database lock.

## **2.3 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 (the Astrazeneca vaccine arm) or at least one dose of the Pfizer COVID-19 vaccine (the Pfizer vaccine arm) and people in England who were not vaccinated with any COVID-19 vaccine during the same time period (concurrent control arm) or during March-Dec 2020 (historical control arm) who were matched to the vaccinated individuals by age, gender, region, and comorbidity. All study participants are 16 years of age or older.

## **2.4 Inclusion Criteria**

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

1. For the vaccinated arms:
  - 16 years of age or older at the index date
  - 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.
- Having been vaccinated with at least 1 dose of the Oxford/Astrazeneca COVID-19 vaccine (the Astrazeneca vaccine arm) or of the Pfizer COVID-19 vaccine (the Pfizer vaccine arm).

## 2. For the control arms

- At least 16 years of age at the index date for the concurrent control individuals or on March 1, 2020 for the historical controls.
- 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.
- Have not received any COVID-19 vaccine (Oxford/Astrazeneca, Pfizer, or Moderna COVID-19 vaccine)

## 2.5 Exclusion Criteria

People who meet the following criteria will be excluded:

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

## 2.6 Participant Follow-up

For the vaccinated and concurrent control arms, study participants will be followed from the index date to the end of follow-up, loss to follow-up, or death whichever is earlier. For historical control arm, study participants will be followed from July 1<sup>st</sup> 2020 to December 31<sup>st</sup> 2020 for COVID-19 related events and March 1<sup>st</sup> to December 31<sup>st</sup> 2021 for all-cause events, loss to follow-up, or death whichever is earlier. When taking into account waves of infection, for COVID-19 related outcomes, the historical control arm will be followed from the designated



date of start of 2<sup>nd</sup> wave of infection by regions in England (but not before July 2020), when COVID-19 test was more complete, to the end of the 2<sup>nd</sup> wave, loss to follow-up, or death whichever is earlier. For all-cause outcomes, the historical control arm will be followed from the designated date of start of 1<sup>st</sup> wave or 2<sup>nd</sup> of infection by regions in England to the end of the corresponding wave, loss to follow-up, or death whichever is earlier. The exact date of start and end of each wave of infection by regions will be determined during the analysis phase.

### **3. VARIABLES AND EPIDEMIOLOGICAL MEASUREMENTS**

#### **3.1 Exposures**

The main exposure in this study is whether an individual was vaccinated with one or two doses of the Oxford/Astrazeneca COVID-19 or the Pfizer COVID-19 vaccine. People without record of any COVID-19 vaccination are considered unexposed.

The COVID-19 vaccination event database contains detailed information on name of the COVID-19 vaccines, specific date of vaccination, and dose sequence (1<sup>st</sup> or 2<sup>nd</sup> dose) of the injection given to each vaccinated individual.

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

#### **3.2 Outcomes**

The primary outcomes of this study are

[REDACTED]

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

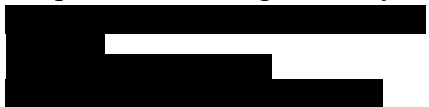
The secondary outcomes of this study are

- Overall hospitalization
- Overall ICU admission
- Overall death

A COVID-19 related event (i.e. hospitalization, ICU admission, death) is defined as:

- An event with an ICD-10 diagnosis of COVID-19 or
- An event within 28 days of a positive RT-PCR for SARS-CoV-2 infection.
  - o A sensitivity analysis for 14 days will also be conducted

Curated hospitalization and ICU admission data including date, frequency, and duration will come from HES until 8 weeks prior to the start of analysis. Raw hospitalization and ICU admission data come from SUS from 8 weeks to 2 weeks prior to start of analysis. Mortality data will come from GDPPR until 2 weeks prior to start of analysis. For cause of death, data from ONS until 6 weeks prior to start of analysis will be used. The HES, SUS, GDPPR, and ONS data may include information on COVID-19 infection through specific COVID-19 diagnosis codes. COVID-19 positive tests come from one or both of the two sources: 1) the SGSS contains information on all first positive tests (with dates) from hospital and community, 2) the NPEX contains test results (positive, negative) for all tests done (with dates) in the community. People with multiple tests in the community, their test dates, and test results are included in the NPEX. Using a combination of both 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 on, restriction of the study period from July 2020





or the 2<sup>nd</sup> wave of infection by regions in England, whichever is later, for determining COVID-19 related events in the historical control arm is needed.

Variables		Data source	Comments
Hospitalization	Date	HES for curated data	6-8 weeks lag for HES  SUS is updated weekly with a maximum of 2 weeks lag
	Frequency	SUS for more recent raw data	
	Duration		
ICU admission	Date	HES APC for combined hospital and ICU data;	weekly with a maximum of 2 weeks lag
	Frequency		
	Duration	HES CC for ICU specific data	
Death	Date	GDPPR for date	2 weeks lag
	Cause of death	ONS for cause of death	6 weeks lag
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	

### 3.3 Other Variables and Covariates

Variables	Categories	Data source	Comment
Age	Different categorizations may be used	GDPPR	BMI is binary as obese or not using HES, SUS, GDPPR or



	-16-<50; 50-<65; 65-<70; 70-<80; 80+ -18-<40; 40-<50; 50+; 65+; 70+; 80+		BMI can be created from weight/height if available;
Gender	M/F		
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 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.	HES and GDPPR	Follow JCVI 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.



Cambridge Multimorbidity Score	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	GPPPR	Prescribed medications at baseline will be grouped into immunosuppressant or not.
Long-term care/assisted living care facility	Yes/No	Not in any database	To be created for people aged 70+ by comparing their zip codes and LTC address; or vaccination at LTC facility.
Frailty score	EFI grouped into - Mild - Moderate - Severe	Derived from GDPPR	Use an algorithm already developed to derive EFI

## 4. STATISTICAL ANALYSIS PLAN

### 4.1 Statistical Methods – General Aspects

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 2.2, in order 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.

The index date of the vaccinated subjects is the date of the 1<sup>st</sup> vaccination. For the concurrent control arm, the index date is generated by assigning a random value to their 2021 follow-up. After this index date is assigned, the following will be applied. Participants in the control arms will be frequency matched to the respective vaccinated arms (Astrazeneca or Pfizer vaccine arm), where matching is based on age group (8 categories: 16-<40, 40-<50, 50-<55, 55-<60, 60-<65, 65-<70, 70-<80, 80+), gender, region (9 regions), and comorbidity status (4 categories: Cambridge Multimorbidity Score by quartile).

The demographic and characteristics of study participants in each vaccine arm and its respective control arms 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 3 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, IQR, median and mean to get insight in these variables, but also in order 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 Rate Ratio and the vaccine effectiveness (VE) which is calculated as  $1 - \text{Rate Ratio}$ . This will



also be provided per age group and per frailty score where further stratification/subgroups is indicated below. Finally, VE will also be provided in shorter periods after dose 1, and between the 2 doses, and presence of comorbidities.

Love plots will be created showing the mean standardized difference between each vaccinated cohort (Astrazeneca or Pfizer vaccine cohort) and the comparator cohorts in participants' characteristics. In addition, summary tables will be provided for the updated characteristics.

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. Other patient characteristics indicated in section 3.3 will be included in the model unless the number of events per covariates is less than 7. In that case, we will use stepwise regression for the additional variables to select these variables into the model if including or excluding them results in change of at least 10% of the point estimate of the vaccination effect.

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

In order to address concern regarding potential confounding due to differences in health care seeking behaviour, we will also duplicate the analysis in a subgroup of people who received an influenza vaccine during the 2020-2021 flu season and before the index date.

[REDACTED]

## 4.2 Stratification/ subgroup analyses

Stratifications/subgroup analyses may include:

- +Age: 16-49, 50-64, 65-69, 70-79, 80+; or 16-39, 40-<49, 50+, 65+, 80+
- +Vaccine dose: people receiving 1 dose only,  $\geq 1$  dose, 2 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 doses: 3-<6, 6-<10, and 10-11, 12, >12 weeks and/or other combinations may be used
- +Comorbidity
  - categories of the Cambridge Multimorbidity Score
  - specific comorbidities known to indicate high risk for covid-19 infections in accordance with JCVI's guidance, 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.
- +Vaccine type (1st dose with Oxford/Astrazeneca vaccine and 2nd dose with a non-Oxford/Astrazeneca vaccine)
- +Long-term care or assisted living care residents (if possible)

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

## 4.3 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-19 related event is defined as an event with a ICD-10 diagnosis of covid-19 or an event within 14 days of a positive RT-PCR for SARS-CoV-2 infection.

#### **4.4 Feasibility assessment for exploratory analyses**

In order 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 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 also can be linked to the Care Quality Commission (CQC) 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.

Bias

##### **4.4.1 Methods to Minimize Bias**

Each comparison group has its pros and cons which is mentioned in the table below.

A successful vaccination campaign in the elderly population can lead to few concurrent controls in the same age group. 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. The advantage of the historical control arm is that it does not have this problem and therefore much less likely to be confounded by age. On the other hand, it is expected that rates of outcome events are time-dependent and is subject to impact of waves of infection and isolation methods such as the use of mask (which may be less early in the pandemic), travel restriction, lockdown (by waves of infection). Since Jan-March 2021 when the COVID-19 vaccine is rolled out is the 3<sup>rd</sup> wave of infection in England, restricting the historical control arm to the 2<sup>nd</sup> wave of infection will help increase validity of the comparison. In addition, early in



the pandemic, testing of the coronavirus was not complete which mostly impacts the historical control arm. Restricting the historical control arm from July 2020 helps alleviate the misclassification of COVID-19 related events.

	Pros	Cons
Concurrent control arm	Most direct comparison; Less impact of potential time-varying rate of events	Potential difference between people who chose to vaccinate vs not;  Few concurrent control people in the older age group may be available leading to different age composition of the study arms
Historical control arm	A similar population with the exposed arm can be selected therefore less bias due to participant's comparability	Event rates may change quickly over time in any direction due to a surge in infection, impact of public health control methods (lockdown, mask requirement etc) and thus mixing with the effect of vaccination

#### 4.4.2 Adjustment for Multiple Comparisons

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

#### 4.4.3 Strengths and Limitations

##### Strengths:

[REDACTED]



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 vaccine effectiveness on the outcomes.

#### Limitations:

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.

## **4.5 Interim Analyses**

In order to quickly generate important vaccine effectiveness data for the first dose, especially in the elderly population, an interim analysis focus on the primary objective is planned to be conducted at the end of May. We expect that by end of May, a significant proportion of the 65 years of age or older population in England will have been vaccinated for at least one dose of the Oxford/Astrazeneca or Pfizer COVID-19 vaccine and had sometime for the vaccine to take effect



on the outcomes that can be captured. Once more data are accumulated for the 2<sup>nd</sup> dose (which is about 3 months after the 1<sup>st</sup> dose), the final analysis is planned for later in the year of 2021. The final analysis will address both the primary and secondary objectives.

#### **4.6 Sample Size and Power Calculations**

The NHS Digital database comprises data from the national registries of vaccination and hospitalization with linkage to the EMR, lab, and mortality data. It is the largest dataset possible for England and covers all regions. As of early March 2021, near 20 millions 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 vaccine effectiveness from the Oxford Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID) in collaboration with Public Health England (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 age 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 millions 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.

### **5. STUDY CONDUCT AND REGULATORY DETAILS**

#### **5.1 Study Conduct**

##### **5.1.1 Study Flow Chart and Plan**

See section 2.1 for the study flow chart



### **5.1.2 Procedures**

N/A for secondary data studies

### **5.1.3 Quality Control**

The Principal Investigators are responsible for ensuring protocol compliance in accordance with AZ standards of quality. The PI 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 Principal Investigators 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.

### **5.1.4 Dissemination**

We will register the study in the ISRCTN registry (<https://www.isrctn.com/>), and list the study on our 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 (PPIE) group in dissemination.

## **5.2 Protection of Human Subjects**

The investigation will be approved by an appropriate ethics board. For this study we will seek research ethics committee (REC) approval through the Integrated Research Application System



(IRAS). IRAS is managed by the Health Research Agency (HRA), given the importance of a timely review we will apply to their fast track review.

**5.2.1 Subject Informed Consent (Primary Data Collection Only)**

N/A

**5.2.2 Confidentiality of Study/Subject Data (Primary Data Collection Only)**

N/A

**5.3 Collection and Reporting of Adverse Events/Adverse Drug Reactions**

N/A since this is a secondary data study

**6. LIST OF 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. Mol Ther. 2014;22(3):668-74.
2. COVID-19 Greenbook chapter 14a:  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/961287/Greenbook\\_chapter\\_14a\\_v7\\_12Feb2021.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/961287/Greenbook_chapter_14a_v7_12Feb2021.pdf)
3. de Lusignan S, Lopez Bernal J, Byford R, 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 Surveill. 2021;7(2):e24341. Published 2021 Feb 19. doi:10.2196/24341
4. 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. *Lancet Infect Dis.* 2020;20(7):816-26.

5. JCVI advice on priority group for COVID-19 vaccination

<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>

6. 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. *Nat Commun.* 2018;9(1):2441.

7. 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. *Lancet.* 2020;395(10224):565-74.

8. Public Health England weekly vaccine effectiveness report, 22 March 2021

9. Vasileiou E, Simpson C, Robertson C, Shi T, Kerr S et al. Effectiveness of first dose of covid-19 vaccines against hospital admission in Scotland: national prospective cohort study of 5.4 million people. Under review, 2021.

10. WHO. (World Health Organization) Coronavirus disease (COVID-19) weekly epidemiological update - 16 March 2021 – Overview.

<https://www.who.int/publications/m/item/weekly-epidemiological-update---16-march-2021>. Published 2021. Accessed 22 March 2021.

11. 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.



## **7. APPENDICES**



## **8. ATTACHMENTS**





## 9. SIGNATURES

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

**ASTRAZENECA SIGNATURE(S)**

Real-world effectiveness of the Oxford/Astrazeneca COVID-19 vaccine

*This Observational Study Protocol has been subjected to an internal AstraZeneca review*  
I agree to the terms of this Study protocol.

**AstraZeneca representative**

 $\langle\langle Name, title \rangle\rangle$ 

Date  
(Day Month Year)

<<Email address and telephone number>>

This document contains confidential information, which should not be copied, referred to, released or published without written approval from AstraZeneca. Investigators are cautioned that the information in this protocol may be subject to change and revision.

[illegible]

## SIGNATURE PAGE

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

<b>Document Name:</b> [REDACTED]		
<b>Document Title:</b>	[REDACTED]	
<b>Document ID:</b>	[REDACTED]	
<b>Version Label:</b>	[REDACTED]	
<b>Server Date</b> (dd-MMM-yyyy HH:mm 'UTC'Z)	<b>Signed by</b>	<b>Meaning of Signature</b>
[REDACTED]	[REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]

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