Analysis plan to investigate potential waning of COVID-19 vaccine protection after the completion of a primary immunisation schedule

Full Project Title	
Version Number	V1
Previous Versions	
Effective Date	25 August 2021
Analyst(s)	Chris Robertson (CR), Thiago Cerqueira-Silva (TCS), Renzo Ortiz (RO)
Co-authors	Vinicius de Araujo Oliveira (VAO), Manoel Barral-Netto (MBN), Mauricio Barreto (MB), Srinivasa Vittal Katikireddi (SVK), Aziz Sheikh (AS), Guilherme Werneck (GW)
Approved	

Version History			
Version	**Date**	**Author**	**Notes**
V1	25/8/2021	SVK, TCS, CR	
V2			
V3			

Contents

1 Introduction

The aim of this study is to determine whether protection from COVID-19 vaccination wanes after the completion of the primary licensed schedule of SARS-CoV-2 vaccines. If possible, we plan to conduct harmonised analyses across different contexts to strengthen confidence that findings are not driven by trends in variants, increase statistical power, and allow robustness of findings when adjusting for differing confounders to be assessed. A collaboration currently allows analyses to be conducted in Scotland and Brazil.

At present, the vaccination programme in Scotland uses two dose regimens of the Pfizer-BioNTech (BNT162b2), Oxford-AstraZeneca (ChAdOx1), and Moderna (mRNA-1273) vaccines.

The vaccination programme in Brazil currently offers the following vaccines:

- Oxford-AstraZeneca – primary dose followed by booster at 12 weeks in Brazil, but the initial trials delivered a booster at 3-4 weeks
- Sinovac CoronaVac/Butantan – booster delivered after three weeks
- Pfizer/BioNTech – booster delivered after 12 weeks in Brazil, with the original trials delivering a booster at 3 weeks
- Janssen – licensed as a single dose.

Our evaluation studies the real-world impacts of being fully vaccinated and assesses whether vaccine effectiveness (VE) is maintained over time. To do this, we will use deidentified individual level linked data. All data and analyses will be hosted within secure analytical environments.

2 Aims and Objectives

2.1 Aims

To assess the relationship between time since the expected onset of maximum protection following completion of a vaccination schedule (14 days post-second dose for two dose regimens) and the risk of severe COVID-19 sequelae (hospitalisation or death). We will also assess the secondary outcomes of confirmed symptomatic COVID-19 disease and intensive care admission.

2.2 Objectives

By vaccine type, we seek to:

a. Estimate VE against the composite outcome of COVID-19 hospitalisation/death as a function of time since 14 days post completion of a COVID-19 vaccination schedule
b. Estimate VE against the separate outcomes of i) COVID-19 hospitalisation; ii) COVID-19 death; iii) confirmed symptomatic COVID-19 disease; and iv) intensive care admission as a function of time since 14 days post completion of a COVID-19 vaccination schedule
c. Assess if waning of VE differs by age group, sex and being in a high-risk COVID-19 group
d. 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 comparing outcomes amongst people who are vaccinated for differing lengths of time. We will consider pursing a test-

negative design case-control study as a secondary study design to estimate the association between length of time after being fully vaccinated and the odds of testing positive.

3.2 Settings

Scotland and Brazil.

3.3 Population

Our main population of interest is adults who have completed a full COVID-19 vaccination schedule (typically two doses) resident in the country.

Our secondary population of interest is all adults who have received either one or two doses of any eligible vaccine. This allows for the risk prior to gaining protection from vaccination to be estimated and therefore allows vaccine effectiveness to be estimated. However, this period is potentially subject to greater time-sensitive behavioural changes affecting COVID-19 risks and may therefore be more susceptible to bias than the main analytical cohort.

For the test-negative design, we will study all fully vaccinated adults who have received a COVID-19 vaccination test. We will restrict analyses to the first COVID-19 test during the follow-up period.

3.4 Data sources

In Scotland, we will draw on the following databases:

- 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.
- Mortality data from National Records Scotland (NRS)

Equivalent datasets will be used for other UK/international countries. In Brazil, we will analyse deterministically linked data provided by the Ministry of Health:

- COVID-19 vaccination (SI-PNI): Dates and types of vaccination administered, primary reason for vaccination, indigenous status, pregnancy status.
- Acute Respiratory Infection Suspected Cases (eSUS-Notifica): register of suspected and confirmed COVID-19 cases
- Severe Acute Respiratory Infection/Illness (SIVEP-Gripe): contains all hospitalisations and deaths due to severe acute respiratory illnesses, including COVID-19.

The above three databases all have whole population coverage, with returns from both public and private healthcare systems. However, data quality may vary geographically (typically data quality is better in the south of Brazil). Testing capacity has also improved over the course of the pandemic, but most improvements occurred prior to the vaccination programmes starting.

In addition to the above datasets, it may be possible in the future to obtain linkage to previous general hospital records for patients treated within the public healthcare system (covering approximately 70% of hospitalisations).

3.5 Inclusion/exclusion criteria

Exclusion criteria for the primary cohort:

- Not fully vaccinated
- Inconsistent vaccination records (i.e. received a second dose with no record of a first dose, received doses from different vaccine types or interval between doses of less than 14 days)
- Previous confirmed infection
- Missing essential covariates (sex, age, municipality)
- Children (under the age of 18 years)

For the secondary cohort, the same exclusion criteria apply except for the first criterion. Instead, we will exclude anyone who has not received any doses of a COVID-19 vaccine.

3.6 Sample size calculations

In Scotland, there have been approximately 3.6M fully vaccinated people to date and 1.5M were vaccinated in May 2021, which should allow up to four months follow-up to assess waning.

Based on a preliminary review of descriptive statistics in Brazil, we anticipate the following numbers of fully vaccinated people will be available for analysis:

- Oxford-AstraZeneca: 863k person-years of follow-up, with 2,390 hospitalisations
- Sinovac CoronaVac/Butantan: 5.3M person-years of follow-up, with 35,049 hospitalisations
- Pfizer-BioNTech: 14M individuals but little follow-up since most people only became fully vaccinated within the two months
- Janssen: 4M individuals but little follow-up since most people only became fully vaccinated within the last two months

If there are not adequate numbers available for statistical analysis for a specific vaccine, we will not progress that analysis and censor individuals who receive that vaccine.

4 Data and Data Validation

4.1 Data variables available

Tables 1 and 2 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
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)		

Table 2: Data items/variables and data sources in Brazil

Data category	Data item	Data source
Demographic	Sex	SI-PNI
	Age	SI-PNI
Socioeconomic	Municipality-level IBP	SI-PNI
	Education	(SIVEP-Gripe)
Race/Ethnicity	Race	SI-PNI (eSUS, SIVEP-Gripe)
Geographic	Region	SI-PNI
	Vaccine type	SI-PNI
Vaccinations	Vaccine dose	SI-PNI
	Vaccination date	SI-PNI
	Primary reason for vaccination: 20 categories	SI-PNI
	Date of RT-PCR/rapid antigen test	eSUS
Laboratory tests	RT-PCR/rapid antigen positive SARS-CoV-2	eSUS
	RT-PCR/rapid antigen negative SARS-CoV-2	eSUS
	Location of test	eSUS
	COVID-19 hospital admission date	SIVEP-Gripe
Secondary care Mortality	COVID-19 intensive care admission date	SIVEP-Gripe
	COVID-19 Death	SIVEP-Gripe
	Need for ventilatory support	SIVEP-Gripe
	Comorbidities	SIVEP-Gripe
	Type of hospital (public vs private)	SIVEP-Gripe
	Number of hospital beds	SIVEP-Gripe
	Number of intensive care beds	SIVEP-Gripe

4.2 Constructed variables

We will create variables for number of PCR tests prior to the second dose of vaccination and having previously had a positive test before vaccination (used for exclusion).

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

Our primary exposure of interest is the length of time since a high level of protection is expected following completion of a vaccination schedule. An individual will enter the study 14 days after becoming fully vaccinated. We will assess the relationship between the number of fortnights since study start and the risk of the outcomes of interest (see below).

Our secondary exposure of interest is vaccination status over time, including length of time since becoming fully vaccinated. This further stage of analysis is required to produce vaccine effectiveness estimates, thereby requiring an estimate of risk during a period when unprotected. Using the cohort of both partially and fully vaccinated people, exposure periods will be classified as time-varying as follows: the control period (0-13 days after first dose), first dose protection (14 days until the receipt of a second dose), second dose control period (0-13 days after second dose) and then fortnightly periods thereafter.

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.

Secondary outcomes will be the single outcomes of: i) COVID-19 hospitalisation, ii) COVID-19 deaths, iii) intensive care admission and iv) RT-PCR confirmed positive test, provided numbers of events allow. 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 more tentative.

5.1.3 Potential confounders

We will consider adjusting for age, sex, socioeconomic position, geography and where possible, comorbidities. More specifically, we anticipate adjusting for the following confounders in Scotland: age (five year bands), sex, socioeconomic position (SEP) 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 (BMI)) and Health Board.

In Brazil, we anticipate adjusting for age, sex, SEP measured by deciles of the Indice Brasileiro de Privacao (IBP) at the municipality level, urban-rural municipality classification, primary

reason for vaccination, race/ethnicity and calendar week will be accounted for within the statistical models.

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 between those at high-risk of serious COVID-19, we will stratify analyses by COVID-19 risk group when data allow. We will also consider stratifying analyses by vaccine type and calendar time (to reflect different dominant variants present in the country).

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 two doses and the length of follow-up available following the second dose by vaccine type.

For the cohort study design, we will fit Poisson regression models with rate ratios (RRs) and 95% confidence intervals (CIs) calculated. If we observe overdispersion, we will consider using Negative Binomial regression models or Poisson with robust standard errors instead. Models will be adjusted for the relevant confounders above, including a set of dummy variables for calendar week. Vaccine effectiveness and 95% Confidence Intervals (CIs) will be calculated as (1-Hazard Ratio)*100.

For the test-negative case-control design, we will start by defining cases as the first occasion during the period of follow-up any individual has a positive SARS-CoV-2 test. We will randomly sample controls testing negative with a date of symptom onset 0-10 days before the test matching on the basis of calendar week, age (five year bands), sex, municipality and primary reason for vaccination. We will then establish the number of weeks post-vaccination for the cases and controls, calculating odds ratios using conditional logistic regression. We will additionally adjust for race/ethnicity. We will also construct a propensity score to be tested among those in the primary cohort and use this as an inverse propensity weight

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 do a trend test on the period post-vaccination from 14-27 days onwards. We will also assess an exploratory change point from 28+, 42+, 56+ days to do a linear trend test post-vaccination. For the latter, we will adopt the US Food & Drug Administration's threshold of achieving a minimum VE of 50% for the point estimate [1, 2].

5.1.6 Sub-group analysis

Subgroup analyses by broad age group (18-64, 65-79, 80+ years), sex and being in a severe COVID-19 risk group. 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

Following previous epidemiological recommendations, we will not correct for multiple testing [3].

5.1.8 Sensitivity analysis

We will consider exploring the impact of alternative approaches to classifying the start of the exposed period (e.g. from day 7 after completion of the vaccination schedule), classifying COVID-19 hospitalisation on the basis of primary diagnosis (rather than any diagnostic

position) and statistical adjustment for people with previous COVID-19 infection (rather than excluding them).

In Brazil, we will additionally consider repeating our main analysis for a cohort of healthcare workers (defined on the basis of primary reason for vaccination) since that provides a cohort with a broader age range and ascertainment of outcomes is likely to be more reliable. We will consider exploring the impact of alternative approaches to classifying the start of the exposed period (from day 7 after completion of the vaccination schedule) and restricting analyses to the regions with the most reliable data (i.e. excluding the Northern regions of Brazil).

5.1.9 Other analysis

We will consider conducting falsification analyses (negative controls) for alternative outcomes (e.g. non-COVID hospitalisations) if data allow.

5.2 Missing data

Missing data will be reported as percentages of total or raw numbers where possible. For covariates which may have a higher proportion of missing data, 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.

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) guidelines. Measures of association will be interpreted based on 95% confidence intervals.

6.2 Dissemination

This statistical analysis plan will be made publicly available on a website prior to conducting the main statistical analyses. The results will be written in a manuscript and submitted to a peer-reviewed journal. We will also seek to provide feedback to relevant COVID-19 advisory bodies (e.g. UK Joint Committee on Vaccination & Immunisation, Brazilian Ministry of Health, WHO) as appropriate. All code will be made publicly available via a GitHub repository.

7 References

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

[2] 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.

[3] Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1(1):43 46.

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		