

Assessment of 115 symptoms for Long COVID (post-COVID-19 condition) and their risk factors in non-hospitalised individuals: a retrospective matched cohort study in UK primary care

Anuradhaa Subramanian

Institute of Applied Health Research, University of Birmingham

Krishnarajah Nirantharakumar (✉ k.nirantharan@bham.ac.uk)

Institute of Applied Health Research, University of Birmingham

Sarah Hughes

University of Birmingham <https://orcid.org/0000-0001-5656-1198>

Puja Myles

Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency

Tim Williams

Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency

Krishna Gokhale

Institute of Applied Health Research, University of Birmingham

Tom Taverner

Institute of Applied Health Research, University of Birmingham

Joht Chandan

Institute of Applied Health Research, University of Birmingham

Kirsty Brown

Institute of Applied Health Research, University of Birmingham <https://orcid.org/0000-0001-9403-1593>

Nikita Simms-Williams

Institute of Applied Health Research, University of Birmingham

Anoop Shah

Institute of Health Informatics, Faculty of Population Health Sciences, University College London

Megha Singh

Institute of Applied Health Research, University of Birmingham

Farah Kidy

Institute of Applied Health Research, University of Birmingham

Kelvin Okoth

Institute of Applied Health Research, University of Birmingham

Richard Hotham

Institute of Applied Health Research, University of Birmingham

Nasir Bashir

University of Bristol <https://orcid.org/0000-0001-7416-7610>

Neil Cockburn

Institute of Applied Health Research, University of Birmingham

Siang Lee

Institute of Applied Health Research, University of Birmingham

Grace Turner

University of Birmingham

Georgios Gkoutos

University of Birmingham

Olalekan Lee Aiyegbusi

University of Birmingham <https://orcid.org/0000-0001-9122-8251>

Christel McMullan

University of Birmingham

Alastair Denniston

University of Birmingham <https://orcid.org/0000-0001-7849-0087>

Elizabeth Sapey

University Hospitals Birmingham NHS Foundation Trust

Janet Lord

University of Birmingham

David Wraith

University of Bristol

Edward Leggett

Clinical Practice Research Datalink, Medicines and Healthcare Products Regulatory Agency

Clare Iles

Clinical Practice Research Datalink, Medicines and Healthcare Products

Tom Marshall

Institute of Applied Health Research, University of Birmingham

Malcolm Price

Institute of Applied Health Research, University of Birmingham

Steven Marwaha

Institute for Mental Health, University of Birmingham

Elin Davies

Aparito

Louise Jackson

Institute of Applied Health Research, University of Birmingham

Jenny Camaradou

Patient and Public Involvement

Melanie Calvert

University of Birmingham <https://orcid.org/0000-0002-1856-837X>

Shamil Haroon

University of Birmingham <https://orcid.org/0000-0002-0096-1413>

Article

Keywords: Long COVID, post-acute COVID-19 syndrome, symptoms, risk factors, primary care

Posted Date: February 15th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1343889/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Nature Medicine on July 25th, 2022. See the published version at <https://doi.org/10.1038/s41591-022-01909-w>.

Abstract

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2) infection is frequently associated with a wide range of persistent symptoms, now referred to as post-COVID-19 condition, or Long COVID. The objectives of this study were to assess which symptoms are associated with confirmed SARS CoV-2 beyond 12 weeks post-infection in non-hospitalised individuals, and the risk factors associated with developing persistent symptoms. We undertook a retrospective matched cohort study between 31st January 2020 and 15th April 2021 using data from a large database of UK-based primary care electronic health records, Clinical Practice Research Datalink (CPRD) Aurum. We selected 486,149 adult patients with a confirmed diagnosis of SARS CoV-2 infection that had not been hospitalised within 28 days of the diagnosis (infected cohort). We propensity score matched them to 1,944,580 patients without a coded record of either confirmed or suspected COVID-19 (uninfected cohort). Outcomes were the presence of 115 separate symptoms at ≥ 12 weeks post-infection, and Long COVID, defined as having at least one of the symptoms included in the World Health Organisation case definition. Separate Cox proportional hazards models were used to estimate adjusted hazard ratios (aHR) for individual symptoms and Long Covid. 62 symptoms were significantly associated with prior exposure to SARS CoV-2 after 12 weeks. The largest adjusted hazard ratios (aHR) were for anosmia (aHR 6.49, 95% CI 5.02 to 8.39), hair loss (3.99, 3.63 to 4.39), sneezing (2.77, 1.40 to 5.50), difficulties with ejaculation (2.63, 1.61 to 4.28), reduced libido (2.36, 1.61 to 3.47), shortness of breath at rest (2.20, 1.57 to 3.08), fatigue (1.92, 1.81 to 2.03), pleuritic chest pain (1.86, 1.41 to 2.46), hoarse voice (1.78, 1.44 to 2.20), and fever (1.75, 1.54 to 1.98). Among the infected cohort, risk factors for Long COVID included younger age (aHR 0.75, 95% CI 0.70 to 0.81, for those aged ≥ 70 years compared to those aged 18 to 30 years), female sex (1.52, 1.48 to 1.56), belonging to an ethnic minority group (1.14 [1.07 to 1.22] for mixed race, 1.21 [1.10 to 1.34] for black ethnic groups, and 1.06 [1.03 to 1.10] for other ethnic minority groups, compared to white ethnic groups), socioeconomic deprivation (1.11 [1.07 to 1.16] for the most compared to the least socioeconomically deprived quintile), smoking (1.12, 1.08 to 1.15), obesity (1.10, 1.07 to 1.14), and a wide range of comorbidities such as COPD. SARS CoV-2 in non-hospitalised individuals is associated with a plethora of symptoms being reported at ≥ 12 weeks post-infection, with a higher risk associated with younger age, female sex, ethnic minority groups, socioeconomic deprivation, smoking, obesity, and several comorbidities.

Introduction

Infection with Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2) causes an acute multisystem illness referred to as COVID-19.¹ It is now recognised that approximately 10% of individuals with COVID-19 develop persistent, and often relapsing and remitting symptoms beyond 4 to 12 weeks post-infection.² The presence of persistent symptoms in a previously infected individual is commonly referred to by a number of terms including post COVID-19 condition, post-acute COVID-19 syndrome, post-acute sequelae of COVID-19 (PASC) and Long COVID.³⁻⁵ The World Health Organisation (WHO) defines it as a condition characterised by symptoms such as fatigue, shortness of breath, and cognitive dysfunction that occurs in individuals with a history of probable or confirmed SARS CoV-2 infection, usually three months from the onset of acute COVID-19 symptoms, lasting for at least two months, and that cannot be explained by an alternative diagnosis.⁶ The UK National Institute for Health and Care Excellence (NICE) makes a distinction between disease occurring from 4 to 12 weeks post-infection (ongoing symptomatic COVID-19) and symptoms persisting beyond 12 weeks (post-acute COVID-19 syndrome).⁴

Long COVID has been associated with a broad range of symptoms and health impacts.^{5,7-9} A systematic review and meta-analysis of Long COVID symptoms showed the ten most prevalent symptoms to be fatigue, shortness of breath, muscle pain, joint pain, headache, cough, chest pain, altered smell, altered taste and diarrhoea.¹⁰ However, previous studies were often based on self-reported symptoms or lacked a control group, making it difficult to make inferences about whether the reported symptoms were due to SARS CoV-2 infection, pre-existing comorbidities, or societal effects related to the pandemic. Furthermore, many previous studies were conducted in hospitalised cohorts,¹¹ and population-level data on the potential breadth of symptoms experienced by non-hospitalised individuals with SARS CoV-2 infection are scarce. Large scale studies leveraging routinely available healthcare data with closely matched control populations are needed to elucidate which symptoms are independently associated with the long-term effects of COVID-19.

There is also a need to gain a better understanding of the risk factors that contribute towards the development of Long COVID, which was highlighted as a research priority on the recently updated NICE guideline on managing the long term effects of COVID-19.⁴ Previous studies suggested that older age, female sex, hospital admission during acute COVID-19 (including the need for oxygen therapy), symptom load (including dyspnoea at presentation and chest pain), abnormal auscultation findings, and the presence of comorbidities such as asthma, were all associated with a higher risk of developing Long COVID.¹²⁻¹⁵ Large scale population-based studies with appropriate control groups are required to assess the long-term symptoms that are specifically attributable to SARS CoV-2 infection and their association with a wide range of demographic and clinical risk factors in non-hospitalised individuals. This is needed to understand the breadth of symptoms that contribute to Long COVID to inform clinical management and help healthcare providers identify population groups at higher risk of reporting persistent symptoms.

In this study we aimed to undertake a large-scale analysis of primary care data from the UK to investigate a comprehensive range of symptoms previously reported by epidemiological studies, patients, and clinicians, to assess their association with confirmed SARS CoV-2 infection at least 12 weeks post-infection in non-hospitalised individuals compared to a propensity score matched uninfected cohort. We also aimed to assess the association between demographic and clinical risk factors, including comorbidities with the development of Long COVID.

Methods

Study design and setting

This analysis was undertaken as part of the National Institute for Health Research (NIHR) and UK Research and Innovation (UKRI) funded Therapies for Long COVID in non-hospitalised individuals (TLC) Study.¹⁶ We conducted a population-based retrospective matched cohort study between 31st January 2020 and 15th April 2021 using data from the Medicines and Healthcare products Regulatory Agency (MHRA) Clinical Practice Research Datalink (CPRD) Aurum. CPRD

Aurum is an anonymised database of primary care medical records of over seven million actively registered patients in general practices that use the EMIS clinical information system.¹⁷ It captures data on patient demographics, diagnoses, symptoms, prescriptions, referrals, and tests. Structured data on diagnoses, symptoms and referrals are recorded using SNOMED CT coding terminology.¹⁸ Selection of SNOMED CT codes for data extraction was done by a team of clinical researchers using an inhouse developed software platform called Code Builder, with systematic searching of existing code lists, reference to the SNOMED CT terminology browser¹⁸, and through clinical knowledge and discussion. Data extraction was performed using the Data extraction for epidemiological research (DEXTER) tool for automated clinical epidemiological studies.¹⁹

Participants

Patients aged 18 years and older with a minimum registration period of 12 months were included in the study. Practices were considered eligible 12 months after they were deemed to be providing research quality data. The infected cohort was defined as patients with a coded record of a positive reverse transcriptase polymerase chain reaction (RT-PCR) or antigen test result for SARS CoV-2 and without a record of hospitalisation 14 days before or 42 days after infection (within 28 days of infection with a ± 14 -day grace period for clinical coding delays) in the primary care record.²⁰ Their index date was assigned as the date of confirmation of SARS CoV-2 infection. SNOMED CT codes for defining COVID-19 are listed in Supplementary Table 1. For each infected patient, a pool of eligible uninfected patients without a record of suspected or confirmed COVID-19 were selected from the database. These patients were assigned the same index date as the index date of the corresponding infected patient to mitigate immortal time bias.²¹

Propensity score matching

To control for confounding, each infected patient was propensity score-matched with up to four uninfected patients using a logistic regression model including the covariates listed in the covariates section below and a calliper width of 0.2. The standardised mean difference (SMD) between the infected and uninfected patients was reported for each variable before and after matching and a variable with SMD > 0.1 after matching was considered to indicate imbalance in baseline characteristics. Kernel density plots were drawn for the two groups before and after matching to check the distribution of propensity scores.

Outcomes and follow-up

We identified 115 relevant symptoms coded within primary care records (Supplementary Table 2) through a systematic review and meta-analysis of Long COVID symptoms¹⁰, a scoping search of Long COVID clinical assessment questionnaires, qualitative interviews with patients, a clinician survey, and refinement of the symptom list using psychometric methods.²² These were grouped into 15 domains: (1) breathing, (2) pain, (3) circulation, (4) fatigue, (5) cognitive health, (6) movement, (7) sleep, (8) ear, nose, and throat, (9) stomach and digestion, (10) muscles and joints, (11) mental health, (12) hair, skin and nails, (13) eyes, (14) reproductive health, and (15) other symptoms.

Our primary outcome definition of Long COVID was predefined as the presence of at least one symptom included in the WHO case definition at ≥ 12 weeks post-infection (Supplementary Table 2).⁶ Our secondary outcome definition of Long COVID was derived post-hoc as the presence of at least one symptom that was statistically associated with SARS CoV-2 infection at ≥ 12 weeks post-infection within this study (Supplementary Table 2).

Patients were followed up from the index date until the earliest of the following end points (patient exit date): (1) recording of symptoms of interest within the time interval studied, (2) death, (3) transfer out of practice, (4) end of general practice data, and (5) study end date (15th April 2021). The follow-up period was split into three time periods from the index date: (1) the first four weeks ("acute COVID-19" among the cases), (2) four to twelve weeks ("ongoing symptomatic COVID-19"), and after twelve weeks (period of "post-COVID-19 condition" or "Long COVID"), in accordance with the current NICE guidelines on managing the long term effects of COVID-19.⁴

Covariates

We extracted data on demographic characteristics (age, sex, ethnic group, socio-economic status [Index of Multiple Deprivation, IMD]), body mass index (BMI), smoking status, and 87 chronic health conditions (Supplementary table 3). We extracted data on 115 symptoms recorded in the period between 12 months and 3 months prior to the index date (at baseline). These variables were used to generate propensity scores for symptom burden to ensure that pre-existing health conditions and symptoms did not differ between the infected and uninfected cohorts.

BMI was categorised as underweight (< 18.5 kg/m²), normal weight (18.5-24 kg/m²), overweight (25-29 kg/m²), and obese (≥ 30 kg/m²). Smoking status was categorised as never smoked, ex-smoker, and current smoker. Ethnic group was categorised as either white, Asian, black, mixed, or other ethnic group. Missing data on ethnic group, socioeconomic status, BMI, and smoking status were denoted by a missing category within the corresponding variable.

Statistical Analysis

Continuous variables were summarised as mean and standard deviation (SD) and categorical variables as frequencies and percentages. A series of Cox proportional hazards regression models were used to provide adjusted hazard ratios (aHRs) for each of the individual symptoms among patients with SARS CoV-2 infection compared to uninfected patients separately during the first four, four to twelve and twelve weeks after the index date, with follow-up initiating

from the index date, four weeks after the index date, and twelve weeks after the index date, respectively until the patient exit date. Patients with a minimum follow-up period of four and twelve weeks were included in the symptom outcome analyses at four to twelve and twelve weeks, respectively. Adjustments were made for age, sex, ethnic group, socioeconomic status, BMI, smoking status, and the specified symptom recorded at baseline between 3 and 12 months prior to the index date. Multiple testing was accounted for by incorporating a Bonferroni correction to adjust the p-value thresholds for statistical significance. Symptoms with significant aHRs after Bonferroni correction in the period twelve weeks after index date were presented in a forest plot.

In a cohort restricted to patients with a positive RT-PCR or antigen test result for SARS CoV-2 (infected cohort) and a minimum of 12 weeks follow-up, adjusted and unadjusted Cox proportional hazards models were used to assess the association between the risk factors described in the covariates section, and the primary (at least one of the symptoms in the WHO case definition for Long COVID) and secondary (at least one of the symptoms statistically associated with SARS CoV-2 infection) outcome definitions of Long COVID. The median follow-up period and interquartile range was reported for patients within each risk factor strata. Hazard ratios were obtained by taking exponentiated coefficients from the Cox proportional-hazards models, and we considered covariates with a p-value <0.05 to be statistically significant. All analyses were performed in Stata IC version 16.

Ethical approval

CPRD obtains annual research ethics approval from the UK's Health Research Authority (HRA) Research Ethics Committee (REC) (East Midlands – Derby, REC reference number 05/MRE04/87) to receive and supply patient data for public health research. Therefore, no additional ethics approval is required for observational studies using CPRD Aurum data for public health research, subject to individual research protocols meeting CPRD data governance requirements. The use of CPRD Aurum data for the study has been approved by the CPRD Independent Scientific Advisory Committee (reference: 21_000423).

Results

Participants

A total of 486,149 non-hospitalised individuals had a coded record of SARS CoV-2 infection (infected cohort), and 8,030,224 had no records of either suspected or confirmed COVID-19 during the study period, which formed the pool of uninfected patients. From the uninfected pool, 1,944,580 individuals were propensity score matched to patients in the infected cohort. Kernel density plots of the propensity scores of the infected and uninfected cohorts, before and after matching is presented in Supplementary Figure 1. The total follow-up time was 0.29 years (IQR 0.24-0.42) for the infected cohort and 0.29 (IQR 0.24-0.41) for the uninfected cohort.

Infected and uninfected patients were well matched in terms of sociodemographic characteristics, smoking status, BMI, comorbidities, and baseline recording of symptoms, indicated by SMD < 0.1 for all variables (Table 1 and Supplementary Table 3). The mean age was 43.8 years (SD 16.9) and 55.3% were female. 64.7% were white, 12.2% were Asian and 4.0% were black, and ethnicity data were missing for 16.2%. 53.8% were overweight or obese (with BMI data missing for 13.0%) and 22.5% were current smokers (with smoking data missing for 4.3%).

The most common comorbidities were depression (22.1%), anxiety (20.3%), asthma (20.1%), eczema (19.5%), and hay fever (18.1%). A full list of comorbidities is provided in Supplementary Table 3. 57% of the SARS CoV-2 infected patients had been diagnosed in 2020 and 43% in 2021. 4.5% of the infected and 4.7% of the uninfected patients had received at least a single dose of a COVID-19 vaccine prior to index date. The most common vaccine prior to index date was the BioNTech-Pfizer (2.8%) followed by Oxford-AstraZeneca (1.7%).

Symptoms

In the 3-to-12-month period prior to the index date, the reporting of symptoms between the infected and propensity score matched uninfected cohorts were similar. Of the 115 symptoms, statistically significant differences between the two groups at baseline were observed only for bowel incontinence and sore throat, after adjustment for age, sex, ethnic group, socioeconomic status, smoking status, and BMI using logistic regression (Supplementary Table 4).

At 12 weeks after the index date, a history of SARS CoV-2 infection was significantly associated with a total of 62 symptoms, compared to propensity score matched uninfected patients, after adjustment for age, sex, ethnic group, socioeconomic status, smoking, body mass index, and presence of symptoms at baseline (Supplementary Table 5). These 62 symptoms spanned 14 out of the 15 domains considered (Figure 1). Of the patients with a minimum of 12 weeks of follow-up, 20,864 out of 384,137 (5.4%) infected patients and 65,293 out of 1,501,689 (4.3%) uninfected patients reported at least one of the symptoms included in the WHO case definition for Long COVID (aHR 1.26, 95% CI: 1.25 to 1.28) (Supplementary Table 5).

The symptoms with the largest adjusted hazard ratios were anosmia (aHR 6.49, 95% CI 5.02 to 8.39), hair loss (3.99, 3.63 to 4.39), sneezing (2.77, 1.40 to 5.50), difficulties with ejaculation (2.63, 1.61 to 4.28), reduced libido (2.36, 1.61 to 3.47), shortness of breath at rest (2.20, 1.57 to 3.08), fatigue (1.92, 1.81 to 2.03), pleuritic chest pain (1.86, 1.41 to 2.46), hoarse voice (1.78, 1.44 to 2.20), and fever (1.75, 1.54 to 1.98).

The association of SARS CoV-2 infection with these symptoms was even larger at 0-4 weeks and 4-12 weeks, and the size of the adjusted hazard ratios reduced with increasing time from the index date. A full list of the adjusted hazard ratios for all 115 symptoms included in the analysis at 0-4 weeks, 4-12 weeks and beyond 12 weeks is presented in Supplementary Table 5, 6, and 7.

Risk factors for Long COVID symptoms

The risk factor analysis included 384,137 SARS CoV-2 infected individuals with a minimum of 12 weeks of follow-up. When using the WHO definition of Long COVID, several sociodemographic and clinical risk factors were significantly associated with the incidence of Long COVID (Table 2 and Supplementary Table 8). Women were at increased risk compared to men (aHR 1.52, 95% CI 1.48 to 1.56). Older age was associated with a higher risk of reporting Long COVID symptoms in the univariate analysis. However, after adjusting for baseline covariates, older age was associated with a lower risk, with those aged 30-39 years having a 6% lower risk (aHR 0.94, 0.90 to 0.97) and those aged ≥ 70 years having a 25% lower risk (aHR 0.75, 0.70 to 0.81) compared to those aged 18-30 years.

There were associations between the risk of Long COVID and certain ethnic minority groups in the multivariable model, with increased risks seen in black ethnic groups (aHR 1.21, 95% CI 1.10 to 1.34), mixed race (aHR: 1.14, 1.07 to 1.22) and other minority ethnic groups (1.06, 1.02 to 1.10), as compared to white ethnic groups. The risk of Long COVID also increased with increasing levels of socioeconomic deprivation, with a 11% risk increase (aHR 1.11, 1.07 to 1.16) in those who were most deprived compared to those least deprived.

Smokers and former smokers were at increased risk of reporting Long COVID symptoms (aHR 1.12, 95% CI 1.08 to 1.15, and 1.08, 1.05 to 1.11, respectively), compared to those who had never smoked. Being overweight or obese was also associated with an increased risk of persistent symptoms, with those who had a BMI of greater than 30 kg/m² having a 10% relative increase in risk of reporting Long COVID symptoms compared to those with a BMI of 18.5-25 kg/m² (aHR 1.10, 1.07 to 1.14).

A wide range of comorbidities at baseline were also associated with an increased risk of Long COVID symptoms. The comorbidities with the largest associations were COPD (aHR: 1.55, 1.47 to 1.64), benign prostatic hyperplasia (1.39, 1.28 to 1.52), fibromyalgia (1.37, 1.28 to 1.47), anxiety (1.35, 1.31 to 1.39), erectile dysfunction (1.33, 1.26 to 1.41), depression (1.31, 1.27 to 1.34), migraine (1.26, 1.22 to 1.30), multiple sclerosis (1.26, 1.03 to 1.53) coeliac disease (1.25, 1.09 to 1.43), and learning disability (1.24, 1.11-1.40). A full list of the adjusted hazard ratios for the included comorbidities is provided in Supplementary Table 8.

When using an alternative definition of Long COVID that consisted of having at least one of the symptoms that were statistically associated with a history of SARS CoV-2 infection ≥ 12 weeks post-infection, the trends described above were largely still observed (Supplementary Table 9). Females; younger age; ethnic minority groups; increasing socioeconomic deprivation; smoking and former smoking; high BMI; and a wide range of comorbidities were all associated with an increased risk of reporting symptoms ≥ 12 weeks post-infection.

Discussion

Main findings

Individuals with confirmed SARS CoV-2 infection were at increased risk of reporting a wide range of symptoms at ≥ 12 weeks post-infection, compared to propensity score matched uninfected patients with no record of suspected or confirmed SARS CoV-2 infection, after accounting for both sociodemographic and clinical characteristics, and the reporting of symptoms prior to infection. The symptoms most associated with SARS CoV-2 infection included some that are already recognized in previous studies,²³ such as anosmia, shortness of breath, chest pain and fever, but also included a range of other symptoms that have previously not been widely reported such as hair loss and sexual dysfunction. Prior SARS CoV-2 infection was independently associated with the reporting to primary care of 20 of the 33 symptoms included in the WHO case definition, and an additional 42 symptoms, beyond 12 weeks from infection. SARS CoV-2 infection was associated with a 26% relative increase in risk of reporting at least one of the symptoms included in the WHO case definition for Long COVID.

Among those with a history of confirmed SARS CoV-2 infection, several risk factors were associated with reporting symptoms 12 weeks or more post-infection. Female sex, younger age, belonging to a black, mixed race or other ethnic minority group, socioeconomic deprivation, smoking, high BMI, and presence of a wide range of comorbidities were associated both with symptoms included in the WHO definition of Long COVID, and with symptoms statistically associated with SARS CoV-2 infection reported 12 weeks or more post-infection.

Strengths & limitations

The strengths of the study include the large sample size, which included 486,149 individuals with a confirmed diagnosis of SARS CoV-2 infection and 1.9 million propensity score matched patients with no prior record of infection. The large sample size provided adequate statistical power to assess differences in the reporting of a wide range of symptoms between the infected and uninfected cohorts and estimation of the association between reporting of symptoms and important sociodemographic and clinical risk factors with a high level of precision.

A key strength of this study is the inclusion of an uninfected comparator group that did not have either suspected or confirmed COVID-19 and had been propensity score matched for sociodemographic factors, previously reported symptoms, and over 80 comorbidities. This enabled us to assess the independent association between exposure to SARS CoV-2 and the reporting of symptoms ≥ 12 weeks post-infection, after accounting for many important confounders.

Another strength is the large number of symptoms included in the analysis, which was based on a previous systematic review of the literature,¹⁰ a scoping review of Long COVID questionnaires, and an extensive consultation with patients and clinicians.²² Symptom code lists were developed rigorously with systematic searches for relevant SNOMED-CT codes with extensive clinical input. We also assessed the outcome of Long COVID using the WHO case definition as well as a novel definition that incorporated symptoms that were statistically associated with a history of SARS CoV-2 infection.

A key limitation of the study is the use of routinely coded healthcare data. Coded symptom data in primary care records is likely to underrepresent the true symptom burden experienced by individuals with Long COVID.²⁴ This could be due to reduced access to primary care (especially during the first wave of the pandemic), patients not consulting their GP about symptoms, or patients underreporting the full extent of their symptoms. In addition, much of a patient's clinical history, in terms of the symptoms reported, are recorded as free text, rather than as SNOMED-CT codes. The symptom data we used for the study thus cannot be used to make inferences about the absolute prevalence of these symptoms. However, since this underrepresentation would be expected to affect both the infected and uninfected patients equally, the data used in the present analysis can still be used to examine relative differences in the reporting of symptoms between infected and uninfected patients. Conversely, with the evolving awareness of Long COVID, it is possible that those with a history of COVID-19 may have been more likely than uninfected patients to access primary care and alert clinicians of their symptoms, which could potentially lead to an inflation of the observed effect sizes.

Another limitation of the study is potential misclassification bias. Community testing for SARS CoV-2 was very limited during the first wave of the pandemic, and many non-hospitalised individuals with COVID-19 were not tested. It is therefore possible that some members of our uninfected cohort had been infected with SARS CoV-2 but had simply not been tested. We attempted to partially account for this bias by excluding individuals from the uninfected pool if they had a coded diagnosis of suspected COVID-19. However, this is unlikely to be 100% sensitive in identifying individuals with unverified COVID-19 from the uninfected pool, which would potentially have the effect of attenuating the observed effect sizes.

Relationship to other studies

Our findings support the results from our previous systematic review and meta-analysis on Long COVID symptoms.¹⁰ That review found the most prevalent symptoms to be fatigue, shortness of breath, muscle pain, joint pain, headache, cough, chest pain, altered sense of smell, altered taste and diarrhoea. Our current analysis was not able to assess symptom prevalence, but rather the relative difference in symptoms between individuals with a history of SARS CoV-2 and closely matched uninfected patients at ≥ 12 weeks post-infection. We similarly identified that anosmia, shortness of breath, fatigue, and chest pain to be symptoms significantly associated with SARS CoV-2 infection. By contrast, we also identified novel symptoms such as hair loss, sneezing, symptoms of sexual dysfunction (difficulties ejaculating and reduced libido), hoarse voice and fever to also be significantly associated. Also, like our review, we found that female sex and the presence of a range of comorbidities were associated with an increased risk of developing persistent symptoms. However, in contrast to our review, the present analysis found that younger age was associated with a higher risk of reporting symptoms at ≥ 12 weeks post-infection. This could partly be due to the adjustment for an extensive range of comorbidities or the differences in the populations studied. Most studies included in our review were based on hospitalised cohorts whereas our present study excluded hospitalised patients. Older patients with COVID-19 were more likely to be hospitalised than younger patients, and therefore be excluded from our study. Older non-hospitalised patients may therefore potentially have had milder disease with lower symptom burden.

An international online cohort study of people with confirmed and suspected long COVID found that respondents reported an average of 56 symptoms across an average of nine organ systems.⁸ This was a comprehensive review of symptom burden in these individuals, but the study lacked a control group and was therefore unable to make strong inferences about the relative contribution of SARS CoV-2 infection to these symptoms over and above pre-existing health conditions or psychosocial effects related to the pandemic. However, like this study, we also found that individuals with a history of confirmed SARS CoV-2 reported a broad range of symptoms, with a total of 62 symptoms being associated at 12 or more weeks post-infection. Importantly, we were able to control for potential confounders, including whether the symptoms of interest were reported prior to infection.

The COVID Symptom Study provided data on self-reported symptoms among participants enrolled on a app.¹² Among those with symptoms persisting 28 days or longer post-infection, key symptoms included fatigue, headache, dyspnoea and anosmia, which were all also significantly associated at ≥ 12 weeks in our cohort. The COVID Symptom Study also found that Long COVID was associated with increasing BMI and female sex, which is in keeping with our findings. However, the study also found that the risk of reporting Long COVID symptoms increased with age, whereas our study observed the opposite trend after adjustment for a comprehensive range of potential confounders. Although the COVID Symptom Study is community-based, it includes individuals with a history of hospitalised and non-hospitalised COVID-19 so the reasons for the discrepant age trend may be due to the same reasons described above.

One of the largest population-based surveys on COVID-19 and Long COVID is the UK Office for National Statistics COVID Infection Survey.²⁵ This estimated that as of 6th December 2021, 1.3 million people living in private households in the UK (2.0% of the population) were experiencing symptoms persisting beyond four weeks from SARS CoV-2 infection, and with 70% experiencing symptoms beyond 12 weeks. Fatigue, shortness of breath, anosmia and difficulty concentrating were the main symptoms reported. The prevalence was greatest in females, those from more socioeconomically deprived areas, people working in health and social care and individuals living with health conditions and disabilities. Our analysis showed similar symptoms, including cognitive effects, as well as similar risk factors. However, we were unable to assess the association between occupational status and reporting of symptoms due to a lack of occupational data in UK primary care records.

Most recently, Whittaker and colleagues undertook an analysis of 456,002 patients with COVID-19 in England using the CPRD Aurum database to determine the rates of GP consultations for post-COVID-19 sequelae.²⁶ This included both hospitalised and non-hospitalised patients and two control groups consisting of patients without COVID-19 and those with influenza before the pandemic. Patients with COVID-19 managed in the community were significantly more likely to consult for loss of taste or smell and other symptoms such as joint pain, anxiety, depression, abdominal pain, and diarrhoea at ≥ 4 weeks post-infection compared to 12 months prior to infection. They also found that patients' GP consultation rates for symptoms, prescriptions and healthcare use were mostly reduced in those who were managed in the community after the first COVID-19 vaccination dose.

We were unable to estimate the effect of vaccination and infection year on Long COVID symptoms in our study due to the very short follow-up period among those vaccinated and infected in the year 2021 [Median 8 (IQR 4-14) and 12 (7-16) days, respectively] compared to those unvaccinated and infected in the year 2022 [33 (16-77) and 64 (31-90) days, respectively]. Furthermore, the majority (81%) of patients vaccinated prior to infection in our cohort were infected with SARS CoV-2 within two weeks of vaccination, which would be before acquiring immunity from vaccination, thus restricting the validity of our data to assess the effects of vaccination on Long COVID.

Implications for practice, policy, and research

Further research is needed to estimate the prevalence of persistent symptoms associated with SARS CoV-2 infection among patients presenting to primary care. Much of the symptom data in primary care records is held in free text entries rather than as clinically coded data. Natural language processing could be used to leverage these textual data to gain more accurate estimates of the prevalence of these symptoms.

62 symptoms reported more than 12 weeks post-SARS CoV-2 infection were found to be associated with COVID-19. It is likely that these symptoms cluster differentially amongst the population, and clustering methods could potentially be used to contribute to the discussion about Long COVID phenotypes. This can allow for assessment of whether clinical outcomes and the underlying pathophysiology differs between these subgroups and potentially develop targeted therapies for the different phenotypic subgroups.

There is also a need to obtain patient-reported data on symptoms and assess the association between symptom burden, quality of life and work capability to ascertain which symptoms have the greatest impact on individuals. Finally, there is a need to understand the natural history of Long COVID by assessing symptom burden serially over time in a population-representative cohort with a history of COVID-19 alongside a matched control population.

Conclusion

Infection with SARS CoV-2 is independently associated with the reporting of 62 symptoms spanning multiple organ systems 12 weeks or longer post-infection. A wide range of both sociodemographic and clinical factors are independently associated with the development of persistent symptoms. Further research is needed to describe the natural history of Long COVID, and characterise symptom clusters, their pathophysiology, and clinical outcomes. Further research is also needed to understand the health and social impacts of these persistent symptoms, to support patients living with long-term sequelae, and to develop targeted treatments.

References

1. Ramos-Casals, M., Brito-Zerón, P. & Mariette, X. Systemic and organ-specific immune-related manifestations of COVID-19. *Nat. Rev. Rheumatol.* 2021 **17**, 315–332 (2021).
2. The prevalence of long COVID symptoms and COVID-19 complications - Office for National Statistics. <https://www.ons.gov.uk/news/statementsandletters/theprevalenceoflongcovidsymptomsandcovid19complications>.
3. Ladds, E. *et al.* Persistent symptoms after Covid-19: qualitative study of 114 “long Covid” patients and draft quality principles for services. *BMC Health Serv. Res.* **20**, 1–13 (2020).
4. Overview | COVID-19 rapid guideline: managing the long-term effects of COVID-19 | Guidance | NICE.
5. Nalbandian, A. *et al.* Post-acute COVID-19 syndrome. *Nat. Med.* 2021 **27**, 601–615 (2021).
6. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1.
7. Del Rio, C., Collins, L. F. & Malani, P. Long-term Health Consequences of COVID-19. *JAMA - Journal of the American Medical Association* vol. 324 1723–1724 (2020).
8. Davis, H. E. *et al.* Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* **38**, 101019 (2021).
9. Groff, D. *et al.* Short-term and Long-term Rates of Postacute Sequelae of SARS-CoV-2 Infection: A Systematic Review. *JAMA Netw. Open* **4**, e2128568–e2128568 (2021).
10. Aiyegbusi, O. L. *et al.* Symptoms, complications and management of long COVID: a review. *J. R. Soc. Med.* **114**, 428–442 (2021).
11. Antonelli, M. *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.* **0**, (2021).
12. Sudre, C. H. *et al.* Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).
13. Jacobs, L. G. *et al.* Persistence of symptoms and quality of life at 35 days after hospitalization for COVID-19 infection. *PLoS One* **15**, (2020).

14. Carvalho-Schneider, C. *et al.* Follow-up of adults with noncritical COVID-19 two months after symptom onset. *Clin. Microbiol. Infect.* **27**, 258–263 (2021).
15. Galal, I. *et al.* Determinants of persistent post-COVID-19 symptoms: value of a novel COVID-19 symptom score. *Egypt. J. Bronchol.* **2021** *151* **15**, 1–8 (2021).
16. Haroon, S. *et al.* Therapies for Long COVID in non-hospitalised individuals - from symptoms, patient-reported outcomes, and immunology to targeted therapies (The TLC Study): Study protocol. *medRxiv* 2021.12.20.21268098 (2021) doi:10.1101/2021.12.20.21268098.
17. Wolf, A. *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int. J. Epidemiol.* **48**, 1740-1740g (2019).
18. SNOMED CT - NHS Digital. <https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct>.
19. Gokhale, K. M. *et al.* Data extraction for epidemiological research (DExtER): a novel tool for automated clinical epidemiology studies. *Eur. J. Epidemiol.* **36**, 165–178 (2021).
20. of Health, D. & Care, S. Coronavirus (COVID-19) Scaling up our testing programmes Department of Health and Social Care. (2020).
21. Yadav, K. & Lewis, R. J. Immortal Time Bias in Observational Studies. *JAMA* **325**, 686–687 (2021).
22. Hughes, S. E. *et al.* Development and validation of the Symptom Burden Questionnaire™ for Long COVID: a Rasch analysis. *medRxiv* 2022.01.16.22269146 (2022) doi:10.1101/2022.01.16.22269146.
23. Lopez-Leon, S. *et al.* More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci. Reports* **2021** *111* **11**, 1–12 (2021).
24. Price, S. J., Stapley, S. A., Shephard, E., Barraclough, K. & Hamilton, W. T. Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case-control study. *BMJ Open* **6**, (2016).
25. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcor>
26. Whittaker, H. R. *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* **375**, e065834 (2021).

Tables

Table 1 Baseline characteristics of cases and propensity matched controls

	Infected cohort (n=486,149)	Uninfected cohort (n=1,944,580)	Standardised differences
Mean age at index date (SD)	44.1 (17.0)	43.8 (16.9)	0.015
Sex, n (%)			
<i>Female</i>	268,367 (55.2)	1,075,963 (55.3)	0.003
<i>Male</i>	217,782 (44.8)	868,617 (44.7)	
Ethnic group, n (%)			
<i>White</i>	313,561 (64.5)	1,258,392 (64.7)	0.004
<i>Asians</i>	59,477 (12.2)	237,133 (12.2)	
<i>Black</i>	19,835 (4.1)	78,501 (4.0)	
<i>Mixed Race</i>	7357 (1.5)	29,614 (1.5)	
<i>Other</i>	6896 (1.4)	26,966 (1.4)	
<i>Missing</i>	79,023 (16.3)	31,3974 (16.2)	
Socioeconomic status quintile, n (%)			0.003
<i>1 (least deprived)</i>	82,538 (17.0)	331,229 (17.0)	
<i>2</i>	86,164 (17.7)	346,054 (17.8)	
<i>3</i>	89,470 (18.4)	358,650 (18.4)	
<i>4</i>	106,578 (21.9)	426,153 (21.9)	
<i>5 (most deprived)</i>	112,656 (23.2)	448,126 (23.0)	
<i>Missing</i>	8743 (1.8)	34,368 (1.8)	
Body mass index (kg/m²), n (%)			
<i><18.5</i>	13,261 (2.7)	52,322 (2.7)	0.001
<i>18.5-25</i>	148,295 (30.5)	590,747 (30.4)	
<i>25-30</i>	138,771 (28.5)	558,287 (28.7)	
<i>>30</i>	121,943 (25.1)	489,389 (25.2)	
<i>Missing</i>	63,879 (13.1)	253,835 (13.1)	
Smoking status			
<i>Never smoked</i>	177,064 (36.4)	714,045 (36.7)	0.009
<i>Ex-smoker</i>	176,899 (36.4)	710,255 (36.5)	
<i>Current smoker</i>	110,848 (22.8)	436,212 (22.4)	
<i>Missing</i>	213,38 (4.4)	84,068 (4.3)	
Comorbidities			
<i>Depression</i>	107,392 (22.1)	428,797 (22.1)	0.001
<i>Anxiety</i>	98,849 (20.3)	395,365 (20.3)	0.000
<i>Asthma</i>	97,509 (20.1)	390,401 (20.1)	0.000
<i>Eczema</i>	94,313 (19.4)	378,604 (19.5)	0.002
<i>Hay fever</i>	87,691 (18.0)	352,090 (18.1)	0.002
<i>Hypertension</i>	73,901 (15.2)	291,389 (15.0)	0.006
<i>Migraine</i>	53,881 (11.1)	215,733 (11.1)	0.000
<i>Osteoarthritis</i>	53,694 (11.0)	211,062 (10.9)	0.006
<i>Fragility fracture</i>	46,608 (9.6)	186,194 (9.6)	0.000
<i>Arrhythmias</i>	34,811 (7.2)	136,280 (7.0)	0.006
Calendar year of index date, n (%)			

2020	275,169 (56.6)	1,077,126 (55.4)	<i>0.024</i>
2021	210,980 (43.4)	867,454 (44.6)	
COVID-19 vaccine status at index date, n (%)			
Vaccine dose 1	21,932 (4.5)	92,355 (4.7)	<i>0.013</i>
Vaccine dose 2	685 (0.1)	5,964 (0.3)	<i>0.035</i>
Oxford-AstraZeneca	8,210 (1.7)	32,183 (1.7)	<i>0.003</i>
BioNTech-Pfizer	12,792 (2.6)	56,559 (2.9)	<i>0.017</i>
Moderna	0 (0.0)	3 (0.0)	<i>0.002</i>

IQR=interquartile range, SD=standard deviation, Socioeconomic status measured using the Index of Multiple Deprivation (IMD); standardised difference of less than 0.1 indicates a relatively small imbalance. Infected cohort included participants with a positive RT-PCR or antigen test for SARS CoV-2. The uninfected cohort included participants with no records of either confirmed or suspected COVID-19.

Table 2 Risk factors associated with the development of Long COVID (WHO definition)

Risk factor	Total numbers per strata [N = 384,137]	Long COVID symptoms [N=29,869] (7.78) n (%)	Follow-up in years Median (IQR)	Unadjusted HR (95% CI)	Adjusted HR* (95% CI)
Sex					
<i>Men</i>	171593	9090 (5.30)	0.09 (0.04-0.21)	Ref	Ref
<i>Women</i>	212544	20779 (9.78)	0.09 (0.04-0.21)	1.86 (1.81-1.90)	1.52 (1.48-1.56)
Age (years)					
<i>18-29</i>	95969	6932 (7.22)	0.09 (0.04-0.23)	Ref	Ref
<i>30 - 39</i>	78302	5805 (7.41)	0.08 (0.04-0.20)	1.13 (1.10-1.18)	0.94 (0.90-0.97)
<i>40 - 49</i>	75349	5784 (7.68)	0.08 (0.04-0.20)	1.14 (1.10-1.18)	0.89 (0.86-0.93)
<i>50 - 59</i>	73262	5485 (7.49)	0.08 (0.04-0.21)	1.07 (1.04-1.11)	0.80 (0.77-0.83)
<i>60 - 69</i>	35932	2790 (7.76)	0.09 (0.04-0.21)	1.09 (1.05-1.14)	0.74 (0.70-0.78)
<i>≥70</i>	25323	3073 (12.14)	0.11 (0.04-0.23)	1.39 (1.33-1.45)	0.75 (0.70-0.81)
Ethnicity					
<i>White</i>	246717	20462 (8.29)	0.09 (0.04-0.21)	Ref	Ref
<i>Asians</i>	47788	3647 (7.63)	0.08 (0.04-0.21)	0.90 (0.82-0.99)	0.99 (0.89-1.09)
<i>Black</i>	15846	1053 (6.65)	0.06 (0.03-0.15)	1.01 (0.91-1.11)	1.21 (1.10-1.34)
<i>Mixed</i>	5976	407 (6.81)	0.08 (0.04-0.19)	0.98 (0.92-1.04)	1.14 (1.07-1.22)
<i>Other</i>	5438	404 (7.43)	0.07 (0.04-0.19)	0.94 (0.91-0.97)	1.06 (1.03-1.10)
<i>Missing</i>	62372	3896 (6.25)	0.09 (0.04-0.22)	0.74 (0.71-0.76)	0.92 (0.88-0.95)
Body mass index (kg/m²)					
<i><18.5</i>	10312	762 (7.39)	0.10 (0.04-0.23)	0.93 (0.86-1.00)	0.93 (0.86-1.00)
<i>18.5-25</i>	117630	8849 (7.52)	0.09 (0.04-0.21)	Ref	Ref
<i>25-30</i>	109707	8612 (7.85)	0.09 (0.04-0.21)	1.06 (1.03-1.09)	1.07 (1.04-1.10)
<i>>30</i>	95799	9233 (9.64)	0.09 (0.04-0.21)	1.29 (1.25-1.33)	1.10 (1.07-1.14)
<i>Missing</i>	50689	2413 (4.76)	0.09 (0.04-0.22)	0.63 (0.60-0.65)	0.91 (0.86-0.95)
Smoking status					
<i>Non-Smoker</i>	141967	9671 (6.81)	0.09 (0.04-0.22)	Ref	Ref
<i>Ex-Smoker</i>	139294	12407 (8.91)	0.09 (0.04-0.21)	1.33 (1.29-1.36)	1.08 (1.05-1.11)
<i>Current Smoker</i>	85765	7072 (8.25)	0.08 (0.04-0.20)	1.31 (1.27-1.35)	1.12 (1.08-1.15)
<i>Missing</i>	17111	719 (4.20)	0.10 (0.04-0.23)	0.61 (0.56-0.65)	0.90 (0.83-0.97)
Socioeconomic status quintile (IMD)	78880	13207 (16.74)	0.09 (0.04-0.21)	2.92 (2.85-2.99)	2.05 (2.00-2.10)

<i>1 (least deprived)</i>					
2	66564	4392 (6.60)	0.09 (0.05-0.21)	Ref	Ref
3	68657	4963 (7.23)	0.09 (0.05-0.21)	1.09 (1.05-1.13)	1.05 (1.00-1.09)
4	70699	5486 (7.76)	0.08 (0.04-0.21)	1.19 (1.14-1.24)	1.10 (1.05-1.14)
5 (most deprived)	84002	6523 (7.77)	0.08 (0.04-0.21)	1.20 (1.16-1.25)	1.07 (1.03-1.11)
Missing	87270	7883 (9.03)	0.09 (0.04-0.22)	1.33 (1.28-1.38)	1.11 (1.07-1.16)
Symptoms recorded before COVID-19	78880	13207 (16.74)	0.09 (0.04-0.21)	2.92 (2.85-2.99)	2.07 (2.02-2.12)
Comorbidities					
COPD	8040	1741 (21.65)	0.09 (0.04-0.21)	2.71 (2.58-2.85)	1.55 (1.47-1.64)
BPH	4961	596 (12.01)	0.09 (0.04-0.22)	1.39 (1.28-1.51)	1.39 (1.28-1.52)
Fibromyalgia	4031	900 (22.33)	0.08 (0.04-0.20)	3.17 (2.97-3.39)	1.37 (1.28-1.47)
Anxiety	77753	10481 (13.48)	0.09 (0.04-0.21)	2.17 (2.12-2.23)	1.35 (1.31-1.39)
Erectile Dysfunction	16678	1551 (9.30)	0.09 (0.04-0.21)	1.15 (1.09-1.21)	1.33 (1.26-1.41)
Depression	83903	11222 (13.37)	0.08 (0.04-0.20)	2.22 (2.17-2.27)	1.31 (1.27-1.34)
Migraine	43043	5597 (13.00)	0.09 (0.04-0.21)	1.88 (1.83-1.94)	1.26 (1.22-1.30)
Multiple Sclerosis	791	98 (12.39)	0.09 (0.04-0.20)	1.52 (1.25-1.85)	1.26 (1.03-1.53)
Coeliac Disease	1669	207 (12.40)	0.09 (0.04-0.21)	1.58 (1.38-1.81)	1.25 (1.09-1.43)
Learning Disability	3283	295 (8.99)	0.07 (0.04-0.19)	1.22 (1.09-1.37)	1.24 (1.11-1.40)
IBS	27492	3691 (13.43)	0.09 (0.04-0.21)	1.84 (1.78-1.91)	1.20 (1.15-1.24)
Endometriosis	5727	800 (13.97)	0.08 (0.04-0.20)	1.92 (1.79-2.06)	1.19 (1.11-1.28)
Low Hb	20039	2683 (13.39)	0.08 (0.04-0.21)	1.78 (1.71-1.85)	1.18 (1.13-1.23)
Deafness	3767	514 (13.64)	0.10 (0.04-0.22)	1.53 (1.40-1.67)	1.16 (1.06-1.27)
Eating disorder	3488	504 (14.45)	0.09 (0.04-0.21)	1.92 (1.75-2.09)	1.16 (1.06-1.27)
Substance Misuse	6449	775 (12.02)	0.08 (0.04-0.20)	1.69 (1.58-1.82)	1.15 (1.07-1.23)
Backpain	5483	718 (13.10)	0.08 (0.04-0.20)	1.76 (1.64-1.90)	1.15 (1.07-1.24)
Asthma	76946	8527 (11.08)	0.09 (0.04-0.21)	1.59 (1.55-1.63)	1.15 (1.12-1.18)
Chronic Sinusitis	6838	873 (12.77)	0.09 (0.04-0.21)	1.63 (1.52-1.74)	1.14 (1.07-1.22)
PCOS	9599	1166 (12.15)	0.08 (0.04-0.19)	1.73 (1.63-1.84)	1.14 (1.07-1.21)

*Adjusted hazard ratios estimated using a multivariable Cox proportional hazards model including age, sex, ethnic group, socioeconomic status, index year, vaccination status, symptoms recorded before COVID-19, and comorbidities.

IMD=Index of Multiple Deprivation, COPD=chronic obstructive pulmonary disease, BPH=benign prostatic hyperplasia, IBS = irritable bowel syndrome, PCOS = Polycystic Ovary Syndrome

Declarations

Author contributions

SH, KN, and MC conceived the research question and idea for the study. SH, KN, A Subramanian, PM and MC agreed the study methods. PM and TW facilitated data acquisition. KG supported data extraction and data management. SEH, MC, SH, OLA, CM, and GT informed the selection of symptoms. TT provided statistical advice. KB, NSW, MS, FK, KO, RH, NB, NC, SL, GT, and SH developed SNOMED-CT code lists for data extraction. A Shah provided advice on study design and checked code lists. A Subramanian performed the statistical analysis with input from SH and KN. A Subramanian and SH drafted the manuscript with input from all co-authors. KB supported the use of the RECORD checklist for drafting the manuscript. KLM and JC provided patient and public involvement for the study. All co-authors reviewed and approved the final draft of the manuscript.

Conflicts of interest

MC is Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation, Director of the Centre for Patient Reported Outcomes Research and is a National Institute for Health Research (NIHR) Senior Investigator. MC receives funding from the National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre, the NIHR Surgical Reconstruction and Microbiology Research Centre and NIHR ARC West Midlands at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, Health Data Research UK, Innovate UK (part of UK Research and Innovation), Macmillan Cancer Support, SPINE UK (UKRI), UCB Pharma, Janssen, GSK and Gilead. MC has received personal fees from Astellas, Aparito Ltd, CIS Oncology, Takeda, Merck, Daiichi Sankyo, Glaukos, GSK and the Patient-Centered Outcomes Research Institute (PCORI) outside the submitted work.

SEH is supported by the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC), West Midlands at the University of Birmingham. SEH has received personal fees from Cochlear Ltd and Aparito Ltd outside the submitted work.

OLA receives funding from the NIHR Birmingham Biomedical Research Centre (BRC), NIHR Applied Research Collaboration (ARC), West Midlands at the University of Birmingham and University Hospitals Birmingham NHS Foundation, Innovate UK (part of UK Research and Innovation), Gilead Sciences Ltd, and Janssen pharmaceuticals, Inc. OLA declares personal fees from Gilead Sciences Ltd, GlaxoSmithKline (GSK) and Merck outside the submitted work.

CM receives funding from the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre, Innovate UK, and has received personal fees from Aparito Ltd outside the submitted work.

ADS is supported by a postdoctoral fellowship from THIS Institute, NIHR University College London Hospitals BRC, grants from NIHR and British Heart Foundation Accelerator Award.

ES has received grant from the Wellcome Trust, MRC, NIHR EME, NIHR HTA, HDR-UK, BLF, EPSRC and Alpha 1 Foundation in the last 36 months. She has been an honorarium for lectures about COVID-19 treatments which are run by GSK, attended a virtual conference at the European Respiratory Society in 2020 which was funded by AstraZeneca, and participated in an advisory board for COPD which is run by Boehringer Ingelheim.

SM has received funding from NIHR (RfPB, PGfAR, HTA and EME streams), UKRI, ESRC and the Midlands Engine. He has attended educational events funded by Psychiatric Genetic Testing, Janssen and Lundbeck in the last 5 years.

PM, TW, CI and EL are employees of CPRD, the data custodians for CPRD Aurum. CPRD is jointly sponsored by the UK government's Medicines and Healthcare products Regulatory Agency and the National Institute for Health Research (NIHR). As a not-for-profit UK government body, CPRD seeks to recoup the cost of delivering its research services to academic, industry and government researchers through research user license fees.

JC receives funding from NIHR on PPI from a study at UCL (NIHR132914) and a study at University Hospitals Bristol (NIHR203304). JC is a lay member on the NICE Covid expert panel and a citizen partner on the COVID END Evidence synthesis global horizon scanning panel. JC declares personal fees from MEDABLE Inc, GlaxoSmithKline, Roche Canada outside of submitted work.

KLM is a Trustee and volunteer at Long COVID SOS. KLM is on Long COVID Advisory Board for Dysautonomia International. KLM is employed by NIHR.

All other co-authors have no relevant competing interests to declare.

Acknowledgements

We would like to thank all the patients on the TLC Lived Experience Advisory Group (LEAP) and Dr Navjeet Mangat for supporting our LEAP member recruitment. We would also like to thank Anita Walker, Karen Jones and Yuangen Lee for providing administrative support for the study.

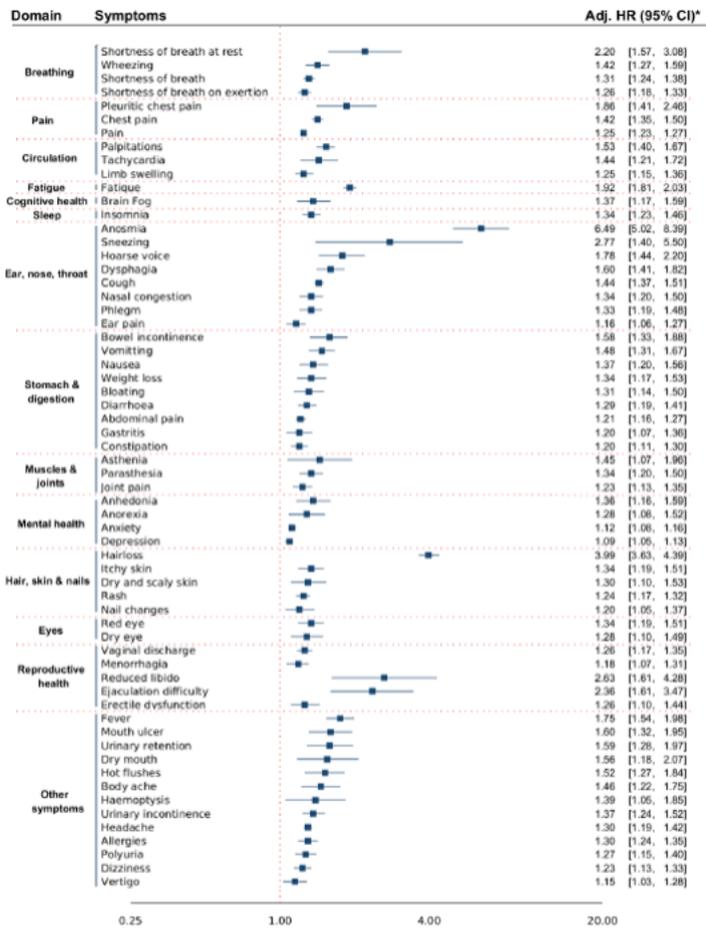
Funding

This work is independent research jointly funded by the National Institute for Health Research (NIHR) and UK Research and Innovation (UKRI) (Therapies for Long COVID in non-hospitalized individuals: From symptoms, patient reported outcomes and immunology to targeted therapies (The TLC Study), COV-LT-0013). The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care or UKRI. The funders had no role in the design and conduct of the study including: the collection, management, analysis, and interpretation of the data; preparation and review of the manuscript.

Data statement

Requests for access to the analysis dataset used in this study should be addressed to the corresponding author at K.Nirantharan@bham.ac.uk. All proposals requesting data access will need to specify how it is planned to use the data, and all proposals will need approval of the study team and CPRD before data release. SNOMED CT code lists for the symptoms and comorbidities can be provided upon reasonable request.

Figures



* Adjusted for age, sex, body mass index, ethnicity, smoking status, deprivation status and symptom of interest at baseline

Figure 1

Symptoms associated with SARS CoV-2 ≥ 12 weeks post-infection

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)
- [flatNirantharakumarepc.pdf](#)
- [flatNirantharakumarrs.pdf](#)