Version history

V0.6	24/7/2020	Added details on models for sensitivity analyses after comments from KK, AW, RD and FZ. Definition of models, tables still needed
V0.7	20/8/2020	Discussion on TRE and adding details after discussion with Rachel Denholm and Farrington and updated comments
V0.8	26/8/2020	Updated all comments and changes from V0.7
V0.9	9/9/2020	Added Ben Bray comments, particularly firming analysis population, comments from meeting of WP 2.5 (7.9.20)
V0.10	11/9/2020	Comments from Abdel Douri, Khamlesh Khunti
V0.11	17/9/2020	Comments from C Smith and C Sudlow, added lay summary
V0.12	16/11/2020	Self-controlled case series analysis added, updated lay summary
V0.13	25/11/20	Further comments from Ashley Akbari, Craig Smith, Fatemeh Torabi, Frank Kee, Kamlesh Khunti, Anne Marie Docherty, Julian Halcox, MJ Macleod, Francesco Zaccardi
V0.14	5/1/21	Further comments from Kate Tilling, Tom Palmer, Jonathan Sterne
V0.15	2/1/21	Removed angina as an exclusion criterion
V0.16	13.5.21	Updated with suggestions from vaccine group
V0.17	28/8/21	Further updates

Lay summary

Coronavirus infection ('COVID') also known as COVID-19, might increase the chance of having a stroke, heart attack or clots in the deep leg veins or lungs ('blood vessel diseases').

During the COVID pandemic, some doctors saw patients who had COVID who also had unusual strokes, clots or heart complaints. This suggested a link between COVID and blood vessel diseases. But no individual doctor saw enough patients to find out if COVID really did increase the risk of blood vessel diseases.

We will study every person alive in NHS England, Scotland and Wales at the beginning of the pandemic in 2020. We will find out how many people had a stroke, heart or other diseases of the blood vessels until the date we begin the analysis.

We will compare the number of people with COVID infection who had a blood vessel disease with the number of people without COVID infection who had a blood vessel disease. Different types of people might have different risks, so we will look at people of different ages, ethnicities and medical history as well. The result will be an estimate of how much COVID increases the risk of different blood vessel diseases.

This information is needed so that people with COVID know whether they need to worry about blood vessel diseases as they recover. If there is an increased risk, then preventative treatments might be needed.

AUTHORS

BHF Data Science Centre

TITLE

SARS-CoV-2 infection and risk of major vascular events

BACKGROUND

Better knowledge about the effect of SARS-CoV-2 infection on short- and long-term risk of MI and stroke is important to estimate the magnitude of any effect, to identify those at greatest risk, and therefore whether preventative treatments (e.g. aspirin) for those with infection are needed.

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. However, the absolute and relative risks of cardiovascular diseases after SARS-CoV-2 infection are uncertain at this stage of the pandemic. If the longer-term risk of cardiovascular diseases is substantial in general or in particular populations, then an effective prevention strategy will be a public health priority for people in the UK and globally.

Any elevated risk of cardiovascular disease after SARS-CoV-2 infection needs to be put into context of risk after other acute viral infections, such as influenza. An estimation of the risk of vascular events of different types, such as myocardial infarction (MI), stroke, heart failure, cardiac arrhythmia, 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 to estimate the risk of different cardiovascular diseases after SARS-CoV-2 infection, and if elevated, identify the period and duration of risk**.**

There are plausible mechanisms by which SARS-CoV-2 infection could increase the risks of cardiovascular diseases: a direct effect on thrombosis, coagulation and fibrinolysis;[1] secondary antiphospholipid antibodies; systemic inflammation and its downstream effects on destabilisation of atheroma and activation of thrombosis; endovascular inflammation ('endotheliopathy');[2,3] or cardiac dysfunction[4,5] leading to cerebral cardio-embolism;[6] or indeed a combination of these mechanisms.

A study in Denmark,[9] which examined infections from the beginning of the epidemic to July 2020 using a self-controlled case series design, suggested that the risk of stroke and MI was elevated, between 3 and 15-fold. The development of new heart failure after infection is infrequent (reported to be ~2.5% in from hospital-based cohorts) but important.[10] Stroke occurs in about 1% of coronavirus admissions, and some studies suggest an increased risk of more severe stroke.[11,12]

Whether any of these factors lead to persistent risk of cardiovascular disorders is important. Some patients experience persistent symptoms after infection ('long-COVID'). 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.

A comparison of the magnitude and time-course of any risk after COVID-19 with the risk of cardiovascular diseases after other viral infections would help to put the risk after COVID-19 in perspective. Acute viral infections, such as influenza, are associated with an increase in the risk of stroke and MI. A number of studies have estimated this risk in the 2 to 4 weeks following infection: laboratory confirmed influenza was associated with 4–6 fold increase of the risk of MI in Ontario; a clinical diagnosis of influenza was associated with a 3–fold risk of stroke in California;[13,14] and a respiratory tract infection with a 5-fold increase in MI risk and 3-fold increase in stroke-risk in the UK (although this study showed no association between influenza and either illness).[15] SARS Co-V2 appears to be vasculo-tropic;[16] and other vasculo-tropic viruses such as varicella zoster and HIV increase the risk of stroke.[17] A retrospective US hospital based cohort suggested that the risk of stroke after COVID-19 is approximately 7-fold greater than after influenza infection (OR, 7.6; 95% CI, 2.3-25.2).[11] A study in US Department of Veterans Affairs hospitals suggested that both cardiovascular and

cerebrovascular disease risk is increased by COVID infection.[18] Therefore, it is plausible that SARS-CoV-2 increases the risk of vascular diseases, possibly to a greater degree than other respiratory infections.

The risk of arterial occlusive events, such MI or ischaemic stroke, and venous events, such as pulmonary thrombosis (whether this is in-situ or embolism is unclear) or deep vein thrombosis, appears to be increased after coronavirus infection. [5,12,19–21] An understanding of the absolute and relative risk of arterial and venous events is an essential prerequisite to defining the need for and the nature of short and longer-term prevention strategies.

Existing studies have necessarily 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 a 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, Scotland and Wales (>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 and its CVD-COVID-UK consortium, 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 is very large with individual-level linked population-wide coverage (including almost all patients with a positive polymerase chain reaction (PCR) for SARS-CoV2 in the UK), so maximising the chance of detecting a modest effect on cardiovascular events of different types; since it draws on data covering the entire population, it is representative of all people in Britain who use healthcare services; it includes data from the beginning of the pandemic in the Britain to the present day; and follow-up can be extended at low cost.

RESEARCH HYPOTHESES

There is a higher risk of venous and arterial thrombosis or embolism after COVID-19 disease than before or without infection.

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
- Hospital episode statistics Admitted Patient Care (HES APC)
- Office for National Statistics (ONS) death registration records
- Community dispensing data

NHS Wales (via SAIL Databank TRE)

- COVID C20 (all people alive and resident in Wales from 01/01/20) and C16 (counterfactual from 01/01/16 to end 2019) total population cohorts 3.2M. Censored by migration out of Wales and death.
- Patient Episode Database for Wales (PEDW).
- Consolidated mortality with 4 separate mortality data sources (Welsh Demographic Service Dataset (WDSD) – population spine weekly flows, ONS - monthly and daily flows (ADDE and ADDD), and records from the MPI (Master Patient Index) - daily flows (CDDS)).
- Pathology data COVID-19 test results (PATD) daily flows from NHS and Public Health Wales(PHW) laboratories and Lighthouse laboratories, including antigen and antibody testing - Pillar 1,2 and 3.

- Primary care (available on 80% that provide data to SAIL on a monthly basis containing all diagnostic, referral, prescribed medication (WLGP). Plus 100% GP daily flow since January 2020 for of COVID-19 specific coding and symptom respiratory codes (GPCD)).
- Community dispensing (WDDS).
- Intensive Care National Audit & Research Centre (ICCD, ICNC).
- Critical Care Data Set (CCDS).
- The ONS Census (CENW), pending approvals.
- Outpatient Database for Wales (OPDW) and Emergency department data set (EDDS)

RESEARCH QUESTION

In people who have had COVID-19 disease compared with people who have not, are there higher rates (expressed as hazard ratios with time since COVID-19 disease) of fatal or non-fatal ischaemic stroke or MI ('arterial); cerebral sinus thrombosis, mesenteric or portal vein thrombosis, PE or DVT ('venous thromboembolism'); and other vascular events before and after adjustment for potential confounders?

STUDY POPULATION

Population for COVID analyses

Follow-up period: 01/01/20 to 07/12/20 (from the beginning of the pandemic to the start of vaccination).

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;
- Alive on 1st January 2020.

Exclude: any patients/records where a COVID-19 infection is recorded prior to 1st January 2020, or linkage is not possible.

EXPOSURES

Any COVID exposure:

- +ve PCR test in SGSS; or
- Primary care COVID-19 diagnosis; or
- Hospital admission using HES APC & SUS and the ICD10 code U07.1

Any COVID exposure with hospitalisation:

- Hospital admission with COVID in primary position, and
- Hospital admission within first 28 days of COVID

Any COVID without hospitalisation

- COVID and no hospitalisation within 28 days

STATISTICAL METHODS

Follow up for each person will begin at the start of the follow up period and be censored at the first of: death; the outcome event, or the end of the follow up period.

We will split follow up time for each person into periods before and after exposure (COVID-19 COVID-19 disease of different severities), and into time periods since exposure (Appendix 2). Where there are few events in any one 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) 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 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, all people exposed to COVID, and a random subset (for example, 10%) of people without the outcome event or COVID, with analyses incorporating inverse probability weights (for example, 10) for data from people without the outcome event, and confidence intervals derived using robust standard errors. All models will be stratified on region (so that risk sets are constructed within region, hence accounting for between-region variation in the baseline hazard). In sensitivity analyses we will confirm that results do not change markedly when we stratify on Lower-layer Super Output Area (LSOA) version 2011.

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/BAME categories, deprivation, anticoagulant prescription, combined oral contraceptive pill prescription, hormone replacement therapy prescription, history of pulmonary embolism or deep vein thrombosis, and 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, history of stroke, and history of MI. For other events, variables to be adjusted for will be selected from the list of potential confounders using a backwards stepwise procedure with p value threshold 0.2. 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.

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

We will estimate post-exposure hazard ratios:

- Following any infection;
- Separately within age groups for all people and in males and females;
- When there are sufficiently many outcome events, by separate calendar periods (e.g. before or after June 2020 between the two waves of COVID-19), and where possible by week post infection.

When there are at least 400 outcome events, we will estimate post-exposure hazard ratios:

- Separately in people with and without a history of cardiovascular disease.
- Separately within subgroups of particular interest (see "effect modifiers") below.

These subgroup analyses will allow for different baseline hazard functions and differential confounding effects within subgroups, because associations between exposures and confounders and between confounders and outcomes may vary between subgroups.

When there are fewer than 400 outcome events, we will conduct sensitivity analyses in which the baseline hazard is additionally stratified by sex, and hazard ratios for males and females are estimated using the data on both sexes but allowing for exposure-sex interactions. Such an approach assumes the effect of covariates on the outcome to be the same in males and females, but avoids fitting a model including a large number of potential confounding variables to datasets with a small number of outcome events.

OUTCOMES

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 where recorded in HES or death records. 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).

Composite events:

- **Arterial events:** first of ischaemic stroke or stroke of unknown type or retinal infarction or myocardial infarction or other arterial thrombosis;
- *Arterial events + thrombocytopenia* (i.e. any arterial event + any thrombocytopenia code present during in same spell), and no prior history of thrombocytopenia;
- **Venous events:** first of pulmonary embolism or lower limb deep venous thrombosis or intracranial venous thrombosis or intracranial venous thrombosis during pregnancy or portal vein thrombus 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 or retinal infarction 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;
- Portal vein thrombosis in HES or ONS death;
- Other deep vein thrombosis in HES or ONS death;
- Thrombosis during pregnancy (excluding intracranial venous thrombosis) in SUS or ONS death;
- Intracranial venous thrombosis during pregnancy in HES or ONS death;
- Intracranial venous thrombosis[22] in HES or ONS death.

Other

- Intracerebral haemorrhage in primary care or HES or ONS death;
- Mesenteric thrombus in HES or ONS death;
- Spinal stroke in HES or ONS death;
- Heart failure in HES or ONS death;
- Angina in HES or ONS death.
- Lower limb fracture in HES or ONS death;
- Transient ischaemic attack in HES or ONS
-

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;
- **Medications**: total number of medications by BNF chapters prescribed within three months prior to the inception 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;
- **COPD** 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 antihypertensives
- **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;
- **COCP** yes/no/unknown (unknown if patient's GP practice not in prescription extract) BNF codes starting '070301' 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;
- **History of stroke** yes/not recorded any stroke type (from 'outcomes above, ischaemic stroke, intracranial haemorrhage, or stroke of unknown type) recorded from start of record to inception; GP or HA;
- **History of MI**: yes/not recorded any MI type (from 'outcomes above) recorded from start of record to inception; GP or HA;
- **History of angina:** yes/not recorded any angina type (from 'outcomes above) recorded from start of record to inception; GP or HA;
- **History of VT**: yes/not recorded and venous event (from 'outcomes above) recorded from start of record to inception; GP or HA;
- **History of thrombophilia**: yes/not recorded from start of record to inception, defined as ICD-10 D68.5 D68.6; acquired or inherited; GP or HA;

- **History of thrombocytopenia** (from outcomes above) yes/not recorded from start of record to inception; GP or HA;
- **History of any outcome type:** (from outcomes above) yes/not recorded from start of record to inception; GP or HA;

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;
- History of CVD yes/not known;

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 I^2 statistics and reporting summary estimates and confidence intervals.

REFERENCES

1. Long B, Brady WJ, Koyfman A, Gottlieb M. Cardiovascular complications in COVID-19. Am. J. Emerg. Med. 2020;

2. Varga Z, Flammer AJ, Steiger P, Haberecker M, Andermatt R, Zinkernagel AS, Mehra MR, Schuepbach RA, Ruschitzka F, Moch H. Endothelial cell infection and endotheliitis in COVID-19. Lancet. 2020;395:1417–1418.

3. Goshua G, Pine AB, Meizlish ML, Chang CH, Zhang H, Bahel P, Baluha A, Bar N, Bona RD, Burns AJ, Dela Cruz CS, Dumont A, Halene S, Hwa J, Koff J, Menninger H, Neparidze N, Price C, Siner JM, Tormey C, Rinder HM, Chun HJ, Lee AI. Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study. *Lancet Haematol*. 2020;7:e575–e582.

4. Madjid M, Safavi-Naeini P, Solomon SD, Vardeny O. Potential Effects of Coronaviruses on the Cardiovascular System: A Review. JAMA Cardiol. 2020;

5. Bonow RO, Fonarow GC, O'Gara PT, Yancy CW. Association of Coronavirus Disease 2019 (COVID-19) with Myocardial Injury and Mortality. JAMA Cardiol. 2020;

6. South K, McCulloch L, McColl BW, Elkind MSV, Allan SM, Smith CJ. Preceding infection and risk of stroke: An old concept revived by the COVID-19 pandemic [Internet]. Int. J. Stroke. 2020 [cited 2020 Sep 22];174749302094381. Available from: http://journals.sagepub.com/doi/10.1177/1747493020943815

7. Halushka MK, Vander Heide RS. Myocarditis is rare in COVID-19 autopsies: cardiovascular findings across 277 postmortem examinations. *Cardiovasc Pathol* [Internet]. 2021 [cited 2020 Nov 3];50:107300. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1054880720301046

8. Siddiqi HK, Libby P, Ridker PM. COVID-19 – A vascular disease. *Trends Cardiovasc Med*. 2020;

9. Modin D, Claggett B, Sindet-Pedersen C, Lassen MCH, Skaarup KG, Jensen JUS, Fralick M, Schou M, Lamberts M, Gerds T, Fosbøl EL, Phelps M, Kragholm KH, Andersen MP, Køber L, Torp-Pedersen C, Solomon SD, Gislason G, Biering-Sørensen T. Acute COVID-19 and the Incidence of Ischemic Stroke and Acute Myocardial Infarction [Internet]. Circulation. 2020 [cited 2021 Jan 7];142:2080–2082. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7682795/

10. Rey JR, Caro-Codón J, Rosillo SO, Iniesta ÁM, Castrejón-Castrejón S, Marco-Clement I, Martín-Polo L, Merino-Argos C, Rodríguez-Sotelo L, García-Veas JM, Martínez-Marín LA, Martínez-Cossiani M, Buño A, Gonzalez-Valle L, Herrero A, López-Sendón JL, Merino JL, Merino JL, Caro-Codon J, Castrejon-Castrejon S, Iniesta AM, Martinez-Cossiani M, Merino C, Martin-Polo L, Martinez LA, Marco I, Garcia-Veas JM, Rodriguez-Sotelo L, Rosillo SO, Lopez-Sendon JL, Rey JR, Rios JJ, Arribas JR, Arnalich F, Prados C, Alvarez-Sala R, Quintana M, García de Lorenzo A, Reinoso F, Rivera A, Torres RM, Garcia-Rodriguez J, Gonzalez-Valle L, Herrero A, Borobia A, Buño A. Heart failure in COVID-19 patients: prevalence, incidence and prognostic implications. *Eur J Heart Fail* [Internet]. 2020 [cited 2020 Nov 4];ejhf.1990. Available from: https://onlinelibrary.wiley.com/doi/10.1002/ejhf.1990

11. Merkler AE, Parikh NS, Mir S, Gupta A, Kamel H, Lin E, Lantos J, Schenck EJ, Goyal P, Bruce SS, Kahan J, Lansdale KN, Lemoss NM, Murthy SB, Stieg PE, Fink ME, Iadecola C, Segal AZ, Cusick M, Campion TR, Diaz I, Zhang C, Navi BB. Risk of Ischemic Stroke in Patients with Coronavirus Disease 2019 (COVID-19) vs Patients with Influenza. *JAMA Neurol* [Internet]. 2020 [cited 2020 Sep 22];Available from: https://jamanetwork.com/

12. Beyrouti R, Adams ME, Benjamin L, Cohen H, Farmer SF, Goh YY, Humphries F, Jäger HR, Losseff NA, Perry RJ, Shah S, Simister RJ, Turner D, Chandratheva A, Werring DJ.

Characteristics of ischaemic stroke associated with COVID-19. *J Neurol Neurosurg Psychiatry* [Internet]. 2020 [cited 2020 May 19];jnnp-2020-323586. Available from: http://jnnp.bmj.com/lookup/doi/10.1136/jnnp-2020-323586

13. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, Katz K, Ko DT, McGeer AJ, McNally D, Richardson DC, Rosella LC, Simor A, Smieja M, Zahariadis G, Gubbay JB. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. *N Engl J Med* [Internet]. 2018 [cited 2020 May 19];378:345–353. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1702090

14. Boehme AK, Luna J, Kulick ER, Kamel H, Elkind MS V. Influenza-like illness as a trigger for ischemic stroke. *Ann Clin Transl Neurol* [Internet]. 2018 [cited 2020 Jun 4];5:456–463. Available from: http://doi.wiley.com/10.1002/acn3.545

15. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination. *N Engl J Med* [Internet]. 2004 [cited 2020 Jun 9];351:2611–2618. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa041747

16. Teuwen L-A, Geldhof V, Pasut A, Carmeliet P. COVID-19: the vasculature unleashed. *Nat Rev Immunol* [Internet]. 2020 [cited 2020 Jun 11];1–3. Available from: http://www.nature.com/articles/s41577-020-0343-0

17. Grau AJ, Urbanek C, Palm F. Common infections and the risk of stroke. Nat. Rev. Neurol. 2010;6:681–694.

18. Xie Y, Bowe B, Maddukuri G, Al-Aly Z. Comparative evaluation of clinical manifestations and risk of death in patients admitted to hospital with covid-19 and seasonal influenza: cohort study. *BMJ* [Internet]. 2020 [cited 2021 Jan 10];371:4677. Available from: http://dx.doi.org/10.1136/bmj.m4677

19. Oxley TJ, Mocco J, Majidi S, Kellner CP, Shoirah H, Singh IP, De Leacy RA, Shigematsu T, Ladner TR, Yaeger KA, Skliut M, Weinberger J, Dangayach NS, Bederson JB, Tuhrim S, Fifi JT. Large-Vessel Stroke as a Presenting Feature of Covid-19 in the Young. *N Engl J Med* [Internet]. 2020 [cited 2020 May 19];382:e60. Available from: http://www.nejm.org/doi/10.1056/NEJMc2009787

20. Morassi M, Bagatto D, Cobelli M, D'Agostini S, Gigli GL, Bnà C, Vogrig A. Stroke in patients with SARS-CoV-2 infection: case series. *J Neurol* [Internet]. 2020 [cited 2020 Sep 18];267:2185–2192. Available from: /pmc/articles/PMC7238403/?report=abstract

21. D'Anna L, Kwan J, Brown Z, Halse O, Jamil S, Kalladka D, Venter M, Banerjee S. Characteristics and clinical course of Covid-19 patients admitted with acute stroke [Internet]. J. Neurol. 2020 [cited 2020 Sep 18];1:3. Available from: https://doi.org/10.1007/s00415-020-10012-4

22. Handley JD, Emsley HCA. Validation of ICD-10 codes shows intracranial venous thrombosis incidence to be higher than previously reported. *Heal Inf Manag J* [Internet]. 2020 [cited 2021 Mar 25];49:58–61. Available from: https://pubmed.ncbi.nlm.nih.gov/30563370/

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* or *portal vein thrombosis* or *intracranial venous thrombosis*

Appendix 2: splitting follow up according to time since the start of follow up and since exposure.

Definitions
Time scale – days since 1/1/2020 (for analyses examining hazard ratios after COVID-19 disease)

Outcome of interest – time to event D (T_D, I_D)

Exposure of interest – binary exposure E (COVID disease) measured at T_E with indicator I_E, parameterised as days since T_E, categorised for example into E1 = 0-13.999 E2=14-27.999; E3=28-41.999; E4=42-55.999; E5=56-70 days (although time interval may change depending on the number of events of different types)

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 example I have set T_C = 300

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)