The burden of long COVID: a multinational cohort analysis of Spanish and UK data including SARS-CoV-2 infections, reinfections, and matched contemporaneous test negative controls

Daniel Prieto-Alhambra (daniel.prietoalhambra@ndorms.ox.ac.uk)
Pharmaco- and Device Epidemiology Group, CSM, NDORMS, University of Oxford https://orcid.org/0000-0002-3950-6346
Kristin Kostka
Pharmaco- and Device Epidemiology Group, CSM, NDORMS, University of Oxford
Elena Roel
I Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol)
Nhung Trinh
PharmacoEpidemiology and Drug Safety Research Group, Department of Pharmacy, University of Oslo
Antonella Delmestri
University of Oxford https://orcid.org/0000-0003-0388-3403
Lourdes Mateu
Infectious Diseases Department, Fight against AIDS Foundation (FLS), Germans Trias i Pujol Hospital, Badalona, Catalonia, Spain
Roger Paredes
Hospital Germans Trias i Pujol & irsiCaixa AIDS Research Institute https://orcid.org/0000-0002-6553-691X
Talita Duarte-Salles
Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain,
Martí Català
Pharmaco- and Device Epidemiology Group, CSM, NDORMS, University of Oxford
Annika Jödicke
Pharmaco- and Device Epidemiology Group, CSM, NDORMS, University of Oxford

Article

Keywords: post COVID complications, epidemiology, primary care data

Posted Date: March 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2642600/v1

Abstract

As limited data was available on the effect of persisting COVID-19 symptoms, we characterised long COVID and identified key symptoms associated with persistent disease. Using primary care data from Spain and UK, we estimated incidence rates of long COVID in the population and among COVID-19 patients over time. Subsequently, we investigated which WHO-listed symptoms were particularly differential for long COVID by comparing their frequency in COVID-19 patients vs matched test-negative controls. Lastly, we compared persistent symptoms after first infections vs. reinfections.

Fortunately, the proportion of COVID-19 cases resulting in long COVID declined over the study period. Risk for altered smell/taste, dyspnoea, and fatigue were consistently higher in long COVID patient vs controls [RR between 5.97-1.09]. All persistent symptoms were less common after reinfection than first infection.

More research is needed into the definition of long COVID, and the effect of interventions to minimise the risk and impact of persistent symptoms.

Introduction

Three years after the world's first reported cases of the novel coronavirus (SARS-CoV-2 and ensuing global pandemic), COVID-19 is still a significant burden to morbidity and mortality globally[1]. As early as May 2020[2], clinicians observed a subset of COVID-19 cases that evolved from an acute viral infection into a long-term condition with a puzzling array of symptoms, subsequently called "long COVID"[3]. Many have sought to standardise how clinicians distinguish long COVID from initial infection and ongoing symptomatic COVID-19[4-6]. Through a Delphi consensus process[7], the World Health Organization (WHO) defined long COVID as a condition that occurs in individuals with a probable or confirmed history of SARS-CoV-2 infection who present at least 3 months from the onset of COVID-19 with new or persisting symptoms for at least 2 months that cannot be attributed to another aetiology. Common persisting symptoms include debilitating fatigue, shortness of breath, memory or cognitive dysfunction, and a variety of other multi-system symptoms[8] that affect day-to-day living. Symptoms can be new onset after initial recovery from first COVID-19 infection or can persist from the infection and may fluctuate or relapse over time. Recent literature suggests significant heterogeneity in how individuals experience symptoms, including the emergence of potential clinical subgroups[9]. Hundreds of single-country analyses have evaluated long COVID patterns in specific care environments, each contributing new understanding to the incompletely understood natural history of long COVID[10 11]. However, few studies have attempted to compare clinical definitions across multiple countries and sources of real-world data.

Quantifying long COVID is challenging because many public health policies combatting the initial disease spread, such as extensive lockdowns, travel bans, and testing protocols, created significant disruption in 'normal' living. Many studies have examined the impact of non-pharmacological interventions on overall population health during the pandemic, including understanding how COVID-19 shaped mental health[12 13]. The result is a muddling of symptomology: some reported long COVID symptoms are non-specific and prevalent in the general population regardless of infection status. As we aim to understand the long COVID symptoms, we must also be critical of symptoms that may reflect pre-existing comorbidities or the wider effects of the pandemic's disruption of day-to-day life.

For this study, we took advantage of large primary care electronic health records datasets from two European countries, namely CPRD AURUM (England, UK) and SIDIAP (Catalonia, Spain), to characterise long COVID and identify key symptoms associated with persistent disease. We estimated age- and sex-specific incidence rates of long COVID in the general population and among people with confirmed COVID-19 over time. We investigated which of the 25 WHO-listed symptoms are more specific to long COVID by comparing the occurrence of each symptom among COVID-19 patients and people who tested negative in the same week. We also compared the occurrence of persistent symptoms after a first infection or after reinfection with SARS-CoV-2.

Results

Characterisation of people with COVID-19 and negative-test comparator cohorts

We included 469,503 and 303,835 COVID-19 infections and 1,643,589 and 1,024,256 first/earliest SARS-CoV-2 negative tests recorded during the study period in SIDIAP and CPRD AURUM, respectively. The study inclusion process is provided in Figure 1. Baseline characteristics of both the COVID-19 infection cohort and first COVID-19 negative test cohort are reported in Table 1. Overall, follow-up was longer in SIDIAP than in CPRD, with for example a median 358 days of follow-up for COVID-19 infections in SIDIAP and 151 days in CPRD. Participants from both databases were mostly young adults (<50 years), female and predominantly unvaccinated. Baseline characteristics for cohorts of first COVID-19 infections, COVID-19 re-infections, and all COVID-19 negative tests are provided in Supplementary Table S1.

Table 1: Baseline characteristics in the COVID-19 infection and first SARS-CoV-2 negative test cohorts, by database

	SIDIAP		CPRD	
	COVID-19 infection	First SARS-CoV-2 negative test	COVID-19 infection	First SARS-CoV-2 negative test
N	469,503	1,643,589	303,835	1,024,256
Days of follow-up (median [IQR])	358 [254, 365]	365 [275, 365]	151 [130, 188]	176 [140, 208]
Test date period (%)				
Sep-Dec 2020	146,975 (31.3)	700,640 (42.6)	244,024 (80.3)	926,529 (90.5)
Jan-Apr 2021	137,728 (29.3)	527,588 (32.1)	59,811 (19.7)	97,727 (9.5)
May-Aug 2021	158,617 (33.8)	289,618 (17.6)		
Sep-Dec 2021	26,183 (5.6)	125,743 (7.7)		
Age (median [IQR])	43 [30, 57]	48 [34, 63]	41 [29, 54]	40 [29, 53]
Age, categories (%)				
≤34	157,778 (33.6)	412,061 (25.1)	112,733 (37.1)	385,521 (37.6)
35-49	138,907 (29.6)	469,681 (28.6)	90,023 (29.6)	308,102 (30.1)
50-64	100,164 (21.3)	395,324 (24.1)	73,746 (24.3)	235,432 (23.0)
65-79	48,249 (10.3)	263,030 (16.0)	21,014 (6.9)	75,058 (7.3)
≥80	24,405 (5.2)	103,493 (6.3)	6,319 (2.1)	20,143 (2.0)
Sex, female (%)	247,262 (52.7)	872,441 (53.1)	166,433 (54.8)	585,794 (57.2)
Vaccination status (%)				
Not vaccinated	376,360 (80.2)	1,357,450 (82.6)	302,394 (99.5)	1,022,635 (99.8)
One dose	40,062 (8.5)	83,652 (5.1)	1,385 (0.5)	1,547 (0.2)
Two doses	52,679 (11.2)	199,534 (12.1)	56 (0.0)	68 (0.0)
Three or more (booster doses	402 (0.1)	2,953 (0.2)	<5	6 (0.0)
COVID-19 PCR test (%)	237,014 (50.5)	1,053,223 (64.1)	283,986 (93.5)	928,768 (90.7)
Comorbidities (%)				
Asthma	35,787 (7.6)	124,741 (7.6)	48,279 (15.9)	176,957 (17.3)
Autoimmune disease	8,513 (1.8)	35,566 (2.2)	6,128 (2.0)	22,070 (2.2)
COPD	12,492 (2.7)	62,252 (3.8)	3,876 (1.3)	16,644 (1.6)
Dementia	7,371 (1.6)	17,482 (1.1)	2,879 (0.9)	9,490 (0.9)
Diabetes	40,129 (8.5)	157,725 (9.6)	21,161 (7.0)	61,211 (6.0)
Heart disease	52,055 (11.1)	220,466 (13.4)	19,038 (6.3)	64,003 (6.2)
Cancer	30,472 (6.5)	147,141 (9.0)	11,977 (3.9)	43,264 (4.2)
Hypertension	82,866 (17.6)	362,131 (22.0)	39,196 (12.9)	123,587 (12.1)
Renal impairment	20,081 (4.3)	84,046 (5.1)	10,126 (3.3)	32,428 (3.2)

SIDIAP = Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària, CPRD = Clinical Practice Research Datalink,

IQR = interquartile range, with q25 and q75 provided.

Long COVID and the distribution of ongoing symptoms across cohorts are reported in Supplementary Tables S2 and S3. Among those with a COVID-19 infection, 20.6% in SIDIAP and 7.8% in CPRD had long COVID, as defined by the presence of at least one of the 25 WHO-listed symptoms recorded ≥90 days after infection (and without a history of the symptoms in the 180 days before COVID-19 infection). In comparison, 20.9% of participants in the first negative test cohort in SIDIAP and 9.7% in CPRD had at least one symptom recorded ≥90 days after the test. Across all cohorts and databases more than 74% of individuals with long COVID symptoms had only one symptom recorded. Overall, the most common symptoms were joint pain, abdominal pain, and anxiety. Sensitivity analyses showed that “ongoing symptoms" (≥28 days) were more frequent than long COVID symptoms (≥90 days), with 25.4% of COVID-19 infections recording at least one symptom ≥28 days after COVID-19 infection or negative test in SIDIAP and 13.1% in CPRD.

Incidence of long COVID across the two databases

Monthly incidence rates per 100,000 person-years of COVID-19 infection and long COVID symptoms ≥90 days after COVID-19 are shown stratified by age and sex in Figure 2 and are shown overall in Figure S2 and Table S4. The incidence rate of COVID-19 in each database mirrored the official rate of SARS-CoV-2 infections in that country. The incidence rate of long COVID in the general population mirrored and followed waves of COVID-19. However, rates of long COVID among individuals with COVID-19 declined over time in both databases.

Risks of long COVID symptoms following COVID-19

We matched 438,159 and 303,010 COVID-19 patients up to 1:3 to 1,059,596 and 765,331 *first* negative test controls in SIDIAP and CPRD, respectively. Baseline characteristics were balanced for the matched cohorts, with standardised mean differences (SMD) <0.1 for all covariates in both databases (Supplementary Table S5). Figure 3 and Table S6 show rate ratios for each pre-defined symptom, comparing COVID-19 to matched test negative cohorts. We found increased risks at ≥90 days for altered smell and taste in both databases, with a rate ratio of 4.73 (95% CI: 4.18 to 5.36) in SIDIAP and 5.97 (4.64 to 7.70) in CPRD. Increased risks of dyspnoea and fatigue/malaise were also seen in both databases, with rate ratios of 1.35 (1.29 to 1.41) in SIDIAP and 1.09 (1.04 to 1.16) in CPRD for dyspnoea and of 1.15 (1.11 to 1.18) in SIDIAP and 1.24 (1.18 to 1.32) in CPRD for fatigue/malaise. Increased risks of menstrual problems, memory issues, chest pain, cough, tachycardia, headache, abdominal pain, pins and needles sensation, joint pain, allergy, blurred vision, dizziness and anxiety were also seen in SIDIAP, with rate ratios ranging from 1.04 (1.02 to 1.07) for anxiety to 1.22 (1.16 to 1.29) for menstrual problems. However, none of these risks were confirmed in CPRD. In CPRD, menstrual problems, headache, joint pain, abdominal pain, neuralgia, gastrointestinal issues, sleep disorder, cough, depression and anxiety were more common in the negative-test cohort than among COVID-19 matched cases, which was the opposite relationship to that seen in SIDIAP.

Results from sensitivity analyses with symptoms ≥28 days after COVID-19 infection or negative test are provided in Supplementary Figure S3 and Table S7. Both databases showed consistent increased risks for altered smell/taste, fatigue, and dyspnoea. In SIDIAP, in addition to the previously mentioned symptoms after ≥90 days, increased risks were also seen for anxiety, dizziness, sleep disorder, and joint pain. In CPRD, increased risks of memory issues, chest pain, tachycardia, and pins and needles sensation were also observed. Results from sensitivity analyses matching to any negative test are included in Supplementary Tables S8 and S9 and Figures S4 and S5 and showed similar findings.

Long COVID symptoms associated with COVID-19 re-infection

We matched 39,492 and 2,126 *first* COVID-19 infections up to 3:1 to 13,253 and 709 *reinfections* in SIDIAP and CPRD, respectively. Baseline characteristics were broadly balanced after matching, with SMD<0.1 for demographics and all co-variates except for dementia in SIDIAP (first infections: 3.4%, reinfections: 6.0%, SMD: 0.119) (Table S10). Figure 4 illustrates that the risk of long COVID symptoms was reduced for all individual symptoms after re-infection, compared to after first infection in both databases (Table S11). Results from sensitivity analyses with symptoms assessed after ≥28 days showed a similar trend (Figure S6, Table S12).

Discussion

Statement of principal findings

This is the first multinational cohort study on the presentation of long COVID, including over 770,000 COVID-19 cases and more than 1.6 million negative-test matched controls from Spain and the UK. We found high proportions of long COVID with persisting symptoms for ≥90 days after infection, of almost 21% of infections in Spain and 8% in the UK. At the population level, waves of long COVID followed each wave of community transmission in the study period, affecting predominantly women and young adults. However, the proportion of people infected with COVID-19 who went on to develop long COVID reassuringly declined over time.

At the patient level, some persisting symptoms appeared more specific and differential of long COVID when compared to matched contemporaneous negative-test controls. Altered smell and taste, fatigue, and dyspnoea were consistently more common in long COVID cases than in controls in both Spain and the UK. Persistent menstrual issues, memory issues, and cognitive dysfunction were substantially increased after COVID-19 infection more than among controls in Spanish data, but not among UK participants.

For the first time, we report a substantial reduction in the risk of persistent symptoms after reinfection compared to first infection. All WHO-listed symptoms of long COVID appeared less common after a reinfection than after a first infection, after matching by age and date of infection.

Research in context

Persistent symptoms were common after COVID-19 infection but were prevalent at almost the same extent as in the general population during the same period. In line with our results, previous studies found a substantial proportion of non-infected people with records of similar symptoms.[14] This finding highlights the challenge in identifying long COVID , which could result in misclassification and difficulties in diagnosing long COVID.

While the prevalence of long COVID remained high in both countries, our results showed a decline in the proportion of people developing persistent symptoms after COVID-19 over time. The REACT-2 study, a representative community survey among adults in England, found a similar trend, with the prevalence of people experiencing at least one symptom >12 weeks following COVID-19 declining from 37.7% between September 2020 and February 2021 to 21.6% in May 2021[15]. This trend may be attributable to the effect of vaccines,[16] previous immunity (i.e., reinfection), and differences in the pre-dominant variant[17]. However, shorter follow-up time available later in the pandemic, potential differences in testing practice shifting from prioritising severely ill people to wider testing for

screening as well as non-systematically collection of symptoms carrying the risk for “reporting exhaustion”, a decreased reporting of the same persisting symptoms, may also account for the observed reduction in long COVID prevalence over time.

Our study adds to previous research focusing on long-term complications following COVID-19 infection and frequencies of persistent symptoms. Hundreds of different symptoms have been reported in relation to COVID-19, of which the WHO highlighted the 25 most characteristic in their Delphi consensus. Previous studies compared some of these symptoms in people with and without COVID-19. Subramanian et al.[14] determined symptoms associated with COVID-19 after 12 weeks by comparing people with confirmed COVID-19 to propensity-score-matched controls without recorded or suspected COVID-19 infection in CPRD. They also found that symptoms with a strong association with SARS-COV-2 included anosmia, shortness of breath at rest, and fatigue. However, although we found that only some symptoms were associated with COVID infection, Subramanian et al. found that the risk for all recorded symptoms was significantly increased after infection.

The effect of reinfection on the severity and persistence of COVID-19 symptoms remains a topic of great interest. Our findings are reassuring, with a clear reduction in risk of persistent symptoms after a reinfection, compared to matched participants with a first infection. Previous studies on this topic are scarce, with ours being the first to study the effect of reinfection on the risk of long COVID symptoms as defined by the WHO. A previous study on post-acute complications and organ system disorders in people following first or reinfection in the US Veterans Health Administration database[18] reported an increased risk for at least one sequela, which was highest during the acute phase but attenuated over time. However, this previous study did not investigate long COVID as an outcome or match or control by age or index date, and it is unclear how the authors controlled for other confounding factors.

Strengths and weaknesses

COVID-19 datasets are burdened with systemic limitations as the pandemic placed significant strain on the global healthcare system. As broad testing was not available in most countries in early 2020, we began our study period in September 2020, excluding the first wave of the pandemic. With widespread issues in testing capacities to meet public demand and the advent of self-administered tests, underreporting of infections is expected across all pandemic waves. Some reinfections may therefore have been misclassified as first infections. Likewise, we expect underreporting of clinical symptoms as people might not have been seen by a clinician, particularly for milder symptoms and during infection peaks. Aside from differences in follow-up length and time of subject inclusions, differences in healthcare, with more virtual clinical work in the UK than in Spanish primary care practice, may explain the observed difference in rates of long COVID between SIDIAP and CPRD. Long COVID is a new condition, with its definition developing over time. With clinical awareness still evolving, reporting bias in recording practice of characteristic symptoms cannot be ruled out. A systematic and comprehensive collection and reporting of long-COVID symptoms would be needed to overcome this limitation.

Our study also has strengths. This is the first multinational database study on this topic. It included two large population-based databases from different European regions with primary-care-based universal public healthcare. CPRD AURUM and SIDIAP provide high-quality data for research and are representative of their respective populations[19 20]. Prior research evaluated individual symptom prevalence associated with long COVID diagnosis, but no other study has compared these estimates between different countries and care settings in a matched cohort, looking at the implications of grouping by first infection, re-infection, and negative test. Our methodology allowed us to ascertain general population averages during pandemic times and quantify overall health status after lockdowns or other public health policies, regardless of COVID-19 status. Despite the many challenges that long COVID patients report facing in gaining clinical recognition of their symptoms, an increasingly consistent clinical presentation is evident in this multi-database view.

Systematic reviews have shown that data harmonisation is fundamental to improving the clinical utility of findings[21]. A strength of our research is the use of a common data model (OMOP CDM) and shared conventions in data harmonization, allowing for consistent representation of clinical information despite heterogeneous source systems.

Conclusions

Waves of long COVID were observable following community transmission during the first two years of the pandemic, affecting predominantly women and young adults. However, the proportion of COVID-19 cases affected by persistent symptoms declined more recently, which could be due to a mixture of growing immunity due to vaccines and natural immunity. Our findings suggest that natural immunity plays a key role as reinfections had a much lower risk of long COVID than first infections. We identified three key symptoms that can help to differentiate people living with long COVID: altered smell and taste, fatigue, and dyspnoea. More work is needed to improve the existing definition of long COVID to enhance future trials into the efficacy of vaccines and antivirals to prevent and/or manage this disease.

Declarations

Author contribution

DPA, MC, and AMJ led the conceptualisation of the study with contributions from ER and KK. KK, DPA, and AMJ lead the phenotyping of long COVID symptoms. AD mapped and curated CPRD data. ER and MC conducted the statistical analyses on the respective databases. DPA, ER, TDS, LM, RP, NT and AMJ clinically interpreted the results. KK, ER, and AMJ wrote the first draft of the manuscript. All authors read, contributed to, and approved the last version of the manuscript. DPA and AMJ obtained the funding for this research.

Conflict of Interest statement

DPA reports grants from the European Medicines Agency and Innovative Medicines Initiative as well as Amgen, Chiesi-Taylor, Lilly, Janssen, Novartis, and UCB Biopharma. His research group has received consultancy fees from Astra Zeneca and UCB Biopharma. Amgen, Astellas, Janssen, Synapse Management Partners and UCB Biopharma have funded or supported training programmes organised by DPA's department.

LM reports grants from Grifols, speaker fees from AstraZeneca and Gilead and participation in Advisory Boards for Gilead.

NT is supported by the Norwegian Research Council (grant no. 288696) and has received internationalization supports from UiO: Life Science and travel grants from the Norwegian Research Council.

All other authors declare no conflict of interest.

Data sharing statement

Analyses were performed locally in compliance with all applicable data privacy laws.

CPRD data were obtained under the CPRD multi-study license held by the University of Oxford after Research Data Governance (RDG) approval. Direct data sharing is not allowed.

Following current European and national law, SIDIAP data are only available for researchers participating in this study. However, researchers from public institutions can request data from SIDIAP if they comply with certain requirements. Further information is available online (https://www.sidiap.org/index.php/menu-solicitudesen/application-proccedure) or by contacting SIDIAP (sidiap@idiapjgol.org).

Ethics approval

The study was approved by the relevant Institution Review Boards: the CPRD's Research Data Governance Process (Protocol No 21_000557), the Clinical Research Ethics committee of Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol) (approval number 4R22/133).

Transparency statement

The lead authors affirm that the manuscript is an honest, accurate, and transparent account of the study being reported, and that no important aspects of the study have been omitted.

Funding

This work is independent research funded by the National Institute for Health and Care Research (NIHR) (COV-LT2-0006). The views expressed in this publication are those of the author(s) and not necessarily those of NIHR or the Department of Health and Social Care. ER was supported by Instituto de Salud Carlos III (ISCIII; Río Hortega 2020, CM20/00174).

Acknowledgement

This study is based in part on data from the Clinical Practice Research Datalink (CPRD) obtained under license from the UK Medicines and Healthcare products Regulatory Agency. We thank the patients who provided these data and the NHS who collected the data as part of their care and support. We also thank the *Institut Català de la Salut (ICS)* and the *Programa d'analítica de dades per a la recerca i la innovació en salut (PADRIS)* for providing access to the different data sources accessible through SIDIAP. All interpretations, conclusions and views expressed in this publication are those of the author(s) alone. We acknowledge English language editing by Dr Jennifer A de Beyer, Centre for Statistics in Medicine, University of Oxford.

References

1. Office for National Statistics. Coronavirus (COVID-19) Infection Survey, UK: 11 November 2022 2022 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/ accessed 01 December 2022.
2. Mahase E. Covid-19: What do we know about "long covid"? *BMJ* 2020;370:m2815. doi: 10.1136/bmj.m2815 [published Online First: 2020/07/16]
3. Perego E, Callard F, Stras L, et al. Why the Patient-Made Term 'Long Covid' is needed. *Wellcome Open Research* 2020;5 doi: 10.12688/wellcomeopenres.16307.1
4. Greenhalgh T, Sivan M, Delaney B, et al. Long covid-an update for primary care. *BMJ* 2022;378:e072117. doi: 10.1136/bmj-2022-072117 [published Online First: 2022/09/23]
5. Michelen M, Manoharan L, Elkheir N, et al. Characterising long COVID: a living systematic review. *BMJ Glob Health* 2021;6(9) doi: 10.1136/bmjgh-2021-005427 [published Online First: 2021/09/29]

6. Crook H, Raza S, Nowell J, et al. Long covid-mechanisms, risk factors, and management. *BMJ* 2021;374:n1648. doi: 10.1136/bmj.n1648 [published Online First: 2021/07/28]
7. World Health Organization. A clinical case definition of post COVID-19 condition by a Delphi consensus 6 October 2021 [Available from: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1 accessed 6 October 2021.
8. Reese JT, Blau H, Bergquist T, et al. Generalizable Long COVID Subtypes: Findings from the NIH N3C and RECOVER Programs. *medRxiv* 2022 doi: 10.1101/2022.05.24.22275398 [published Online First: 2022/06/07]
9. Kenny G, McCann K, O'Brien C, et al. Identification of Distinct Long COVID Clinical Phenotypes Through Cluster Analysis of Self-Reported Symptoms. *Open Forum Infect Dis* 2022;9(4):ofac060. doi: 10.1093/ofid/ofac060 [published Online First: 2022/03/11]
10. Deer RR, Rock MA, Vasilevsky N, et al. Characterizing Long COVID: Deep Phenotype of a Complex Condition. *EBioMedicine* 2021;74:103722. doi: 10.1016/j.ebiom.2021.103722 [published Online First: 2021/11/29]
11. Raveendran AV, Jayadevan R, Sashidharan S. Long COVID: An overview. *Diabetes Metab Syndr* 2021;15(3):869-75. doi: 10.1016/j.dsx.2021.04.007 [published Online First: 2021/04/24]
12. Penninx B, Benros ME, Klein RS, et al. How COVID-19 shaped mental health: from infection to pandemic effects. *Nat Med* 2022;28(10):2027-37. doi: 10.1038/s41591-022-02028-2 [published Online First: 2022/10/04]
13. Hossain MM, Tasnim S, Sultana A, et al. Epidemiology of mental health problems in COVID-19: a review. *F1000Res* 2020;9:636. doi: 10.12688/f1000research.24457.1 [published Online First: 2020/10/24]
14. Subramanian A, Nirantharakumar K, Hughes S, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nat Med* 2022;28(8):1706-14. doi: 10.1038/s41591-022-01909-w [published Online First: 2022/07/26]
15. Whitaker M, Elliott J, Chadeau-Hyam M, et al. Persistent COVID-19 symptoms in a community study of 606,434 people in England. *Nat Commun* 2022;13(1):1957. doi: 10.1038/s41467-022-29521-z [published Online First: 2022/04/14]
16. UK Health Security Agency. The effectiveness of vaccination against long COVID: A rapid evidence briefing 2022 [accessed February 2022.
17. Magnusson K, Kristoffersen DT, Dell'Isola A, et al. Post-covid medical complaints following infection with SARS-CoV-2 Omicron vs Delta variants. *Nat Commun* 2022;13(1):7363. doi: 10.1038/s41467-022-35240-2 [published Online First: 2022/12/01]
18. Bowe B, Xie Y, Al-Aly Z. Acute and postacute sequelae associated with SARS-CoV-2 reinfection. *Nat Med* 2022;28(11):2398-405. doi: 10.1038/s41591-022-02051-3 [published Online First: 2022/11/11]
19. Recalde M, Rodriguez C, Burn E, et al. Data Resource Profile: The Information System for Research in Primary Care (SIDIAP). *Int J Epidemiol* 2022;51(6):e324-e36. doi: 10.1093/ije/dyac068 [published Online First: 2022/04/14]
20. Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019;48(6):1740-40g. doi: 10.1093/ije/dyz034 [published Online First: 2019/03/13]
21. O'Mahoney LL, Routen A, Gillies C, et al. The prevalence and long-term health effects of Long Covid among hospitalised and non-hospitalised populations: A systematic review and meta-analysis. *EClinicalMedicine* 2023;55:101762. doi: 10.1016/j.eclinm.2022.101762 [published Online First: 2022/12/08]
22. CPRD. Release Notes: CPRD Aurum June 2021. 2021 doi: 10.48329/pyc2-we97
23. Burn E, Fernández-Bertolín S, Voss EA, et al. Establishing and characterising large COVID-19 cohorts after mapping the Information System for Research in Primary Care in Catalonia to the OMOP Common Data Model. *medRxiv* 2021 doi: 10.1101/2021.11.23.21266734
24. The Book of OHDSI. Chapter 4 The Common Data Model [Available from: https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html2022.
25. Rao G, Schuemie M, Ryan P, et al. CohortDiagnostics: Diagnostics for OHDSI Cohorts 2022 [Available from: https://ohdsi.github.io/CohortDiagnostics/2022.
26. National Institute for Health and Care Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN), and Royal College of General Practitioners (RCGP). COVID-19 rapid guideline: managing the longterm effects of COVID-19 2022 [Available from: https://www.nice.org.uk/guidance/ng188/resources/covid19-rapid-guideline-managing-the-longterm-effects-of-covid19-pdf-51035515742 accessed 03.11.2022

Methods

Data sources, study design, and study period

We conducted a population-based descriptive cohort study using primary care electronic health records from England, UK and Catalonia, Spain.

Primary care electronic health records in England were obtained from the Clinical Practice Research Datalink (CPRD) AURUM, which comprises 20% of the population in the UK.[20 22] Spanish data were obtained from the Information System for Research in Primary Care (SIDIAP; www.sidiap.org) database, which captures more than 75% of the population living in Catalonia, a region in the northeast of Spain.[19] SIDIAP was linked to hospital discharge records from public and private hospitals in Catalonia (*Conjunt Mínim Bàsic de Dades d'Alta Hospitalària, CMBD-AH*).[23] Both databases include information on demographics, clinical diagnoses, and laboratory tests, including SARS-CoV-2 reverse transcription polymerase chain reaction (RT-PCR) tests. SIDIAP also captures SARS-CoV-2 antigen tests performed at public healthcare facilities. Although information on SARS-CoV-2 antigen testing may appear in CPRD, the counts are expected to be incomplete.

The databases were standardised to the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM),[24] allowing the same analytical code to be applied without sharing individual data.

The study period spanned from 1 September 2020 to the end of data availability, i.e. June 2021 for CPRD and March 2022 for SIDIAP, where data was censored to avoid misclassification due to changes in COVID testing policies.

Study population

We defined three non-mutually-exclusive COVID-19 cohorts – (1) all COVID-19 cases, (2) first SARS-CoV-2 infections, and (3) SARS-CoV-2 reinfections – and two negative-test comparator cohorts – (1) first/earliest SARS-CoV-2 negative tests and (2) all SARS-CoV-2 negative tests.

COVID-19 cases were identified using positive SARS-CoV-2 antigen and RT-PCR tests, using the test date as the index date. COVID-19 was defined as infections without a record of SARS-CoV-2 infections in the previous 42 days. First infections were defined as SARS-CoV-2 infections without any prior history of COVID-19. Reinfections were defined as SARS-CoV-2 infections that were not identified as a first infection.

The two negative-test comparator cohorts were identified using negative SARS-CoV-2 antigen and RT-PCR tests, using the test date as the index date. Individuals included in these cohorts were required to have a SARS-CoV-2 negative test result without a clinical COVID-19 diagnosis or positive SARS-CoV-2 test result before the index date and up to 120 days after the index date. First SARS-CoV-2 negative tests were defined as SARS-CoV-2 negative tests without any prior history of a negative test. SARS-CoV-2 negative tests were defined as records of a negative test without a record of a prior negative test 42 days before the index date. Concept lists used to define the COVID-19 and test-negative cohorts are available from https://github.com/oxford-pharmacoepi/LongCOVIDWP1A.

All cohorts included individuals aged ≥18 years with ≥180 days of data visibility available before the index date. Individuals with an influenza clinical diagnosis or positive test result for influenza 42 days before or on the index date were excluded. To ensure sufficient follow-up to develop long-COVID-related symptoms, we only included individuals with ≥120 days of follow-up, i.e., with an index date ≥120 days before the end of data availability. All cohorts were followed until the occurrence of the first of the event of interest, death, new COVID-19 infection, or a record of a COVID-19 clinical diagnosis, influenza infection (positive test result or clinical diagnosis), one year of follow-up, or end of data availability. In SIDIAP, cohorts were also censored on 28 March 2022, as national guidelines no longer recommended testing all suspected COVID-19 cases after that date.

Long COVID symptoms

We identified long COVID symptoms defined by the WHO clinical case definition of "post COVID-19 syndrome"[7] based on SNOMED codes in the OMOP CDM mapped respective datasets. Twenty-five symptoms were included: abdominal pain, allergy, altered smell and/or taste, anxiety, blurred vision, chest pain, cognitive dysfunction, cough, depression, dizziness, dyspnoea, fatigue or malaise, gastrointestinal issues (acid reflux, constipation, or diarrhoea), headache, intermittent fever, joint pain, memory issues, menstrual problems, muscle spasms or pain, neuralgia, pins and needles sensation, post-exertional fatigue, sleep disorder, tachycardia, and tinnitus and hearing problems. Separate code lists were developed for each symptom and reviewed independently by three clinicians (https://github.com/oxford-pharmacoepi/LongCOVIDWP1A). Quality checks were conducted to systematically identify missing codes (CohortDiagnostics R package)[25].

The WHO definition was then operationalised to identify long COVID in primary care data. Long COVID was defined as having at least one record of any of the pre-defined symptoms between 90 and 365 days after the date of COVID-19 infection and no record of that symptom 180 days before the index date. Figure S1. illustrates this definition.

For the negative-test cohorts, we anchored the algorithm at the date of the negative test to compare the proportion of people with symptoms. In sensitivity analyses, we also reported "ongoing symptomatic COVID-19"[26], defined as having at least one record of one of the symptoms ≥28 days after the index date.

Statistical analyses

We developed a common analytical code, which was run locally in OMOP CDM mapped CPRD AURUM and SIDIAP. All results are reported separately by database. We described and compared baseline characteristics (age groups [≤34, 35-49, 50-64, 65-79 and ≥80], sex, calendar time [trimester], COVID-19 vaccine status [unvaccinated and number of vaccine doses received], and co-morbidities) for people with COVID-19 infection and negative-test comparator cohorts. We compared the proportion of people with long and ongoing COVID-19 symptoms (≥90 and ≥28 days) across the five cohorts. We calculated monthly incidence rates per 100,000 person-years for COVID-19 and long COVID in the general population and among people with COVID-19.

To understand which of the pre-specified symptoms would be more differential for long COVID, we matched people with COVID-19 infections and negative controls (first negative tests, and any negative test, respectively) by 5-year age group, SARS-CoV-2 test (antigen or PCR), and index week (ratio 1:3). Rate ratios with 95% confidence intervals for each symptom are presented in forest plots. We similarly matched people with first and re-infections (ratio 1:3) and compared rate ratios for long and ongoing COVID symptoms.

Analyses were conducted in R (version 4.3.1). All analytical code is available at https://github.com/oxford-pharmacoepi/LongCOVIDWP1A.

Patient and Public Involvement

A patient and public representative was involved in planning the overarching project and helped to contextualise the study results using their patient perspective.

Figures

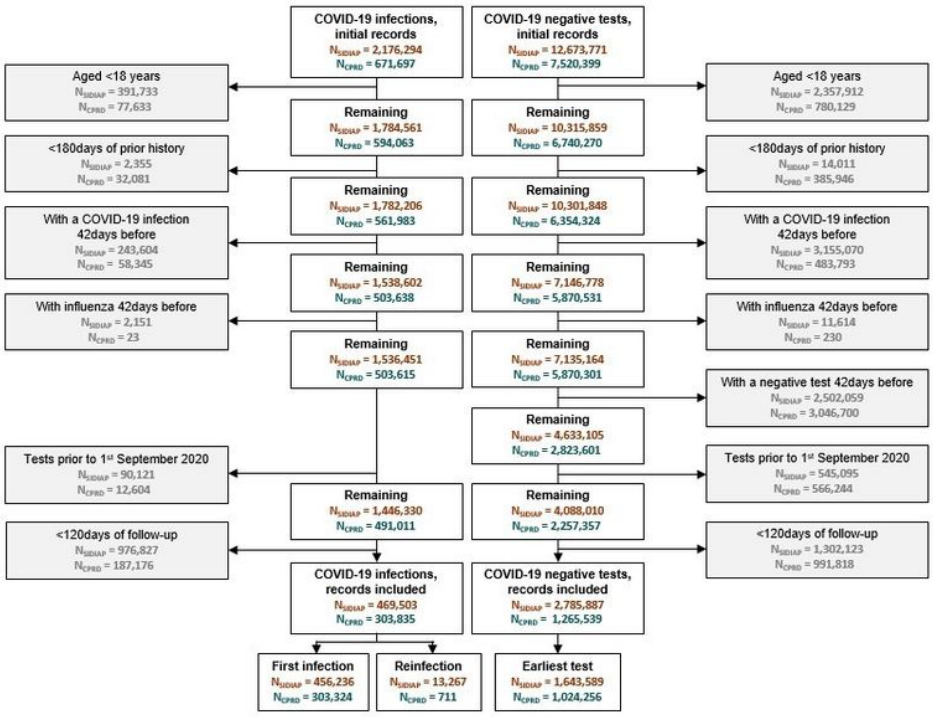

Figure 1

Selection of participants from the SIDIAP and CPRD databases

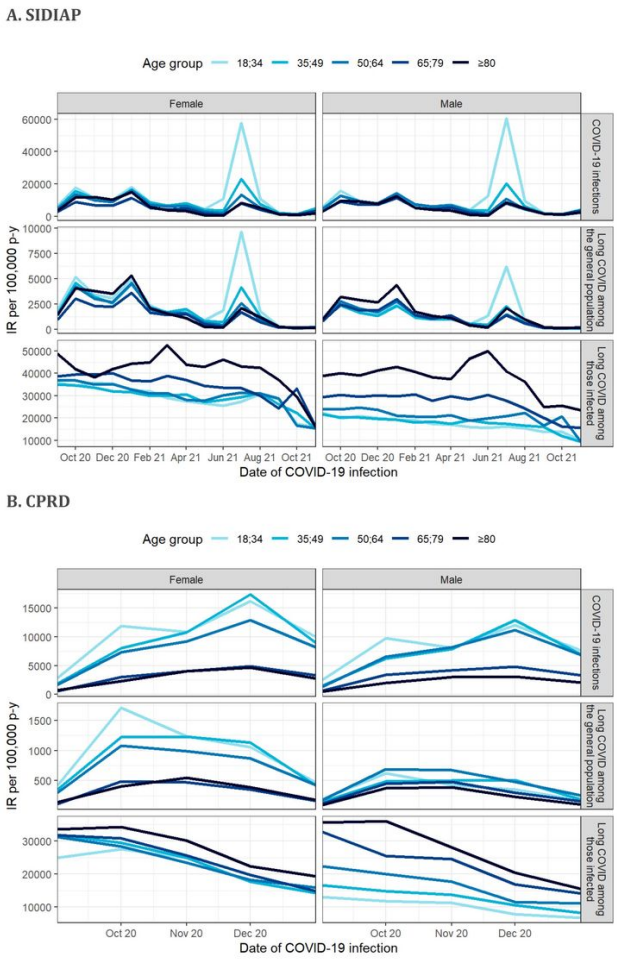

Figure 2

Incidence rates of COVID-19 and long COVID symptoms ≥90 days after infection in the general population and among infected people over time, stratified by sex and age group. IR: incidence rate, p-y: person-years.

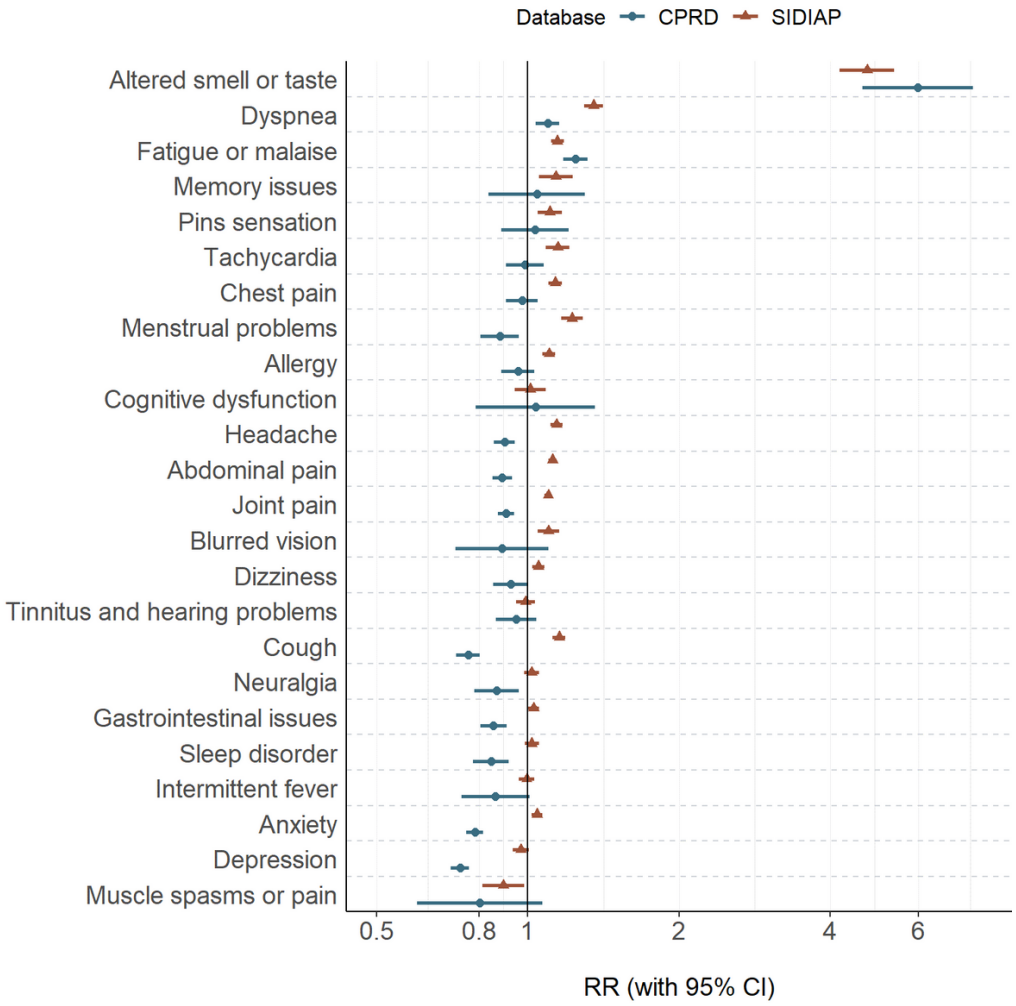

Figure 3

Rate ratios (RRs) for pre-defined long COVID symptoms, comparing 1:3 matched COVID-19 infections to first SARS-CoV-2 negative tests

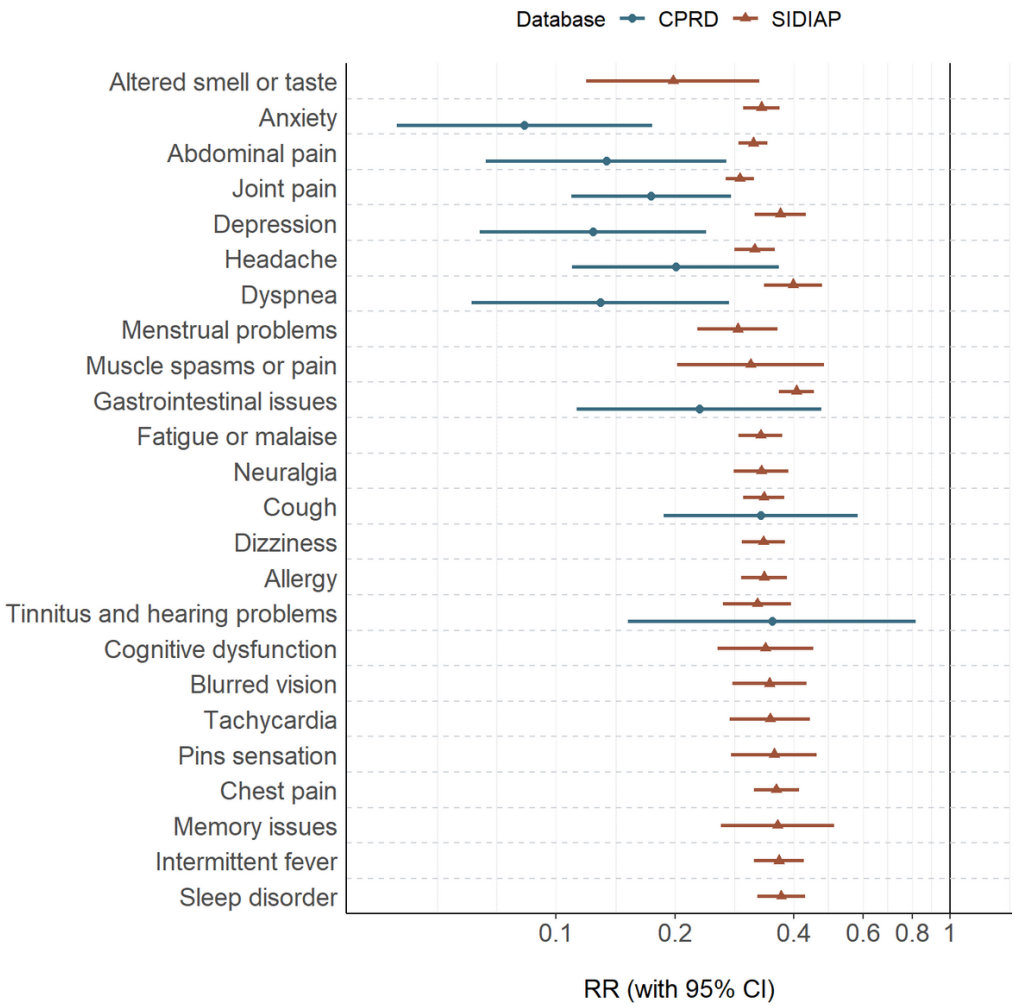

Figure 4

Rate ratios (RRs) for pre-defined long COVID symptoms comparing 1:3 matched re-infections to first infections.

Supplementary Files

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

- Supplement.docx