Analysis Plan for Programmatic evaluation of COVID-19 vaccination against hospitalisations in adults

Full Project Title	EAVE II
Version Number	V1
Previous Versions	N/A
Effective Date	05 March 2021
Analyst(s)	*Chris Robertson (CR), Eleftheria Vasileiou (EV), Emily Moore (EM), Rachel Mulholland (RM)
Co-authors	Colin Simpson (CS), Ting Shi (TS), Steven Kerr (SK)
Target journal	

* Corresponding author

Version History			
Version	**Date**	**Author**	**Notes**
V1	05.03.2021	EV, SK	First version sent to team

Contents

1 Introduction

The aim of this study is to investigate the association between receiving first and second doses of the COVID-19 vaccines that are currently licensed in the UK (Pfizer-BioNTech, Oxford-Astra-Zeneca, Moderna) as part of the vaccination programme in Scotland, and COVID-19 hospital admissions. We will use a data platform containing pseudonymised, linked data from GP records, vaccination records, secondary care, and virological testing with coverage for the whole of Scotland. All data and analysis will be hosted within a trusted research environment at Public Health Scotland (PHS) or the National Safe Haven.

2 Aims and objectives

2.1 Aims
To investigate the association between receiving one and two doses of the Pfizer-BioNTech, Oxford-Astra-Zeneca and Moderna COVID-19 vaccines as part of the vaccination programme in Scotland, and COVID-19 hospitalisation.

2.2 Objectives
We seek to:
   a. Investigate the association between receiving one and two doses of the Pfizer-BioNTech, Oxford-Astra-Zeneca and Moderna vaccines as part of the vaccination programme in Scotland, and hospital admission within 28 days of an RT-PCR positive test for COVID-19, or admission with ICD-10 code for COVID-19 (see appendix).

3 Study Design

3.1 Study design
Open prospective cohort study. Data from multiple sources will be linked deterministically using the Community Health Index (CHI). CHI is a unique identifier used for all health contacts in Scotland.

3.2 Setting
Scotland.

3.3 Population
Individuals resident in Scotland (~5.4 million people).

3.4 Data sources
● Primary care data: Routinely collected records from 940 General Practices across Scotland containing information on demographics, clinical history, and vaccination status.
● Vaccination data: Data on vaccines administered in national vaccination centres through the Turas Vaccination Management Tool (TVMT).
● Secondary care data: Hospital admissions through the Scottish Morbidity Record (SMR) and Rapid Preliminary Inpatient Data (RAPID).

- Laboratory data: RT-PCR SARS-CoV-2 test data, available through the Electronic Communication of Surveillance in Scotland (ECOSS) database.

3.5 Inclusion/exclusion criteria

Exclusion criteria:
    RT-PCR test positive prior to index date
    Age 17 or less at the index date

After applying the exclusion criteria, the dataset consists of ~4.4 million people.

3.6 Sample size calculation
From our first paper on vaccine effectiveness against COVID-19 hospitalisation [1], we estimated VE for combined vaccine status at 28-34 days post vaccination as 0.84, with a standard deviation of 0.06. Assuming our VE estimates are asymptotically normally distributed, this gives virtually 100% power to detect a VE ≥ 0.5.

3.7 Type of study
We will interrogate our results to determine the degree to which they can be interpreted as predictive or causal.

4 Data and data validation

4.1 Data variables available
Table 1 lists the groupings of variables available for this study by data source. Exposure data are described in the Vaccinations category. Outcome data are described in the Secondary care and Laboratory tests categories. The rest of the categories contain data on potential confounding variables and effect modifiers.

Table 1: Data items/variables and data sources

Data category	Data item	Data source
Demographic	Sex	GP practice
	Age	GP practice
Socioeconomic	SIMD	GP practice
Other characteristics	Body Mass Index (BMI)	GP practice
	Smoking	GP practice
	Blood Pressure	GP practice
Geographic	Urban Rural Index (UR6), Health Board, council area	GP practice
Type of residence	Private housing, care home or social housing	GP practice
Clinical diagnoses	Underlying conditions (e.g., asthma, cardiac disease etc.)	GP practice
Vaccinations	Vaccine type	GP practice, TVMT
	Vaccine dose	GP practice, TVMT
	Vaccination date	GP practice, TVMT
	RT-PCR SARS-CoV-2 test result	ECOSS

Laboratory tests	Date of RT-PCR SARS-CoV-2 test	ECOSS
	Genome sequencing data	PHS
Secondary care	Hospital admission	SMR, RAPID
	Admission ICD code	SMR

Abbreviations: Scottish Index of Multiple Deprivation (SIMD), Body Mass Index (BMI), Electronic Communication of Surveillance in Scotland (ECOSS), Reverse-transcription polymerase chain reaction (RT-PCR); Turas Vaccination Management Tool (TVMT), Public Health Scotland (PHS), Scottish Morbidity Record (SMR), Rapid Preliminary Inpatient Data (RAPID), International Classification of Diseases (ICD).

4.2 Constructed variables
None.

4.3 Consistency and error checking
We will check for implausible values for all variables used in the analysis, and decide rules for cleaning them.

5   Statistical analyses

5.1   Objective a. Investigate the association between receiving one and two doses of the Pfizer-BioNTech, Oxford-Astra-Zeneca and Moderna vaccines and hospital admission within 28 days of an RT-PCR positive test for COVID-19, or admission with ICD-10 code for COVID-19.

5.1.1 Exposures of interest
Derived from date and type of vaccine received. Types are first/second doses of Pfizer-BioNTech, Oxford-Astra-Zzeneca or Moderna vaccines. Exposure categories are as follows:

1) 0-13 days after dose 1 or no vaccine record.
2) ≥14 days after dose 1 and before dose 2.
3) 0-6 days after dose 2;
4) ≥7 days after dose 2

5.1.2 Outcomes of interest
Time to hospital admission with a RT-PCR confirmed positive test for SARS-CoV2 in the 28 days prior to admission, or with ICD-10 code for COVID-19.

5.1.3 Potential confounders
Age, sex, council area, socio-economic status (SES) measured by quintiles of the Scottish Index of Multiple Deprivation (SIMD) (1 refers to most deprived and 5 refers to least deprived), residential settlement measured by the urban/rural 6 fold classification (1 refers to large urban areas and 6 refers to small remote rural areas), Health Board, comorbidities commonly associated with COVID-19 illness (asthma, chronic kidney disease, liver cirrhosis, chronic neurological condition, heart failure, diabetes (type 1 and type 2), dementia, coronary heart disease), risk factors (smoking status, blood pressure, body mass index), and nursing home residential status.

Behavioural variables could be important confounders. For example, people may have a tendency to change the degree to which they social distance after receiving the vaccine, or in response to government guidelines. We will not have access to these variables.

5.1.4 Potential effect modifiers
We will consider time intervals of different lengths, sex, age in bands and vaccine exposure cateogries as effect modifiers.

5.1.5 Analytical techniques
The parameters of a Cox proportional hazards model will be estimated using the corresponding Poisson regression adjusting for an offset representing the time at risk. The Cox models will be used to estimate the cumulative incidence of death in the vaccinated and unvaccinated groups. Vaccine effectiveness and 95% confidence intervals (CIs) will be calculated according to VE = (1-Rate Ratio)*100. We will carry out an analysis that uses the full cohort, and analyses that use subsets selected by matching procedures.

In the full cohort analysis, stratification variables for time intervals will be included in the model. A polynomial/spline will be fitted to the resulting discrete set of VE estimates. We will conduct hypothesis testing on the resulting vaccine effectiveness fit as a function of time. The reason for doing it this way as opposed to including a polynomial/spline in time since vaccination in the model is that this would require adding rows to the data for every individual and every time, which will make the dataset impractically large to work with when using the full cohort.

We will conduct a case control analysis using propensity scores, matching both with other individuals who were hospitalised on the same date, and also other individuals that tested positive with the same specimen date. For the matching procedure, individuals who have a COVID hospitalisation with no confirmatory test will be treated as if they had a positive test 21 days prior to hospitalisation. This will not affect many individuals. We will consider incorporating the following variables in the propensity score matching algorithm: age, sex and council area. We will use up to 10 controls per case.

We will also carry out a matched analysis among those who are vaccinated in the early part of the vaccination period using propensity score matching. We will consider incorporating the following variables in our propensity score matching algorithm: age, sex, geography, comorbidities, risk factors, SES, number of previous PCR tests, SES, presence in hospital prior to vaccination and urban-rural settlement. Vaccinated individuals will be matched with an individual who is unvaccinated at the time. Follow up will be to the earliest of: time of hospitalisation in either vaccinated or unvaccinated control, time of vaccination of unvaccinated control, or the terminal time period if neither of these events occur.

In the matched analyses, the model will include a categorical variable for vaccination status, multiplied by a polynomial/spline in time since vaccination. This is possible because the dataset that will be used is significantly smaller than in the full cohort.

We will also look at descriptive analysis of trends in hospitalisation rates and vaccination rates by age group.

5.1.6 Sub-group analysis
Subgroup analyses by vaccine type, age and sex will be considered.

5.1.7 Corrections for multiple testing
None.

5.1.8 Sensitivity analysis

We will consider exploring the use of different time intervals following administration of the vaccine to define exposure. We will report all exposure categories 0-7 and 7-14 days to understand the impact of any programmatic effect.

■　 5.1.10 Other analysis

We will consider conducting a falsification analyses (negative controls) using a fictional date of exposure that is two months prior to receiving the first vaccine dose in order to estimate the influence of variables that are not directly associated with immunological effects of the vaccine (e.g. behaviour changes associated with the rollout programme) on our estimates.

5.2 Missing data

Missing data will be reported as percentages of total or raw numbers. We will carry out a complete cases analysis.

5.3 Statistical software

All analyses will be carried out using R/RStudio, version 3.6.1.

6 Reporting results

6.1 Reporting guidelines and conventions

Results will be reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and REporting of studies Conducted using Observational Routinely-collected Data (RECORD) (via the COVID-19 extension) guidelines. 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. Statistical estimates will be reported with 95% confidence intervals.

1　 6.2 Dissemination

The analysis will be written in a manuscript and submitted to a peer reviewed journal. We will also seek to provide near real-time reports on vaccine safety, effectiveness and uptake for the various vaccines to the funders and government COVID-19 advisory bodies as appropriate. All code will be made publicly available via a GitHub repository.

7 References

1. Vasileiou E, Simpson CR, Robertson C, et al. Effectiveness of First Dose of COVID-19 Vaccines Against Hospital Admissions in Scotland: National Prospective Cohort Study of 5.4 Million People. Pre-print 2021 SSRN.

8 Appendix

○ **Table S1. ICD-10 codes**

Code	Description	Category
U07.1	COVID-19, virus identified	U07.1
U07.2	COVID-19, virus not identified	U07.2
Source: https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf		