Protocol: Long-term kidney outcomes after SARS-CoV-2 infection

Viyaasan Mahalingasivam, Kathryn E. Mansfield, Dorothea Nitsch, Juan-Jesus Carrero, Krishnan Bhaskaran, John Tazare, Alex Walker, Elizabeth Williamson, Brian MacKenna, OpenSAFELY collaborative, National Core Studies Team... Laurie A.Tomlinson

Date: January 2022

1. Lay summary

1.1 Background

People who require hospital admission for COVID-19 usually have severe lung infection which means they can struggle to maintain oxygen levels. However, other organs can also be affected, with the kidneys being among the most common. Around a third of people admitted to hospital with COVID-19 develop kidney damage (known as acute kidney injury or AKI); of these, around 10% develop such severe AKI that they require short-term dialysis (treatments to clear out toxins and excess fluid from their bodies). The reasons for this AKI are yet to be definitively understood.

It is not unusual for AKI to complicate any severe illness (such as life-threatening infections or trauma), especially in people who have other medical problems such as diabetes, chronic kidney disease (CKD) or heart failure. We do not know whether AKI related to COVID-19 just reflects this or whether it is because of processes directly related to COVID-19 infection itself.

While AKI is often recoverable, it can sometimes lead to worsening long-term kidney function. Chronic kidney disease is diagnosed when kidney function declines to less than 60% of what it should be. In some cases, chronic kidney disease can lead to permanent kidney failure requiring long-term dialysis (often involving attending a dialysis unit three-times a week for four hours each time), or a kidney transplant. Even without kidney failure, chronic kidney disease is an important risk factor for several other complications including heart disease and death. The complications caused by chronic kidney disease place considerable stress on the NHS and it is important that we know what to expect to be able to ensure that services are planned with sufficient capacity to cope. For example, growing numbers requiring dialysis has a devastating impact on quality of life for those people and will require investment in the workforce and infrastructure. It can also mean more people requiring hospital admissions for heart attacks.

One way of determining what happens to kidney function after COVID-19 is to analyse readily-available health records collected from day-to-day patient care. A recent study of this sort in US military veterans suggested that people who survived COVID-19 are more likely to develop kidney failure and lose 50% of their kidney function compared to levels

before they had COVID-19. We plan to investigate this in greater detail using electronically coded NHS medical records from England using a secure, Trusted Research Environment platform called OpenSAFELY (www.opensafely.org). OpenSAFELY comprises 40% of the general population in England (24 million people).

This study has been funded by the National Institute for Health Research (NIHR). The study was approved by the Health Research Authority and the London School of Hygiene & Tropical Medicine Ethics Board.

1.2 Methods

Our study will include adults not already on dialysis and not already having received a kidney transplant. We will compare people with and without COVID-19 for differences in the rates of the following kidney complications:

1. Newly-diagnosed kidney failure (i.e. new dialysis or kidney transplant, or loss of kidney function down to less than 15% of what it should be),
2. At least 50% loss in kidney function from before they developed COVID-19,
3. Further AKI (suggesting their kidneys are more susceptible to future damage).

We will be able to determine whether outcomes vary based on the severity of COVID-19 (i.e. based on whether people required hospitalisation, intensive care, a ventilator, dialysis or had AKI), based on time (e.g. during peaks, after widespread use of effective anti-COVID-19 drugs such as steroids), and based on their levels of vaccination.

In our statistical analysis, we will account for factors such as age, sex, ethnicity, socioeconomic deprivation, smoking, BMI, and pre-existing medical problems such as diabetes.

1.3 Strengths and limitations

As kidney failure is a relatively rare occurrence in the general population, very large studies are needed to be able to investigate whether or not a group is at increased risk compared to other groups. Because of the size of the population in OpenSAFELY, we are uniquely placed to be able to look for this in the context of COVID-19. Furthermore, we can expect the data to be comprehensive as it comes from GPs and hospitals within a national healthcare system with universal coverage.

As with other studies of this nature, we anticipate challenges associated with real-world data collection. We will be as rigorous as possible to address these limitations by undertaking multiple analyses with slight differences. For example, we will compare people with COVID-19 (including people hospitalised and not hospitalised) to different groups of people without COVID-19:

1. A historical group comprising individuals followed-up from exactly 2 years before the person with COVID-19. The reason for using a historical comparator group is because not everyone with COVID-19 gets diagnosed, especially when access to testing has been limited.

2. A contemporary group comprising individuals followed-up from the same time as the person with COVID-19. The reason for using a contemporary comparator as well is to account for changes over time during the pandemic such as differences in access to healthcare (e.g. likely to have the same access to blood tests).

A further historical analysis will be restricted only to people hospitalised with COVID-19. In this analysis, we will compare this group to people hospitalised with other lung infections from February 2018 onwards up to the start of the pandemic. This will help us understand whether any differences in kidney outcomes are specifically due to COVID-19 or due to having any lung infection severe enough to require hospital admission.

This particular study will not investigate outcomes in kidney transplant recipients. One of the reasons for this is because blood test results in OpenSAFELY come from GP records rather than hospitals where many transplant patients have their tests done. We do intend to work around this with a different study design.

2. Background

In this study, we will investigate long-term adverse kidney outcomes after surviving recorded severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and its disease-state, coronavirus disease 2019 (COVID-19).

COVID-19 hospitalisation is associated with acute kidney injury (AKI). Analysis of data from over 40,000 patients from Britain hospitalised for COVID-19 found that 32% developed AKI with associated risks including pre-existing CKD and black ethnicity; 3% required renal replacement therapy (RRT), the treatment for severe AKI (Sullivan 2021). The development of AKI in mild or asymptomatic SARS-CoV-2 infection remains unknown.

The pathophysiology of AKI in COVID-19 is yet to be understood. AKI is not unusual in the context of acute illness, especially on a background of comorbidity (e.g., diabetes, chronic kidney disease [CKD], heart failure), while factors such as increasing age and socioeconomic deprivation also contribute (Phillips 2018). Mechanisms in COVID-19 may include sepsis, hypovolaemia, immune-mediated processes, drugs (e.g., antibiotics and antivirals,) and direct viral infection. Acute tubular injury has been demonstrated in post-mortems of people who died following COVID-19, and reports suggest hypercoagulability and thrombotic microangiopathy in tissues including the kidneys (Batlle 2020). Further, autopsy microdissection studies have detected SARS-CoV-2 in the kidney (Puelles 2020), and ultrastructural virus-like particles within podocytes and tubular epithelial cells (Su 2020). As SARS-CoV-2 binds to angiotensin-converting-enzyme-2, expressed on proximal tubular epithelial cells and podocytes in the kidneys (as well as lung tissue), there may be replication of the virus within the kidney itself (Batlle 2020).

AKI is a risk factor for the development of chronic kidney disease (CKD) (Hsu 2016), and the potential mechanisms described above in COVID-19 (e.g., immune-mediated process, viral infection) could independently result in long-term sequelae. In addition, CKD is an independent risk factor for severe outcomes in COVID-19 raising the possibility of accelerated renal decline in those most at risk (Williamson 2020). People are diagnosed with CKD when their estimated glomerular filtration rate (eGFR) (calculated from serum creatinine level, a routine biochemical blood test) is <60ml/min/1.73m^2 or urinary albumin creatinine ratio is >3mg/mmol for at least three months, or if they have a structural kidney abnormality (e.g. polycystic kidneys). Trajectories in eGFR can vary and progressive decline to <15ml/min/1.73m^2 results in end-stage renal disease (ESRD) (also known as “kidney failure”) with chronic RRT required (in the form of dialysis or kidney transplantation) to sustain life; almost a third of ESRD in the UK is secondary to diabetes with other causes including glomerulonephritis, hypertension and polycystic kidney disease (UK Renal

Registry 2021). In addition to ESRD, CKD is also associated with increased all-cause mortality, cardiovascular mortality and morbidity, and AKI (Hsu 2016, KDIGO 2013). Complications of CKD include anaemia, hyperkalaemia, hypervolaemia, altered pharmacokinetics, and secondary hyperparathyroidism (KDIGO 2013). CKD is common, affecting 8.5% of the UK population (Stevens 2007), and is associated with a substantial burden on healthcare costs: the cost to NHS England was estimated at over £1.4 billion in 2009-10 including excess hospitalisation and cardiovascular complications amounting to over £220m, and £550m attributable to dialysis therapy (Kerr 2012).

For these reasons, we are concerned about the consequences in terms of long-term mortality, morbidity and healthcare costs for the substantial number of people who survive SARS-CoV-2 infection.[3] Using observational data to investigate long-term kidney outcomes after SARS-CoV-2 infection and COVID-19 is methodologically challenging due to biases related to the broader impacts of the pandemic as well as inherent limitations when investigating renal outcomes (McDonald 2016, Mansfield 2021). Recent analysis of national routinely-collected electronic health record (EHR) data up to 30 April 2021 from the US Department of Veterans Affairs concerningly found that people who survived SARS-CoV-2 infection at 30 days had an increased risk of AKI, eGFR decline and ESRD compared to contemporary controls; this included people who were not hospitalised with SARS-CoV-2 infection with increased risk of adverse outcomes in those who were hospitalised with concurrent AKI (Bowe 2021). However, this study was limited to a comparison with a contemporaneous population of people without known SARS-CoV-2 infection who may have been susceptible to misclassification. We aim to further explore renal complications post COVID-19 using national-level data from England with both contemporary and historical comparisons.

3. Objectives

Overall, our aim is to investigate kidney outcomes after recorded SARS-CoV-2 infection.

Our specific objectives are:

1. To describe the relative and absolute rates of ESRD (primary outcome) after recorded SARS-CoV-2 infection compared to both contemporary and historical groups, stratified by:
    a. Hospitalisation for COVID-19
    b. Critical care admission for COVID-19
    c. Mechanical ventilation for COVID-19

    d. Acute RRT for COVID-19
    e. AKI during hospitalisation for COVID-19
    f. Calendar time of SARS-CoV-2 infection
    g. Vaccination status

2. To describe the relative and absolute rates of 50% reduction in eGFR, a composite outcome of ESRD or 50% reduction in eGFR, and incident AKI (secondary outcomes), after recorded SARS-CoV-2 infection compared to both contemporary and historical groups.
3. To describe the relative and absolute rates of these outcomes after recorded SARS-CoV-2 infection compared to both contemporary and historical groups stratified by:
    a. Baseline eGFR
    b. Diabetes
    c. Age
    d. Sex
    e. Ethnicity
    f. Socioeconomic deprivation
    g. Rural/urban living
4. To describe the relative and absolute rates of these outcomes after hospitalised COVID-19 compared to historical hospitalised groups, stratified by:
    a. Critical care admission for COVID-19
    b. Mechanical ventilation for COVID-19
    c. Acute RRT for COVID-19
    d. AKI during hospitalisation for COVID-19
    e. Calendar time of COVID-19

4. Methods

We will use routinely collected EHR data from primary care practices using TPP SystmOne software, covering approximately 40% of the population in England, linked at the individual level to NHS Secondary Uses Service (SUS) data on hospitalisations, and Second Generation Surveillance System (SGGS) data on SARS-CoV-2 test results. We will conduct cohort studies investigating kidney outcomes in people after recorded SARS-CoV-2 infection compared to both contemporary (2020-2021 general population) and historical (2018-2019 general population) groups.

4.1 Data source

Primary care records managed by TPP, a GP software provider, were linked to other data sources through OpenSAFELY, a data analytics platform created by our team on behalf of NHS England to address urgent COVID-19 research questions (https://opensafely.org) (Williamson 2020). OpenSAFELY provides a secure software interface allowing the analysis of pseudonymized primary care patient records from England in near real-time within the EHR vendor's highly secure data centre, avoiding the need for disclosive pseudonymized patient data to be transferred off-site. This, in addition to other technical and organisational controls, minimizes any risk of re-identification. Similarly, pseudonymized datasets from other data providers are securely provided to the EHR vendor and linked to the primary care data. The dataset analysed within OpenSAFELY is based on 24 million people currently registered with GP surgeries using TPP SystmOne software. It includes pseudonymized data such as coded diagnoses, medications, and physiological parameters. No free text data are included. Further details on our information governance can be found here: information governance and ethics.

4.2 Information governance

NHS England is the data controller; TPP is the data processor; and the key researchers on OpenSAFELY are acting on behalf of NHS England. This implementation of OpenSAFELY is hosted within the TPP environment, which is accredited to the ISO 27001 information security standard and is NHS IG Toolkit compliant (NHS Digital Data Security Standards, NHS Digital Data Security & Protection Toolkit); patient data have been pseudonymized for analysis and linkage using industry standard cryptographic hashing techniques; all pseudonymized datasets transmitted for linkage onto OpenSAFELY are encrypted; access to the platform is through a virtual private network (VPN) connection; the researchers hold contracts with NHS England and only access the platform to initiate database queries and statistical models; all database activity is logged; and only aggregate statistical outputs leave the platform environment following best practice for anonymization of results such as

statistical disclosure control for low cell counts (NHS Digital Anonymisation standard for publishing health and social care data). The OpenSAFELY research platform adheres to the data protection principles of the UK Data Protection Act 2018 and the EU General Data Protection Regulation (GDPR) 2016. In March 2020, the Secretary of State for Health and Social Care used powers under the UK Health Service (Control of Patient Information) Regulations 2002 (COPI) to require organizations to process confidential patient information for the purposes of protecting public health, providing healthcare services to the public and monitoring and managing the COVID-19 outbreak and incidents of exposure (Department of Health and Social Care Coronavirus (COVID-19): notification to organisation to share information). Together, these provide the legal bases to link patient datasets on the OpenSAFELY platform. GP practices, from which the primary care data are obtained, are required to share relevant health information to support the public health response to the pandemic and have been informed of the OpenSAFELY analytics platform.

4.3 Study designs

We will compare people after SARS-CoV-2 infection to both historical and contemporary cohorts in order to address epidemiological challenges to be able to triangulate meaningful conclusions.

Testing for SARS-CoV-2 during the early stages of the pandemic was restricted due to a lack of infrastructure (NHS Test and Trace: the journey so far, Iacobucci 2020); therefore many people who were infected (both symptomatic and asymptomatic) will not have been coded as such and so will be misclassified in a contemporary comparison. Using historical comparator groups will allow us to definitively ensure that the comparator group was not infected with SARS-CoV-2 during follow-up. Some comparisons will also be matched on date (i.e., to start follow-up on the same date two years before), which will allow us to account for consistent seasonal variations (Tang 2021, Poole 2020, Iwagami 2018, Goto 2020).

Contemporary comparator populations will account for potential biases we anticipate for any historical comparison. Using a contemporary comparator will also allow us to finely adjust for changes over time during the pandemic (with time since the start of the pandemic as the underlying timescale), such as the emergence of SARS-CoV-2 variants (Tracking SARS-CoV-2 variants 2021), changes in testing policy (NHS Test and Trace: the journey so far, Iacobucci 2020), changes in treatment of COVID-19 (RECOVERY Collaborative Group 2021, RECOVERY Collaborative Group 2021), any environmental differences in air pollution due to lockdowns (Al-Aly 2020), and differences in ascertainment of outcomes due to reduced access to healthcare (including blood bottle shortages in 2021) (Mansfield 2021; Rimmer 2021).

4.4 Study populations

Our overall study population will be all adults (18+ years) registered with primary care practices using TPP SystmOne software (Figure 1) since 1 February 2016. We will exclude all individuals with missing age, sex, or Index of Multiple Deprivation (IMD) as these are likely to indicate poor data quality. All individuals included will need to have at least three months of follow-up available before cohort entry to ensure reliable capture of baseline health status; we have been reassured by TPP that three months is sufficient for records to be updated and this will optimise power.

4.4.1 Exposure group

Our exposure group will comprise people with recorded SARS-CoV-2 infection identified using data from SGSS, primary care records or SUS who remain alive 28 days after the first recording of infection (i.e. 28 days after date of testing positive in SGSS, coding in primary care records, or admission on SUS) (index date) (Figure 1).

Anyone coded in primary care or SUS to be receiving chronic dialysis or kidney transplant or with eGFR <15 ml/min/1.73m^2 (i.e. anyone with ESRD or CKD Stage 5) at the time of first recording of SARS-CoV-2 infection will be excluded.

Our secondary analysis will investigate individuals who remain alive 28 days after the date of first hospital admission coded for COVID-19 in SUS. Anyone coded in primary care or SUS to be receiving chronic dialysis or kidney transplant or with eGFR <15 ml/min/1.73m^2 (i.e. anyone with ESRD or CKD Stage 5) at the time of hospital admission will be excluded.

4.4.2 Historical comparator groups

For our main analysis, historical comparator groups will comprise individuals from the general population under follow-up in OpenSAFELY, each matched five to one to each individual from the recorded SARS-CoV-2 group on age (within 2 years), sex, NHS Sustainability and Transformation Partnership (STP, geographical areas configured for regional reorganisation available in our data), IMD decile (a measure of socioeconomic deprivation based on postcode), and date (i.e. 2 years before the first recording of SARS-CoV-2 infection in the exposed case) (Figure 1). Anyone coded in primary care or SUS to be receiving chronic dialysis or kidney transplant or with eGFR <15 ml/min/1.73m^2 before this date will be excluded (i.e. anyone with ESRD or CKD Stage 5). We will match without replacement (i.e. individuals in the historical comparator group will not be matched to more than one individual with SARS-CoV-2 infection) in order to optimise statistical precision.

In our secondary analysis, we will compare individuals who remain alive 28 days after admission to hospital with COVID-19 with those who remained alive 28 days after admission to hospital for pneumonia (including influenza) from 1 February 2018 up to two years before the end of follow-up for the exposed group (unmatched) (Figure 1). A previous study done using OpenSAFELY found similarities in outcomes between people discharged after hospitalisation with COVID-19 and a historical comparator group discharged after pneumonia (Tazare 2021 - submitted). Individuals who survived COVID-19 and hospitalisation from pneumonia in 2018-2019 will be followed-up in both the exposed and unexposed groups.

4.4.3 Contemporary comparator group

For our main analysis, contemporary comparator groups will comprise individuals from the general population under follow-up in OpenSAFELY, individually matched five to one to each individual from the recorded SARS-CoV-2 group on age (within 2 years), sex, NHS STP, IMD decile, and date of first recording of SARS-CoV-2 infection (Figure 1). Anyone coded in primary care or SUS to be receiving chronic dialysis or kidney transplant or with eGFR <15 ml/min/1.73m^2 before the first recording of SARS-CoV-2 infection will be excluded (i.e. anyone with ESRD or CKD Stage 5). We will again match without replacement. Individuals from this group will be censored if they themselves are recorded with SARS-CoV-2 infection and further follow-up will commence in the exposed group from 28 days after recording if they remain alive (Figure 1).

4.4.4 Follow-up start

By using index dates 28 days after first recording of SARS-CoV-2 infection (i.e., date of testing in SGSS, date of primary care coding, or date of admission on SUS), we will be able to look back and further stratify the exposure by measures of SARS-CoV-2 severity (i.e. community, hospitalisation, critical care admission etc.).

For analyses comparing people after SARS-CoV-2 infection to a historical comparator group from 2018-2019, the matched historical comparators will be followed-up from exactly two years before the index date.

For analyses comparing people hospitalised for COVID-19 with an unmatched historical group of individuals hospitalised for pneumonia from 2018-2019, we will follow all eligible individuals from 28 days after their admission date for pneumonia.

For analyses comparing people after SARS-CoV-2 infection to a matched contemporary comparator group, we will follow all eligible individuals from the index date.

4.4.5 Follow-up end

In all analyses we will follow individuals until the earliest of: record of specific kidney outcome under investigation, ESRD, death, no longer registered with GP practice, or study end. Individuals from the matched contemporary comparator group will also be censored at the date of first recorded SARS-CoV-2 infection and will transfer to the exposed group if they survive 28 days after this date.

For analyses using matched historical 2018-2019 comparators, the study end for people with COVID-19 will be the last available date for data at the time of extraction, and for their matched comparators it will be January 31 2020 at the latest. Our choice of end of follow-up for people with COVID-19 is to be consistent with the available follow-up time for people in the historical comparator group, as we will no longer be able to be able to follow people in the historical comparator group after February 2020 (as we will no longer be sure that those in our comparator group did not have COVID-19 after the start of the initial UK COVID-19 outbreak in February 2020). By allowing the same duration of potentially available follow up for people with and without COVID we will be able to ensure our results are not biased by allowing additional follow up for people with COVID-19 (and consequently allowing more time for them to develop the outcomes under investigation) compared to those without. By allowing the same duration of potentially available follow up for people with and without SARS-CoV-2 infection, we will be able to ensure our results are not biased by allowing additional follow up for people pre-pandemic.

For analyses with the contemporary comparator group, the study end will be the latest available date of fully linked data available on the day the initial cohort is extracted.

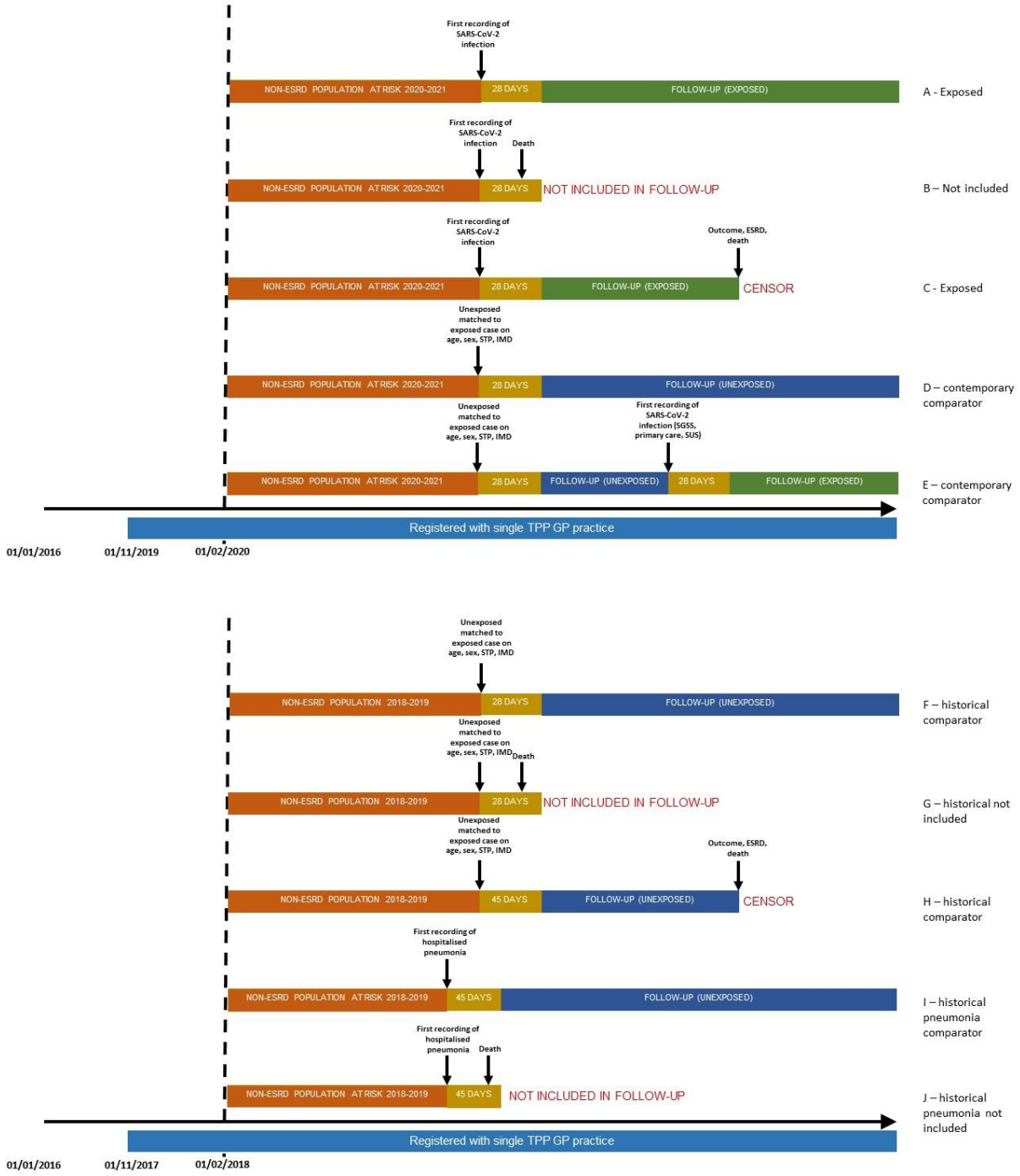

Figure 1. Graphical illustrations of potential study scenarios. **A-C: Examples of individuals with recorded SARS-CoV-2 infection**. A was followed-up from 28 days after SARS-CoV-2 infection until the study conclusion; B died within 28 days of SARS-CoV-2 infection so was not followed-up; C commenced follow-up from 28 days after SARS-CoV-2 infection until they were censored (i.e. for outcome event, death, or ESRD). **D-E: Examples of matched contemporary comparators**. D was matched to an individual with SARS-CoV-2 infection and followed-up from 28 days after matching until the study conclusion; E was matched to an individual with SARS-CoV-2 infection and followed-up from 28 days after matching until they were first recorded with SARS-CoV-2 infection themselves when they were censored from further follow-up as an unexposed comparator and then

followed-up as an exposed case from 28 days after infection until the study conclusion. **F-G: Examples of matched historical comparators from 2018-2019**. F was matched to an individual with SARS-CoV-2 infection and followed-up from 28 days after matching until the study conclusion; G died within 28 days of matching so was not followed-up; H commenced follow-up from 28 days after matching until they were censored (i.e. for outcome event, death, or ESRD). **I-J: Examples of unmatched historical comparators hospitalised for pneumonia from 2018-2019**. I was followed-up from 28 days after admission for pneumonia until the study conclusion. J died within 28 days of admission for pneumonia so was not followed-up.

4.5 Study measures

Further details regarding the definition of all study measures are available here.

4.5.1 Exposure

We will identify recorded SARS-CoV-2 infection using data from SGSS, primary care records and SUS. We will include all codes indicative of SARS-CoV-2 infection. SGSS contains the date of an individual's earliest positive SARS-CoV-2 test results from Pillar 1 (NHS and Public Health England laboratories) and Pillar 2 (commercial partners) between 1 February 2020 to the most recent available time point (i.e. the COVID-19 pandemic era).

Individuals meet criteria for inclusion in the COVID-19 exposed group if they:

- Tested positive for SARS-CoV-2 infection in SGSS
- Had a primary care code indicative of SARS-CoV-2 infection
- Discharged from hospital with an International classification of Diseases Version 10 (ICD-10) code in SUS indicative of COVID-19

The index date for those exposed to COVID-19 will be from 28 days after first recording of SARS-CoV-2 infection (i.e., after the earliest date of the record of a positive test in SGSS, primary care coding, or admission in SUS).

Looking at all identified SARS-CoV-2 infections will allow us to investigate how many adverse kidney outcomes can be prevented by preventing SARS-CoV-2 infection altogether through comparisons with people from the general population without known infection. We will also be able to investigate whether outcomes vary based on infection severity (hospitalisation, critical illness, mechanical ventilation and AKI [including requiring RRT]) as well as calendar time and vaccination status through further stratification based on morbidity coding (ICD-10 codes) from SUS. We will categorise people as i) unvaccinated if had not received any vaccine dose at least 7 days before first recording of SARS-CoV-2 infection, ii) single dose vaccinated if they had received one vaccine done at least 7 days

before first recording of SARS-CoV-2 infection, iii) double vaccinated if they had received two vaccine doses at least 7 days before first recording of SARS-CoV-2 infection, and iv) more than double vaccinated if they had received more than two vaccine doses (including booster doses) at least 7 days before first recording of SARS-CoV-2 infection.

Hospitalised COVID-19 cases will be further investigated through a separate comparison with a historical group. Additional levels of exposure for those hospitalised will include critical care admission, mechanical ventilation, acute RRT and AKI, determined from ICD-10 and procedural codes in SUS.

4.5.2 Outcomes

4.5.2.1 Primary outcome

We will identify incident ESRD through codes for dialysis or kidney transplant using either primary care codes, or ICD-10 or procedural codes from SUS, or eGFR <15ml/min/1.73m^2 in primary care not within 14 days of hospitalisation.

4.5.2.2 Secondary outcomes

1) 50% reduction in eGFR

   This analysis will exclude anyone without a baseline eGFR measurement. We will be cautious when interpreting our analysis of eGFR outcomes as this is unlikely to be generalisable to the general population and only to people who had an indication for baseline eGFR testing (e.g. comorbidities, drugs).

   Baseline eGFR will be defined as the mean primary care eGFR calculated from serum creatinine using the CKD-EPI equation from 18 months before the index date up to 14 days before the index date (or 14 days before hospital admission date for hospitalised COVID-19 cases) (Levey 2009).

   Primary care eGFR measurements will be followed-up monthly from the index date, excluding measurements within 14 days of any hospital admission until criteria for the outcome endpoint is met (50% reduction in eGFR from the baseline eGFR). The OpenSAFELY platform is currently only able to extract a predefined number of measures for any variable (i.e., it is not possible to extract all available measures); we will therefore extract and use the last eGFR measure for every calendar month in follow-up.

   We also consider assessing changes in eGFR trajectory using slopes.

2) Composite outcome: ESRD or 50% reduction in eGFR

We will investigate for a composite outcome of ESRD or a 50% reduction in eGFR to mitigate the former being a competing risk of the latter.

This analysis too will exclude anyone without a baseline eGFR measurement.

3) AKI

   We will identify episodes of hospitalisation with AKI through ICD-10 codes in any diagnostic position in SUS after the index date.

4) Death

   We will describe rates of death as this is an important competing risk for outcomes. Date of death will be obtained using primary care records, which compared to the Office for National Statistics (ONS) (the gold standard), has previously been demonstrated to have 78% exact agreement and 99% agreement within 30 days (Gallagher 2019). In case such accurate date coding practices have been affected by the pandemic, we will be able to link to the ONS date of death for 2020 data (we do not have access to ONS date of death linkage before 2019).

4.5.3 Covariables

Using primary care codes and ICD-10 codes in SUS, we will describe the distribution of the baseline covariables in the exposed and comparator groups. We will consider the following factors for inclusion in the regression analysis: demographic (age, sex), geographic region, rural/urban living, deprivation, ethnicity, existing comorbidities, smoking and body mass index (BMI).

For the matched historical comparison groups, we will extract covariable data as of 28 days before the index date (i.e. on the date of first recording of infection) on OpenSAFELY.

4.5.3.1 Demographic

Age and sex will be included in all analyses. For descriptive purposes, we will classify age in the following categories: 18-39, 40-49, 50-59, 60-69, 70-80 and 80 years and over. In regression analyses, we will include a spline function for age to account for variation in SARS-CoV-2 infection by age.

4.5.3.2 Social

It is important to account for geographical variation in SARS-CoV-2 infection rates. We will define an individual's geographical region based on NHS STP administrative regions of

England grouped into East of England, London, Midlands, North East and Yorkshire, North West, South East and South West. Separately, we will also classify individuals based on whether they live in rural or urban settings using residential postcodes grouped according to the Office for National Statistics Rural-Urban Classification for Output Areas 2011; urban conurbation, urban city and town, or rural. This is to factor in potentially differential access to healthcare and risk of recorded and unrecorded SARS-CoV-2 infection.

We will define deprivation using quintiles of IMD (English indices of Deprivation 2019). IMD is assessed across seven domains of deprivation: income, employment, education, health, crime, barriers to housing and services, and living environment. The IMD indices rank each "lower layer super output area", derived from residential postcodes in England, from most deprived to least deprived.

We will define ethnicity using self-reported ethnicity recorded in primary care and collapsed into four high-level categories of: white, south Asian, black, and other and mixed (Mathur 2021). We expect that ethnicity will be missing for approximately 20% of our study population (Mathur 2021, Bhaskaran 2021). For analyses adjusting for ethnicity, we will do a complete-case analysis (only including those with known ethnicity) with a follow-up sensitivity analysis.

4.5.3.3 Clinical

We will describe several clinical conditions including: diabetes mellitus, hypertension, ischaemic heart disease, heart failure, atrial fibrillation, stroke, peripheral vascular disease, non-haematological cancer, haematological cancer, chronic respiratory disease, chronic liver disease, dementia, other chronic neurological diseases, systemic lupus erythematosus, vasculitis, rheumatoid arthritis, HIV and sickle cell disease.

We will describe baseline eGFR for each individual based on the mean eGFR recoded from 18 months to 14 days before the index date, excluding measures recorded in the 14 days before or after an acute hospital admission including the index admission (acute illness may cause AKI resulting in misclassification of baseline eGFR and CKD stage) . As with follow-up eGFR, it is not possible to obtain infinite serial eGFR measurements from OpenSAFELY and so the final eGFR measurement of each calendar month before the index date (up to 18 months before the cohort start date) will be recorded for each individual. eGFR will be calculated using the CKD-EPI equation using serum creatinine measurements without adjustment for ethnicity (Levey 2009). Baseline eGFR will be classified as follows:

1. ≥105 ml/min/1.73m^2

2. 90-104 ml/min/1.73m^2
3. 75-89 ml/min/1.73m^2
4. 60-74 ml/min/1.73m^2
5. 45-59 ml/min/1.73m^2 (CKD stage 3A)
6. 30-44 ml/min/1.73m^2 (CKD stage 3B)
7. 15-29 ml/min/1.73m^2 (CKD stage 4)
8. No eGFR measurement

We will also describe albuminuria and/or proteinuria. In addition, we will describe coded CKD as clinical management may vary (e.g. greater routine follow-up) between people diagnosed and subsequently coded as having CKD compared to people without recognised CKD.

The electronic frailty index (eFI) is a validated tool for older people that calculates a frailty score based on an individual's accumulation of deficits (e.g. atrial fibrillation, CKD, falls) that can be derived from primary care codes (Clegg 2016). The tool is used widely in primary care after NHS England contracted practitioners to identify all individuals with moderate or severe frailty (NHS England 2017). The eFI has been built into TPP for clinical use and we will request baseline eFI calculations for each individual aged ≥65 years based on their accumulation of deficits and will be used to categorise as: fit (≤0.12), mild frailty (0.13-0.24), moderate frailty (0.25-0.36) and severe frailty (>0.36).

We will define smoking status based on most recent primary care coding as non-smokers, ex-smokers or current smokers. We will assume that individuals with no record of smoking status are non-smokers as smoking is less common than not smoking (approximately 14% of the UK adult population were smokers in 2019) (Office for National Statistics 2020). However, we will run a sensitivity analysis restricting to those with known smoking status.

BMI will be calculated using the most recent height and weight closest to the index date, then categorised as: underweight (<18.5kg/m^2), normal weight (18.5-24.9kg/m^2), overweight (25.0-29.9kg/m^2), obese (≥30.0kg/m^2) and undocumented (assumed to be normal in main analyses) (Bhaskaran 2013; Williamson 2020).

We will describe the number of individuals who were being prescribed drugs from the following classes in primary care: insulin, other antidiabetic drugs, immunosuppressants and anticoagulants.

4.6 Statistical analysis

We will initially describe the baseline characteristics of people after SARS-CoV-2 infection and those from the matched historical and contemporary comparator groups.

4.6.1 Main analyses

Due to the competing risk of death, we will use Fine and Gray regression models (using calendar time as the underlying timescale) to estimate subdistribution hazard ratios (SHRs) and 95% confidence intervals (CIs) to compare the rates of each outcome for individuals with SARS-CoV-2 infection to individuals in historical and contemporary comparator groups (Tazare 2021). Deviations from proportional subdistribution hazards will be further investigated using time-period specific SHRs (Dey 2020, Li 2015).

We will run sequential regression models:

1) Minimally adjusted for age (using spline functions) and sex,
2) Additionally adjusting for ethnicity, IMD, smoking and BMI,
3) Additionally adjusting for clinical comorbidities.

We will use robust standard errors to account for clustering by GP practice as well as on individuals as there may be overlap between those in the recorded SARS-CoV-2 infection groups and the historical comparison group (although the time periods are non-overlapping). We will assume people with missing BMI are not obese and people without a documented smoking status are non-smokers (Williamson 2020).

4.6.2 Secondary analyses

All analyses will be stratified by hospitalisation, critical care admission, mechanical ventilation, acute RRT and AKI after COVID-19, as well as calendar time of SARS-CoV-2 infection (i.e. February – June 2020, July – September 2020, October – November 2020, December 2020 – February 2021, and March – December 2021) and vaccination status.

A priori, we intend to fit interaction terms to obtain hazard ratio estimates stratified by age group, ethnicity, quintile of IMD, frailty group (for those aged 65 years and over only), diabetes, baseline eGFR, coded CKD, and calendar time of infection.

For people who survive 28 days after admission for COVID-19, we will undertake an unmatched analysis of historical comparator populations followed-up from 1 February 2018, restricted to those who survived 28 days after a hospital admission for pneumonia (with or without influenza). Because this analysis will not be matched, there may be

residual confounding due to seasonality which we will address by adjusting for calendar month.

Stratified SHRs will also be reported for critical care admission, mechanical ventilation, acute RRT and AKI after COVID-19.

We will describe Kaplan-Meier graphs for survival after COVID-19. We will obtain *p*-values using log rank tests.

4.6.3 Sensitivity analyses

4.6.3.1 ESRD

In case of possible residual misclassification of AKI as ESRD (where an individual has an eGFR <15ml/min/1.73m^2 in primary care not within 14 days of hospitalisation), we will undertake a sensitivity analysis restricting only to people with new codes for dialysis or transplantation (i.e., incident RRT).

4.6.3.2 eGFR reduction

In case a difference in 50% reduction in eGFR is not identified, we will also investigate for 40% and 30% reduction in eGFR. If there was a causal association between SARS-CoV-2 infection and reduction in renal function, we would anticipate that the strength of association would be on a gradient.

4.6.3.3 Incident eGFR <30ml/min/1.73^2

As healthcare services have been impacted heavily by the pandemic, there may not have been sufficient routine testing of patients to be able to detect a reduction in eGFR. As patients who are highest risk may be prioritised, we will also investigate for incident eGFR <30ml/min/1.73m^2 not within 14 days of hospitalisation or new codes for dialysis or kidney transplant in either primary care records, or ICD-10 or procedural codes from SUS.

4.6.3.4 eGFR equation

In our main analyses, we will calculate eGFR using the CKD-EPI equation from 2009 (Levey 2009). This equation has recently been updated excluding adjustment for ethnicity (Inker 2021). We will assess how many individuals change CKD stage from the original CKD-EPI equation after switching to the 2021 equation. If there are marked differences in the number of people with CKD and within each CKD stage, we will undertake a sensitivity analysis using the 2021 equation instead for eGFR.

4.6.3.5 Age

As our main analyses will adjust for age in broad age bands, to ensure that we have fully accounted for age as a potential confounder, we will repeat analyses with age as the underlying timescale (allowing us to finely account for age).

4.6.3.6 Ethnicity

For our main analyses, we will be undertaking a complete case analysis of only individuals who have recorded ethnicity data. We will undertake a sensitivity analysis using multiple imputation for ethnicity.

4.6.3.7 Smoking

In our main analyses we will assume that individuals with missing smoking status were non-smokers. To test this assumption, we will repeat our analyses restricting to those with known smoking status.

4.6.3.8 BMI

In our main analyses we will assume that individuals with missing BMI data have normal BMI. To test this assumption, we will repeat our analyses restricting to those with known BMI.

4.7 Software and Reproducibility

Data management will be undertaken using Python, with analysis carried out using Stata 16.1. Code for data management and analysis and code lists used to define study measures (exposures, outcomes, covariates) will be archived online [link project github repo].

5. Power calculations

Another study of 89,000 survivors after SARS-CoV-2 infection found increased incidence of ESRD (our primary outcome) with a hazard ratio of 2.96 (95% confidence interval 2.49-3.51). To date, there have been over 6.4 million people with recorded SARS-CoV-2 infection in England; as OpenSAFELY comprises 40% of the population, we anticipate that our sample size will be over 2 million people including around 20,000 who were hospitalised with sufficient power to detect a difference in outcome compared to historical and contemporary groups.

6. Strengths and limitations

Similar to all observational studies using routinely collected electronic health record data, the internal validity of our study will potentially be limited by selection and information bias, and residual confounding. We will aim to conduct quantitative bias analyses for unmeasured confounding.

Accurate capture of the majority of the variables used in the study (exposures, outcomes, covariables) relies on individuals having access to healthcare. We are aware that early in the COVID-19 pandemic, individuals appear to have limited their GP contacts and it is not yet clear whether GP contacts have recovered to pre-pandemic levels (Mansfield 2021). Our analysis will, at least in part, be able to mitigate this, as the underlying timescale for analyses with the contemporary comparator group will be time since the initial UK COVID-19 outbreak, allowing us to be able to finely adjust for changes over time in outcome ascertainment.

Ascertainment of decline in kidney function as an outcome may vary between people who have survived COVID-19 and their comparators due people who survived COVID-19 having more kidney function testing, which may result in an overestimation of any effect on decline in renal function; however, by making an additional historical comparison with hospitalised pneumonia comparators we may be able to assess the extent of this bias. It is not currently possible to calculate the number of kidney function tests done within a period of time using OpenSAFELY.

It is likely that a proportion of individuals with SARS-CoV-2 infection will be missed in the contemporary comparison, particularly early in the pandemic when testing was limited to specific groups (NHS Test and Trace: the journey so far, Iacobucci 2020). This may lead us to misclassify some matched individuals as not having infection, potentially meaning that we underestimate the true impact of SARS-CoV-2 on kidney outcomes. In order to address this, we will also undertake matched comparisons with recent historical populations pre-pandemic unaffected by SARS-CoV-2 infection.

Investigating all recorded SARS-CoV-2 infections and comparing them to comparator groups from the general population may also be problematic. Those hospitalised with COVID-19 may be more likely to experience adverse kidney outcomes because of hospitalisation for an acute illness or hospitalisation for pneumonia more specifically when compared to the general population, therefore exaggerating the relationship with SARS-CoV-2 infection specifically. In order to address this, we are undertaking secondary analyses disaggregating hospitalised COVID-19 and comparing this group to a historical hospitalised population admitted for pneumonia. A previous study conducted on

OpenSAFELY found that outcomes in patients discharged after COVID-19 or historical pneumonia were similar (Tazare 2021 - publication pending).

As this is an observational study we will not be able to exclude residual confounding as a reason for any associations we see. Our modelling strategy may also be limited due to the current lack of clarity around the potential causal pathological mechanisms linking SARS-CoV-2 and kidney outcomes. However, our effect estimates will provide initial insights into potential temporal links which may provide vital information for follow-up studies investigating what may drive those links.

7. Administrative

7.1 Ethics

The study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the London School of Hygiene and Tropical Medicine (LSHTM) ethics board (reference 21863).

7.2 Funding

[Placeholder for OpenSAFELY overall funding]

[Enter details of local funding] NCS?

VM is funded by the National Institute for Health Research (Award ID NIHR301535)

7.3 Patient and public involvement

This protocol has been developed in collaboration with our Patient and Public Involvement partners.

7.4 Conflict of interests

The authors have declared no competing interest.

8. References

1. Phillips D, Holmes J, Davies R, Geen J, Williams JD, Phillips AO. The influence of socioeconomic status on presentation and outcome of acute kidney injury. *QJM* 2018; **111**: 849–57.
2. Sullivan, M. K. et al. Acute kidney injury in patients hospitalised with COVID-19 from the ISARIC WHO CCP-UK Study: a prospective, multicentre cohort study. Nephrol Dial Transplant, doi:10.1093/ndt/gfab303 (2021)
3. Batlle D, Soler MJ, Sparks MA, *et al*. Acute Kidney Injury in COVID-19: Emerging Evidence of a Distinct Pathophysiology. *J Am Soc Nephrol* 2020; **31**: 1380–3.
4. Puelles VG, Lütgehetmann M, Lindenmeyer MT, *et al*. Multiorgan and Renal Tropism of SARS-CoV-2. *N Engl J Med* 2020; **383**: 590–2.
5. Su H, Yang M, Wan C, *et al*. Renal histopathological analysis of 26 postmortem findings of patients with COVID-19 in China. *Kidney Int* 2020; **98**: 219–27.
6. Hsu RK, Hsu C-Y. The Role of Acute Kidney Injury in Chronic Kidney Disease. *Semin Nephrol* 2016; **36**: 283–92.
7. Williamson EJ, Walker AJ, Bhaskaran K, *et al*. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020; **584**: 430–6.
8. Registry UKR. UK Renal Registry 23rd Annual Report Data to 31/12/2019. 2021.
9. Outcomes KDI. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. 2013.
10. Stevens PE, O'Donoghue DJ, de Lusignan S, *et al*. Chronic kidney disease management in the United Kingdom: NEOERICA project results. *Kidney Int* 2007; **72**: 92–9.
11. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. *Nephrol Dial Transplant* 2012; **27 Suppl 3**: iii73–80.
12. Mansfield KE, Mathur R, Tazare J, *et al*. Indirect acute effects of the COVID-19 pandemic on physical and mental health in the UK: a population-based study. *The Lancet Digital Health* 2021; published online Feb. DOI:10.1016/s2589-7500(21)00017-0.
13. Bowe B, Xie Y, Xu E, Al-Aly Z. Kidney Outcomes in Long COVID. *J Am Soc Nephrol* 2021; published online Sept 1. DOI:10.1681/ASN.2021060734.
14. BETA – Data Security Standards. https://digital.nhs.uk/about-nhs-digital/our-work/nhs-digital-data-and-technology-standards/framework/beta---data-security-standards (accessed July 9, 2021).
15. Data Security and Protection Toolkit. https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit (accessed July 9, 2021).
16. ISB1523: Anonymisation standard for publishing health and social care data. https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/isb1523-anonymisation-standard-for-publishing-health-and-social-care-data (accessed July 9, 2021).

17. Department of Health and Social Care. Coronavirus (COVID-19): notification to organisations to share information. 2020; published online April 1. https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information (accessed July 9, 2021).
18. NHS Test and Trace: the journey so far. https://www.health.org.uk/publications/long-reads/nhs-test-and-trace-the-journey-so-far (accessed Aug 17, 2021).
19. Iacobucci G. Covid-19: What is the UK's testing strategy? *BMJ* 2020; **368**: m1222.
20. Tang JW, Bialasiewicz S, Dwyer DE, *et al.* Where have all the viruses gone? Disappearance of seasonal respiratory viruses during the COVID-19 pandemic. *J Med Virol* 2021; **93**: 4099–101.
21. Poole S, Brendish NJ, Clark TW. SARS-CoV-2 has displaced other seasonal respiratory viruses: Results from a prospective cohort study. *J Infect* 2020; **81**: 966–72.
22. Iwagami M, Moriya H, Doi K, *et al.* Seasonality of acute kidney injury incidence and mortality among hospitalized patients. *Nephrol Dial Transplant* 2018; **33**: 1354–62.
23. Goto S, Hamano T, Ogata S, Masakane I. Seasonal variations in cause-specific mortality and transition to renal replacement therapy among patients with end-stage renal disease. *Sci Rep* 2020; **10**: 2325.
24. Tracking SARS-CoV-2 variants. World Health Organization. 2021; published online Jan 9. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed Jan 9, 2021).
25. RECOVERY Collaborative Group, Horby P, Lim WS, *et al.* Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* 2021; **384**: 693–704.
26. RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet* 2021; **397**: 1637–45.
27. Al-Aly Z, Bowe B. Air Pollution and Kidney Disease. Clin. J. Am. Soc. Nephrol. 2020; **15**: 301–3.
28. Rimmer A. What has caused the NHS blood tube shortage, and how is it affecting doctors and patients? *BMJ* 2020; 374:n2174
29. Gallagher AM, Dedman D, Padmanabhan S *et al.* The occurrence of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations. *Pharmacoepidemiol Drug Saf* 2019; 28:563-9
30. English indices of deprivation 2019. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 (accessed Aug 25, 2021).
31. Mathur R, Rentsch CT, Morton CE, *et al.* Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. *Lancet* 2021; **397**: 1711–24.
32. Bhaskaran K, Rentsch CT, MacKenna B, *et al.* HIV infection and COVID-19 death: a population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *The Lancet HIV* 2021; **8**: e24–32.
33. Dey T, Mukherjee A, Chakraborty S. A Practical Overview and Reporting Strategies for

Statistical Analysis of Survival Studies. *Chest* 2020; **158**: S39–48.
34. Li J, Scheike TH, Zhang M-J. Checking Fine and Gray subdistribution hazards model with cumulative sums of residuals. *Lifetime Data Anal* 2015; **21**: 197–217.