EAVE II

Analysis Plan to investigate effectiveness of first dose and second dose of vaccine over time in adults

Full Project Title	EAVE II
Version Number	V1
Previous Versions	N/A
Effective Date	05 March 2021
Analyst(s)	Rachel Mulholland* (RM)
Co-authors	Srinivasa Vittal Katikireddi* (SVK), Ting Shi (TS), Eleftheria Vasileiou (EV), Steven Kerr (SK)
Approved	Chris Robertson (CR), Aziz Sheikh (AS)

*corresponding author

Version History			
Version	**Date**	**Author**	**Notes**
V1	05.03.2021	EV, SK	First version sent to team
V2	09.03.2021	SVK	Second version sent to team
V3	15.03.2021	AS	Approved

Contents

Introduction 2

Aims and Objectives 3
- 2.1 Aims 3
- 2.2 Objectives 3

Study Design 3
- 3.1 Study design 3
- 3.2 Setting 3
- 3.3 Population 4
- 3.4 Data sources 4
- 3.5 Inclusion/exclusion criteria 4
- 3.6 Sample size calculations 4

Data and Data Validation 4
- 4.1 Data variables available 4
- 4.2 Constructed variables 5
- 4.3 Consistency and error checking 5

Statistical analyses 5
- 5.1 Objective a) Assess the effectiveness of the vaccines **Error! Bookmark not defined.**
  - 5.1.1 Exposures of interest 5
  - 5.1.2 Outcomes of interest 6
  - 5.1.3 Potential confounders 6
  - 5.1.4 Potential effect modifiers 6
  - 5.1.5 Analytical techniques 6
  - 5.1.6 Sub-group analysis 7
  - 5.1.7 Corrections for multiple testing 7
  - 5.1.8 Sensitivity analysis 7
  - 5.1.9 Other analysis 8
- 5.2 Missing data 8
- 5.3 Statistical software 8

Reporting Results 8
- 6.1 Reporting guidelines and conventions 8
- 6.2 Dissemination 8

References 8

Appendix 9

1 Introduction

The aim of this study is to determine whether protection from COVID-19 vaccination wanes after the first and second SARS-CoV-2 vaccine doses. At present, the vaccination programme in Scotland uses two dose regimens of the Pfizer-BioNTech (BNT162b2) and the Oxford-AstraZeneca (ChAdOx1) vaccines. It will in due course also include the Moderna vaccine.

We theorise the vaccination programme as a natural experiment which has impacts that result from behavioural and other responses that occur as a consequence of broader programmatic effects – for example, being asked not to attend for vaccination if symptomatic and being reminded about protective behaviours to avoid infection – and the direct effects of the vaccine itself. Our evaluation studies the real-world impacts of the overall vaccination programme in Scotland. We will test the hypothesis that vaccine effectiveness (VE) estimates against COVID-19 hospitalisation and death decrease over time. To do this, we will use pseudonymised individual level linked routinely collected primary, laboratory and vaccination healthcare data across Scotland. All data and analyses will be hosted within Public Health Scotland or the Scottish National Safe Haven.

2 Aims and Objectives

2.1 Aims

To assess the relationship between time since receiving the first vaccine dose and second dose (of the Pfizer-BioNTech and Oxford-AZ vaccines and in due course Moderna vaccine) compared to unvaccinated and partially vaccinated people, and effectiveness of SARS-CoV-2 vaccines against COVID-19 hospitalisation and death.

2.2 Objectives

We seek to:

a. Estimate VE against the composite outcome of COVID-19 hospitalisation/death as a function of time since first SARS-CoV-2 vaccine dose
b. Estimate VE against the separate outcomes COVID-19 hospitalisation and death as a function of time since first COVID-19 vaccine dose
c. Assess if VE as a function of time since first vaccine dose differs by vaccine type, age group and sex
d. Estimate VE for first/second dose timing amongst different age and risk groups (to be completed when adequate numbers for statistical analysis)
e. Conduct statistical tests for reductions in VE over time against these outcomes.

3 Study Design

3.1 Study design

The primary study design is an open prospective cohort study with risk-set matching to emulate a ‘target trial’ comparing outcomes amongst first dose vaccinated (partially vaccinated) to people who have not yet been vaccinated [1-3]. Similarly, outcomes amongst second dose vaccinated (fully vaccinated) will be compared to two control groups, people who have not received any vaccine (unvaccinated) and people who have received only one vaccine (partially vaccinated). Analyses on waning of the second dose will be conducted when numbers allow. To maximise statistical power, we will conduct secondary analyses using the entire study population.

3.2 Setting

Scotland

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

3.4 Data sources
- Primary care data: General Practices (n=940) for information on demographics, other population characteristics and vaccination data.
- Vaccination centre data: Vaccines administered in national vaccination centres and data available via the Turas Vaccination Management Tool (TVMT).
- Secondary care data: Hospital admissions through the Scottish Morbidity Record (SMR) and Rapid Preliminary Inpatient Data (RAPID).
- Laboratory test data: RT-PCR laboratory confirmed SARS-CoV-2 infection and data available via the Electronic Communication of Surveillance in Scotland (ECOSS) database.

3.5 Inclusion/exclusion criteria

Exclusion criteria:

- Age 17 or less at the study start date (8 December 2020)
- Deceased prior to the study start date.

3.6 Sample size calculations
In this study, we are providing sample size calculations based on the Scottish testing and vaccination data due to being the largest UK nation with national coverage data.
From our first paper on VE against COVID-19 hospitalisation [4], 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 of ≥ 0.5. There is a possibility the study will be underpowered for estimating the association between receiving the vaccine and the secondary outcome of death.

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, mortality and laboratory tests categories. The rest of categories contain data on potential confounding factors and effect modifiers.

Table 1: Data items/variables and data sources

Data category	Data item	Data source
Demographic	Sex	GP
	Age	GP
Socioeconomic	SIMD	GP
Other characteristics	Body Mass Index (BMI)	GP
	Smoking	GP
	Blood Pressure	GP
Geographic	Urban Rural Index (UR6); Health Board	GP
Type of residence	Private housing, care home or social housing	GP
Clinical diagnoses	Underlying conditions (e.g., asthma, cardiac disease, immunosuppression etc.)	GP
Vaccinations	Vaccine type	GP, TVMT

	Vaccine dose	GP, TVMT
	Vaccination date	GP, TVMT
Laboratory tests	RT-PCR positive SARS-CoV-2	ECOSS
	RT-PCR negative SARS-CoV-2	ECOSS
	Date of RT-PCR test	ECOSS
Secondary care	Hospital admissions	RAPID/SMR
Mortality	Death with COVID-19 as the main cause according to death certificate, or death within 28 days of a positive RT-PCR test for COVID-19	NRS
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); Scottish Morbidity Record (SMR); National Records of Scotland (NRS)		

4.2 Constructed variables

None

4.3 Consistency and error checking

We will check for patterns of missingness and implausible values (e.g. date of second vaccine dose being earlier than the first) for all analytical variables being used, with a record maintained of reasons for exclusion of any records from analysis. In the case where a variable of interest has high levels of missingness, we will consider using alternative variables that are closely related as a proxy for these missing data.

5 Statistical Analyses

5.1.1 Exposures of interest

For the first vaccine dose (partial vaccination), an individual will be defined as exposed from day 14 after receiving the first vaccine dose between the period of 8 December 2020 and until the end of follow up (currently 5 March 2021). For the primary analysis, controls who have not yet received a COVID-19 vaccine will be matched 1:1 on the basis of sociodemographic characteristics, risk factors and prior health status. Follow-up of the exposure period will be censored for both the recipient and the control participant if the control becomes vaccinated (with any vaccine i.e. including the Moderna vaccine) or if the recipient receives a second (booster) vaccination. Controls who become vaccinated will then be eligible to be included in the analysis within the intervention group. Analyses will be stratified by vaccine type i.e. Pfizer-BioNTech and Oxford-AstraZeneca vaccine (and in due course Moderna vaccine).

For the second vaccine dose (full vaccination), two comparators will be used. An individual will be defined as exposed from day 7 after receiving the second vaccine dose during the study period. For the analysis comparing effectiveness of full vaccination against no vaccination, controls who have not yet received a COVID-19 vaccine will be matched on the basis of sociodemographic characteristics, risk factors and prior health status. For the analysis comparing effectiveness of full vaccination against partial vaccination, controls who have only received one vaccine will be matched. The exact matching ratio will be determined based on the ratio of available exposed to control people within the dataset. Follow-up of the exposure period will be censored for both the recipient and the control participant if the control goes on to meet the criteria for being classified as exposed (i.e. receiving a first dose for the unvaccinated control group and receiving a second dose for the partially vaccinated control group). Fully vaccinated controls who become vaccinated will then be eligible to be included in the analysis within the intervention group. Analyses will be stratified by vaccine type and length of time between vaccine doses.

To increase statistical power in secondary analyses, we will conduct analyses of whole population data. Exposure status will be defined as a time-varying exposure, with individuals being defined as exposed from 14 days after vaccination for first doses and 7 days for

second doses. Again, we expect part of the effects of the vaccination programme to arise from behavioural responses (e.g. behavioural advice received with the invitation letter).

5.1.2 Outcomes of interest

The primary outcome will be a composite outcome of time to COVID-19 hospitalisation or death. COVID-19 hospitalisation will be defined as a RT-PCR confirmed positive test for SARS-CoV-2 in the 28 days prior to admission, or with ICD-10 code for COVID-19 (in any diagnostic position). COVID-19 deaths will be defined as COVID-19 as the underlying ICD-10 cause of death recorded on the death certificate (see Appendix), or death from any cause within 28 days of a positive RT-PCR test for SARS-CoV-2 infection.

Secondary outcomes will be the single outcomes of: a) COVID-19 hospitalisation (as defined above), b) COVID-19 deaths, and c) RT-PCR confirmed positive test. We anticipate the RT-PCR confirmed SARS-CoV-2 infection results to be more susceptible to bias arising from differential ascertainment and therefore anticipate treating these results as exploratory.

5.1.3 Potential confounders

Age, sex, 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), household size, number and types of 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 Health Board. We will also include nursing home status as a potential confounder when data are available.

5.1.4 Potential effect modifiers

Stratification into different population groups by age group (18-64, 65-79, 80+ years) and sex will be performed. Given that VE may differ amongst previously infected individuals (since the first dose may in effect be considered a booster dose), we will stratify analyses by previous confirmed infection status when adequate numbers are available for meaningful statistical analysis. As noted above, we will also stratify analyses by vaccine type.

5.1.5 Analytical techniques

We will commence analysis by conducting descriptive analyses to visually inspect trends in vaccination delivery, age-specific COVID-19 hospitalisations and COVID-19 deaths, including by age group and sex. This will include inspecting the number of people who have received no doses, one dose and two doses and the length of time between the receipt of one dose and two doses.

To create control groups for our primary analysis, we will emulate a ‘target trial’. We will do this by conducting 1:1 risk-set matching to identify individuals who had not yet become exposed (i.e. remained unvaccinated) on the date an exposed person received their first vaccine dose. We will do this using a propensity score matching algorithm and will consider incorporating the following characteristics: sex, age, geography, comorbidities, risk factors, number of previous SARS-CoV-2 tests, SES, presence in hospital pre-vaccination and urban-rural settlement. The adequacy of the matching will be assessed by checking for imbalance of the individual covariates across exposure groups.

We expect that any effects before 14 days for first dose vaccination will largely reflect programmatic effects (e.g. being instructed not to attend vaccination if symptomatic and reinforcement of behavioural advice to reduce COVID-19 related risks). Follow-up will therefore start from day 14 after the date of first dose vaccination (and day 7 after second

dose) for both the exposed and control person. It will end on the first of: experiencing the outcome of interest, death (from any cause) or end of follow up period. Unvaccinated controls who become vaccinated will be eligible to become considered as exposed, after follow up is censored for the exposed-unexposed pair at the date of vaccination of the control.

For the matched cohort analysis, we will draw Kaplan-Meier curves to inspect cumulative incidence. We will then fit Cox proportional hazards models and conduct Poisson regression for our outcomes of interest. Given we are particularly interested in the potential for vaccine waning, we will fit a flexible model over time (using a fractional polynomial or smooth spline with person-time defined in days) to explore the timing of any waning of vaccine effectiveness.

Given the relatively small number of anticipated COVID-19 deaths during the follow up period, the secondary analysis focusing on mortality will be based on data from the whole cohort. In addition, secondary analyses will be conducted for hospitalisation and having confirmed SARS-CoV-2 infection.

Vaccination status (unvaccinated, one dose and two doses) will be defined as a time-varying exposure. Poisson regression adjusted for an offset representing the time at risk, with rate ratios (RRs) and 95% confidence intervals (CIs) will be calculated. Vaccine effectiveness and 95% Confidence Intervals (CIs) will be calculated as (1-Rate Ratio)*100. Models will be adjusted for relevant confounders, including age, sex, SEP, geography, time period and comorbidities. Stratification variables for each week post-vaccination will be included in the Poisson regression. A polynomial/spline will be fitted to the resulting discrete set of weekly 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 Poisson regression is that this would require adding rows to the data for every individual and every time, which would make the dataset impractically large to work with when using the full cohort.

We will assess vaccine waning using two different approaches: first, looking for statistical evidence of reducing effectiveness; and second, by assessing whether effectiveness achieves a minimum acceptable level. For the former, we will assess the curvature in the polynomial/spline weekly effect on the VE estimates, to evaluate whether the VE estimates peak and return to null. For the latter, we will adopt the US Food & Drug Administration's threshold of achieving a minimum VE of 50% for the point estimate [5, 6].

5.1.6 Sub-group analysis

Subgroup analyses by vaccine type, age group and sex will be performed. We will consider conducting sub-group analysis by time period too, especially if there is evidence of different circulating variants predominating.

5.1.7 Corrections for multiple testing

N/A

5.1.8 Sensitivity analysis

We will consider exploring the impact of alternative approaches to classifying the start of the exposed period (from day 0 of vaccination for both first and second doses, and day 21 after first dose), classifying COVID-19 hospitalisation on the basis of primary diagnosis (rather than any diagnostic position) and using a doubly robust estimator (i.e. including variables for confounder adjustment within regression models that already incorporate a propensity score).

5.1.9 Other analysis
We will consider conducting falsification analyses (negative controls) for alternative time periods (e.g. repeating analyses using time periods two months prior to first vaccination dose) to check the comparability of our exposure groups.

5.2 Missing data
Missing data will be reported as percentages of total or raw numbers where possible. Previous analyses have demonstrated that little missing data exist for our key variables of interest. For covariates which may have a higher proportion of missing data (such as body mass index), we will either use records with no item missingness or use a missing category.

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. Measures of association will be reported with 95% confidence intervals.

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] Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. J Clin Epidemiol. 2018;96:12-22.

[2] Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.

[3] Watson D, Spaulding AB, Dreyfus J. Risk-Set Matching to Assess the Impact of Hospital-Acquired Bloodstream Infections. Am J Epidemiol. 2018;188(2):461-6.

[4] Effectiveness of First Dose of COVID-19 Vaccines Against Hospital Admissions in Scotland: National Prospective Cohort Study of 5.4 Million People https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3789264

[5] US Food and Drug Administration. Development and Licensure of Vaccines to Prevent COVID-19. 2020.

[6] Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KRW, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. The Lancet Infectious Diseases. 2021;21(2):e26-e35.

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		