VERSION HISTORY

V1	26/10/2021	**Internal development** VW, RD, JS developed from post-covid-unvaccinated protocol with comments from YW, AMW, RT and GC.
V2	04/11/2021	**Circulation to wider project team**

AUTHORS

CONVALESCENCE Study Team

TITLE

Understanding the risk of adverse health events following severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in the era of delta among the fully vaccinated and the electively unvaccinated

LAY SUMMARY

To be completed

BACKGROUND

To be completed

RESEARCH HYPOTHESES

In the era of the delta variant of SARS-CoV-2, rates of adverse health events differ with and without SARS-COV-2 infection.

RESEARCH QUESTIONS

1. Among vaccinated individuals in the era of the delta variant of SARS-CoV-2, are there higher rates (expressed as hazard ratios with time since SARS-CoV-2 infection) of an incident outcome in those with and without SARS-CoV-2 infection, before and after adjustment for potential confounders?

2. Among electively unvaccinated individuals (i.e., individuals eligible for vaccination that have chosen not to receive it) in the era of the delta variant of SARS-CoV-2, are there higher rates (expressed as hazard ratios with time since SARS-CoV-2 infection) of an incident outcome in those with and without SARS-CoV-2 infection, before and after adjustment for potential confounders?

DATA SOURCES

This research will be conducted using OpenSafely and requires the following data sources:

- Primary care data (TPP)
- Secondary Use Services (SUS) hospital data

- Second Generation Surveillance System (SGSS) for Pillar 1 and Pillar 2 SARS-COV-2 infection laboratory testing data
- Hospital Episode Statistics Admitted Patient Care (HES APC)
- Office of National Statistics (ONS) death registry
- Intensive Care National Audit & Research Centre (ICNARC)

KEY DATES

The study will start on 01-06-2021 (i.e., study start date), which is the date that delta was thought to be ubiquitous, and end on the last date of data collection.

STUDY POPULATION

Patients will be included if they meet ALL the following criteria:

- Alive on the first day of follow-up
- Known age between 18 and 110 inclusive on the first day of follow-up
- Known sex
- Known deprivation
- Registered in an English GP with TPP software for at least 6 months prior to the study start date

Patients will be excluded if they meet ANY the following criteria:

- SARS-CoV-2 infection recorded prior to the start of follow-up [Note: these individuals are required for a sensitivity analysis and so should not be removed at the data extraction stage]
- For the vaccinated analysis, patients will be excluded if they do not have a record of two vaccination doses received more than 3 weeks apart prior to the study end date
- For the unvaccinated analysis, patients will be excluded if they have a record of one or more vaccination doses at the start of follow-up

FOLLOW-UP

Vaccinated individuals

Follow-up will start at the latest of the following dates:

- Two weeks after their second vaccination
- Study start date

Follow-up will end at the earliest of the following dates:

- Death
- Outcome event (see: Outcomes)
- Study end date

Unvaccinated individuals

Follow-up will start at the latest of the following dates:

- 12 weeks after they became eligible for vaccination (see: https://docs.google.com/spreadsheets/d/1Epre2Cv_4UVTwHJ6ccJN7QwDRq9pGGyQoWJoWolODZQ/edit?usp=sharing)
- The start of the delta pandemic era

Follow-up for all individuals will end at the earliest of the following dates:

- Vaccination
- Death
- Outcome event (see: Outcomes)
- Last date of data collection

EXPOSURES

SARS-COV-2 infection

Exposure to SARS-COV-2 infection will be defined as the first date of a confirmed COVID event post index date. Exposures can be recorded in any of the following data sources:

Data source	Definition
SGSS	Date of positive SARS-COV-2 PCR antigen test
Primary care	Date of confirmed diagnosis code
HES APC	Start date of episode with confirmed diagnosis in any position
SUS	Start date of episode with confirmed diagnosis in any position
ICNARC	Start date of episode with confirmed diagnosis in any position
ONS death registry	Date of death with SARS-COV-2 infection listed as primary or underlying cause

SARS-COV-2 infection severity

Individuals with a hospital admission record that includes confirmed SARS-CoV-2 infection in the primary position within 28 days of first SARS-CoV-2 infection will be defined as ‘SARS-CoV-2 infection with hospitalisation’. All other individuals will be defined as ‘SARS-CoV-2 infection without hospitalisation’.

OUTCOMES

Each outcome will be defined as the first event occurring within the follow-up period. Outcomes can be recorded in any of the following data sources:

Data source	Definition
Primary care	Date of diagnosis or prescription code
HES APC	Start date of episode with confirmed diagnosis in any position
SUS	Start date of episode with confirmed diagnosis in any position
ONS death registry	Date of death with SARS-CoV-2 infection listed as primary or underlying cause

The events of interest are:

Category	Event
Cardiovascular	Acute myocardial infarction
Cardiovascular	Ischaemic stroke
Cardiovascular	Pulmonary embolism
Cardiovascular	Deep vein thrombosis

Cardiovascular	Transient ischaemic attack
Cardiovascular	Subarachnoid haemorrhage and haemorrhagic stroke
Cardiovascular	Heart failure
Cardiovascular	Angina
Cardiovascular	Arterial thrombosis events
Cardiovascular	Venous thromboembolism events

Outcomes for the categories respiratory, renal, mental health, gastrointestinal, and diabetes are still under discussion and will be added shortly.

POTENTIAL CONFOUNDERS

We will consider the following potential confounders, which will be defined using the most recent data prior to index date:

Confounder	Type	Definition	Data sources
Sex	Categorical	Male, Female	Primary care
Age	Continuous	Modelled as age in years using a restricted cubic spline with 3 knots at the 10th, 50th and 90th percentiles	Use OpenSafely command 'patients.age_as_of(…)'
Ethnicity	Categorical	Black or Black British, Asian or Asian British, Other Ethnic Groups, Mixed, Missing or Unknown, White	Use OpenSafely command 'patients.categorised_as(helpers.generate_ethnicity_dictionary(16),…)'
Deprivation	Continuous	Index of Multiple Deprivation 2019	Use OpenSafely command 'patients.categorised_as(helpers.generate_deprivation_ntile_dictionary(10),…)'
Region	Categorical	East of England, London, Midlands, North East and Yorkshire, North West, South East, South West, Scotland, Wales	Primary care
Consultation rate	Continuous	Number of primary care contacts in the year prior to index date	Use OpenSafely definition from replication, i.e., 'cov_n_disorder=patients.with_gp_consultations(between=["index_date - 12 months", "index_date"], returning="number_of_matches_in_period",return_expectations={"int": {"distribution": "normal", "mean": 10, "stddev": 3},"incidence": 1, },)'
Number of regular medications	Categorical	To be confirmed	To be confirmed
Smoking status	Categorial	Current, former, never	Primary care
Obesity	Binary	1 if BMI>=30 or coded	Primary care, HES APC

		diagnosis for obesity; 0 otherwise	
Ever acute myocardial infarction	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever stroke	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever other arterial embolism	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever venous thromboembolism events	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever heart failure	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever angina	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever dementia	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever liver disease	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever chronic kidney disease	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever cancer	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever hypertension	Binary	1 if diagnosis or prescription present; 0 otherwise	Primary care, HES APC
Ever diabetes	Binary	1 if diagnosis or prescription present; 0 otherwise	Primary care, HES APC
Ever depression	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Ever chronic obstructive pulmonary disease	Binary	1 if diagnosis present; 0 otherwise	Primary care, HES APC
Lipid lowering medications	Binary	1 if prescription present; 0 otherwise	Primary care
Antiplatelet medications	Binary	1 if prescription present; 0 otherwise	Primary care
Anticoagulation medications	Binary	1 if prescription present; 0 otherwise	Primary care
Combined oral contraceptive pill	Binary	1 if prescription present; 0 otherwise	Primary care
Hormone replacement therapy	Binary	1 if prescription present; 0 otherwise	Primary care

CODELISTS

See: https://docs.google.com/spreadsheets/d/1xyIMCQkTsMUGbh-11f6kvR8kjSYZ9pq-nE8fqKEhEh0/edit?usp=sharing

MAIN ANALYSES

Initial descriptive statistics will be used to describe the demographic and clinical characteristics of the baseline cohort, overall and for the subgroups hospitalised and non-hospitalised with SARS-CoV-2 infection.

We will split follow up time for each person into periods before and after SARS-CoV-2 infection, and into time periods since diagnosis defined in days (time periods: [0,14), [14,28), [28,56), [56,84), [84,182), [182,365); see: Appendix 1: splitting follow-up time). We will tabulate numbers of outcome events (see: Outcomes), person-years of follow-up and rates of events before and with time since exposure. If any of these time periods contains no events, we will collapse the time periods after SARS-CoV-2 infection into [0,28) and [28,365) prior to analysis.

We will fit Cox regression models with calendar time scale using the start of study date as the origin. This will ensure that all analyses account for changes with calendar time in rates of the outcome event. Using this approach, we will estimate hazard ratios for events of different types before and after exposure, and by time since exposure.

For computational efficiency, Cox models will be fitted to datasets including: all people with the outcome event and a random subset of people without the outcome event equal to ten times the number of people with the outcome event. Analyses will incorporate inverse probability weights (e.g., for a N% sample of the people without the outcome, weight = N) for data from people without the outcome event. Confidence intervals will be derived using robust standard errors.

Potential confounders (see: Potential confounders) will be based on data recorded on or before the start of follow-up in each analysis. We will exclude potential confounders from any analysis when there are fewer than 3 disease events at any level. All models will be stratified by region so that risk sets are constructed within region, hence accounting for between-region variation in the baseline hazard.

We will estimate: (i) age, sex and region adjusted; and (ii) maximally adjusted HRs. We will exclude potential confounders with ≤2 disease events at any level. We will construct a propensity score that combines all the covariates into a single metric and adjust for this (using a restricted cubic spline), in addition to individual covariates to obtain maximally adjusted HRs. We will examine the fit of the restricted cubic splines used for age and propensity score.

We will analyse outcomes for which there are at least 400 events after SARS-CoV-2 infection. We will apply the same criterion to subgroup analyses.

For each exposure-outcome combination, we will also estimate the absolute excess risk after SARS-CoV-2 infection. To do this, we will calculate the average daily incidence of each outcome over time before or in the absence of SARS-CoV-2 infection across the whole follow up period, separately in subgroups defined by age and sex. We will multiply these by the maximally adjusted age- and sex-specific HR for that day to derive the incidence on each day after SARS-CoV-2 infection. Using a life table approach, we will then calculate age- and sex-specific cumulative risks over time with and without SARS-CoV-2 infection, subtracting the latter from the former to derive the absolute excess risks over time after SARS-CoV-2 infection, compared with no SARS-CoV-2 infection. Overall absolute excess risk will be estimated from a weighted sum of the age- and sex-specific excess risks, weighted by the proportions of individuals in age and sex strata within the SARS-CoV-2 infected population during the follow-up period.

Proposed outputs for this project are included as Appendix 2: proposed outputs.

SENSITIVITY ANALYSES

Subgroup analyses

We will repeat the main analysis to estimate stratified post-exposure hazard ratios as detailed below:

- Subgroups according to severity (hospitalised / non-hospitalised) and, where appropriate, by use of intensive care
- Subgroups according to age group (18-39 / 40-59 / 60-79 / 80-110)
- Subgroups according to sex (male / female)
- Subgroups according to ethnicity (White / Asian or Asian British / Black or Black British / Mixed / Other Ethnic Groups)
- Subgroups according to prior history of outcome (prior history of outcome / no prior history of outcome)

Prior infection analysis

We will repeat the main analysis in individuals who had a COVID infection prior to the start of the study.

MISSING DATA

Individuals with missing age, sex, or deprivation are excluded from the analysis by the study definition. We will include a missing category for ethnicity. All other covariates are defined using the presence versus absence of specific codes in the EHRs, so have no identifiable missing values. We will not use multiple imputation.

APPENDIX 1: SPLITTING FOLLOW-UP TIME

Consider the following definitions:

- Time scale: days since start of study
  - Outcome of interest: time to event D measured at T_D with indicator I_D in days
  - Exposure of interest: binary exposure E measured at T_E with indicator I_E, parameterised as days since T_E. This will be categorised, for example, into: E1 = [0,14); E2 = [14,28); E3 = [28,56); E4 = [56,84); E5 = [84,182); E6 = [182,365), …
  - Administrative censoring time: set as day T_C

For individuals without exposure and without event then T_D=T_C, I_D = 0, T_E=T_C, I_E=0 (e.g., individual 1 in table below)

For individuals without exposure and with event at time *t* then T_D=*t*, I_D = 1, T_E=*t*, I_E=0 (e.g., individual 2 in table below)

For individuals with exposure at T_E and without event then: (1) split follow-up time at T_E, and (2) split follow-up time>T_E at T_E+14; T_E+28; T_E+42; T_E+56 and then censor at earliest of T_E+70 or T_C (e.g., individual 3 in table below)

For individuals with exposure at T_E and event at T_D, then first (1) split follow-up time at T_E, and then (2) split follow-up time>T_E at T_E+14; T_E+28; T_E+42; T_E+56 and then censor at earliest of T_E+70 or T_D (e.g., individual 4 in table below)

In the following example, T_C = 300 days. T0 and T1 represent the start and finish of the time period (for instance, the first row for a person is their pre-exposure time period so T0 represents the start of the study and T1 the time of exposure). E1-E5 are indicator variables for the time period post infection (for instance, as before, the first row for a person is their pre-exposure time period so all indicators are zero and then each subsequent row represents a different time period).

id	T_E	T_D	T_C	T0	T1	I_E	I_D	E1	E2	E3	E4	E5
1	300	300	300	0	300	0	0	0	0	0	0	0
2	47	47	300	0	47	0	1	0	0	0	0	0
3	35	300	300	0	35	0	0	0	0	0	0	0
3	35	300	300	35	49	1	0	1	0	0	0	0
3	35	300	300	49	63	1	0	0	1	0	0	0
3	35	300	300	63	77	1	0	0	0	1	0	0
3	35	300	300	77	91	1	0	0	0	0	1	0
3	35	300	300	91	105	1	0	0	0	0	0	1
4	105	136	300	0	105	0	0	0	0	0	0	0
4	105	136	300	105	129	1	0	1	0	0	0	0
4	105	136	300	129	136	1	1	0	1	0	0	0

Cox model in R: Coxph(Surv(T0, T1, I_D) ~ E1+E2+E3+E4+E5)

APPENDIX 2: PROPOSED OUTPUTS

The proposed outputs from this protocol will be like those included in the paper 'Association of COVID-19 with arterial and venous vascular diseases: a population-wide cohort study of 48 million adults in England and Wales'. Listed here for convenience:

Table 1. Number of patients analysed and, in paratheses, the risk per 100,000 of hospitalized and non-hospitalized SARS-CoV-2 infection

Table 2. Numbers of arterial thrombotic, venous thromboembolic and other vascular events before and after SARS-CoV-2 infection

Figure 1. Hazard ratios (log scale) for different arterial thrombotic, and venous thromboembolic and other vascular events after SARS-CoV-2 infection by time since diagnosis.

Figure 2. Hazard ratios (log scale) for arterial thrombotic events after SARS-CoV-2 infection by time since diagnosis, overall and stratified by whether hospitalised with SARS-CoV-2 infection, prior history of an arterial event, age, sex and ethnicity.

Figure 3. Hazard ratios (log scale) for venous thromboembolic events after SARS-CoV-2 infection by time since diagnosis, overall and stratified by whether hospitalised with SARS-CoV-2 infection, prior history of an arterial event, age, sex and ethnicity.

Figure 4. Absolute increase in risk of arterial thrombotic and venous thromboembolic events over time after SARS-CoV-2 infection, compared with no SARS-CoV-2 infection.

Supplementary Table 1. Derivation of major outcomes in OpenSafely.

Supplementary Table 2. Derivation of covariates.

Supplementary Table 3. Hazard ratios compared with no SARS-CoV-2 infection, according to time since SARS-CoV-2 infection. All results are maximally adjusted unless otherwise stated.

Supplementary Figure 1. Hazard ratios (log scale) for different after SARS-CoV-2 infection by time since diagnosis, stratified by where hospitalised with SARS-CoV-2 infection.

Supplementary Figure 2. Increases in absolute risk of arterial (upper plots) and venous events (lower plots) by time since diagnosis.