Descriptive cohort analysis of Long COVID and vaccination status

Version history	Date	Comment
1	2022-12-01	Initial draft created
2	2022-01-27	First draft shared
3	2022-03-20	Edits following first draft comments

Aims & Objectives

Aim
Describe the demographics and clinical characteristics of individuals with primary care coded Long COVID by vaccination status

Objectives
1. Calculate the overall rate of Long COVID diagnoses and referrals in TPP general practices between November 2020 and January 2023, and stratified by vaccination status, demographic and clinical characteristics
2. Describe the temporal trends of Long COVID cases in OpenSAFELY and compare these to the incidence of cases/hospitalisations from COVID-19
3. Describe the pathway to Long COVID cases in OpenSAFELY by analysing the proportion that tested, tested positive or attended hospital prior to recording Long COVID
4. Describe vaccination coverage in those with Long COVID before vaccines were available and compare to national average vaccination coverage

Lay summary
As of January 2023 the Office for National Statistics (ONS) estimated that 2.1 million people in the UK reported experiencing persistent symptoms following COVID-19. In the absence of effective treatments for Long COVID, it is important to understand the possible effect of vaccination on reducing the burden of Long COVID.

Previous research using OpenSAFELY has found fewer records of Long COVID in primary care data when compared to the ONS estimates of the amount of Long COVID in the UK. Our study aims to provide a description of people with and without a record of Long COVID, broken down by vaccination status of individuals. For those with a Long COVID diagnosis before vaccinations were available, we will also describe the proportion vaccinated in this group.

This descriptive study will provide a detailed description of those with and without a record of Long COVID in electronic health records to guide and motivate future observational research of the condition.

Background

A small but notable proportion of the people infected with SARS-COV-2 continue to report ongoing symptoms lasting more than 12 weeks (“long COVID”). The Office for National Statistics estimates that 2.1 million people (3.3% of UK population) have ongoing coronavirus (COVID-19) symptoms as of January 2023 (1). Using a similar definition of Long COVID, a cross-sectional study from the USA found 7.3% of respondents had persistent COVID-19 symptoms (2), a study of US army veterans found an estimated 7% with “post acute sequelae of COVID-19” (3). However, there remains large uncertainty around research into the condition as demonstrated by a longitudinal cohort in the Netherlands where researchers found increases in persistent symptoms following SARS-COV-2 infection in the whole cohort such that only one eighth of patients with at least one Long COVID symptom

could be attributed to having COVID-19 (4). As of November 2020, diagnostic and referral codes have been available for General Practitioners (GPs) in the UK if a patient presents with persistent COVID-19, enabling research in registry data of GP recorded “Long COVID”.

The burden of Long COVID is severe for patients with the condition which has ignited research investment to define the disease and discover effective treatments (5). However, Long COVID is not a single entity which makes conducting rigorous research challenging. One hope is that the increase in SARS-COV-2 vaccination in the UK population will reduce the burden of Long COVID, since the majority of the UK population has had at least three doses of SARS-COV-2 vaccine (6). However, the association between SARS-COV-2 vaccine and Long COVID is very uncertain and mostly reliant on self-reported outcome definition from survey instruments (4,7–10). GP data is a valuable tool for observational research however the discrepancy between GP recorded Long COVID, and estimates from survey data mean that Long COVID data in GP records need to be better described before more formal vaccine analysis can be robustly performed.

Diagnosis of Long COVID is difficult because of a lack of clear biochemical or radiological features, and there are potentially several phenotypes with different presentations (11). Other comorbidities may occur after SARS-COV-2 infection that may be misclassified as Long COVID when a more specific diagnosis is possible. Diagnosing Long COVID is therefore challenging in primary care, however it is unclear why the rate of Long COVID codes in primary care are lower (12) than would be expected compared to estimates from self-reported Long COVID surveys such as the ONS CIS. There remain several unanswered questions about Long COVID cases in primary care records. Firstly, did people with a record of Long COVID or referral to a Long COVID clinic have a severe SARS-COV-2 infection that led to hospitalisation? Secondly, how did the incidence of Long COVID in primary care change over time, and does this correlate with patterns of SARS-COV-2 infections and COVID-19 hospitalisations in the general population? Finally, how have those with Long COVID since 2020 responded to vaccination programmes by receiving the vaccine given the lack of evidence about the impact of vaccination in those with Long COVID? What is the vaccination coverage in this group?

In this study, we will use data from OpenSAFELY-TPP to describe the demographic and clinical characteristics of people with and without Long COVID diagnoses, focussing on various definitions of “vaccinated” including prevalence of booster vaccines, the overall number of doses and type of vaccine. We will take advantage of the longitudinal nature of GP records to present the rates of Long COVID diagnosis over the course of the COVID-19 pandemic and describe the trends in comparison to other outcomes including infection and hospitalisation with SARS-COV-2.

Methods

Dataset description

Primary care records managed by the GP software provider, TPP will be linked to ONS death data 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). 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 large volumes of potentially disclosive pseudonymized patient data to be transferred off-site. This, in addition to other technical and organisational controls, minimises 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 under "Information governance".

Study design and population

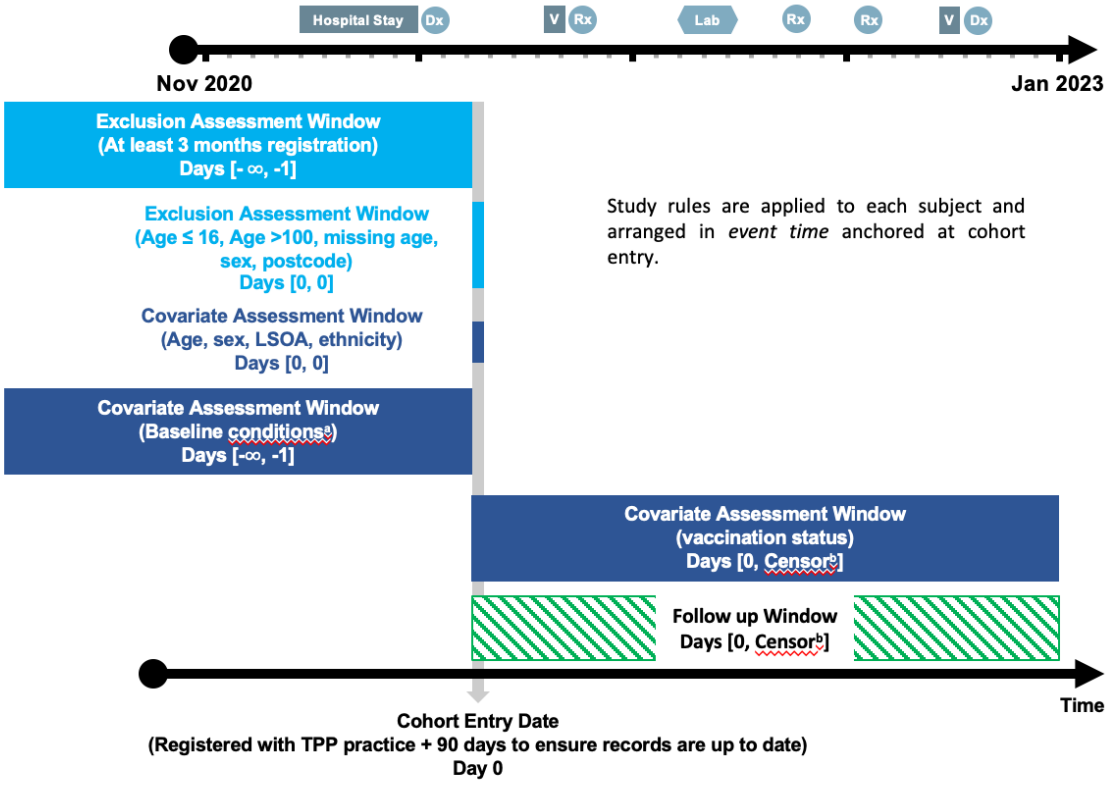
This study will include two separate population based retrospective cohorts of adults registered with a TPP general practice. These will be descriptive cohorts only so there will be one group of individuals without a comparison arm in each cohort, descriptive comparisons will only be made across pre-specified stratifiers.

The primary cohort (Figure 1) (13) will include all individuals registered with an eligible general practice from the 1st November 2020 onwards (the date that Long COVID codes became available in GP software). We will define their demographic and clinical characteristics at baseline. We will then follow these individuals until the earliest of:

- Long COVID diagnosis or referral (see outcome)
- End of registration with general practice
- End of study period (January 2023)
- Death

Figure 1: Primary cohort for the descriptive study of Long COVID in OpenSAFELY

Individual-patient data is documented as encounters from various sources, including diagnoses/procedures (Dx/Px), drug dispensings (Rx), laboratory tests (Lab), visits (V), or hospital stays. It is arranged in *calendar time*.



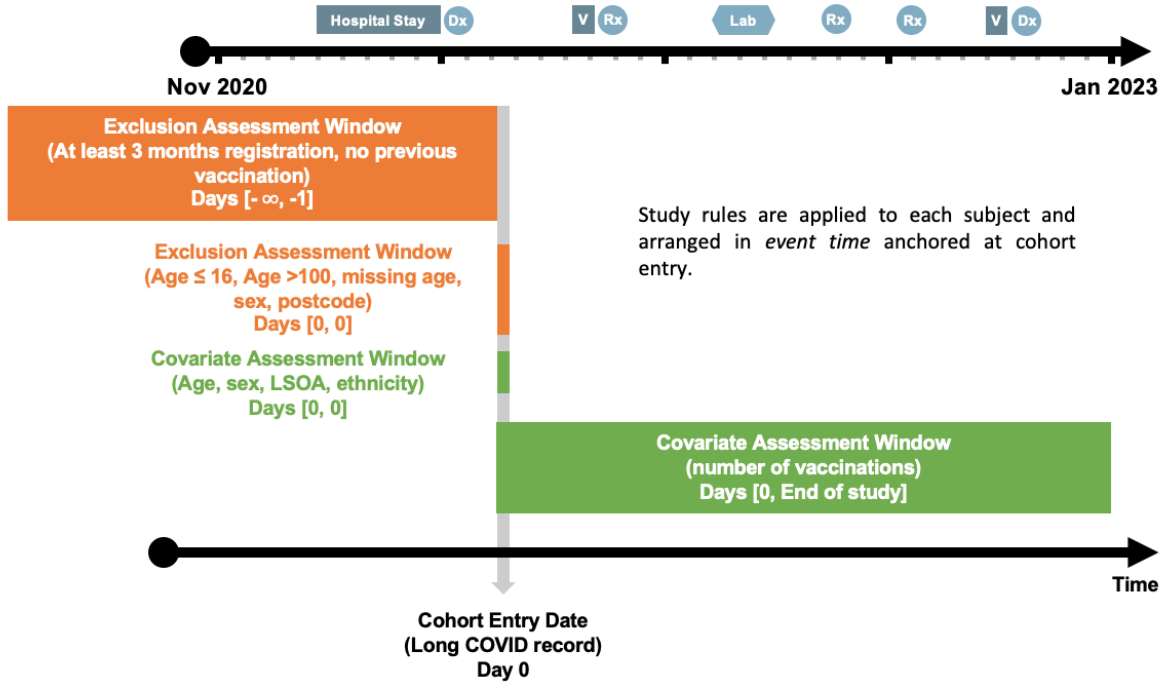
a. Baseline conditions included: post-viral fatigue, High risk category for developing complication from COVID-19, number of comorbidities
b. Earliest of: outcome of interest (Long COVID), end of registration, death, end of the study period

LSOA: Lower super output area

Our primary cohort ends follow-up at the time of Long COVID diagnosis (or referral) as the outcome under investigation. However, to meet objective 4, a secondary cohort that begins follow-up from the time of Long COVID diagnosis (or referral) will be developed. The secondary cohort will include all patients with a record of Long COVID before they had a vaccination. Individuals will enter the cohort at the time of a diagnosis (or referral) for Long COVID. Individuals will be excluded if they have already received a vaccination at this point since this will reduce the number of further vaccinations they can receive and affect their likelihood of accepting further vaccination. We will include all individuals who meet this criteria from 1st November 2020 to 22nd October 2021, when the first booster vaccination was delivered (14). We will include any Long COVID record in the patient's history although we expect only a minority of records to be backdated prior to November 2020 when the codes were made available (12). We will then follow up this cohort until January 2023 and measure the number of vaccine doses these individuals receive over the study period. We will calculate the 'risk' of vaccination in this cohort and any loss to follow up (e.g. patients transferring from a TPP GP practice) will be excluded in the calculation of the 'risk'.

Figure 2: Secondary cohort. Assessing vaccination coverage in people with a record of Long COVID

Individual-patient data is documented as encounters from various sources, including diagnoses/procedures (Dx/Px), drug dispensings (Rx), laboratory tests (Lab), visits (V), or hospital stays. It is arranged in *calendar time*.



LSOA: lower super output area

Inclusion criteria

Primary cohort: aged between 18 and 100 years old and registered with a TPP practice during the study period (Nov 2020 - Jan 2023).

Secondary cohort: aged between 16 and 100 years old and registered with a single TPP practice and a record of Long COVID before any record of vaccination during the study period (prior to 2021-10-22).

Exclusion criteria

Individuals will be excluded from either cohort if they meet any of the following conditions:

- Less than 3 months registered at their current GP at baseline. This is to ensure that records are up to date and accurate at enrolment to the cohort
- Multiple records with multiple TPP GPs in the study period. It is unclear how to handle multiple registrations in the cohort therefore excluding these individuals, which we expect to be a small minority
- Missing age, or a recorded age that is under 16 years or over 100 years on 1st November 2020, missing sex, or missing postcode

In addition, for the primary cohort, if an individual has a record of Long COVID prior to their registration start date they will be excluded.

In addition, for the secondary cohort, individuals who have a contraindication for COVID-19 vaccination will be excluded (15,16).

Study measures

1. Rates of Long COVID per 100,000 person years
2. ‘Risk’ of at least two vaccine doses administered following Long COVID

Exposure

Primary cohort: N/A. All individuals eligible for inclusion in the cohort will be considered “at-risk” of Long COVID.

Secondary cohort: individuals with a record of Long COVID as defined under Outcome will be considered exposed and included in the study. There will not be a comparator unexposed arm in this descriptive study.

Outcome

Primary cohort - Long COVID outcomes summary

Below is a summary Table of the three different definitions of Long COVID to be used in this study:

Long COVID outcome	Definition
1 - primary outcome	Any Long COVID diagnosis or referral code
2 - Long COVID by code type	1 - Long COVID diagnosis 2 - Long COVID referral
3 - Long COVID by prior COVID-19 status	1 - Long COVID record in primary care following hospitalisation with COVID-19 (we will also analyse hospitalisations *for* COVID-19 if it is possible in linked data) 2 - Long COVID record in primary care following a positive SARS-COV-2 test (SGSS) 3 - Long COVID record in primary care with no further information Groups 1 and 2 will be a subset of group 3

Primary cohort - primary outcome - Long COVID

The primary outcome will be Long COVID as defined by 15 SNOMED-CT codes as used previously (12).

Primary cohort - Long COVID referral (Rx) versus diagnosis (Dx)

We will further describe Long COVID using in more detail for secondary analyses. Firstly we will divide the Long COVID codelist into two categories:

1. Long COVID Dx (17)

1325161000000102	Post-COVID-19 syndrome
1325181000000106	Ongoing symptomatic disease caused by severe acute respiratory syndrome coronavirus 2

2. Long COVID Rx for assessment and/or treatment (18,19)

1325021000000106	Signposting to Your COVID Recovery
1325031000000108	Referral to post-COVID assessment clinic
1325041000000104	Referral to Your COVID Recovery rehabilitation platform
1325051000000101	Newcastle post-COVID syndrome Follow-up Screening Questionnaire
1325061000000103	Assessment using Newcastle post-COVID syndrome Follow-up Screening Questionnaire
1325071000000105	COVID-19 Yorkshire Rehabilitation Screening tool
1325081000000107	Assessment using COVID-19 Yorkshire Rehabilitation Screening tool
1325091000000109	Post-COVID-19 Functional Status Scale patient self-report
1325101000000101	Assessment using Post-COVID-19 Functional Status Scale patient self-report
1325121000000105	Post-COVID-19 Functional Status Scale patient self-report final scale grade
1325131000000107	Post-COVID-19 Functional Status Scale structured interview final scale grade
1325141000000103	Assessment using Post-COVID-19 Functional Status Scale structured interview
1325151000000100	Post-COVID-19 Functional Status Scale structured interview

Primary cohort - Long COVID conditional on preceding COVID-19 outcomes

We will also define Long COVID outcomes dependent on previous SARS-COV-2 history. We expect data to be missing on SARS-COV-2 tests taken and positive results so this will be restricted to a secondary analysis. All records of Long COVID as defined previously will be divided into the following groups:

1. Long COVID record in primary care following hospitalisation with COVID-19
2. Long COVID record in primary care following a positive SARS-COV-2 test (SGSS)
3. Long COVID record in primary care with no further information

Definition of hospitalisation with COVID-19. Data on Hospital Episode Statistics are available including the ICD-10 codes recorded during non-emergency hospital admission. Hospitalisation with COVID-19 will be identified from Hospital Episode Statistics Admitted

Patient Care (HES-APC) data available in OpenSAFELY-TPP. The following ICD-10 codes (29) will be used to define COVID-19 hospitalisations (at any point in the hospitalisation record, i.e., not primary diagnosis only):

icd10_code	
U071	covid19 virus identified
U072	covid19 virus not identified

Although ICD-10 code U072 refers to SARS-COV-2 *not* being identified, this code is used when COVID-19 is suspected clinically or epidemiologically but the laboratory test is inconclusive or negative (see this discussion).

Hospitalisation for COVID-19 will be defined the same way but only include records in the "primary diagnosis" of the hospitalisation record.

All previous SARS-COV-2 test results and/or hospitalisation records at least 12 weeks prior to the Long COVID record will be considered when classifying Long COVID outcomes. In the case of hospitalisations the 12 weeks window will begin from the date of discharge not admission.

Secondary cohort - SARS-COV-2 vaccination

For the secondary cohort, the outcome being measured is vaccination against SARS-COV-2. Vaccine administration details are recorded in the National Immunisation Management Service (NIMS) and electronically transmitted to every individual's GP record on a daily basis and is available via the `vaccinations` table in OpenSAFELY-TPP.

In addition and to ensure that all vaccinations are recorded, vaccine data will be supplemented with SNOMED-CT clinical codes for administered vaccines if there is no record from NIMS. This supplementary data will include codelists defined in previous OpenSAFELY research (20). The codelists used to capture vaccinations in clinical records will be:

- Astra Zeneca vaccination medication code (21)
- Moderna vaccination medication code (22)
- PFIZER/BIONTEC vaccination medication code (23)
- First COVID vaccination administration codes (24)
- Second COVID vaccination administration codes (25)
- administration of 3rd COVID vaccine (26)
- administration of 4th COVID vaccine (27)
- administration of 5th COVID vaccine (28)

The outcome used in the analysis will be the total number of vaccine doses an individual received at the end of the study period categorised as 1, 2, or 3+.

Comparison outcomes

We want to compare the rates of Long COVID in primary care to other metrics of the COVID-19 pandemic to examine any differences in temporal trends. The rates of Long COVID will be defined using the primary outcome as above. These will be calculated by calendar month over the study period and compared to monthly rates of:

1. SARS-COV-2 positive tests - available from the Second Generation Surveillance System (SGSS) table in OpenSAFELY-TPP. SGSS contains information on patients receiving a swab test for SARS-CoV-2, from Pillar 1 (NHS and PHE labs) and Pillar 2 (commercial partners).
2. Hospitalisation with COVID-19. This will be defined as in the section "Long COVID conditional on preceding COVID-19 outcomes" above.

Stratifiers

Our pre-specified variables that we will calculate stratum specific rates in will be as described as below.

Stratifier	**Definition**
Number of vaccination doses	Time updated stratification variable where individuals move from 0 to 6 doses as they become vaccinated (as defined above)
Homologous/heterologous initial vaccine schedule	Vaccines distributed in the UK were categorised as either mRNA (Pfizer. Comirnarty, Moderna, Spikevax) or not. Recipients of vaccines will be categorised as homologous (if both the first two vaccines received were of the same type (mRNA or non-mRNA) and heterologous if not.
Age	Age at study entry
Sex	Male or female, excluding individuals with missing data
Region	Geographical area as categorised by 9 English regions categorised as North East, North West, Yorkshire and the Humber, East Midlands, West Midlands, East, London, South East, South West
Ethnicity	Five categories, obtained from 16 (White, Black, South Asian, Mixed, Other). Assessed from primary care records and supplemented with HES APC records (30)
Deprivation	Index of multiple deprivation (IMD) quintile based on lower super output area
History of post-viral fatigue	Binary variable. Codelist is available (31) from previous Long COVID research
High risk category for developing complication from COVID-19	Binary variable. Codelist is available (32) from previous vaccine research (33) -ref-
Level of multimorbidity	Categorised as 0, 1, 2+. Comorbidities are assessed at study entry. Relevant comorbidities will be defined based on previous research of risk factors for Long COVID in OpenSAFELY (34). A previous code 6 months to 5 years before March 2020 for one or more

Stratifier	Definition
	of: diabetes; cancer; haematological cancer; asthma; chronic respiratory disease; chronic cardiac disease; chronic liver disease; stroke or dementia; other neurological condition; organ transplant; dysplasia; rheumatoid arthritis, systemic lupus erythematosus or psoriasis; or other immunosuppressive conditions. Those with no relevant code for a condition will be assumed not to have that condition. Number of conditions were categorised into “0”, “1”, and “2 or more”..

Statistical analysis

Main analysis

Primary cohort

Estimate the rate of Long COVID per 100,000 person years and 95% confidence intervals using Poisson regression. For stratified estimates each stratifier will be included as a single predictor variable in the regression and stratum specific estimates presented.

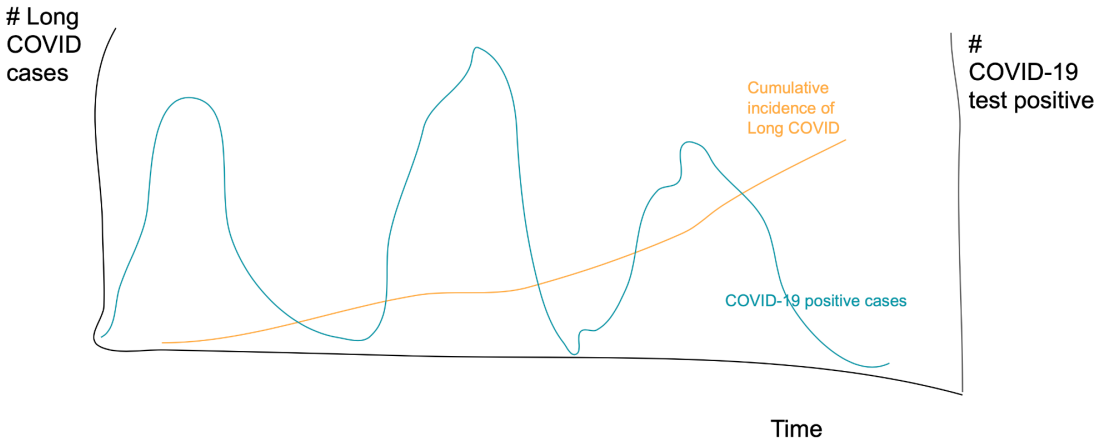
Age- and sex-adjusted models: estimates are likely to be strongly influenced by age and sex so we will include these as additional variables to estimate minimally-adjusted rate ratios for the effect of vaccination and other stratifiers on Long COVID.

Secondary cohort
The proportion of eligible individuals (those with Long COVID first) that receive at least two doses of COVID-19 vaccination will be presented with the exact 95% confidence interval.

Secondary analysis

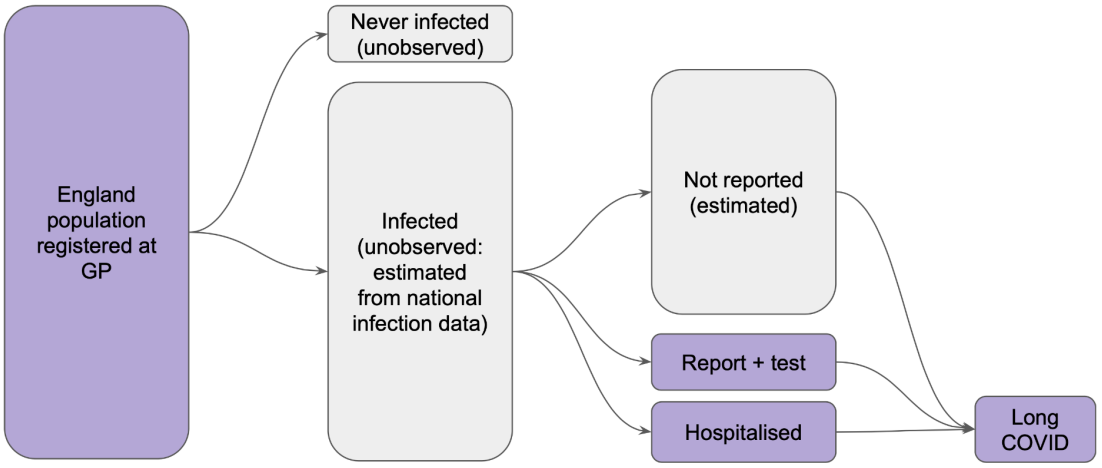
1. Data visualisation of pandemic trends

The monthly rate of Long COVID records in primary care will be plotted over the full study period (Nov 2020 - Jan 2023). The weekly rate of COVID-19 positive tests from SGSS data, and COVID-19 hospitalisations from HES-APC. This will be a purely descriptive exercise to qualitatively assess whether records of Long COVID in primary care follow the dynamics of the COVID-19 epidemic in the UK. Rates stratified by age bands will also be presented (<40, 41-50, 51-60, 61-70, 71+).



2. Describe the COVID-19 continuum for the primary cohort

Record the flow of individuals in the primary cohort from at-risk to Long COVID record, using the varying definitions of Long COVID and whether the individual had tested or been hospitalised with COVID-19 prior to Long COVID. This will be supplemented with SGSS and HES-APC data to plot the flow of people. This will be affected over time by availability of testing and circulating SARS-COV-2 variants, so if possible this will be repeated for different phases of the pandemic depending on the dominant variant in circulation.



3. Compare Long COVID vaccine coverage to national averages

The vaccination coverage in the secondary cohort - people with Long COVID prior to vaccination - will be estimated as a risk with the number of people with either 1, 2, or 3+ vaccine doses at the end of the study as the numerator, and the number of people eligible for inclusion in the cohort as the denominator.

We will compare the vaccination coverage in this group to national averages available from the ONS. Given the expected differences in the age distribution of this cohort we will standardise our estimates using the English population age distribution as the standardised population to make a more equal comparison between the two groups.

Sample size and precision

Our sample size is fixed by the number of individuals attending TPP practices so we calculated the precision of estimated rates from our study according to the formula $n = \frac{\mu}{\sigma^2}$ where n is the sample size, μ is the expected rate and σ is the standard error (36). Assuming that the the 95% confidence interval from our estimated rate will be approximately 2σ we report here the expected precision of estimates from this study.

For overall estimates, if we assume that there are 20 million individuals with TPP data available in our primary cohort and that they are followed for 2 years there are 40 million person years of follow up. If we conservatively assume that there are 40,000 Long COVID records in TPP over the full study period based on previous research (37), we will be able to estimate a rate of 100 Long COVID events per 100,000 person years with ± 1 precision.

We expect Long COVID incidence to vary over time. Early in the pandemic when fewer events were recorded, our precision will be lower but still acceptable. Taking December 2020 as an example with expected low incidence, if there are 200 events (12) and 20 million people at-risk for the month then we would estimate a rate of 12 per 100,000 person years with ± 1.7 precision.

In stratified estimates the sample size in each stratum will be lower and precision of monthly estimates will be lower. Focussing on the number of vaccination doses and taking April 2021 as an example, UK data shows that approximately 4 million people had received two vaccine doses by 1st April (14) so if 40% of people are covered by TPP practices there will be 1.6 million individuals in this group in our study. If these are largely older individuals with fewer records of Long COVID and we assume there are 80 events in April 2021 in this group then we would estimate a rate of 60 per 100,000 person years with ± 13.4 precision.

Limitations

This study will be limited from drawing any causal or clinical conclusions about the role of vaccinations in the prevention of Long COVID because it is designed to be descriptive. The aim is to describe interesting characteristics of the population with a primary care code of Long COVID and temporal trends of how coding and incidence of Long COVID has changed over the past 2 years.

There is likely to be a large amount of non-differential Long COVID misclassification (38). Given the large disparity between previous estimates of Long COVID risk in OpenSAFELY and estimates from ONS CIS (10,12) then it is likely many people who self-report having Long COVID will not have a record from their GP. It is also possible that people with a code for Long COVID in their GP records do not have the condition (e.g., they have an alternative undiagnosed condition that is misrecorded as Long COVID). This is a limitation in studying a newly emerging condition and using routine data as clinical practice evolves rapidly in real-time and will be acknowledged accordingly in any publication. We plan to learn more about the reasons for recording Long COVID in primary care through this study, especially whether it is due to prior severe COVID-19 (see secondary analysis 2 above) to help inform

future research using routine data. Finally, it is likely that some people will recover from Long COVID during this study period. There is a SNOMED code for recovery but we think it is unlikely that people will consult their GP if they feel recovered after Long COVID so we expect this code to be rare. However, in our primary cohort follow-up ends when Long COVID is recorded in the patient's record so the duration of Long COVID symptoms is outside the scope of this work.

There are other limitations on using the codelists described to capture Long COVID. The code was not available until November 2020 and some people will develop the condition before this. It is possible for clinicians to backdate diagnoses but we do not expect this to be common. We will also not capture people who recover from Long COVID or symptoms during the study period.

When drawing comparisons between the rate of Long COVID records and COVID-19 disease incidence our comparisons will be limited by testing availability and behaviour so generalising our results to the population will be limited. In addition, uncertainty about the timing of Long COVID recording removes the possibility to make inferences about the variant of SARS-COV-2 that caused the Long COVID.

There will be loss to follow up in our analysis of vaccine coverage in people who record Long COVID before vaccination. If this selection bias is differential (e.g., if unvaccinated are more likely to die) then our estimate of vaccine coverage in people with Long COVID may be an overestimate.

Administrative

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. 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. 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. Together, these provide the legal basis 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.

Ethics

This research is part of the OpenPROMPT study “Quality-of-life in patients with long COVID: harnessing the scale of big data to quantify the health and economic costs” which has ethical approval from HRA and Health and Care Research Wales (HCRW) (IRAS project ID 304354).

References

1. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics [Internet]. [cited 2023 Jan 11]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/5january2023
2. Robertson MM, Qasmieh SA, Kulkarni SG, Teasdale CA, Jones H, McNairy M, et al. The epidemiology of long COVID in US adults. Clin Infect Dis. 2022 Dec 21;ciac961.
3. Xie Y, Bowe B, Al-Aly Z. Burdens of post-acute sequelae of COVID-19 by severity of acute infection, demographics and health status. Nat Commun. 2021 Nov 12;12(1):6571.
4. Ballering AV, Zon SKR van, Hartman TC olde, Rosmalen JGM. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. The Lancet. 2022 Aug 6;400(10350):452–61.
5. Owens B. How “long covid” is shedding light on postviral syndromes. BMJ. 2022 Sep 21;o2188.
6. Vaccine Coverage | OpenSAFELY: Reports [Internet]. [cited 2023 Jan 27]. Available from: http://reports.opensafely.org/reports/vaccine-coverage/
7. Taquet M, Dercon Q, Harrison PJ. Six-month sequelae of post-vaccination SARS-CoV-2 infection: A retrospective cohort study of 10,024 breakthrough infections. Brain Behav Immun. 2022 Jul 1;103:154–62.
8. Antonelli M, Penfold RS, Merino J, Sudre CH, Molteni E, Berry S, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. Lancet Infect Dis. 2022 Jan 1;22(1):43–55.
9. Mizrahi B, Sudry T, Flaks-Manov N, Yehezkelli Y, Kalkstein N, Akiva P, et al. Long covid outcomes at one year after mild SARS-CoV-2 infection: nationwide cohort study. BMJ. 2023 Jan 11;380:e072529.
10. Ayoubkhani D, Bermingham C, Pouwels KB, Glickman M, Nafilyan V, Zaccardi F, et al. Trajectory of long covid symptoms after covid-19 vaccination: community based cohort study. BMJ. 2022 May 18;377:e069676.
11. Lancet T. Understanding long COVID: a modern medical challenge. The Lancet. 2021 Aug 28;398(10302):725.
12. Walker AJ, MacKenna B, Inglesby P, Tomlinson L, Rentsch CT, Curtis HJ, et al. Clinical coding of long COVID in English primary care: a federated analysis of 58 million patient records *in situ* using OpenSAFELY. Br J Gen Pract. 2021 Nov;71(712):e806–14.
13. Schneeweiss S, Rassen JA, Brown JS, Rothman KJ, Happe L, Arlett P, et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. Ann Intern Med. 2019 Mar 19;170(6):398–406.
14. Vaccinations in the UK | Coronavirus in the UK [Internet]. 2023 [cited 2023 Jan 27]. Available from: https://coronavirus.data.gov.uk/details/vaccinations

15. COVID-19: the green book, chapter 14a [Internet]. GOV.UK. 2022 [cited 2023 Feb 15]. Available from: https://www.gov.uk/government/publications/covid-19-the-green-book-chapter-14a
16. OpenCodelists: COVID-19 Vaccination - Contraindication or Allergy [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/opensafely/covid-19-vaccination-contraindication-or-allergy/59f7a1a0/
17. OpenCodelists: NICE managing the long-term effects of COVID-19 [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/opensafely/nice-managing-the-long-term-effects-of-covid-19/64f1ae69/
18. OpenCodelists: Referral and signposting for long COVID [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/opensafely/referral-and-signposting-for-long-covid/12d06dc0/#full-list
19. OpenCodelists: Assessment instruments and outcome measures for long covid [Internet]. Available from: https://www.opencodelists.org/codelist/opensafely/assessment-instruments-and-outcome-measures-for-long-covid/79c0fa8a/#full-list
20. Curtis HJ, Inglesby P, Morton CE, MacKenna B, Green A, Hulme W, et al. Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records *in situ* using OpenSAFELY. Br J Gen Pract. 2022 Jan;72(714):e51–62.
21. OpenCodelists: Astra Zeneca vaccination medication code [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/azdrx/v1.5.3/
22. OpenCodelists: Moderna vaccination medication code [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/modrx/v1.5.3/
23. OpenCodelists: PFIZER/BIONTEC vaccination medication code [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/pfdrx/v1.5.3/
24. OpenCodelists: First COVID vaccination administration codes [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/covadm1/v.1.5.3/
25. OpenCodelists: Second COVID vaccination administration codes [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/covadm2/v.1.5.3/
26. OpenCodelists: administration of 3rd COVID vaccine [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/covadm3_cod/v.1.5.3/
27. OpenCodelists: administration of 4th COVID vaccine [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/covadm4_cod/v.1.5.3/
28. OpenCodelists: administration of 5th COVID vaccine [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/covadm5_cod/v.1.5.3/
29. OpenCodelists: COVID Identification [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/opensafely/covid-identification/2020-06-03/
30. OpenCodelists: ethnicity (SNOMED) [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/opensafely/ethnicity-snomed-0removed/2e641f61/
31. Clinical coding of long COVID in English primary care: a federated analysis of 58 million patient records in situ using OpenSAFELY [Internet]. OpenSAFELY; 2022 [cited 2023 Feb 15]. Available from: https://github.com/opensafely/long-covid/blob/d44629d020e2687171264d6b42bb81a6aa9

2f07e/codelists/user-alex-walker-post-viral-syndrome.csv
32. OpenCodelists: High Risk from COVID-19 code [Internet]. [cited 2023 Feb 15]. Available from: https://www.opencodelists.org/codelist/primis-covid19-vacc-uptake/shield/v1/#full-list
33. Curtis HJ, Inglesby P, MacKenna B, Croker R, Hulme WJ, Rentsch CT, et al. Recording of 'COVID-19 vaccine declined': a cohort study on 57.9 million National Health Service patients' records in situ using OpenSAFELY, England, 8 December 2020 to 25 May 2021. Eurosurveillance. 2022 Aug 18;27(33):2100885.
34. Thompson EJ, Williams DM, Walker AJ, Mitchell RE, Niedzwiedz CL, Yang TC, et al. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. Nat Commun. 2022 Jun 28;13(1):3528.
35. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020 Aug 20;584(7821):430–6.
36. Kirkwood BR, Sterne JAC, Kirkwood BR. Essential medical statistics. 2nd ed. Malden, Mass: Blackwell Science; 2003. 501 p.
37. Clinical coding of long COVID in English primary care: a federated analysis of 58 million patient records in situ using OpenSAFELY | OpenSAFELY: Reports [Internet]. [cited 2023 Jan 27]. Available from: http://reports.opensafely.org/reports/clinical-coding-of-long-covid-in-english-primary-care-a-federated-analysis-of-58-million-patient-rec/
38. Amin-Chowdhury Z, Ladhani SN. Causation or confounding: why controls are critical for characterizing long COVID. Nat Med. 2021 Jul;27(7):1129–30.