Prevalence of long COVID symptoms in Haryana, India: a cross-sectional follow-up study

Dhruva Chaudhry,[a,*] Shweta Khandelwal,[b] Chandrika Bahadur,[c] Benjamin Daniels,[d] Malay Bhattacharyya,[e] Raman Gangakhedkar,[f] Sapna Desai,[g,h] and Jishnu Das,[d,h] Long COVID India study group[i]

[a]Dept of Pulmonary & Critical Care Medicine, Pt BDS Post Graduate Institute of Medical Sciences (PGIMS), Rohtak, Haryana, India
[b]Public Health Foundation of India, Gurugram, Haryana, India
[c]Sustainable Development Solutions Network, Delhi and New York, USA
[d]Georgetown University, Washington, DC, USA
[e]Indian Statistical Institute, Kolkata, India
[f]Indian Council of Medical Research (ICMR), Delhi, India
[g]Population Council Institute, Delhi, India

The Lancet Regional Health - Southeast Asia 2024;25: 100395

Published Online xxx
https://doi.org/10.1016/j.lansea.2024.100395

Summary

Background Emerging research indicates growing concern over long COVID globally, although there have been limited studies that estimate population burden. We aimed to estimate the burden of long COVID in three districts of Haryana, India, using an opportunity to link a seroprevalence study to follow-up survey of symptoms associated with long COVID.

Methods We used a population-based seroprevalence survey for COVID-19 conducted in September 2021 across Haryana, India. Adults from three purposively selected districts (Rohtak, Gurugram, and Jhajjar) were eligible to participate; 2205 of 3213 consented to participate in a survey on health status. Trained investigators administered a structured questionnaire that included demographic characteristics, self-reported symptoms of illness in the last six months before the survey, mental health, and history of COVID-19.

Findings Unadjusted regression estimates indicated positive correlations between symptomatic complaints and COVID-19 exposure, suggesting lingering effects of COVID-19 in this population. The overall physical morbidity index was higher among those who tested positive for COVID-19, as was the incidence of new cases. However, both morbidity and incidence became statistically insignificant after adjustment for multiple comparisons. Cough emerged as the only statistically significant individual persistent symptom. Sex-stratified analyses indicated significant estimates only for physical morbidity in women.

Interpretation This study is one of the first from India that uses a large population-based sample to examine longer term repercussions of COVID infections. The burden of long COVID should primarily be addressed in clinical settings, where specialised treatment for individual cases continues to evolve. Our analyses also provide insight into the size and nature of studies required to assess the population-level burden of long COVID.

Funding This paper was produced under the auspices of the Lancet COVID 19 Commission India Task Force, which was supported financially by the Reliance Foundation. The Lancet COVID 19 Commission was set up in July 2020 and submitted its final report by October 2022. This report by the India Task Force was prepared during the same period.

Keywords: COVID-19; Seroprevalence; Long COVID; Haryana; India; Community survey

*Corresponding author.
E-mail address: dhruvachaudhry@yahoo.in (D. Chaudhry).
[h]Contributed equally.
[i]Usha Gupta[j]; Veena Singh[j]; Shalu Garg[j]; Liza Bhardwaj[k]; Vinod Chayal[l]; Varun Arora[l]; Pawan Kumar Singh[a]; Nizamuddin Khan[g]; Suneela Garg[m], Sheela Godbole[n]. [j]Director General Health Services (DGHS), Haryana; [k]Civil Hospital Panchkula, Haryana; [l]Department of Community Medicine, PGIMS, Rohtak; [m]Maulana Azad Medical College & Associated Hospitals, Ministry of Health & Family Welfare, Govt. of India, Delhi; [n]ICMR-National AIDS Research Institute, Pune India.

Research in context

Evidence before this study

To guide our work in India, we first undertook an (unpublished) review of literature to assess the prevalence of region-specific long COVID (LC) among adults worldwide. Our rapid assessment of available literature published between Jan 1, 2019 and July 31, 2022 was sourced from nine databases: WHO COVID-19 Global literature on coronavirus disease, PubMed, Ovid Medline, Ovid Embase, Cochrane, ScienceDirect, Scopus, Web of Science and CINAHL. Secondary research papers (including systematic reviews and meta-analyses), commentaries, individual case reports or case series, pre-prints, editorials, theoretical or discussion papers, opinion papers or non-peer-reviewed abstracts or conference publications, news articles, grey literature, and animal studies were excluded. Studies published in a language other than English were excluded from the review and we utilised the STROBE framework.

Most studies published were from Europe, the Americas and few from Asia (six studies from India). Many of these studies report evidence on LC across multiple organ groups such as chest pain, cognitive impairment, dizziness and nausea etc. Prospective, short-term studies amongst COVID-19 diagnosed cases have helped define the symptoms of LC, as well as suggest associated risk factors. The majority of the studies related to COVID-19 have varied in sample sizes, loss-to-follow-up rates, and have used a variety of comparator groups, from none at all to finely grained comparisons with patients who were diagnosed with influenza during the same time or patients who were in intensive care units prior to COVID-19.

Added value of this study

This study examined long COVID symptoms that persisted over six months in Haryana, India. Our two-staged study linked seroprevalence data to self-reported symptoms in a representative population sample, with measures to limit response bias. It expands the largely clinical evidence base by providing population-level estimates, as well as contributes findings from India, where there has been limited research on LC. Our findings indicate evidence for persistent cough associated with previous COVID infection, with higher risk amongst women. However, our study was insufficiently powered to detect associations between prior COVID exposure and other conditions. If further studies of LC prevalence in the population are required, significantly larger sample sizes will need to be deployed, and a study design that can compare multiple categories of symptoms and associated persistent morbidity in seropositive individuals will be needed.

Implications of all the available evidence

In the absence of administrative data, understanding the population-level burden of LC requires very large population surveys that capture a range of morbidity data with clear comparison groups. Population-based estimates of symptoms six months after infection with the delta variant in India suggest that public health strategies in settings with low prevalence of persistent symptoms can focus on clinical or individual treatment strategies that account for potentially higher risk amongst women. In light of wide vaccination coverage and multiple waves of COVID-19 exposure, understanding LC further calls for continued investment in targeted studies to examine its clinical manifestations and inform appropriate interventions.

Introduction

Three years into the COVID-19 pandemic, evidence from a range of settings suggests that many patients report suffering from the long-term consequences and complications of COVID-19 infections.[1,2] In the subset of patients who did not need hospitalisation, some symptoms have persisted beyond the 4-week acute COVID-19 period and all COVID-19 patients are at risk of post-acute sequelae of COVID-19 or post-acute COVID syndrome, more commonly referred to as 'long COVID (LC)'.[3,4] Epidemiological studies to examine the prevalence, risk factors and experience of LC use varying terminology to define LC [Panel 1].

The authors conducted a review of LC studies published through July 31, 2022. Many of these studies report evidence of LC affecting multiple organ groups example chest pain, cognitive impairment, dizziness, and headache. For instance, a meta-analysis of 41 studies reported that 43% of patients who reported COVID-19 suffered from some sort of post-acute sequelae (pooled prevalence 0.43 [95% CI 0.39–0.46]).[5] The same meta-analysis reports an estimated prevalence of post-COVID-19 conditions of 0.34 (95% CI 0.25–0.46) among those who were not hospitalised. Another study (pre-print) estimates that globally, in 2020 and 2021, 144.7 million (95% uncertainty interval [UI] 54.8–312.9) people suffered from one or more of the three symptom clusters (fatigue, respiratory and cognitive) of LC.[6] This study also estimated that nearly 40 million people, in India alone, experienced and reported symptoms related to LC. Factors associated with risk of LC include biological sex (females are at greater risk), minority status, socioeconomic deprivation, smoking, and a wide range of comorbidities including obesity, hypertension, and diabetes, based on self-reported symptoms amongst patients.[7] For those who were hospitalised and/or endured extended stays in intensive care unit (ICU), these symptoms were more prevalent and more severe.[8] A recent study in Bangladesh by Afroze and colleagues[9] noted that manifestations of post-acute COVID syndrome (or LC) decreased significantly over time in the hospitalised participants, while

Panel 1: Definitions of Long COVID (LC) and included symptoms.

a) NICE definition: The British National Institute for Health and Care Excellence (NICE) describes LC as the persistence of symptoms for at least 12 weeks after onset.[3]
b) WHO definition: The WHO defines LC symptoms as those that last for a slightly shorter period (2 months).[4]

Common symptoms of LC reported include fatigue, shortness of breath, cognitive dysfunction but also others and generally have an impact on everyday functioning. Less severe systemic manifestations of LC include fatigue, intermittent low-grade fever, breathlessness, cough, chest pain, palpitations, dysfunction in taste and smell, headache, gastroenterological disturbance, joint pain, myalgia, weakness, insomnia, pins and needles sensation, diarrhea, rash, hair loss, imbalance, and inability to walk, neurocognitive issues like memory and concentration problems and deteriorated quality of life. Other less frequent but potentially more severe symptoms may include increased resting heart rate, myocarditis, stroke, arrhythmia, diabetes mellitus, reduced pulmonary capacity and pulmonary fibrosis, hepatic dysfunction, renal failure, hearing loss, anxiety, depression, insomnia, and mood disorders.

some findings, such as the prevalence of dyspnea, tachycardia, depression and anxiety disorder, and post-traumatic stress disorder remained unchanged among the non-hospitalised group. Peripheral neuropathy increased significantly over time in both groups. Another study from north India highlights that hypothyroidism and vaccination post-recovery from COVID-19 have been found to be strong determinants of LC.[10]

Short-term prospective studies amongst COVID-19 diagnosed cases have helped define the symptoms of LC, as well as suggest associated risk factors. However, studies have varied in sample sizes, loss-to-follow-up rates, and studies have used a variety of comparator groups, from none at all to finely grained comparisons[11] with patients who were diagnosed with influenza during the same time or patients who were in ICU prior to COVID-19.[12]

Moreover, studies with large sample sizes have primarily come from the UK and the USA and have used electronic health records to study the trajectories of millions of patients. The lack of such administrative data in low-income and middle-income countries raises the issue of global imbalance in the research on COVID-19.[2] In India for instance, five prospective cohort studies (3 from south India and 2 from Delhi)[13–17] have estimated the prevalence of LC thus far. All the studies relied on telephone interviews for self-reported data collection of symptoms from previously hospitalised patients and two studies collected some additional in-person data on complex symptoms (e.g. heart or brain function related). None of the five studies were population based or recruited a comparison group. Sample sizes varied from 57 to 1234 adults and loss-to-follow-up (LTFU) ranged from 13% to 100%. These studies suggest that fatigue was the most reported post-acute sequalae of COVID-19, with prevalence varying from 5% to 32% across the studies. Another study, from north India among 2760 health workers, with a primary outcome of vaccine effectiveness against COVID-19 during the second wave in India, reported that ~27% of persistent health issues (follow up of more than 2 months) were related to infection with COVID-19 (post-vaccination).[10]

Towards developing a better understanding of LC symptoms in a population-based study, this study aimed to estimate the population prevalence of persistent LC symptoms in Haryana, India. We used in-person surveys as a follow-up with individuals who had participated in a seroprevalence survey in April 2022. Because the studies conducted to date in India come from relatively small samples without comparators, we view this paper as an exploratory study on measuring the population burden of long COVID and include a discussion of implications for future research.

Methods

Setting

This cross-sectional study was conducted in three districts of Haryana, India. Haryana is one of India's wealthier states and is ranked 11th in the national index of health performance by the NITI Aayog, the Government of India's public policy think-tank.[18] The state's population of ~26 million is 65% rural, with 81% literacy amongst men and 60% amongst women. The second wave of COVID-19 in Haryana reported spikes up to 15,786 daily new cases on May 4, 2021, after which cases began to decline (Fig. 1). Approximately ~80% of total positive cases were asymptomatic based on the government report [19]. In this context, seroprevalence surveys may provide an accurate estimate of the population exposed to infection with SARS-CoV-2, including asymptomatic individuals. In September 2021, the Government of Haryana conducted a population-based seroprevalence study through testing anti-spike antibodies. Vaccination coverage across Haryana's districts at that time point are shown in Supplementary Fig. S1.[19]

Survey

We selected three districts, Rohtak, Jhajjar and Gurugram using non-probabilistic sampling, to conduct a follow-up survey to estimate prevalence of symptoms associated with LC. These districts were selected for logistical convenience for the lead institute based in Rohtak. In these three districts, 2597 of 3213 (81%; 95%

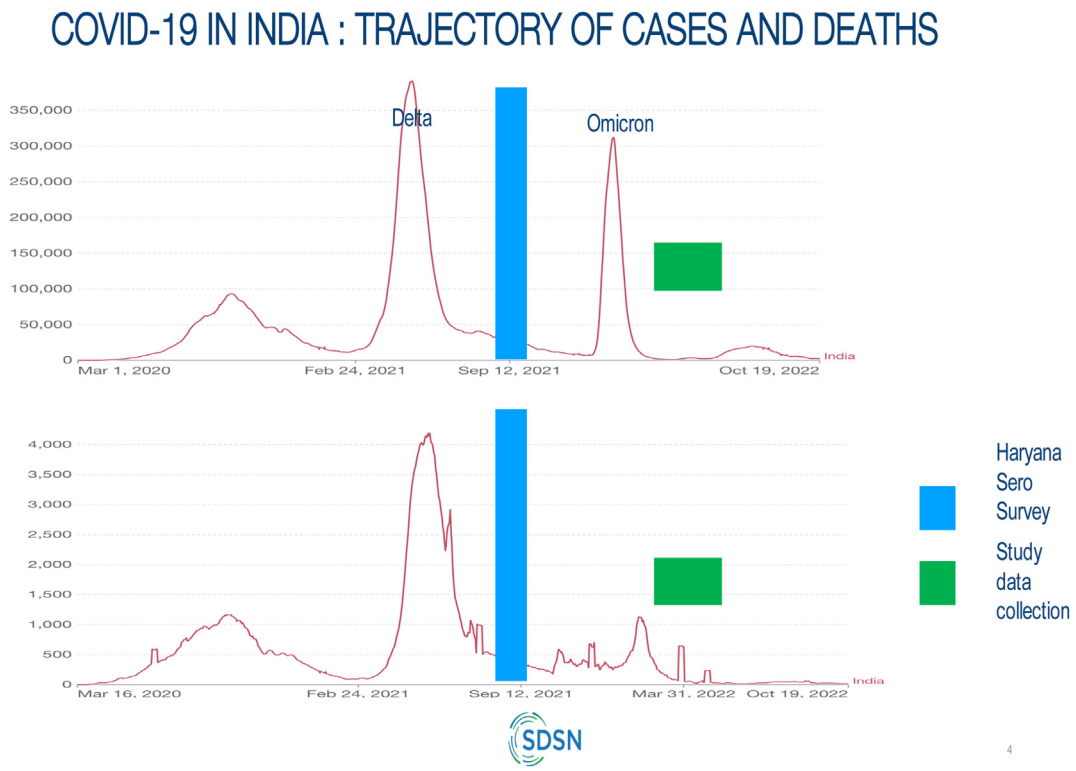

Fig. 1: COVID-19 in India: trajectory of cases and deaths. Data from Our World in Data (ourworldindata.org). Source: Our world in data (ourworldindata.org).

CI: 79–82) individuals tested seropositive in September 2021 (Rohtak: 87.3% [95% CI: 84.8–89.5]; Jhajjar: 72.8% [95% CI: 69.9–75.5]; Gurugram: 82.9% [95% CI: 80.9–84.8]). By basing our sample on those with known seropositivity, the original idea behind our survey was to use the seronegative individuals as a potential comparison group.

For the follow-up survey, we contacted all adults who participated in the seroprevalence survey, irrespective of COVID-19 status, except for pregnant and lactating women. The survey was conducted in April–May 2022. All participants provided written or thumbprint consent to participate in a survey on their health status. Trained investigators administered a structured questionnaire that included demographic characteristics, self-reported symptoms of illness in the last six month before the survey, mental health and COVID-19 history including vaccination.

One issue that all self-reported surveys face is that the recipient's reports change depending on their belief of the nature of the survey.[20] This follow-up survey was therefore de-linked from the seroprevalence survey in two ways. First, questions on COVID-19 were presented at the end of the survey, so that respondents' COVID-19 related answers could not have biased their responses to the general health questions. Second, we recruited a new team of investigators who were not linked to the seroprevalence team. Investigators and analysts were blinded to seroprevalence status until the final analysis. The study was reviewed and granted ethics approval (BREC/21/98) dated Oct 5, 2021 by the Biomedical Research Ethics Committee (BREC) Pandit Bhagwat Dayal Sharma, Post Graduate Institute of Medical Sciences (PGIMS/UHS), Rohtak institutional review board.

Exposure and outcome measures

This survey investigated the possible long-term effects of COVID-19 exposure, including positive antibodies (SARS-CoV-2 anti-spike antibodies using Chemiluminescent immunoassay assay kits [Siemens] were tested during the seroprevalence study), self-reported positive COVID-19 test history, and self-reported hospitalisation. The study checked physical morbidity, mental health and depression, and occurrence of new diseases. In the physical morbidity group, we include a set of eight symptoms: fever, anosmia (loss of smell), fatigue or body pain, digestion, ear/nose/throat (ENT) or respiratory or influenza like illness symptoms, heart, cough, and urinary symptoms (Supplementary Table S1). The mental health and depression outcomes include a standard battery of 12 questions from the General Health Questionnaire (GHQ-12)[21] regarding mental

health, such as difficulty sleeping, difficulty concentrating, feelings of worthlessness and lack of pleasure in daily activities. Each was rated on a Likert-scale from 1 to 4, where 1 is the most positive answer (for example, "I am not having sleeping difficulties at all") and 4 the most negative ("I am having much more difficulty sleeping than usual"). Patel and colleagues[22] reported an area under curve (AUC) of 0.94 for the GHQ in India for an International Classification of Diseases Version 10 (ICD-10) depressive episode and GHQ in India had the most accurate performance of the five questionnaires they examined. For new diseases, individuals were asked about new heart conditions (including coronary artery disease and/or heart failure), diabetes, hypertension, and pulmonary conditions. "Any new disease" is an indicator variable that takes the value 1 if the respondent reports a new disease and 0 otherwise. The timeframe for this variable was past 6 months.

Statistical methods

We report the prevalence of LC symptoms disaggregated by sex, and we estimate the associations between exposure to COVID-19 infection and potential LC symptoms. To do so, we used principal components to produce summary measures of morbidity and mental health, and a binary indicator was used to summarise new diseases. Multivariable logistic regression was used to adjust for age, sex, and an index of household assets (index comprises of the possession of the following: pressure cooker and colour television). As indicators of exposure, we use (a) seropositivity, (b) self-reported past COVID-19 test positivity, and (c) hospitalisation among those who tested positive. The latter two regressions were also adjusted for seropositivity status.

Due to many potential outcomes in each of the three groups (as well as the group of three summary measures), we evaluate the precision of regression estimates using two common adjustments to standard analytical *t*-statistics and p-values. First, we used the Bonferroni correction within each group of outcome variables to calculate adjusted confidence intervals for all point estimates. Second, instead of using critical values at the individual-outcome level, we assessed the statistical significance of all point estimates using the Benjamini and Hochberg procedure,[23] which bounds the false discovery rate, defined as controlling the expected proportion of false positives within each outcome family at 5%. We also report analytical (unadjusted) standard errors and p-values.

Additional analyses

One of the aims of this exploratory study is to assess the statistical suitability of our sample size for the effect sizes we observe and therefore provide guidance for further potential studies in India on LC. In order to do so, we characterise the suitability of the data set for the statistical hypotheses posed in the regression models by presenting minimum detectable effects (MDEs) using the methodology outlined in Bloom[24] and a desired power of 80%. For each regression, we also evaluated its performance against statistical benchmarks using a simulation approach that tests performance for the dataset using regression at the current sample size up to 20 times the sample size by duplicating the dataset repeatedly. In addition to measures of statistical power and detectable effects that take the data as given and assume representativeness, we provide statistical assessments of potential bias and power issues under various potential scenarios regarding loss-to-follow-up.

Role of funding source

The funders had no role in study design, data collection, data analysis, interpretation or writing of the report.

Results

Of 3213 potential participants identified, 2205 participated in the survey (69%) (Supplementary Fig. S1). The most common reason for loss of participants to follow-up was absence of the individual from the house at the time of the survey and follow-up attempts. The sampled population had a median age 42 years. Sixty-one percent of respondents were women. Sixty-three percent lived in rural areas. A higher proportion of men completed at least primary school (85%) compared to women (63%) (Table 1).

Of 2205 adults, 1816 (82%; 95% CI: 81–84) were seropositive in the original survey and was 5% lower among those who were lost-to-follow-up, reaching 781 of 1008 (77%; 95% CI: 75–80). In our sample, 136 of 2203 respondents indicated they had tested positive for COVID-19 at some point (6%; 95% CI: 5–7), of whom 14 (10%; 95% CI: 6–17) indicated they had been hospitalised for that incident. Eight of 49 (16%) of men who tested positive were hospitalised, compared to 6 of 87 (7%) of women. Almost all respondents (2166 of 2203; 98%) had received at least one COVID-19 vaccine, and 2025 of those 2166 (93%) of these received a second dose. Finally, 7 (0.2%) participants reported a second case of COVID-19, and 3 (0.1%) reported three cases (not used in further analysis).

Of 136 respondents with a positive COVID test, 1704 (82%) were also seropositive. A similar proportion of respondents who did not report testing positive for COVID were also seropositive (112/2067; 82%). Amongst 1816 seropositive respondents, 112 (6%) reported a positive COVID test; and of 387 non-seropositive respondents, 24 (6%) reported a previous positive COVID test. Altogether, 1707 (77%) were seropositive. Of these, 5% had tested positive (n = 112); 24 (1%) tested positive but were not seropositive; and 363 (16%) had neither COVID-19 indicator. The majority of reported positive tests were in March–April 2021 (SARS-CoV-2 delta variant period), although

Characteristics (mean, SD)	Total (n = 2205)	Women (n = 1341)	Men (n = 864)
Median Age [IQR]	42.0 [32.0,55.0]	42.0, [33.0, 54.0]	42.0, [32.0.57.0]
Rural residence	0.63 [0.48]	0.65 [0.48]	0.61 [0.49]
Primary Education	0.72 [0.45]	0.63 [0.48]	0.85 [0.36]
Secondary Education	0.16 [0.36]	0.13 [0.34]	0.20 [0.40]
Asset index	0.00 [1.92]	−0.02 [1.89]	0.03 [1.96]
Vaccine first dose	0.98 [0.13]; n = 2203	0.98 [0.13]; n = 1339	0.98 [0.12]; n = 864
Vaccine second dose	0.93 [0.25]; n = 2166	0.95 [0.22]; n = 1315	0.92 [0.28]; n = 851
% seropositive	0.81 [0.39]; n = 3213	0.82 [0.38]; n = 1341	0.82 [0.38]; n = 864
% reported a positive test	0.06 [0.24]; n = 2203	0.06 [0.25]; n = 1339	0.06 [0.23]; n = 864
% hospitalised	0.10 [0.31]; n = 136	0.07 [0.25]; n = 87	0.16 [0.37]; n = 49

This table reports the mean and standard deviation of key demographic characteristics, first in the full sample and then disaggregated by gender. For age, median and IQR are reported rather than mean and SD. Indentation indicates conditionality on the prior characteristic. Where N is not equal to the full survey sample, N is reported separately.

Table 1: **Demographic characteristics and COVID-19 status.**

some infections were as early as January 2020, and some as late as June 2022 (Supplementary Fig. S2).

Health and wellness status

Table 2 reports prevalence of illness symptoms, mental health concerns, and new diseases in the population.

Association between health status and COVID-19 exposure

Fig. 2 presents the main associations between our outcome measures and the three different measures of exposure to COVID-19. Supplementary Tables S2 and S3 report adjusted and unadjusted odds ratios by sex, and Supplementary Fig. S4 illustrates the unadjusted relationships between test positivity and seropositivity against the illness index.

There are three main findings. First, across all comparator groups, linear regression estimates indicate positive correlations between symptomatic complaints and COVID-19 exposure. The overall physical morbidity index is higher among those who tested positive for COVID-19 (0.3, unadjusted 95% CI: [0.04 to 0.56]), as is the incidence of new diseases (0.06, unadjusted 95% CI: [0.00 to 0.12]). However, both become statistically insignificant when adjusted for the multiple comparison using the three indices (indicated as [F1] on Fig. 2). Cough and loss of smell constitute substantial parts of the overall difference, and all components except generic respiratory or influenza-like illness conditions (ENT) were observed to move in the same direction among those exposed to COVID-19.

Second, cough is the only individual persistent symptom with evidence of a correlation after adjusting for multiple hypothesis testing (cough +9.7 p.p., adjusted 95% CI: [0.3–19.0]). All other individual symptoms (indicated as [F2] on Fig. 2) are not statistically distinguishable from null.

Third, statistical correlations emerged only for women (Supplementary Tables S2 and S3), and only for physical morbidity. No correlations emerged for men after adjusting for multiple hypothesis testing. There is no evidence that the date of testing had any effect on these estimates, although the sample of those who tested positive is very small to draw inferences. The strongest result was that the risk of persisting anosmia decreased by 0.02% per day elapsed since testing positive.

Adjustments for loss-to-follow-up

These estimates do not account for the 31% loss-to-follow-up (LTFU) between the serosurvey and the health survey. Given the high LFTU and lack of personal data collected in the serosurvey, we report estimates using seropositive status across missing observations. Linear regression assuming various incidences of cough (with seropositivity fixed from the serosurvey) across the missing observations finds an association ranging between ±30 percentage points (p.p.).

Power

We report two ancillary analyses that can inform future study design. First, we estimate a minimum detectable effect-size (MDE) close to the coefficient estimate only for the morbidity index, cough, and anosmia; and only in the comparison between people who tested positive versus those who did not. In all other cases, the MDE is significantly higher than our actual coefficient estimate, indicating that, in repeated studies of the same size and outcome variables, the estimates would have large sample-dependent variability and would not provide reliable estimates. For instance, the MDE for the new disease index is ±9.7 p.p., which means that given our sample size, a true effect that size or larger would be detected in at least 80% of replication studies. However, we estimated the effect as 5.9 p.p.

Second, required sample sizes with 80% power to detect the estimated effect, either for individual components or for multiple outcomes, are presented in Table 3.

	Total (N = 2205)	SD	Women (N = 1341)	SD	Men (N = 864)	SD
Illness index (PCA) components (Mean)	0.00	1.52	0.24	1.62	−0.38	1.27
Fever	0.34	0.47	0.38	0.49	0.28	0.45
Smell	0.08	0.27	0.10	0.31	0.05	0.21
Tiredness	0.36	0.48	0.45	0.50	0.22	0.41
Digestion	0.26	0.44	0.28	0.45	0.22	0.42
ENT	0.37	0.48	0.41	0.49	0.30	0.46
Heart	0.15	0.35	0.18	0.38	0.10	0.29
Cough	0.19	0.39	0.18	0.39	0.19	0.39
Urination	0.10	0.30	0.12	0.33	0.06	0.24
Depression index (PCA) components (% reporting 3 or 4)	0.00	2.27	0.22	2.39	−0.34	2.04
Concentrate	0.16	0.37	0.19	0.39	0.13	0.33
Sleep	0.13	0.34	0.17	0.38	0.07	0.25
Useful	0.07	0.26	0.08	0.28	0.06	0.24
Decisions	0.09	0.29	0.11	0.31	0.07	0.26
Strain	0.20	0.40	0.24	0.43	0.13	0.33
Difficult	0.11	0.32	0.13	0.34	0.08	0.27
Enjoy	0.14	0.34	0.16	0.37	0.10	0.30
Problems	0.12	0.32	0.14	0.34	0.09	0.29
Unhappy	0.17	0.37	0.21	0.41	0.10	0.30
Confidence	0.08	0.27	0.09	0.29	0.06	0.24
Worthless	0.05	0.23	0.06	0.24	0.04	0.19
Happy	0.12	0.33	0.14	0.35	0.09	0.29
New disease components (Mean)	0.14	0.34	0.17	0.37	0.09	0.29
Heart related	0.01	0.11	0.01	0.10	0.02	0.13
Diabetes	0.02	0.14	0.02	0.14	0.02	0.12
Hypertension	0.06	0.23	0.07	0.26	0.03	0.18
Pulmonary	0.01	0.07	0.01	0.08	0.00	0.06
Other new disease	0.07	0.25	0.09	0.28	0.04	0.20

This table reports summary statistics (mean, SD) for the health status elements recorded in the health survey, first in the full sample and then disaggregated by gender. The illness index and the depression index are principal-components summaries of their component elements, standardized to mean zero in the sample. The new disease index is equal to one if any new health condition was reported and zero otherwise. All index components are binary (yes/no) responses. ENT = Ear, nose, throat; also known as respiratory or influenza like illness; PCA = Principal components analysis score.

Table 2: **Self-reported health status, April–May 2022.**

Third, for the morbidity index, a power of 80% is achieved at 2 times the current sample size, as well as for cough and in comparisons between those who tested positive for COVID-19 and those who did not. All other specific symptoms required 4–20 times the study sample to be considered well-powered. The depression index achieves only 22% at 20 times the current sample size and therefore requires very large samples to confirm, assuming our estimates are accurate.

If a study considers multiple outcomes as hypotheses, required sample sizes become substantially larger (Table 3). Survey attrition at the levels observed in this study would further increase the risk of bias and sample size requirements.

Discussion

This study is one of the first, to our knowledge, to estimate the population prevalence of LC symptoms in India. The study's unique design of linkage to a prior seroprevalence survey allowed for follow-up and analysis blinded to initial seroprevalence status. While over 80% of the population had COVID-19 antibodies, 6% reported testing positive for COVID-19 at any point. This gap is likely due to high prevalence of asymptomatic infection, vaccination-induced antibodies (Supplementary Fig. S2), as well as low testing and potentially, stigma associated with reporting a positive test.

A significant feature of our sample is that there is no correlation between seropositivity and previously having tested positive for COVID-19. This suggests that using previously measured seropositivity in design or analysis of LC will be difficult in general. Specifically, it will be very challenging to time any follow-up survey such that no new infections have taken place since the serosurvey. As all the symptoms were asked before any questions on COVID-19 testing, we are confident that our estimates reflect actual differences in morbidity among those who tested positive and others.

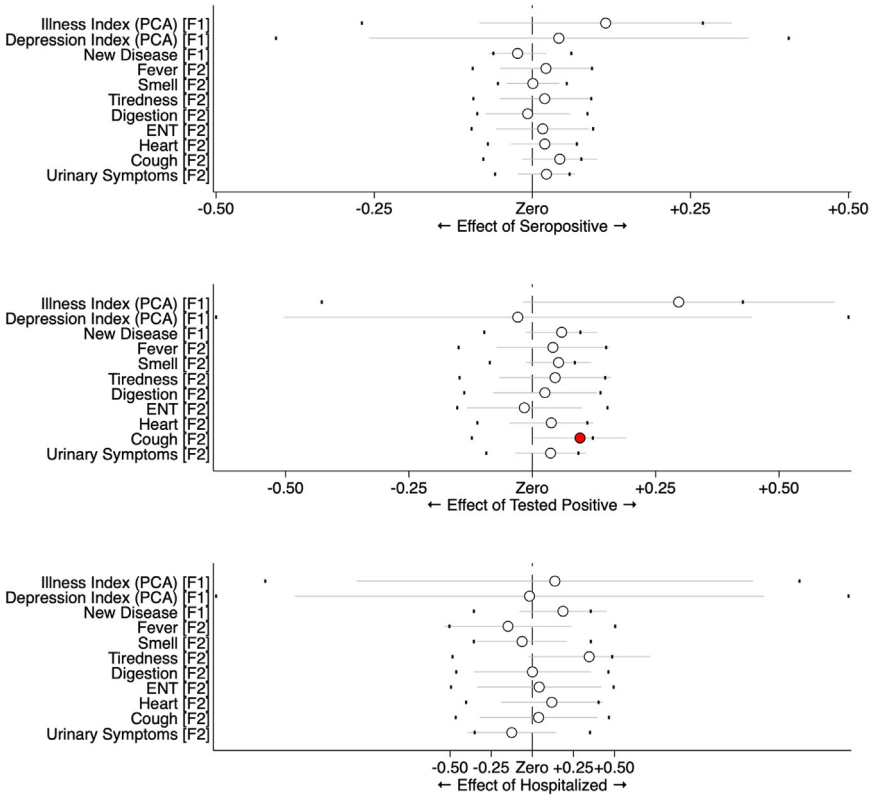

Fig. 2: COVID-19 Status associations with current health. Estimated differences across a range of outcomes in the population for three different measures of exposure to COVID-19. In the top panel, all individuals are included, and the figure illustrates estimated regression coefficients for the effect of seropositivity. In the second panel, all individuals are included, and the figure illustrates estimated regression coefficients for the effect of reporting a positive COVID-19 test. In the third panel, all test-positive individuals are included, and the figure illustrates estimated regression coefficients for hospitalisation. All regressions are controlled for age, gender, and assets, and the second and third panels are controlled for seropositivity. Confidence intervals (95% CI) are adjusted using the Bonferroni correction by family (F1, F2). Coloured markers indicate significant results after the Benjamini-Hochberg procedure with a false discovery rate of 0.05. Black markers indicate 80% minimum detectable effect sizes for each regression. ENT = Ear, nose, throat; also known as respiratory or influenza like illness.

Our analyses indicate that statistically significant reported long-term persistent effects of COVID-19 in a population amongst those who reported a positive test were limited to cough, which was reported by one in seven respondents who tested positive. Since the test-positive share is much lower than the 82% who were seropositive, we assume that this group experienced symptomatic and perhaps severe infections. This could indicate that long-term effects were more discernible among patients that had a symptomatic episode of COVID-19 infections. Moreover, we find associations were more pronounced in women, a growing area of inquiry on the sex-specific effects of COVID-19 and gendered differences in disease experience.[25]

Our study design departs from previous studies in three important ways. First, we use a population-based sample rather than a specific cohort or clinical setting; Second, the study limited reporting bias by recruiting a population sample, rather than a sample based on COVID-19 status; and third, our study covers a six-month period (September 2021–March 2022) which allows us to identify persistent LC. In general, most studies have shorter follow-up times, ranging from as low as 28 days–180 days. Our study's main contribution is provision of estimates of LC from a population-based, relatively large sample sized, two-staged study with robust quality checks with moderate to high follow-up and participation rates. Our analyses also provide critical insight into the size and nature of studies required to assess the population-level burden of LC.

Based on this study, it does not appear that symptoms associated with LC after the period of delta variant varied based on seroprevalence status. We also found limited evidence of continuing severe symptoms for those who had tested positive, although there is clear evidence of lingering coughs. These results provide cautious optimism that persistent LC has limited

Without multiple hypothesis corrections	Seropositive		Test positive		Hospitalised	
	Effect size	Required multiple (N = 2203)	Effect size	Required multiple (N = 2203)	Effect size	Required multiple (N = 136)
Illness index (PCA)	0.12	5	0.30	2	0.14	20+
Depression index (PCA)	0.04	20+	−0.03	20+	−0.02	20+
New disease (Proportion)	−0.02	6	0.06	2	0.19	3
Fever	0.02	12	0.04	8	−0.15	7
Smell	0.00	20+	0.05	2	−0.06	19
Tiredness	0.02	14	0.05	7	0.35	2
Digestion	−0.01	20+	0.03	19	0.00	20+
ENT	0.02	20+	−0.02	20+	0.04	20+
Heart	0.02	8	0.04	6	0.12	7
Cough	0.04	2	0.10	1	0.04	20+
Urination	0.02	5	0.04	4	−0.13	5
With multiple hypothesis corrections	**Seropositive**		**Test positive**		**Hospitalised**	
	Effect size	Required multiple (N = 2203)	Effect size	Required multiple (N = 2203)	Effect size	Required multiple (N = 136)
Illness index (PCA)	0.12	6	0.30	3	0.14	20+
Depression index (PCA)	0.04	20+	−0.03	20+	−0.02	20+
New disease (Proportion)	−0.02	7	0.06	3	0.19	4
Fever	0.02	19	0.04	13	−0.15	12
Smell	0.00	20+	0.05	3	−0.06	20+
Tiredness	0.02	20+	0.05	11	0.35	2
Digestion	−0.01	20+	0.03	20+	0.00	20+
ENT	0.02	20+	−0.02	20+	0.04	20+
Heart	0.02	13	0.04	9	0.12	11
Cough	0.04	4	0.10	2	0.04	20+
Urination	0.02	7	0.04	7	−0.13	8

This table reports regression results (point estimates) obtained from the regressions of health outcomes on (a) seropositivity; (b) test positivity; and (c) hospitalisation for those with positive COVID tests. All regressions are controlled for age, gender, and assets, and the second and third panels are controlled for seropositivity. The "Required Multiple" columns report the number of times larger than the current sample that would be required for the 80% minimum detectable effect of our regressions to reach the estimated effect size. The first panel does so with all hypotheses treated as independent; the second does so with multiple hypothesis testing corrections. See Fig. 2 for a graphical illustration of the effect sizes, their uncertainties, and the minimum detectable effects at the current sample size. ENT = Ear, nose, throat; also known as respiratory or influenza like illness, PCA = Principal components analysis score.

Table 3: **Effect sizes and sample size multiples required for 80% statistical power.**

prevalence and intensity over six months, in this population sample.

A more worrying conclusion, however, is that future studies will have to invest substantially higher resources to capture LC symptoms accurately at the population level. There are three aspects to this that are particular to LC studies.

First, at the effect sizes we have detected, samples for future studies will have to be up to 20 times larger to be sufficiently well powered (for instance, to capture the impact on depression). At sample sizes smaller than this, there is a very high likelihood that effects that are estimated to be significant are exaggerated or even the wrong sign. For instance, following the method proposed by Gelman and Carlin,[26] if a study is carried out with a sample size of 4400 individuals (twice the current sample size) and a significant effect is found for the depression index, the likelihood that the estimated effect has the wrong sign is 6.1%; and statistically significant results are exaggerated on average by a factor of 5.4.

Second, the LTFU in current studies from India ranges from 13% to 100%. At these rates, it is unclear that we can statistically account for these losses in the estimates through methods that use bounds. We will therefore require additional assumptions to interpret these estimates. An alternative is to use double-sampling techniques, whereby a random subsample of those who were LTFU in the first stage are followed up again with significantly higher resource deployment.

Third, it is unclear which comparator group should be used to attribute symptoms to COVID-19 infections rather than other illnesses or vaccination.[10] Studies can choose tighter comparison group,[27] for instance, by comparing individuals who were hospitalised with flu versus those hospitalised with COVID-19, but then the study becomes relevant only for a very small fraction of the population. Prospective studies that are relevant to a large proportion of the population—which are those who tested positive or were later detected to be seropositive—will always face the problem that the longer

the post-survey is conducted, the more likely that the comparison group will become contaminated due to COVID-19 exposure. This is an urgent problem as available studies on LC are primarily short-term prospective cohorts of patients who have tested positive for COVID-19. Although a few have followed patients at multiple time points after discharge or during the follow-up period, the symptoms are self-reported and not clinically endorsed and therefore subject to recall bias.[28] A very small number of studies with large sample size and more extended follow-up periods have been published from developed country settings.[29,30]

We recognise several limitations to our study design. First, this study was conducted after the period of delta variant in India. Hence, we cannot comment on the after-effects of omicron variant-related LC symptoms or after either two or three doses of vaccinations.[30] Second, our study follow-up period of six months only allowed us to track persistent symptoms, rather than the range of LC effects closer to infection. Third, like previous studies in the literature, our findings on LC prevalence are also based on self-reported symptoms by study participants.[2] No clinical examination or tests were performed, which may have contributed to high participation rates but missed medical conditions (e.g. rheumatologic conditions in LC symptomatology). Our focus on a wide range of conditions prevented in-depth investigation into multiple symptoms within a group, such as both body and joint pain, as well as mental or cognitive effects of COVID-19. Fourth, we cannot comment on clinical episodes of LC, and, unlike most previous studies, we have not checked specifically the hospitalised cases where there may be more significant after-effects.[27,31] Fifth, we could not account for vaccination-induced seropositivity in our analyses. In light of our analyses of required sample size, it is possible that our estimates of an association with cough and physical morbidity are over-estimates, which require further validation in larger population sizes.

We acknowledge that there is consistent evidence that LC can be severe and persistent at an individual level.[8] Accordingly, cohort-based studies remain important to isolate the experience of LC amongst patients that have been infected with COVID-19 and to study their experience from a clinical perspective.[32] Our findings suggest that the burden of LC after infection by delta variant in India is amenable to treatment clinical settings, where specialised treatment for individual cases continues to evolve, rather than indicating the need for population-level interventions. Our study clearly indicates that the estimation of LC prevalence requires significantly larger sample sizes. Future studies will need investment in study designs equipped to compare multiple categories of symptoms and associated persistent morbidity in seropositive individuals, with comparison groups will be increasingly harder to identify, given vaccination coverage and multiple waves of COVID-19 exposure.

Contributors

DC is the PI of the study and takes full responsibility for the study. SK oversaw and managed the study in field and wrote the first draft. CB was the Chair of the Lancet COVID-19 India Taskforce and contributed to the writing. BD carried out the analysis and wrote the results section. MB, RG provided inputs in the study design, reviewed the draft of the paper and provided inputs to improve it. SD and JD contributed equally in designing the study, data analysis and provided critical feedback to improve the manuscript draft. UG, VS, SG and LB were involved in the Haryana State seroprevalence study and facilitated the access to data from the state seroprevalence study for carrying out present study and related analysis. VC, VA and PKS managed the on-site supervision and provided inputs in study design. NK helped review the data collection tool, led the quality control aspect in data collection and analysis. SuG and ShG provided feedback on the study design and reviewed the manuscript.

Data sharing statement

Anonymised participant data and study protocol will be shared by the corresponding author after corresponding author's [DC] institutional approval and following a reasonable submitted request.

Declaration of interests

All the authors have declared no conflict of interest.

Acknowledgements

We thank the participants and their families for sparing the time to take part in this study. The Lancet COVID-19 Commission and the valuable insights from all its members are deeply acknowledged in making this study a high quality one. Efforts of all administrative staff of PGIMS Rohtak, field workers, etc are sincerely acknowledged. We also thank the Haryana government and their officials in the health facilities in the 3 districts- Rohtak, Jhajjar and Gurugram for facilitating smooth conduct of the study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lansea.2024.100395.

References

1 Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol*. 2023;21(3):133–146. https://doi.org/10.1038/s41579-022-00846-2.
2 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. https://doi.org/10.1016/j.eclinm.2022.101762.
3 NICE. *Overview COVID-19 rapid guideline: managing the long-term effects of COVID-19 guidance NICE, the scottish intercollegiate guidelines network (SIGN) and the royal college of general practitioners (RCGP)*; 2021. https://www.nice.org.uk/guidance/NG188. Accessed August 23, 2023.
4 WHO. *Post COVID-19 condition (long COVID)*. WHO; 2022. https://www.who.int/europe/news-room/fact-sheets/item/post-covid-19-condition. Accessed December 1, 2023.
5 Lopez-Leon S, Wegman-Ostrosky T, Perelman C, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci Rep*. 2021;11(1):16144. https://doi.org/10.1038/s41598-021-95565-8.
6 Hanson SW, Abbafati C, Aerts JG, et al. A global systematic analysis of the occurrence, severity, and recovery pattern of long COVID in 2020 and 2021. *medRxiv*. 2022. https://doi.org/10.1101/2022.05.26.22275532.
7 Fernández-de-Las-Peñas C, Martín-Guerrero JD, Pellicer-Valero ÓJ, et al. Female sex is a risk factor associated with long-term post-COVID related-symptoms but not with COVID-19 symptoms: the LONG-COVID-EXP-CM multicenter study. *J Clin Med*. 2022;11(2):413. https://doi.org/10.3390/jcm11020413.
8 Rees EM, Nightingale ES, Jafari Y, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. *BMC Med*. 2020;18(1):270. https://doi.org/10.1186/s12916-020-01726-3.

9 Afroze F, Arafat SM, Ahmed CM, et al. Features and risk factors of post-COVID-19 syndrome: findings from a longitudinal study in Bangladesh. *Lancet Reg Health Southeast Asia*. 2023;11:100134. https://doi.org/10.1016/j.lansea.2022.100134.

10 Kaur U, Bala S, Joshi A, et al. Persistent health issues, adverse events, and effectiveness of vaccines during the second wave of COVID-19: a cohort study from a tertiary hospital in north India. *Vaccines (Basel)*. 2022;10(7):1153. https://doi.org/10.3390/vaccines10071153.

11 Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: a 6-month retrospective cohort study of 273,618 survivors of COVID-19. *PLoS Med*. 2021;18(9):e1003773. https://doi.org/10.1371/journal.pmed.1003773.

12 Spanos M, Shachar S, Sweeney T, et al. Elevation of neural injury markers in patients with neurologic sequelae after hospitalization for SARS-CoV-2 infection. *iScience*. 2022;25(8):104833. https://doi.org/10.1016/j.isci.2022.104833.

13 Anjana NKN, Annie TT, Siba S, Meenu MS, Chintha S, Anish TSN. Manifestations and risk factors of post COVID syndrome among COVID-19 patients presented with minimal symptoms - a study from Kerala, India. *J Family Med Prim Care*. 2021;10(11):4023–4029. https://doi.org/10.4103/jfmpc.jfmpc_851_21.

14 Areekal B, Sukumaran ST, Andrews AM, et al. Persistence of symptoms after acute COVID-19 infection- an experience from a tertiary care centre in south India. *J Clin Diagn Res*. 2021;16(6):LC05–LC08. https://doi.org/10.7860/JCDR/2021/49600.14967.

15 Chopra N, Chowdhury M, Singh AK, et al. Clinical predictors of long COVID-19 and phenotypes of mild COVID-19 at a tertiary care centre in India. *Drug Discov Ther*. 2021;15(3):156–161. https://doi.org/10.5582/ddt.2021.01014.

16 Naik S, Haldar SN, Soneja M, et al. Post COVID-19 sequelae: a prospective observational study from Northern India. *Drug Discov Ther*. 2021;15(5):254–260. https://doi.org/10.5582/ddt.2021.01093.

17 P S, Madhavan S, Pandurangan V. Prevalence, pattern and functional outcome of post COVID-19 syndrome in older adults. *Cureus*. 2021;13(8):e17189. https://doi.org/10.7759/cureus.17189.

18 Bank W, Aayog N, Welfare MoHaF. *Healthy States, Progressive India: HEALTH INDEX ROUND IV 2019-20*; 2021. https://www.niti.gov.in/sites/default/files/2021-12/NITI-WB_Health_Index_Report_24-12-21.pdf. Accessed August 23, 2023.

19 Haryana IHD. *3rd round of sero-prevalance of anti-SARSCoV-2 IgG antibodies in Haryana, India: a population based cross sectional study*; 2021. http://www.nhmharyana.gov.in/WriteReadData/userfiles/file/CoronaVirus/SSR-3%20Final%20Booklet%20with%20cover%20page_compressed.pdf. Accessed August 23, 2023.

20 Zwane AP, Zinman J, Van Dusen E, et al. Being surveyed can change later behavior and related parameter estimates. *Proc Natl Acad Sci USA*. 2011;108(5):1821–1826. https://doi.org/10.1073/pnas.1000776108.

21 Goldberg DP, Gater R, Sartorius N, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol Med*. 1997;27(1):191–197. https://doi.org/10.1017/s0033291796004242.

22 Patel V, Araya R, Chowdhary N, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med*. 2008;38(2):221–228. https://doi.org/10.1017/s0033291707002334.

23 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B*. 1995;57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

24 Bloom HS. Minimum detectable effects: a simple way to report the statistical power of experimental designs. *Eval Rev*. 1995;19(5):547–556. https://doi.org/10.1177/0193841x9501900504.

25 Bai F, Tomasoni D, Falcinella C, et al. Female gender is associated with long COVID syndrome: a prospective cohort study. *Clin Microbiol Infect*. 2022;28(4):611.e9–611.e16. https://doi.org/10.1016/j.cmi.2021.11.002.

26 Gelman A, Carlin J. Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. *Perspect Psychol Sci*. 2014;9(6):641–651. https://doi.org/10.1177/1745691614551642.

27 Hastie CE, Lowe DJ, McAuley A, et al. Outcomes among confirmed cases and a matched comparison group in the Long-COVID in Scotland study. *Nat Commun*. 2022;13(1):5663. https://doi.org/10.1038/s41467-022-33415-5.

28 Yoo SM, Liu TC, Motwani Y, et al. Factors associated with post-acute sequelae of SARS-CoV-2 (PASC) after diagnosis of symptomatic COVID-19 in the inpatient and outpatient setting in a diverse cohort. *J Gen Intern Med*. 2022;37(8):1988–1995. https://doi.org/10.1007/s11606-022-07523-3.

29 Sandeep B, Mona A, Rebecca W, et al. Long term health consequences of COVID-19 in hospitalized patients from north India: a follow up study of upto 12 months. *medRxiv*. 2021. https://doi.org/10.1101/2021.06.21.21258543.

30 Sun LL, Wang J, Wang YS, et al. 15-Month health outcomes and the related risk factors of hospitalized COVID-19 patients from onset: a cohort study. *Front Med*. 2022;9:854788. https://doi.org/10.3389/fmed.2022.854788.

31 Huang L, Li X, Gu X, et al. Health outcomes in people 2 years after surviving hospitalisation with COVID-19: a longitudinal cohort study. *Lancet Respir Med*. 2022;10(9):863–876. https://doi.org/10.1016/s2213-2600(22)00126-6.

32 Nittas V, Gao M, West EA, et al. Long COVID through a public health lens: an umbrella review. *Public Health Rev*. 2022;43:1604501. https://doi.org/10.3389/phrs.2022.1604501.