**Lay Summary**

We plan to investigate whether people with pre-existing chronic lung diseases who develop COVID-19 have a higher risk of heart attacks, stroke, and blood clots in the lungs compared with people with COVID-19 without pre-existing lung disease. People with chronic lung diseases will include those with asthma, chronic obstructive pulmonary disease, bronchiectasis, cystic fibrosis, and interstitial lung disease.

An international research prioritisation exercise was undertaken by members of this team in collaboration with Asthma UK-British Lung Foundation. We asked people with underlying lung diseases what the priorities should be for further research and investigation. The research question for this research was then developed from this consultation.

We will collaborate with researchers to use UK-wide de-identified patient data to answer this question. These data sit in secure and safe research environments with several data sources linked together, including primary care, hospital data, death registry data, COVID-19 testing and vaccination data.

This research is of public benefit as we will answer important questions we know patients and the public want answered. We will add value to these data, sharing expertise and data on lung disease. Findings will help better understand complications following COVID-19 in people with lung disease, if we should be treating them differently or even if some treatments taken for lung disease protect from some COVID-19 complications. Our work could be used to develop disease specific risk models’ clinicians can use to predict an individual's risk, communicate the risk effectively or design a personalised follow-up schedule, which reassures the patient, or even prioritise strategies prior to patients getting infected.

**Authors:** JK Quint, H Whittaker, C Kallis, AM Wood on behalf of the project steering committee

**Title:** SARS-CoV-2 infection and risk of major vascular events in people with chronic respiratory diseases

**Background:** Infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus - which causes the illness COVID-19 - induces a prothrombotic and proinflammatory state that may increase the risk of serious cardiovascular diseases. Whilst absolute and relative risks of cardiovascular diseases after SARS-CoV-2 infection have been investigated among the general population, whether this risk is greater in people with underlying chronic respiratory disease is unknown. Given that generally people with chronic respiratory diseases are at increased risk of cardiovascular diseases compared with the general population, particularly following the time of acute pulmonary infection, and that cardiovascular disease is a common cause of mortality in this population, it is important to better understand these risks. An estimation of the risk of vascular events of different types, such as myocardial infarction (MI), stroke, deep vein thrombosis (DVT) and pulmonary embolism (PE) will make an important contribution to our understanding of the clinical relevance of any long-standing endothelial or systemic inflammation.

Therefore, this project aims toestimate the risk of different cardiovascular diseases after SARS-CoV-2 infection among people with chronic respiratory diseases (i.e. asthma, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), bronchiectasis and cystic fibrosis (CF)). Whether any of these factors lead to persistent risk of cardiovascular disorders is important and we hypothesise they are at higher risk as these chronic respiratory diseases are all associated with an increased risk of cardiovascular disease both when people are stable and following an acute infection (ref Rothnie, Navaratnam, Morgan, Feary).

Some patients experience persistent symptoms after SARS-CoV-2 infection (‘long-COVID’). Among people who have had COVID-19 and are managed in the community, this is around 8-9% of the population who have symptoms persisting after 9 months (ref Whittaker). The cause of this illness is unclear, but is likely to be due to a combination of COVID-related inflammation and end organ damage, with an important psychosocial component. However, whether or not there is a more prolonged increase in risk of cardiovascular disease of different types after infection is uncertain but highly plausible based on descriptions of vascular inflammation. With new variants constantly emerging, it is also important to understand how risk has changed with virus mutation over time and how risk has changed after subsequent vaccinations. .

Existing studies have been hospital-based and relatively small. Because the increased risk of any one of the vascular diseases after SARS-CoV-2 infection is likely to be small, a very large study that includes as many people as possible with SARS-CoV-2 infection is needed to provide reliable information on future risk, and risk estimates during subsequent waves of the pandemic, in a variety of people. A cohort study is the most appropriate research design, but creating a new population-based cohort that is large enough to detect an effect and that approaches and recruit participants to answer this question would be both very time-consuming and expensive, and prone to loss to follow up.

Here we propose a cohort study leveraging established, secure access to linked, nationally collated NHS electronic healthcare record (EHR) data sources covering the entire populations of England, Wales and, if possible, Scotland (>65 million people). We will use all available prospectively collected SARS-CoV-2 infection data linked to hospitalisations, deaths and GP consultations to identify previous and new cardiovascular events. The project is supported by the [BHF Data Science Centre](https://www.hdruk.ac.uk/help-with-your-data/bhf-data-science-centre/), its [CVD-COVID-UK](https://www.hdruk.ac.uk/projects/cvd-covid-uk-project/) consortium, and BREATHE, and has been funded by a HDR UK Data and Connectivity grant, with access to the relevant, linked data and secure analysis environments already in place.

The advantage of the proposed study design in comparison to the existing studies is that: it builds on work already undertaken by the BHF data Science centre CVD-COVID-UK consortium using very large, individual-level linked population-wide data (including almost all patients with a positive polymerase chain reaction (PCR) for SARS-CoV2 in the UK). It is representative of all people in Britain who use healthcare services; it includes data from the beginning of the pandemic Britain to the present day; and follow-up can be extended at low cost.

**RESEARCH HYPOTHESES**

People with pre-existing respiratory disease who develop COVID-19 have a higher risk of future cardiovascular and venous thromboembolic events over follow-up compared with COVID-19 patients without pre-existing respiratory disease

**DATA SOURCES**

***NHS Digital TRE for England (up to latest release)***

* Primary care data (GP Data for Pandemic Planning and Research via General Practice Extraction Service, GPES)
* Second Generation Surveillance System (SGSS) COVID-19 infection laboratory testing data
* Pillar 2 Antigen
* Hospital episode statistics Admitted Patient Care (HES APC)
* SUS: Secondary Uses Service
* Vaccination Events
* Office for National Statistics (ONS) death registration records
* CHESS: COVID-19 Hospitalisation in England Surveillance System
* Medicined Dispensed in Primary Care (NHS BSA)
* Secondary Care Prescribed Medicines (EPMA)

***NHS Wales (via SAIL Databank TRE)***

* GPCD: Welsh Longitudinal General Practice (Daily COVID codes only)
* WLGP: Welsh Longitudinal General Practice
* PEDW: Patient Episode Database for Wales
* PATD: Pathology Data COVVID-19 [Daily (Laboratory Information Management System [Pillar 1&2 NHS/Lighthouse Labs Results & Pillar 3 Antibody Results])](https://web.www.healthdatagateway.org/dataset/f5f6d882-163d-4ef1-a53e-000fba409480)
* CVSP: COVID-19 Shielded People List
* CVVD: Covid Vaccination Dataset
* ADDD: Annual District Death Daily (OS Deaths)
* WDDS: Wales Dispensing Datasets
* ADDE - Annual District Death Extract (ONS Deaths)
* CDDS - Consolidated Death Data Source
* WDSD - Welsh Demographic Service Dataset *(for cohort curation)*
* CENW - Office of National Statistics Census (2011) *(for ethnicity)*

***Scotland (via eDRIS)***

* Primary care
* General Acute Inpatient and Day Case - Scottish Morbidity Record (SMR01)
* [COVID-19 laboratory and lighthouse testing (ECOSS)](https://web.www.healthdatagateway.org/dataset/5542cadc-11a4-4957-bbdd-8a6c46bcfb87)
* Deaths
* Dispensed/prescribed/paid
* *Yuras*Vaccination*Management Tool (*TVMT)

**RESEARCH QUESTION**

Does pre-existing respiratory disease modify the association between COVID-19 and future cardiovascular, cerebrovascular, and venous thromboembolic events?

**STUDY POPULATION**

**Population for COVID analyses**

Follow-up period: 01/01/20 to most recent data available

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

* An age of ≥18 can be calculated on 1st January 2020;
* Known sex;
* Have a record in the primary care extract;
* Have a valid linkage identifier;
* Alive on 1st January 2020.
* Have a record of a COVID-19 diagnosis from 01/01/2020. This will include a any of the following options: i) a +ve PCR test in SGSS; ii) a primary care COVID-19 diagnosis; iii) a hospital admission using HES APC & SUS and the ICD10 code U07.1. We will categorise people as:
* those hospitalised with COVID: hospital admission with COVID in primary position and hospital admission within first 28 days of COVID.
* those not hospitalised with COVID: COVID diagnosis but not hospitalisation within 28 days.

Exclude:any patients/records where a COVID-19 infection is recorded prior to 1st January 2020, or linkage is not possible, or if patients have a history of heart failure, myocardial infarction, stroke, or VTE.

**EXPOSURES**

The exposure will be pre-existing respiratory disease defined as a binary variable (0 for no pre-existing disease, and 1 for pre-existing disease). Pre-existing respiratory disease will include a history of the following diseases prior to COVID-19 diagnosis: asthma, COPD, bronchiectasis, cystic fibrosis, and ILD. (Code lists to be made and uploaded to Github).

**STATISTICAL METHODS**

**Comparison of cardiovascular (CVD) and venous thromboembolic (VTE) outcomes between COVID-19 patients with and without pre-existing respiratory disease, hospitalised or not for COVID-19**

Population: People in the UK 18+ years old, diagnosed with COVID-19 from the introduction of mass testing onwards, in primary or secondary care.

Exposure: Composite exposure group for pre-existing respiratory diseases; asthma, chronic obstructive pulmonary disease (COPD), cystic fibrosis, bronchiectasis and interstitial lung diseases (ILD) recorded prior to COVID-19 diagnosis. Patients with a history of myocardial infarction (MI), stroke or VTE recorded prior to start of follow-up will be excluded as they may be at a higher risk of outcomes We will also conduct a sensitivity analysis including these patients to understand how pre-existing CVD and respiratory disease with COVID-19 are associated with future CVD risk compared to patients without pre-existing respiratory disease

Follow up for each person will begin at the start of the follow up period which will be patient’s COVID-19 diagnosis date. Patient follow-up will be censored at the first of: death; the outcome event, or the end of the follow up period.

Statistical analyses: Follow-up starts from the first record of COVID-19. We will split follow up time for each person into periods since COVID-19. Where there are few events in any follow up period, the post-exposure periods will be collapsed. Outcome events of interest are listed below. We will tabulate numbers of outcome events, person-years of follow-up and rates of events before and with time since exposure.

We will fit survival models (Cox models or parametric survival models such as Weibull models) in which time zero is defined as the calendar date of the start of follow up. 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 different outcome events by time since exposure.

After identification of our study population described above, we start our analysis by including age and sex as confounding variables. If feasible, general practice will be included as a random effects variable in a 2-level survival model. Subsequently, we will consider matching people with pre-existing respiratory disease 1:3 to people without pre-existing respiratory disease on age, sex and general practice/region. People will be followed up from their date of COVID-19 diagnosis until the most recent available data or beforehand if people die, are lost to follow-up or have the CVD/VTE outcome of interest, or at first COVID-19 vaccination date. Time to event analysis (Cox regression) will be implemented adjusting for confounders such as age, gender, obesity, and other comorbidities, socioeconomic deprivation, and medication use.

As patients will be censored at first COVID-19 vaccination, we will repeat analyses from patient’s first COVID-19 vaccination to the second vaccination, and from second vaccination to third vaccination (i.e., booster vaccination).

Potential confounding factors (factors that predict both exposure and outcome, listed below) will be based on data recorded on or before the start of follow up in each analysis. We will estimate (i) crude; (ii) age and sex adjusted and (iii) maximally adjusted HRs (when there are sufficiently many outcome events). The following risk factors for venous thrombotic events will be included as confounders: sex, age, ethnicity in ONS, deprivation, anticoagulant prescription, combined oral contraceptive pill prescription, hormone replacement therapy prescription, history of coronavirus infection; and the following confounders for arterial thromboembolic events: sex, age, ethnicity, deprivation, diabetes, hypertension, smoking, anti-platelet prescription, blood pressure lowering prescription, lipid lowering prescription, anticoagulant prescription, and history of coronavirus infection. Variables selected in any model will be adjusted for in all models. The same selected variables will be adjusted for in models for all outcomes.

We will explore development of risk prediction models depending upon results.

**Missing data**

All analyses will be “complete-case” analyses (note that many potential confounders are defined using the presence versus absence of specific codes in the EHRs, so by definition have no missing values). We will not use multiple imputation or indicators for missing data.

**Subgroup analyses**

When there are at least 400 outcome events, we will estimate post-exposure hazard ratios separately within subgroups of particular interest (see “effect modifiers”) below.

We will estimate post-exposure hazard ratios:

* Separately within age groups for all people and in males and females;
* Separately by ethnicity
* Separately pre and post delta variant (pre and post march 2021)
* Separately for those with and without pre-existing CVD.
* Separately for before and after first vaccination, before and after second vaccination, and before and after third vaccination/booster vaccination.

These subgroup analyses will allow us to test for any effect modification by age, gender, ethnicity, medication use, and previous CVD as it is possible that the relationship between pre-existing respiratory disease and CVD outcomes can vary by these variables. These analyses will also allow us to understand whether this relationship varies by COVID-19 variant. As COVID-19 if ever changing it is important to understand if and how this might change the association between pre-existing respiratory disease and CVD outcomes.

**OUTCOMES (see appendix for code lists)**

Each outcome is defined as the first event of that type following the start of follow up in one of the following data sources: HES, primary care or ONS death registry. For the primary analyses, we will use events in the primary position were recorded in HES or death records, and in sensitivity analyses of the main results, events in any position. Some outcomes (largely venous) do not appear in the primary care data extract within the NHS Digital Trusted Research Environment, because of this the relevant codes will be confirmed to be available or not from the primary care extracts available in the respective UK Trusted Research Environments, and if required will only be ascertained in HES or death records.

Events will be defined as fatal if they are followed by death of any cause within 28-days, or are only recorded as fatal (i.e. reported only in death records). We will examine all events and fatal events in separate analyses. Lower limb fractures are included as an outcome that is unlikely to be affected by infection.

**Composite events:**

* **Arterial events:** first of ischaemic stroke or stroke of unknown type or myocardial infarction or other arterial thrombosis;
* **Venous events:** first of pulmonary embolism or lower limb deep venous thrombosis or other deep vein thrombosis;

**Individual events of different types**

**Arterial**

* Incident myocardial infarction in primary care or HES or ONS death;
* Incident ischaemic stroke or stroke of unknown type in primary care or SUS or ONS death;
* Incident non-stroke non-MI arterial embolism in primary care or SUS or ONS death.

**Venous**

* Pulmonary embolism in HES or ONS death;
* Deep vein thrombosis in HES or ONS death;

**Other**

* Life threatening cardiac arrhythmias or sudden cardiac death HES or ONS death;
* Lower limb fracture in HES or ONS death;

Date of onset defined as: of date of start of SUS or hospital admission spell with event; OR date of General Practitioner (GP) consultation with event; OR death with event (whichever comes first).

**POTENTIAL CONFOUNDERS**

Defined up to the start of follow up (defined henceforth as 1st January 2020), with a look back in GP and hospital admission (HA) for each person:

* **Sex:** categorical, GP;
* **Age in years** continuous, at inception; GP;
* **Ethnicity** 5 categorical, most recent recorded prior to inception; GP and if missing from GP data then from HA;
* **Deprivation** continuous, most recent recorded prior to inception; HA or GP;
* **Region**: East of England, London, Midlands, NE and Yorkshire, North West, South East, South West, Scotland, Wales, most recent residence prior to inception; GP;
* **Consultation rate**: number of primary care contacts in the year prior to inception; GP;
* **Other CVD medications**: yes/no; prescriptions/dispensing in year prior to index date; GP
* **Diabetes**: yes/not recorded from start of record to inception; GP or HA or medication;
* **Depression**: yes/not recorded from start of record to inception; GP or HA;
* **Obesity**: yes/not recorded from start of record to inception or BMI>30; GP
* **Cancer** yes/not recorded from start of record to inception; GP or HA;
* **CKD** yes/not recorded from start of record to inception; GP or HA;
* **Liver disease**: yes/not recorded from start of record to inception; GP or HA;
* **Major Surgery** yes/not recorded from start of record to inception; GP or HA;
* **Hypertension** yes/not recorded from start of record to inception; GP or HA or medication;
* **Dementia** yes/not recorded from start of record to date of linkage; GP or HA;
* **Smoking** current/ex-/never/unknown most recent prior to inception; GP;
* **Antiplatelet** yes/no/unknown (unknown if patient’s GP practice not in prescription extract) BNF codes starting ‘0209' with at least one prescription within three months prior to the inception date; dispensing;
* **BP lowering** yes/no/unknown (unknown if patient’s GP practice not in prescription extract) using DMD list of anti-hypertensives
* **Lipid lowering** yes/no/unknown (unknown if patient’s GP practice not in prescription extract) BNF codes starting ‘0212’ at least one prescription within three months prior to the inception date; dispensing;
* **Anticoagulant** yes/no/unknown (unknown if patient’s GP practice not in prescription extract) BNF codes starting ‘020802’ (exclude: '0208020I','0208020W') with at least one prescription within three months prior to the inception date; dispensing;
* **HRT** yes/no/unknown (unknown if patient’s GP practice not in prescription extract) BNF codes starting ‘0604011’ with at least one prescription within three months prior to the inception date; dispensing;

**EFFECT MODIFIERS (see definitions in list of confounders above)**

* Age within age group categories
* Sex: categorical;
* Ethnicity: categorical;
* Medication: yes/no for each of antiplatelet, BP lowering, lipid lowering, anticoagulant;

**META-ANALYSIS ACROSS NATIONS**

We will use inverse-variance weighted meta-analysis to pool estimates of post-COVID hazard ratios from common models across nations, examining heterogeneity using I2 statistics and reporting summary estimates and confidence intervals.

**Appendix 1 : outcome definitions**

|  |  |  |  |
| --- | --- | --- | --- |
| Composite events, composed of more than one phenotype | | | |
| Arterial event | First of *ischaemic stroke* or *stroke of unknown type* or *myocardial infarction* *or retinal infarction* or *other arterial embolism* | | |
| Venous event | first of *pulmonary embolism* or *deep venous thrombosis* or *other deep vein thrombosis* | | |
|  |  | | |
| Individual events | **Phenotype** | **Code** | **description** |
| Arterial | Incident myocardial infarction | I21\* | Acute myocardial infarction |
| Arterial | Incident myocardial infarction | I22\* | Subsequent myocardial infarction |
| Arterial | Incident myocardial infarction | I23\* | Certain current complications following acute myocardial infarction |
| Arterial | Incident myocardial infarction | [primary](https://github.com/BHFDSC/phenotype-development/blob/master/phenotypes/BHFCVDCOVID/AMI_SNOMED.csv) care codes, type=1 | SNOMED codes |
| Arterial | Ischaemic stroke | I63\* | Cerebral infarction, excluding I63.6 |
| Arterial | Ischaemic stroke | [primary](https://github.com/BHFDSC/phenotype-development/blob/master/phenotypes/BHFCVDCOVID/stroke_IS_SNOMED.csv) care codes type=1 | SNOMED codes |
| Arterial | Stroke of unknown type | I64\* | Stroke, not specified as haemorrhage or infarction |
| Arterial | Stroke of unknown type | [Primary care codes =1](https://github.com/BHFDSC/phenotype-development/blob/master/phenotypes/BHFCVDCOVID/stroke_NOS_SNOMED.csv) | SNOMED codes |
| Arterial | Stroke, subarachnoid haemorrhage | I60\* | Nontraumatic subarachnoid haemorrhage |
| Arterial | Other arterial embolism | I74\* | Arterial embolism and thrombosis |
|  |  |  |  |
| Venous | Pulmonary embolism | I26.0 | Pulmonary embolism without mention of acute cor pulmonale |
| Venous | Pulmonary embolism | I26.9 | Pulmonary embolism with mention of acute cor pulmonale |
| Venous | Deep vein thrombosis | I80\* | Phlebitis and thrombophlebitis of other sites |
| Venous | Other deep vein thrombosis | I82.0 | Budd Chiari Syndrome |
| Venous | Other deep vein thrombosis | I82.2 | Embolism and thrombosis of vena cava |
| Venous | Other deep vein thrombosis | I82.3 | Embolism and thrombosis of renal vein |
| Venous | Other deep vein thrombosis | I82.8 | Embolism and thrombosis of other specified veins |
| Venous | Other deep vein thrombosis | I82.9 | Embolism and thrombosis of unspecified vein |
|  |  |  |  |
| Fractures | Lower limb fracture | S720 | Fracture of neck of femur |
| Fractures | Lower limb fracture | S721 | Pertrochanteric fracture |
| Fractures | Lower limb fracture | S723 | Fracture of shaft of femur |
| Fractures | Lower limb fracture | S724 | Fracture of lower end of femur |
| Fractures | Lower limb fracture | S727 | Multiple fractures of femur |
| Fractures | Lower limb fracture | S728 | Fractures of other parts of femur |
| Fractures | Lower limb fracture | S729 | Fracture of femur, part unspecified |
| Fractures | Lower limb fracture | S820 | Fracture of patella |
| Fractures | Lower limb fracture | S821 | Fracture of upper end of tibia (and fibula) |
| Fractures | Lower limb fracture | S822 | Fracture of shaft of tibia (and fibula) |
| Fractures | Lower limb fracture | S823 | Fracture of lower end of tibia (and fibula) |
| Fractures | Lower limb fracture | S824 | Fracture of fibula alone |
| Fractures | Lower limb fracture | S825 | Fracture of medial malleolus |
| Fractures | Lower limb fracture | S826 | Fracture of lateral malleolus |
| Fractures | Lower limb fracture | S827 | Multiple fractures of lower leg |
| Fractures | Lower limb fracture | S828 | Fractures of other parts of lower leg |
| Fractures | Lower limb fracture | S829 | Fracture of lower leg, part unspecified |
| Fractures | Lower limb fracture | S920 | Fracture of calcaneus |
| Fractures | Lower limb fracture | S921 | Fracture of talus |
| Fractures | Lower limb fracture | S922 | Fracture of other tarsal bone(s) |
| Fractures | Lower limb fracture | S923 | Fracture of metatarsal bone |
| Fractures | Lower limb fracture | S927 | Multiple fractures of foot |
| Fractures | Lower limb fracture | S929 | Fracture of foot, unspecified |
| Fractures | Lower limb fracture | T12 | Lower limb fracture |
| Fractures | Lower limb fracture | T025 | Fractures involving multiple regions of both lower limbs |
| Fractures | Lower limb fracture | T023 | Fractures involving multiple regions of one lower limb |
|  |  |  |  |
| Life threatening arrhythmias |  | I472 | Ventricular tachycardia |
| Life threatening arrhythmias |  | I490 | Ventricular fibrillation and flutter |
| Life threatening arrhythmias |  | I460 | Cardiac arrest with successful resuscitation |
| Life threatening arrhythmias |  | I461 |  |
| Life threatening arrhythmias |  | I469 | Cardiac arrest, unspecified |
| Life threatening arrhythmias |  | I47.0 | Re-entry ventricular arrhythmia |