The Healthcare Utilisation of People with Long COVID

Lay summary

After recovering from coronavirus infections, some people experience symptoms such as fatigue, shortness of breath, or problems with memory or concentration. This is known as long COVID. According to a survey published in December 2022 by the Office for National Statistics, two million people, nearing 3.3% of the total UK population, had long COVID. However, the impact on the healthcare system has yet to be quantified. Our study aims to investigate the healthcare utilisation of people with long COVID and factors associated with increased healthcare resource use.

We will identify people with long COVID using their electronic health records and analyse their primary and secondary resource use. In addition, we will explore the potential factors associated with increased healthcare utilisation among people with long COVID and compare their healthcare utilisation to historical records before the pandemic.

Background

After being infected with SARS-CoV-2, some common symptoms, such as fever, cough, shortness of breath, and changing or losing the sense of smell/taste, usually resolve in four weeks; however, some people have persistent symptoms for longer than four weeks. The National Institute for Health and Care Excellence (NICE) defines people with signs and symptoms of COVID-19 lasting from four to 12 weeks as "ongoing symptomatic COVID-19" and those who have continuous symptoms for more than 12 weeks as "post-COVID-19 syndrome." According to NICE guidelines, ongoing symptomatic COVID-19 and post-COVID-19 syndrome refer to long COVID. The common symptoms of long COVID include weakness, general malaise, fatigue, concentration impairment (known as "brain fog"), and breathlessness (1). According to the Office for National Statistics, about two million people, in December 2022, approximately 3.3% of the UK population, were reported to have long COVID symptoms (2).

Due to various long-term symptoms after the acute phase of COVID-19, people may increase their healthcare utilisation. A retrospective matched cohort study in the US estimated the healthcare utilisation before and after testing positive for SARS-CoV-2 using electronic health records (EHRs) from eight healthcare systems. They used the difference-in-difference method to compare individuals with COVID-19 to people without testing positive and reported a 4% increase in healthcare utilisation six months after testing positive for SARS-CoV-2 (ratio in rate ratio = 1.04, 95% confidence interval(CI): 1.03-1.05) (3). Another study using CPRD in the UK analysed the healthcare contact rate after receiving COVID-19 diagnoses. They reported that the primary care consultation rate among people who had COVID-19 was 18% higher than one year before being diagnosed with COVID-19 (hazard ratio (HR): 1.18, 95% CI=1.17-1.19), and the overall healthcare use also increased by 15% than historical records (HR=1.15, 95% CI=1.14-1.15) (4). A study followed 1,190 COVID-19 survivors after they were discharged from a hospital in China, indicating that people with long COVID symptoms at two years were three times more likely to visit outpatient clinics (odds ratio(OR): 2.82, 95% confidence interval:1.99–4.0), and had a 60% increased risk of being re-hospitalised (OR:1.64, 95%CI:1.12–2.41) (5).

Although some studies indicate an increase in healthcare utilisation caused by COVID-19 symptoms and sequelae, it is unclear that healthcare utilisation would increase among people with clinically diagnosed long COVID who were not hospitalised for their acute COVID-19 illness. In addition, the financial burden of the healthcare utilisation of people with long COVID in the community has yet to be quantified. Therefore, our study aims to investigate the healthcare utilisation of long COVID in patients not hospitalised for COVID-19 and the factors associated with increased healthcare resource use.

Conceptual framework

The conceptual framework of our study is shown in **Figure 1.** We hypothesise that healthcare resource use would increase among people with long COVID due to persistent symptoms and worsening underlying health conditions. In contrast, people without long COVID would maintain their baseline healthcare use. Consequently, we plan to compare people with long COVID to those without long COVID to estimate additional healthcare utilisation.

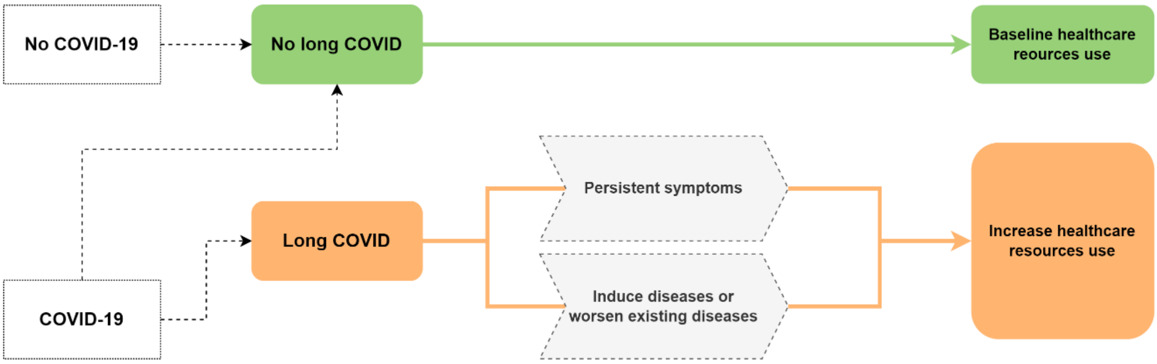


Figure 1. The conceptual framework of the long COVID healthcare utilisation study. We would like to assess the association between long COVID and possible increased healthcare resource use. Elements including “COVID-19”, “no COVID-19”, “persistent symptoms”, and “induce diseases or worsen existing diseases” are considered unmeasured factors in this study.

Aim

To investigate the healthcare utilisation of people with long COVID and the factors associated with the change in healthcare resource use.

Objectives and research questions

1. To identify people with long COVID from OpenSAFELY–TPP.
    - Who had long COVID diagnoses in their EHR?
2. To assess the healthcare utilisation and the costs of individuals with long COVID.
    - What are the frequencies of primary care contacts, A&E visits, primary care prescriptions and secondary care referrals of people with long COVID?
    - Are healthcare utilisation and costs higher among people with long COVID than among people without long COVID?
3. Have people with long COVID been high healthcare resource users in their historical records, compared to people without long COVID before the COVID-19 pandemic?
    - Can the difference in healthcare utilisation between people with/without long COVID be explained by baseline health-seeking behaviours before the pandemic?
4. To explore possible factors associated with increased healthcare utilisation among individuals with long COVID.
    - What are the demographic, socioeconomic and clinical factors associated with increased healthcare utilisation among people with long COVID?

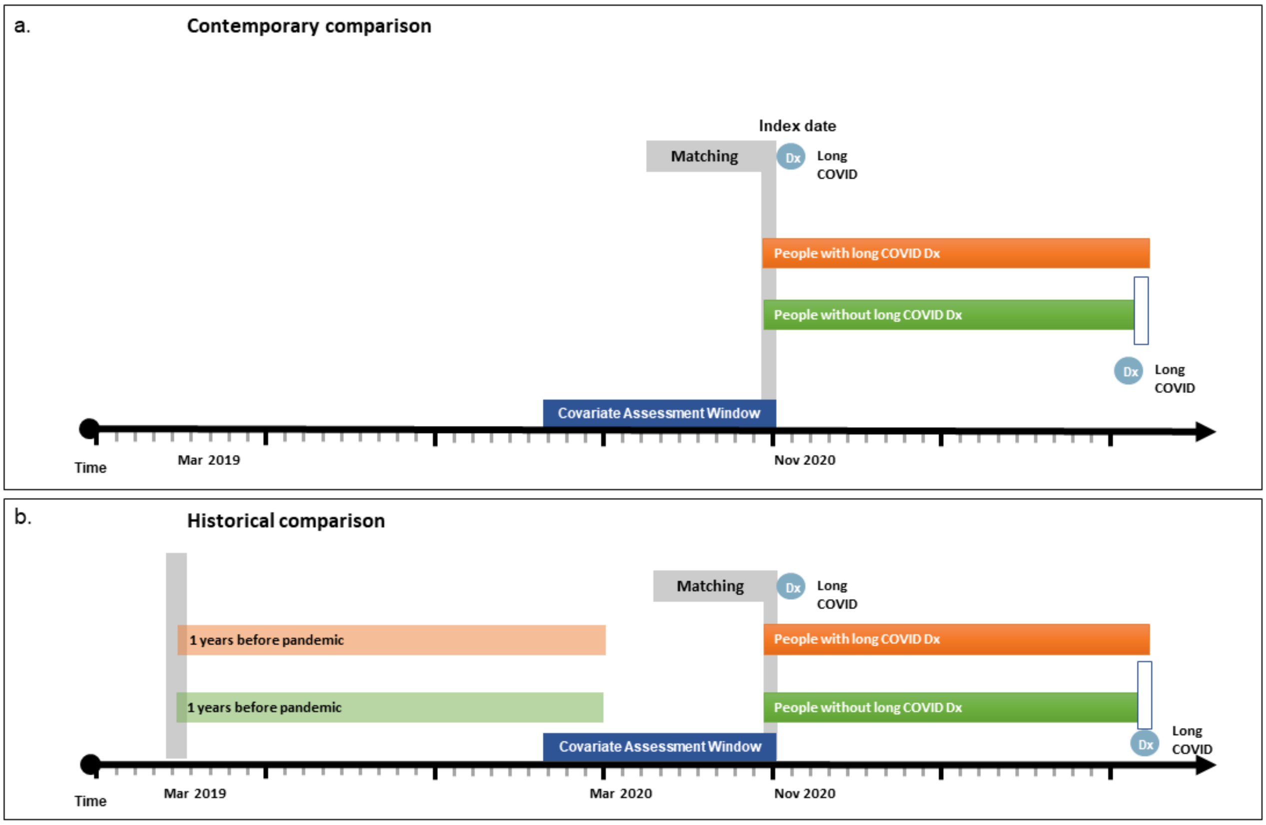
Methods

Data source: OpenSAFELY platform

Primary care records managed by the GP software provider TPP were 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 pseudonymised 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 pseudonymised patient data to be transferred off-site. This, in addition to other technical and organisational controls, minimises any risk of re-identification. Similarly, pseudonymised 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 pseudonymised data such as coded diagnoses, medications and physiological parameters. No free text data are included.

Study design

We will undertake a matched cohort study and a series of analyses at different time frames to address our research questions. First, we will perform a contemporary, between-person analysis. We will identify patients with and without long COVID and



compare their subsequent healthcare utilisation (**Figure 2a**). Second, we will undertake a within-person comparison of healthcare utilisation in the two years prior to the pandemic with up to one year after long COVID diagnoses by performing a difference-in-difference analysis to compare the change in utilisation between those with and without long COVID (**Figure 2b**).

Figure 2. Study design overview. a. A between-persons contemporary comparison. People with long COVID diagnoses will be matched with people without long COVID by sex, age and region. We will follow up with both groups from the index date, the date that the exposure group received a long COVID diagnosis until people died or deregistered from the practice. The comparator group will also be censored if they got long COVID; 2. a historical comparison. For people in both groups with historical registration data, we will match them by sex, age and region and analyse their historical healthcare utilisation.

Inclusion and exclusion criteria

Figure 3 is a diagram showing the selection of participants in our study. For the contemporary assessment, we will include adults (aged over 18) registered at GP practices using TPP software on 1 November 2020 who had been registered for at least three months. We exclude all patients registered at practices that have not used at least one long Covid code since November 2020 to prevent introducing bias due to unusual coding practices (6).

For the historical comparison, we will include people from the contemporary comparison group registered at GP practices using TPP software between November 2018 and November 2019. People who did not have historical records on 1 November 2018 or are under 18 years old will be excluded (**Figure 3**, **Figure 4**).

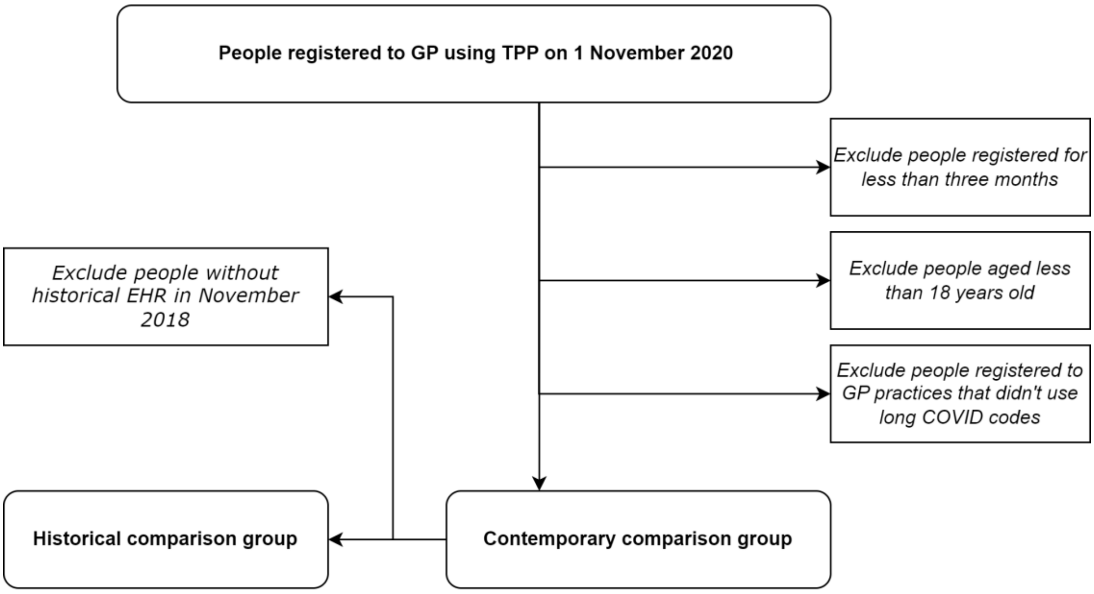


Figure 3. The diagram of participants' selection of contemporary and historical comparison groups.

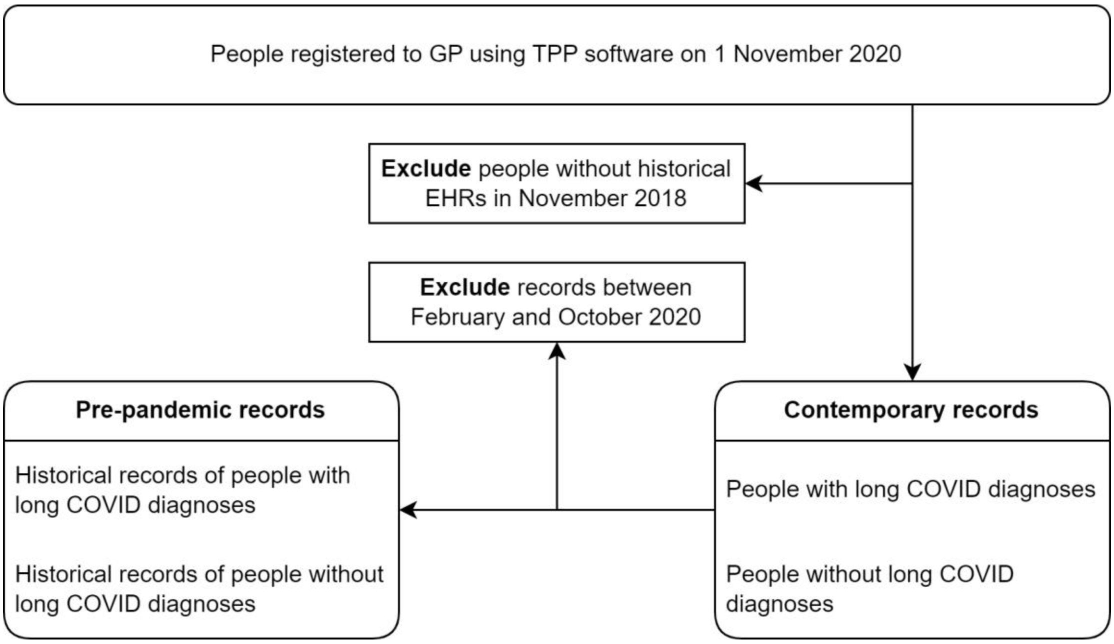


Figure 4. The diagram of the participants' selection of the historical comparison groups.

Exposure and comparator group definition

The exposure group will be people with recorded long COVID diagnoses. We will identify people with long COVID using a list of SNOMED-CT codes for long COVID (6).

The comparator group will be people without recorded long COVID diagnoses, identified by matching their sex, age and region to the exposure group (**Figure 2**).

Outcomes

We will first evaluate the monthly rate of healthcare resource utilisation. The healthcare utilisation will include 1. consultation with GPs; 2. all-cause hospital admission; 3. accident and emergency (A&E) visit; 4. all-cause hospital outpatient visits. Second, we will estimate the approximate cumulative cost of healthcare utilisations, including GP consultations, GP prescriptions, hospitalisation costs, A&E costs, and hospital outpatient costs. The estimated cost of a GP consultation is #41 (7), and we will estimate the prescription cost by its BNF chapter (8). The costs of secondary care, including hospitalisation, A&E, and outpatient clinic visits, will be obtained from the Secondary Uses Services (SUS). We will display the raw and relative costs.

Covariates

Covariates will be decided through literature. We will include demographic characteristics, including age, sex, and ethnicity; socioeconomic factors such as region and Index of Multiple Deprivation (IMD); underlying chronic diseases; and the number of COVID-19 vaccination doses (any vaccine) before the index date. The underlying chronic diseases include asthma, obesity/overweight, previous psychiatric conditions, and other comorbidities (9). Other comorbidities will be defined as the level of multimorbidity, including any of the chronic diseases listed in **Table 1**.

Disease
Non-haematological cancer
Haematological cancer [1]
Chronic respiratory disease
Chronic cardiac disease
Chronic liver disease
Stroke or dementia
Other neurological conditions [2]
Organ transplant
Rheumatoid arthritis
Systemic lupus erythematosus
Psoriasis
Other immunosuppressive conditions [3]

1. Having haematological cancers six months before index date; 2. Such as Huntington's disease, multiple sclerosis, motor neuron diseases, and other neurological diseases; 3. Including other permanent and temporary immunosuppressive diseases.

Follow-up

The index date will be the time when the exposed individuals received a long COVID diagnosis after November 2020. We will start follow-up for exposure and comparator groups from the index date and censor at 1. Date of death; 2. end of GP registration; 3. receipt of a resolved long COVID SNOMED code (1326351000000108) among the exposed group; 4. receipt of a long COVID diagnosis among the control group; 5. 31st January 2023. For the historical comparison, eligible participants will be followed from March 2019 to March 2020.

Statistical analysis

Descriptive analysis

We will compare the distribution of demographic factors, underlying comorbidities, and socioeconomic factors in the exposure group and comparators. The mean and standard deviation of continuous variables will be compared, and the categorical variables will be assessed using Chi-square statistics. We will compare the characteristics of the included and excluded participants.

Contemporary comparisons

In the contemporary comparison, we expect that the distribution of healthcare visits would be right skewed, containing a lot of zeroes (10). We will show the proportion of zero healthcare visits in the supplementary material. Due to the skewness nature of the data, we will use a two-part hurdle model, consisting of a binomial model estimating the probability of non-zero healthcare visits and a truncated GLM model to estimate the healthcare utilisation rate among exposure and comparator groups, with potential confounders adjusted. The truncated GLM model will be a positive Poisson regression or positive negative binomial model, depending on the dispersion of the data.

We will describe the healthcare utilisation rate 12 months after the index date. The outcomes will be visualised using a forest plot showing the probability of non-zero healthcare visits and the rate ratio of the truncated GLM model. We will estimate the average healthcare visits among exposure and comparator groups by applying a prediction function, multiplying the probability of non-zero healthcare visits and the average healthcare visits estimated by the truncated GLM model. The average healthcare visits estimated by the model will be visualised using a stacked bar plot consisting of different healthcare sectors, including primary care and secondary care, including A&E visits, outpatient clinic visits, and hospitalisation.

For the healthcare cost analysis, we will also use a two-part model consisting of a binomial model estimating the probability of non-zero healthcare cost and a gamma GLM model to assess the association between cost and exposure variables. A forest plot will show the non-zero probability of using a healthcare service and the cost ratios between exposure and comparator groups. We will also apply the prediction function to estimate the average costs in exposure and comparator groups and visualise the average cost using a stacked bar chart, stratifying by different healthcare sectors (**Figure 6**).

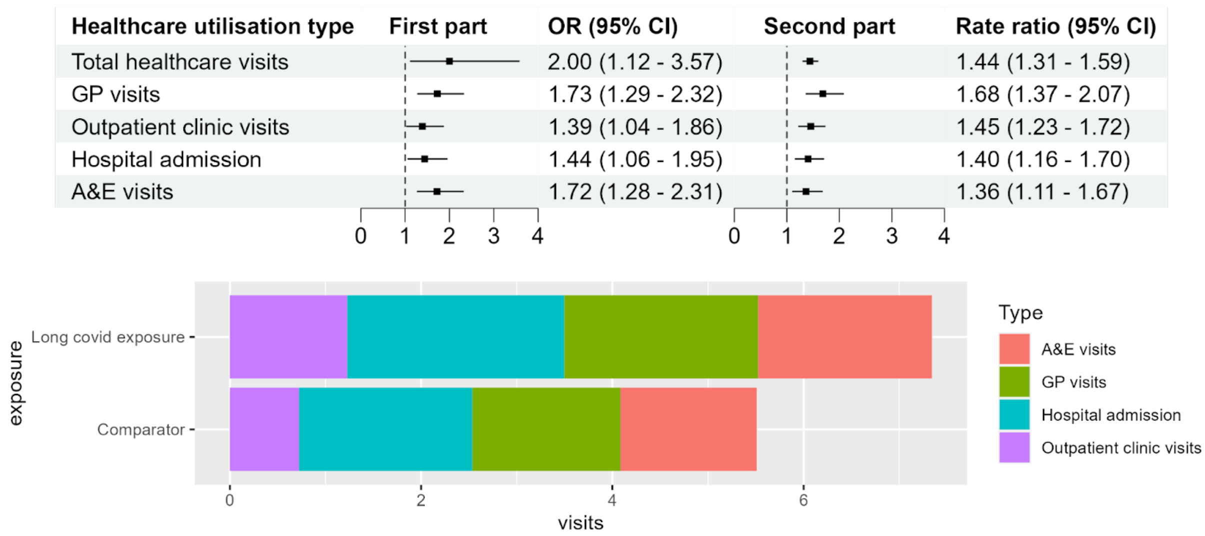


Figure 6. Planned figures for the outcomes (generated by using a dummy dataset)

In addition, we will also stratify our main analyses by sex, age categories, and previous hospitalisation due to COVID.

Historical comparisons

Among people with historical data, we will compare the pre-pandemic historical healthcare utilisation to the contemporary records to estimate the healthcare utilisation change after having long COVID. Because the health-seeking behaviour changed drastically from the beginning of the pandemic (11), we will exclude the historical records between February 2020 and October 2020 to diminish the healthcare utilisation change due to the pandemic (**Figure 4**). We will explore the data using the difference-in-difference (DID) model, a quasi-experimental method evaluating the change of the outcomes over time between treatment (exposure) and non-treatment (comparator) groups. In brief, the DID model first compares the difference before and after the long COVID index date within the exposure and the comparator groups, then further calculates the difference between these two values. By subtracting the effect before long COVID in each group, the results will not be influenced by time-invariant confounders (12) . In our analysis, we will first visualise the monthly healthcare utilisation rate trend before the COVID-19 pandemic (November 2018 – November 2019) and after index dates. Then we will apply a DID regression model to estimate the rate difference between the healthcare utilisation within the long COVID exposure group and comparator group and the between-group difference.

Analysing the factors associated with increased healthcare utilisation

Among the subgroup of people with long COVID and with historical data, we will further analyse the factors associated with the change in healthcare utilisation before and after the pandemic. We will first calculate the healthcare utilisation rate difference between one year before the pandemic and one year after the index date. Then we will apply a Poisson regression model to investigate factors associated with healthcare utilisation change, including sex, age, region, IMD, number of COVID-19 vaccination doses, and other clinical factors. We will further stratify the analysis by previous hospitalisation events due to COVID-19.

Sensitivity analysis

The sensitivity analysis	**Justification**
Rerun the primary analysis among people having previous evidence of COVID-19 and match the case with control by sex, age, region, and the diagnostic month.	Some people without long COVID diagnoses had not been infected. Confining our analysis among people with evidence of prior COVID-19 could further balance the chance of getting long COVID in the cohort, which will help us accurately estimate the impact of long COVID.
Exclude people who do not have any GP visits one year before the index date from the analysis.	Because in EHR studies, the outcomes can only be captured if people visit healthcare providers. People with outcomes without visiting their GP will not be identified. Excluding

	people who did not visit their GP can help us more precisely compare the additional healthcare utilisation caused by the exposure.
Examine the parallel trend assumption of difference in difference analysis.	We will test the parallel trend assumption by visualising the historical monthly healthcare utilisation rate in the exposure and control groups.

Bias and limitation

- **Misclassified long COVID exposure and comparators**:
    - The diagnostic codes for long COVID were not available until November 2020, and the prevalence of long COVID diagnosis in the EHR database was underreported (6, 13). Therefore, it is likely that the long COVID exposure group will be underestimated, and some people with long COVID will be misclassified as a comparator, which would bias the result estimation.
- The analysis cannot include all healthcare utilisation, such as dental services or nursing care services.
- Barriers to accessing healthcare services, such as geographic practice patterns and other socioeconomic factors, are not considered in our study. Although we will adjust for IMD, residual confounding and competing effects cannot be ruled out.
- Selection bias:

    - We will only include people who have been registered for three months in the database. However, people with active long COVID symptoms but registered for less than three months will not be included in our analysis.
    - People who did not have historical EHR will be excluded from the historical comparison, which may introduce selection bias. We will compare the included and excluded participants to examine their differences.
- Residual confounding:
    - Time-varying covariates such as vaccine status cannot be fully adjusted
- Violating the parallel trend assumption of the difference-in-difference model would bias the results.

References:

1. Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. BMJ Global Health. 2021;6(9):e005427.
2. Office for National Statistics. Coronavirus (COVID-19) latest insights 2023 [updated 23/01/23. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/infections#long-covid.
3. Tartof SY, Malden DE, Liu I-LA, Sy LS, Lewin BJ, Williams JTB, et al. Health Care Utilization in the 6 Months Following SARS-CoV-2 Infection. JAMA Network Open. 2022;5(8):e2225657-e.
4. Whittaker HR, Gulea C, Koteci A, Kallis C, Morgan AD, Iwundu C, et al. GP consultation rates for sequelae after acute covid-19 in patients managed in the community or hospital in the UK: population based study. BMJ. 2021;375:e065834.
5. Huang L, Li X, Gu X, Zhang H, Ren L, Guo L, et al. Health outcomes in people 2 years after surviving hospitalisation with COVID-19: a longitudinal cohort study. The Lancet Respiratory Medicine. 2022.
6. 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;71(712):e806.
7. Jones KWH, Birch, S., Castelli, A., Chalkley, M., Dargan, A., Forder, , J. G, M., Hinde, S., Markham, S. Ogunleye, D. Premji, S., Roland, D. . Unit Costs of health and Social Care 2022 2023 13 July 2023.
8. NHS Business Services Authority. Prescription Cost Analysis – England – 2021/22 2022 [updated 9 June 2022. Available from: https://www.nhsbsa.nhs.uk/statistical-collections/prescription-cost-analysis-england/prescription-cost-analysis-england-202122.
9. 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;13(1):3528.
10. Belotti F, Deb P, Manning WG, Norton EC. Twopm: Two-Part Models. The Stata Journal. 2015;15(1):3-20.
11. Mansfield KE, Mathur R, Tazare J, Henderson AD, Mulick AR, Carreira H, 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;3(4):e217-e30.
12. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CMJ. Difference-in-Differences. 2016. In: Impact Evaluation in Practice [Internet]. Washington, DC: Inter-American Development Bank and World Bank. 2. [129]. Available from: https://openknowledge.worldbank.org/handle/10986/25030.

13. Meza-Torres B, Delanerolle G, Okusi C, Mayor N, Anand S, Macartney J, et al. Differences in Clinical Presentation With Long COVID After Community and Hospital Infection and Associations With All-Cause Mortality: English Sentinel Network Database Study. JMIR Public Health Surveill. 2022;8(8):e37668.