

RESEARCH

Open Access



Characterisation, symptom pattern and symptom clusters from a retrospective cohort of Long COVID patients in primary care in Catalonia

Gemma Torrell¹, Diana Puente^{2,3,4*}, Constanza Jacques-Aviñó², Lucia A. Carrasco-Ribelles^{4,5,6}, Concepció Violán^{3,4,5,6,7}, Tomás López-Jiménez^{2,3}, Veronica Royano^{8,9}, Alba Molina Cantón⁹, Laura Medina-Perucha², Israel Rodríguez-Giralt¹⁰ and Anna Berenguera^{2,3,4}

Abstract

Background Around 10% of people infected by SARS-COV-2 report symptoms that persist longer than 3 months. Little has been reported about sex differences in symptoms and clustering over time of non-hospitalised patients in primary care settings.

Methods This is a descriptive study of a cohort of mainly non-hospitalized patients with a persistence of symptoms longer than 3 months from the clinical onset in co-creation with the Long Covid Catalan affected group using an online survey. Recruitment was from March 2020 to June 2021. Exclusion criteria were being admitted to an ICU, < 18 years of age and not living in Catalonia. We focused on 117 symptoms gathered in 18 groups and performed cluster analysis over the first 21 days of infection, at 22–60 days, and ≥ 3 months.

Results We analysed responses of 905 participants (80.3% women). Median time between symptom onset and the questionnaire response date was 8.7 months. General symptoms (as fatigue) were the most prevalent with no differences by sex, age, or wave although its frequency decreased over time (from 91.8 to 78.3%). Dermatological (52.1% in women, 28.5% in men), olfactory (34.9% women, 20.9% men) and neurocognitive symptoms (70.1% women, 55.8% men) showed the greatest differences by sex. Cluster analysis showed five clusters with a predominance of *Taste & smell* (24.9%) and *Multisystemic* clusters (26.5%) at baseline and *_Multisystemic* (34.59%) and *Heterogeneous* (24.0%) at ≥3 months. The *Multisystemic* cluster was more prevalent in men. The *Menstrual* cluster was the most stable over time, while most transitions occurred from the *Heterogeneous* cluster to the *Multisystemic* cluster and from *Taste & smell* to *Heterogeneous*.

Conclusions General symptoms were the most prevalent in both sexes at three-time cut-off points. Major sex differences were observed in dermatological, olfactory and neurocognitive symptoms. The increase of the *Heterogeneous* cluster might suggest an adaptation to symptoms or a non-specific evolution of the condition which can hinder its detection at medical appointments. A carefully symptom collection and patients' participation in research may generate useful knowledge about Long Covid presentation in primary care settings.

*Correspondence:

Diana Puente
dpuente@idiapjgol.org

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords SARS-CoV-2, Long COVID, Post-COVID-19 syndrome, Cluster analysis, Primary health care, Participatory research

Background

From March 2020 onwards, many people infected with SARS-CoV-2 who were never hospitalised during the acute phase of the disease presented with persisting symptoms three or more months after symptom onset. At the beginning of pandemic, little attention was paid to mild or moderate symptoms. There was only a single story about what COVID-19 was: a potentially deadly respiratory disease [1]. People with mild or moderate COVID-19 who developed persistent symptoms were invisible in the eyes of the health system and their immediate surroundings. They gathered through social media in a number of countries to raise awareness about their condition in the scientific community (who were sceptical about its existence) and began to produce knowledge about it [2] before the first scientific study was published [3]. Thus, the first studies were created based on self-reported data. This condition, referred to as Long COVID by patients [4] and renamed as Post-COVID Condition by the World Health Organisation (WHO) [5], has been estimated to affect 10–50% of people infected with SARS-CoV-2 depending on the initial clinical spectrum of infection [6–8].

Long COVID has been described as a multisystemic condition [9] with many fluctuating symptoms at different levels of intensity over time which causes different levels of episodic (or long-term) impairment on a person's ability to do normal day-to-day activities [9, 10]. Long COVID constitutes a long-term condition or evolution of COVID-19 independent of the severity of the acute disease [11]. However, the mechanisms related to the persistence of symptoms are unknown being the main hypothesis investigated: persistence of virus, chronic inflammation with blood clotting, existence of autoantibodies, microbiota dysbiosis, tissue damage and dysfunctional neurological signalling [12–18]. Other studies have found that low cortisol levels may be a biomarker for Long COVID [19]. Although there are many ongoing studies trying to find a specific biomarker for Long COVID, as yet there is no consistent evidence available.

Long COVID has been described to be more prevalent in women than in men and at about middle age [2, 9, 20, 21]. Specifically, some articles point out that Body Mass Index (BMI), female sex, increasing age and having comorbidities [22] are risk factors for Long COVID. Other studies report that the presence of five symptoms such as fatigue, headache, dyspnoea, hoarse voice and myalgia at the first week of the disease can also be risk factors for Long COVID

[20, 22, 23]. As some studies point out, however, gender differences may not only be related to differences in the prevalence and symptomatology of the condition but also to broader social and cultural factors that affect how individuals are perceived and treated by others [24].

Some studies have described symptoms, categorised them in domains, grouped them in clusters and then observed their evolution over time, suggesting the existence of different phenotypes which can help to identify the mechanisms involved and also different care needs [21, 25–27].

Long COVID symptoms can be identified through the reporting of symptoms recorded by health professionals in the EHR or by symptoms self-reported by people affected by Long COVID through public participation, as this study does [28].

Some studies have identified different trajectories of the evolution of post-COVID-19 conditions. For example, one study identified three trajectories: “high persistent symptoms,” “rapidly decreasing symptoms,” and “slowly decreasing symptoms” [29]. Another study found that COVID-19 symptoms persisted for 1 year after illness onset, even in some individuals with mild disease, and that female sex and obesity were associated with symptoms persistence [30].

There are studies that have identified the evolution of symptoms and trajectories over time. However, little is known about the study of symptom evolution since the onset of the disease. Access to this information is only possible in studies conducted since the beginning of the SARS-CoV-2 pandemic.

Thus, much has been described about the symptoms of Long COVID but there is still much to learn about the evolution of persistent COVID-19 symptoms, also known as post-COVID conditions (PCCs) or Long COVID.

This study aims to add knowledge about Long COVID symptoms and their evolution over time and to highlight the co-participatory research work between patients and primary care professionals.

Methods

Design

It consists in a retrospective cohort of adults.

Study population

This study was co-created with people belonging to the Long COVID group in Catalonia [31] that involved

participants with Long COVID symptoms in Catalonia (Spain).

Inclusion criteria were being ≥ 18 years old, living in Catalonia and having symptoms that lasted more than 3 months after suspected or confirmed (by a positive Polymerase Chain Reaction or Rapid Antigen Test) SARS-COV-2 infection and agreeing to participate and confirming their availability to answer surveys. People who had been hospitalised in an ICU (Intensive Care Unit) were excluded. The 3 month inclusion criterion was based on the available information provided by Greenhalgh et al. in August 2020 [32].

Recruiting was performed through people belonging to the Catalan Long COVID group through social media (Twitter, blog, WhatsApp group) and by snowball sampling. It was publicised through a webinar for primary care professionals (doctors, nurses, social workers) working for health providers in Catalonia to recruit more participants.

Recruiting was opened on 3rd December 2020 and closed on 30th June 2021. However, cases that were diagnosed during the first and second wave were also collected.

People were asked to report their symptoms at the first 21 days from symptom onset (baseline), at 22–60 days and at ≥ 3 months from the initial diagnosis. These cut-off points were based on available studies in 2020, about the average time for recovery from mild COVID-19 and the cut-off point used by patient led reports [33].

Data source

This paper looks at the recruitment questionnaire of this study and the variables related to sociodemographic data, clinical data and symptoms out of 40 variables included in the questionnaire that supply information about various domains (not included in this analysis) such as quality of life, use of the health system and others.

The variables were collected by a self-reported questionnaire initially performed by people affected based on their own questions about their condition and finally worked out together with a primary care doctor and a research group from the Institut Universitari d'Investigació en Atenció Primària (IDIAPJ Gol).

A group belonging to the *Col·lectiu d'Afectades i Afectats persistents per COVID-19 a Catalunya* [31] participated in the design of the study and two of them in the discussions of the results, sharing their experiences and points of view and enriching each part of the project.

Data were hosted on the REDCap (Research Electronic Data Capture) platform, allowing participants to enter their data while retaining anonymity and protection. REDCap is a secure, web-based software platform designed to collect data for research studies providing: 1)

an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages, and 4) procedures for data integration and interoperability with external sources [34, 35].

Variables

The main variable was symptoms. In total, 117 symptoms were collected; their attributes were YES/NO.

Symptoms were gathered by systems and creating a new variable for each system: dermatological, ophthalmological, urological, sexually related, menstruation related, general (including fatigue and fever), rheumatologic, neurological (including headache and insomnia), digestive, gynecological, neurocognitive, cardiac, respiratory, upper airway, ear, nose, and throat (ENT), disautonomic, olfactory and altered taste and smell based on clinical intuition. All of them were stratified by sex (women, men), age (18–34, 35–49, 50–64, ≥ 65 years) and wave. Information about what symptoms each system contains can be found in the Supplementary Data 1 (SD1).

Co-variables were date of self-reporting of the initial questionnaire, and sociodemographic data such as sex, date of birth, weight, and height. Clinical data related to date of symptom onset, type of symptoms, previous comorbidities, previous treatments and diagnostic tests were also included.

The dates of the pandemic waves were gathered from Ministerio de Sanidad data published in the reports by the Red Nacional de Vigilancia Epidemiológica (RENAVE) establishing the following periods: first wave from 13th March 2020 to 21st June 2020, second wave from 22nd June 2020 to 6th December 2020 and third wave from 7th December 2020 to 14th March 2021 [36].

Symptom perception evolution was self-reporting, and its variable was created through six graphics and definitions constructed by patients themselves and following the trends of the symptoms they had been experiencing and noting down in a diary since the beginning of these symptoms (Fig. S1).

Data analysis

An initial descriptive analysis of the included population was performed using mean (standard deviation) and median (interquartile range) for quantitative variables and percentages for categorical variables. To assess differences between sex and age, the t-test or the U Mann-Whitney test for quantitative variables and the Chi-squared test for qualitative variables were performed. A Trend test was performed to assess differences between symptoms by system at the three cut-offs (Table

S1 and Fig. S2). Stratified analysis for symptom length at < 21 days (baseline), 22–60 days and ≥ 3 months days was performed.

To identify clusters of symptoms by Long COVID system, PCAmix [37] transformation of the data was performed prior to applying fuzzy c-means to reduce dimensionality. In this reduction, symptoms by systems, age and sex of the individuals were considered, leaving a total of four dimensions after applying the Karlis-Saporta-Spinaki criterion [37, 38]. Fuzzy c-means is a soft clustering technique that relates the symptoms by system of each individual at each time point (i.e., < 21 days, 20–60 days, and ≥ 3 months) to a different cluster through membership probability [39]. Having each participant's time point assigned to a cluster made it possible to draw each individual's course in terms of patterns of system affection due to Long COVID over time. The number of clusters (from 2 to 8) and degree of fuzziness (from 1.1 to 1.8, per 0.1) were chosen through validation indices calculated 100 times in order to account for the random nature of the clustering initialisation. Once the clusters had been identified, symptoms by system at each time point were assigned to the cluster for which they had the highest membership probability. The clusters were described through the calculation of observed/expected ratios (OE ratios), which compares the prevalence of the symptom by system in each cluster with that in the study population. In addition, exclusivity was calculated as the percentage of records presented by each system divided by the total number of records with that system in the study population. A system with an OE > 2.5% or an exclusivity > 30% was considered as characteristic of the

cluster and used to name the cluster. This approach has already been used in other studies [38, 40–43]. R v 4.0.2. was used to conduct the clustering analysis.

Results

From 1258 respondents, we excluded those who had less than 3 months from the beginning of symptom onset to the enrolment date (n=298), those who were missing a symptoms variable (n=5) and those who reported an end date of symptoms of less than 3 months from the symptom's onset (N=47) (Fig. 1).

Finally, 905 respondents who had symptoms for 3 or more months from symptom onset (80.3% women, 19.0% men and 0.7% non-binary) were included. Median age was 46.0 years, 57.1% had comorbidities and 51.8% reported not taking any chronic treatment. Median Body Mass Index (BMI) was 24.2%, a third of respondents were non-smokers (32.7%) and 37.1% did physical exercise 2–3 times a week before SARS-COV-2 infection. 3.3% (30 from the total of 905) of participants (4.6% of men and 3.02% of women) reported an end date of their symptoms, which showed a median of 184 days (p25-p75 of 156.2 days to 389.2 days) since the onset of symptoms higher in men (184 days) than in women (183 days). Characteristics of the self-reported cohort are presented in Table 1 and characteristics of the “end date of symptoms cohort” are in Table S2.

A total of 117 symptoms were collected, analysed by sex and period (Table S3, S4, S5) and subsequently gathered in 18 groups of symptoms to facilitate the analysis. Analysing the symptoms individually by time period, we found that the median number of symptoms per

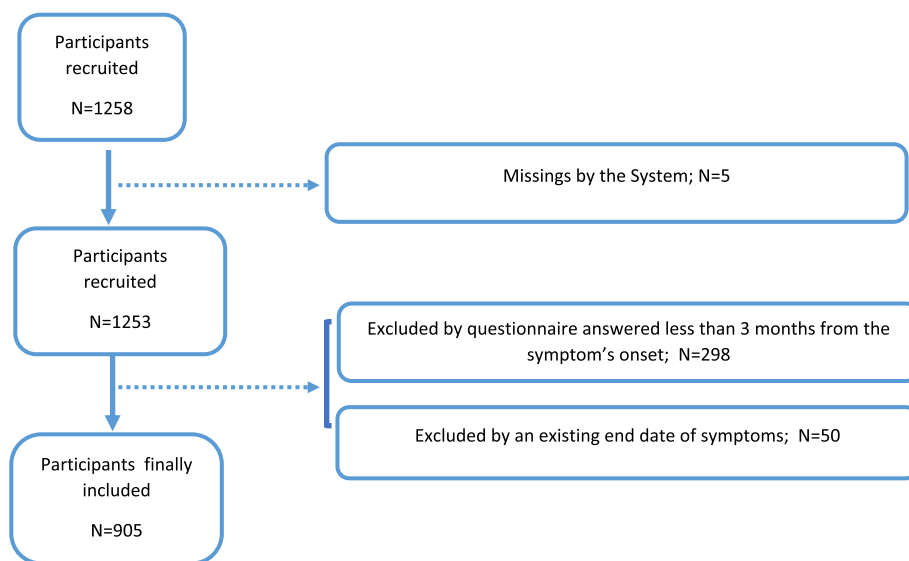


Fig. 1 Flow chart of the study population

Table 1 Socio-demographical and clinical characteristics of the self-reported cohort

| Characteristics | Total N (%) | Female N(%) | Male N(%) | p-value |
|---|------------------|------------------|------------------|--------------|
| Gender | 905 (100) | 727 (80.3) | 172 (19.0) | |
| Age (years) | | | | |
| Median (P25-P75) | 46.0 (40–54) | 46.0 (39–53) | 49.0 (42–56) | |
| Source of income | | | | |
| Contract worker | 696 (76.9) | 571 (78.5) | 120 (69.8) | 0.046 |
| Independent worker with contribution | 70 (7.7) | 46 (6.3) | 24 (14.0) | 0.003 |
| Informal worker (no contract nor contribution) | 19 (2.1) | 16 (2.2) | 3 (1.7) | 0.873 |
| Unemployment benefit or other benefits | 25 (2.8) | 18 (2.5) | 7 (4.1) | 0.475 |
| Unemployed without benefit nor social benefit | 9 (1.0) | 7 (1.0) | 2 (1.2) | 0.943 |
| Household chores or caregiver | 13 (1.4) | 12 (1.7) | 1 (0.6) | 0.546 |
| Student | 45 (5.0) | 34 (4.7) | 11 (6.4) | 0.553 |
| Other | 74 (8.2) | 58 (8.0) | 15 (8.7) | 0.711 |
| Health worker^a | | | | 0.000 |
| No | 611 (67.8) | 462 (63.9) | 145 (84.3) | |
| Yes | 290 (36.1) | 261 (36.1) | 27 (15.7) | |
| Previous physical activity^a | | | | 0.326 |
| Everyday | 152 (16.9) | 123 (17.0) | 28 (16.5) | |
| 2–3 times a week | 334 (37.1) | 268 (37.0) | 65 (38.2) | |
| < 2–3 times a week | 202 (22.4) | 163 (22.5) | 39 (22.9) | |
| No practice | 213 (23.6) | 171 (23.6) | 38 (22.4) | |
| Smoking habit^b | | | | |
| Smoker | 78 (8.6) | 68 (9.4) | 9 (5.3) | 0.180 |
| Non-smoker | 296 (32.7) | 241 (33.1) | 54 (31.6) | 0.649 |
| Ex-smoker | 228 (25.2) | 157 (21.6) | 69 (40.4) | 0.000 |
| BMI (kg/m²) | | | | |
| Median (P25-P75) | 24.2 (21.5–27.8) | 23.8 (21.2–27.3) | 25.3 (23.1–28.7) | |
| Comorbidities | | | | |
| No | 388 (42.9) | 287 (40.9) | 88 (51.2) | 0.046 |
| Yes | 517 (57.1) | 430 (59.1) | 84 (48.8) | |
| Previous treatments | | | | |
| Yes | 436 (48.2) | 363 (49.9) | 70 (40.7) | 0.093 |
| No | 469 (51.8) | 364 (50.1) | 102 (59.3) | |
| Hospitalization^a | | | | |
| Yes | 142 (15.7) | 102 (14.0) | 39 (22.7) | |
| Positive at any time by PCR or TAR^a | | | | 0.533 |
| Never positive | 437 (48.5) | 345 (47.6) | 89 (52.4) | |
| Sometime positive | 464 (51.5) | 380 (52.4) | 81 (47.6) | |
| WAVE^c | | | | 0.981 |
| First wave | 442 (60.0) | 361 (60.1) | 78 (60.0) | |
| Second wave | 253 (34.3) | 205 (34.1) | 45 (34.6) | |
| Third wave | 41 (5.6) | 34 (5.7) | 7 (5.4) | |
| Fourth wave | 1 (0.1) | 1 (0.2) | 0 (0) | |
| Vaccinated (probable)^d | | | | 0.903 |
| < 27/12/2020 | 882 (97.5) | 708 (97.4) | 168 (97.7) | |
| > 27/12/2020 | 23 (2.5) | 19 (2.6) | 4 (2.3) | |
| Median number of symptoms by period (p25-p75) | | | | |
| < 21 days | 24 (15–37) | 25 (16–38) | 20 (11–33) | |
| 22–60 days | 20 (9–32) | 21 (10–34) | 16 (7–28) | |
| ≥ 3 months | 16 (7–28) | 17 (8–29) | 12 (6–22) | |

^a 4 missings^b 1 missing^c 168 missings^d Based on the vaccination first date in Spain

participant was 24 at baseline, 20 at 22–60 days and 16 after 3 months, being higher in women at the three cut-offs than in men.

Symptoms

As shown in Fig. 2, percentages of grouped symptoms are presented at baseline, 22–60 days and ≥ 3 months showing that most of the symptoms’ system frequency decreased over time, some remained almost the same, such as dermatological, disautonomic, urological and ENT, and others such as menstrual, sexual, gynaecological, and neurocognitive increased.

General (including tiredness or fatigue, dysthermia, fever, general malaise, inappetence, weight loss, muscle pain, oral herpes) and neurologic symptoms were the most frequently reported by all respondents at all time cut-off points.

By sex, at baseline the most frequent groups of symptoms in both sexes were the general (92.8% in women, 87.2% in men), followed by the neurologic ones in women (88%) and the respiratory (79%) ones in men. The big difference observed between sexes at all cut-offs was in dermatologic symptoms followed by olfactory symptoms, both of which were more frequent in women than in men (Table S6). The evolution of symptoms by system is shown in Fig. 3.

By age, we found was that olfactory symptoms were widely reported at baseline for the 18–34 years group (68.5%), more than at any other age ways and that respondents

aged 50–64 years old reported a major frequency of respiratory symptoms (83.5%) than other ages respondents). The most frequent symptoms reported at ≥ 3 months at age 50–64 were neurological (81.5%), while the most common symptoms in other age groups were the general ones. A significant finding is that the frequency of general symptoms at the three cut offs points was lower in those aged + 64 years than in any other age range (68.1%) (Table S7).

By wave, general symptoms were the most reported for the three waves at the three-time cut-off points for baseline, 22–60 days and ≥ 3 months, while neurocognitive symptoms increased their prevalence among the first and second waves in the three-time cut-off points. Olfactory symptoms were more frequent in the second (58.9%) and third (63.4%) waves in the first 21 days from symptom onset and their prevalence decreased by more than 10% over time in all waves at ≥ 3 months (Table S8).

We analysed symptoms by microbiological diagnostic testing and found no significant differences in symptoms between participants who had a positive RAT or PCR and those who did not, except for olfactory alterations that were more common during the first 21 days in those who had a positive test (63.6%) than in those who did not (50.8%), taste and smell alterations (53.9% of those who had a positive test and 39.3% of those who hadn’t) (Table S9).

The self-reported symptom evolution of participants was included in the questionnaire. Figure 4 shows the representation of the self-perceptions of participants

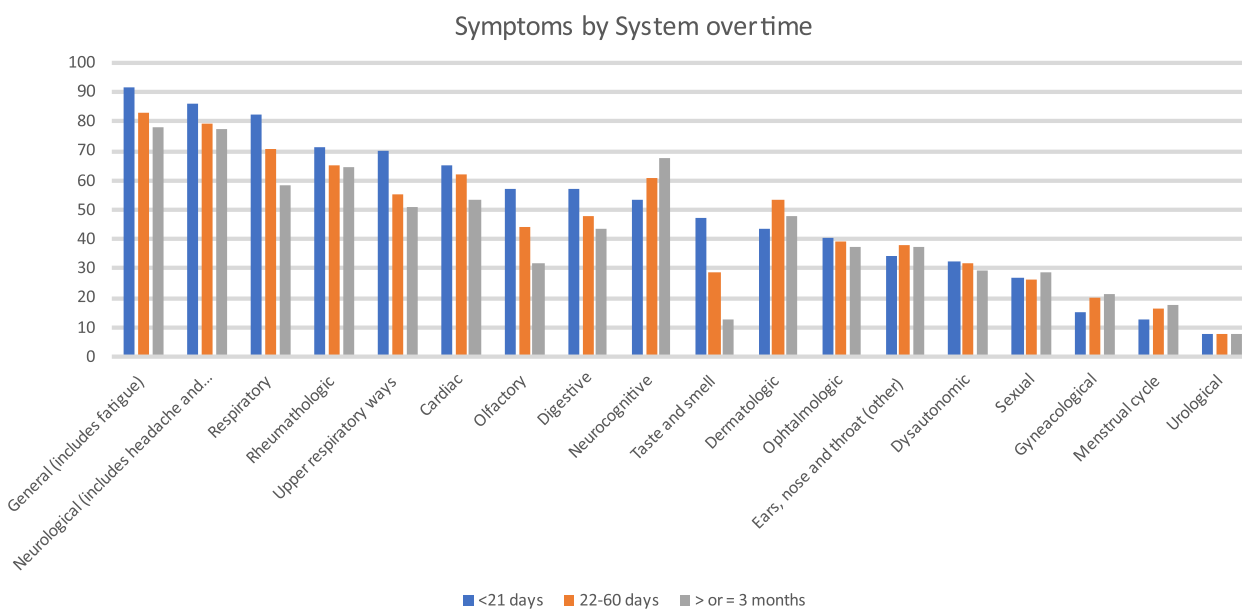


Fig. 2 Evolution of symptoms grouped by system over time

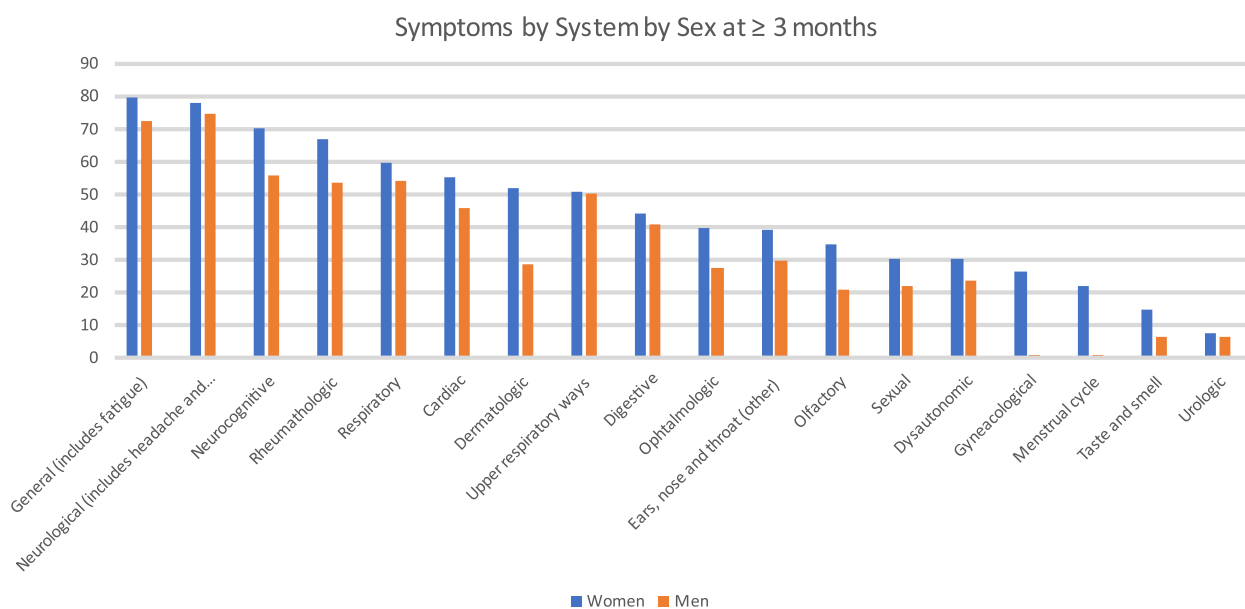


Fig. 3 Symptoms by system by sex at ≥ 3 months

on symptom evolution over time. For both sexes and at all ages, the most frequent evolution was “Symptoms were of high intensity for the first 3–4 weeks and then persist, intensifying, in a cyclical way without disappearing completely” (36.8% in women and 31% in men) (Fig. 4D). The second most frequent evolution was the one with no identified pattern (20.7% in women and 22.0% in men) by people affected at any age (Fig. 4F), except for the 50–64 age group where the second most frequent evolution was high symptom intensity followed by a progressive decrease in their intensity until disappearance (Fig. 4E).

Clusters of symptoms

Five clusters were identified and named according to the systems most predominantly affected based on the OE ratio and exclusivity of each cluster (Fig. 5): *Multisystemic*, *Multisystemic – predominantly dysautonomic*, *Heterogeneous*, *Taste & smell*, and *Menstrual & sexual alterations*.

The explained variance and the loadings of the PCAMix transformation can be found at supplementary data, Fig. S3.

Multisystemic and *Multisystemic – predominantly dysautonomic* were the most common clusters, gathering 29.8 and 21.1% of the records during the follow-up period, respectively. *Heterogeneous*, a cluster in which no single system is predominantly affected, gathered 18.5% of the records. It was followed by *Menstrual & sexual alterations* (15.6%), and *Taste & smell* (15.0%).

Taste & smell and *Multisystemic* were the most common clusters at the beginning of the condition, while *Heterogeneous* and *Multisystemic* were more common after 3 months (see Fig. 6). The prevalence of all clusters except *Taste & smell* and *Multisystemic – predominantly dysautonomic* increased over time (see Fig. 6).

Some clusters were more stable over time than others. For example, 76.1% of participants who started with *Menstrual & sexual alterations* remained in this same cluster >60 days, while only 12% of participants in *Taste & smell* stayed in it and 32 and 33.8% of them changed to *Heterogeneous* and *Multisystemic*, respectively. Participants gathered in *Multisystemic* mainly either remained in the same cluster (47.5%) or transitioned to *Heterogeneous* (29.2%). Similarly, participants with *Multisystemic – predominantly dysautonomic* affection mostly either transitioned to *Multisystemic* (33.8%) or remained in the same cluster (41.2%), while participants with a *Heterogeneous* affection either remained in it (43%) or transitioned to *Multisystemic* (35.1%) (see Fig. 7).

Discussion

This study presents the evolution of persistent COVID-19 symptoms at three-time cut-off points in a cohort of 905 people in Catalonia. The key findings are as follows: 1) The pattern of symptom evolution observed at the three cut-off points (baseline, 22–60 days and ≥ 3 months) was a decrease in the frequency of many of the symptoms (digestive, upper respiratory tract, olfactory, ophthalmologic, respiratory, cardiac, rheumatologic, general,

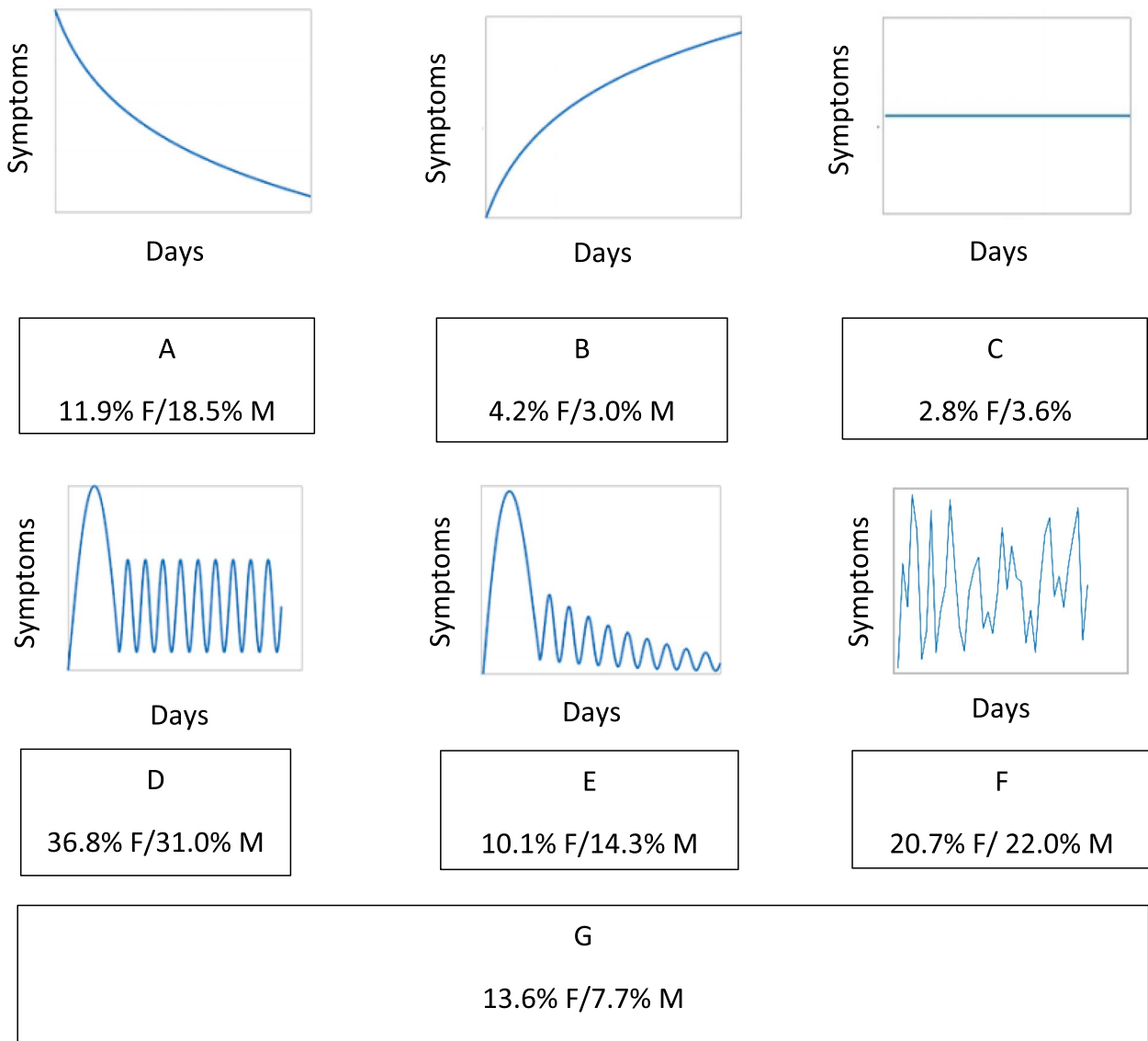


Fig. 4 Representation of the perceptions of participants on their symptom's evolution during time. F = female; M = men. Percentages refers to the frequency of each graphic in each sex. **A** Symptoms were very intense at first 3–4 weeks and progressively decrease; **B** Symptoms increase their intensity for the first 3–4 weeks and haven't decrease its intensity; **C** Symptoms have maintained same intensity from the beginning since nowadays; **D** Symptoms were of high intensity for the first 3–4 weeks and then persist, intensifying, in a cyclical way, without disappearing completely; **E** Symptoms were of high intensity for the first 3–4 weeks and after that, decreased their intensity fluctuating, until they disappear; **F** Symptoms intensity don't follow any pattern that I can identify; **G** No graphic represents my perception of my symptom's evolution over time

neurologic, disautonomic and taste and smell). 2) Neurocognitive, dermatological, ENT symptoms, gynaecological, sexual menstrual symptoms increased. 3) Urologic symptoms remained stable. 4) The most frequent clusters at baseline were *Taste & smell* and *Multisystemic*. 5) The most frequent cluster at ≥ 3 months was *Multisystemic*. We have examined the progression of COVID-19 symptoms towards long COVID-19, enabling the execution of a pertinent clinical investigation for the management of

individuals in care and providing insights into the clinical course of long COVID-19.

The data are similar to other studies reviewed. They show a predominance of women younger than men, who had more comorbidities, the most frequent being allergy, and report no previous treatments [9, 21, 44–47]. However, the women interviewees did not smoke and had an average “normal weight” BMI. These last two characteristics differ from those reported by other researchers [2, 21, 22].

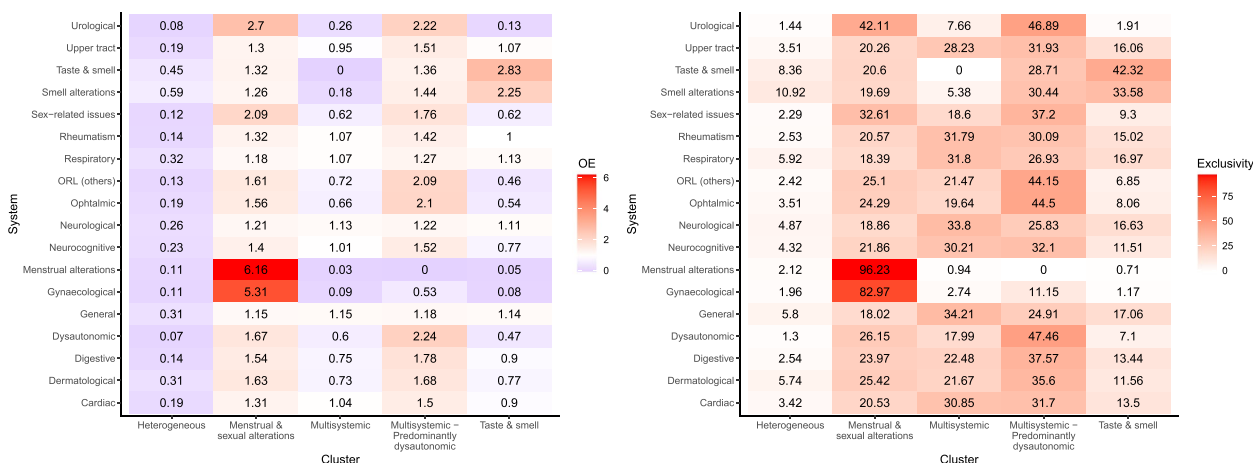


Fig. 5 Groups of symptoms included in each cluster by OE and Exclusivity

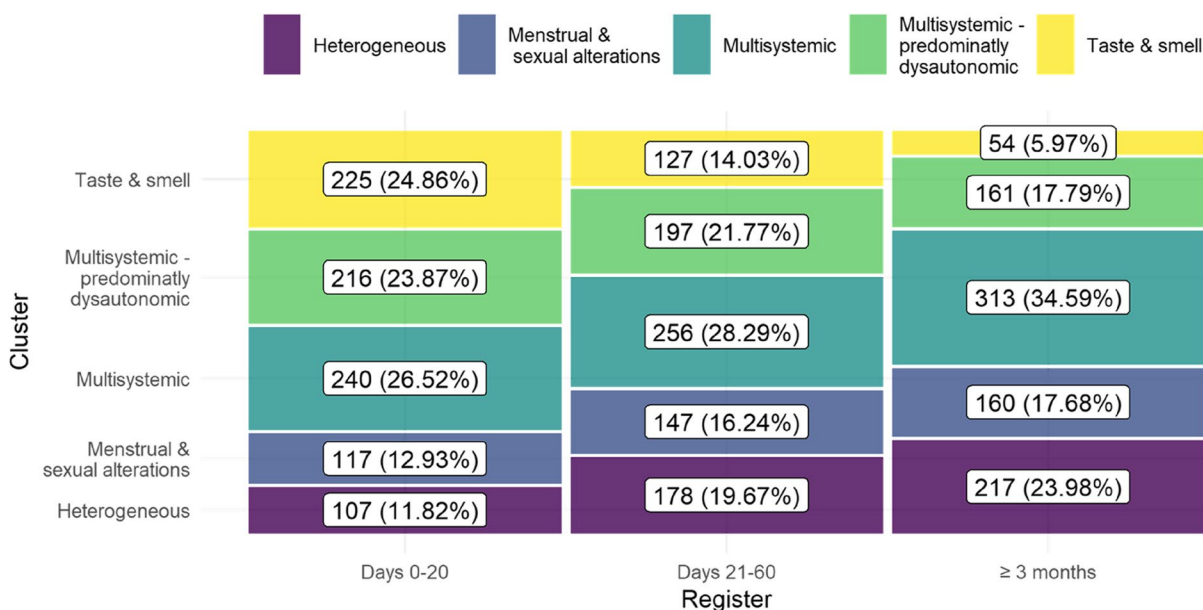


Fig. 6 Prevalence of clusters over time. The percentage reports the prevalence on each time period

Most of the women in our cohort caught the disease during the first wave (60.1%) and had a positive diagnostic test (PCR or RAT) at some point in its course (52.4%).

Beyond 3 months of symptom onset, respondents reported a mean of 16 symptoms with a higher number of symptoms in women (17 symptoms) than in men (12 symptoms). This is similar to data from other studies which reported means of 13.76 and 55.9 symptoms per patient [2, 21, 48].

Some studies suggest that greater involvement in women may be related to a different expression of angiotensin converting enzyme 2 (ACE-2) or transmembrane

protease serine 2 (TMPRSS2) receptors or to lower production of proinflammatory cytokines such as interleukin-6 (IL-6) in women after a viral infection [49]. However, the sex difference in our cohort might be due to greater involvement of women than men. It is known that women may be more able to express symptoms or allow themselves to express them more than men, whereas men are more restricted in expressing symptoms in order to conform to hegemonic masculinity patterns [50–54]. We also consider that the higher frequency of women’s participation in this study may have to do with the fact that women tend to look after their health more, as has

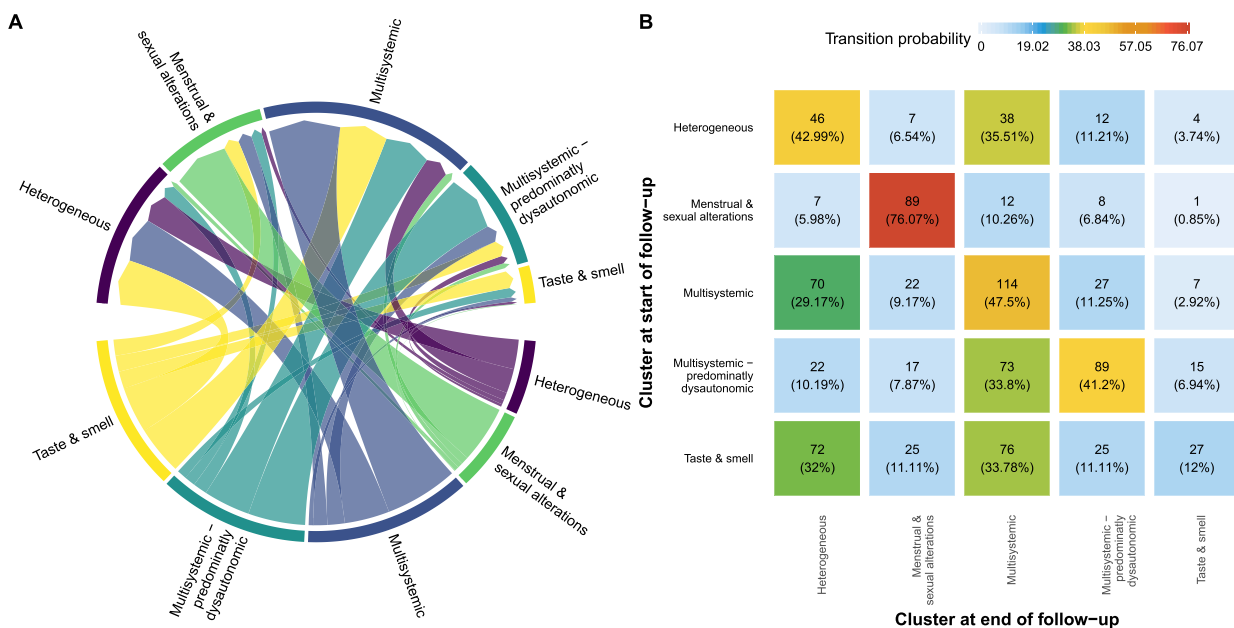


Fig. 7 Transitions and cluster evolution over time. **A** shows the transitions from the cluster at the start of the follow-up (bottom) to the cluster at the end of follow-up (top). **B** shows the transition matrix of these transitions, reporting the percentage of individuals that changed from one initial cluster (rows) to a final cluster (columns)

been described in a number of studies [55]. The higher frequency of symptoms which are more difficult to refer for consultation, such as fatigue or brain fog, may mean that they are underestimated, especially in women (gender bias) when treating women with persistent symptoms which would not be found when treating a man reporting the same symptoms.

General symptoms predominated in our cohort in both sexes in the first 21 days and in the cut off 22–60 days. Neurocognitive symptoms were more common in women. These results are similar to those reported in studies conducted in other countries [46, 56]. After 3 months, general symptoms were the most frequent symptoms in women and neurological in men, but neurological symptoms were the seconds in frequency reported by women, most likely related to continued headache. These results are close to the ones found by Ballering et al., who describes as a core Long COVID symptoms those that in our cluster analysis will correspond to Multisystemic cluster and Multisystemic-predominantly dysautonomic cluster [57]. Neurocognitive symptoms were predominant, especially in the 35–49 and in the 50–64 year ages groups, along with general and neurologic symptoms, which is consistent with the studies reviewed [2, 21, 26, 44, 56, 58]. Furthermore, differences between men and women in the frequency of dermatological symptoms are striking across all time

cut-off points in the study where they are more frequent in women. Some researchers point to a potential relationship between dermatological symptoms and systemic inflammation and between systemic inflammation and neurocognitive symptoms [59]. Olfactory symptoms were also more present in women than in men and persisted more over time in this group as reported in published meta-analyses [60, 61].

Most of our cohort was infected in the first and second waves. It is noticeable that the frequency of olfactory symptoms during the first 21 days increased in the second and third waves compared to the first. A study following a cohort of individuals who experienced COVID-19 in Norway indicates that 16.6% of those infected during the first wave still had olfactory- and taste-related symptoms 1 year later [62]. Another study [27] including anosmia and dysosmia as part of the central neurological cluster indicated that this neurological cluster was the largest cluster in both the alpha and delta variants [27].

From a clinical point of view, it is important to know which clusters may be found in the acute phase of SARS-COV-2 infection and which patterns those initial symptoms and clusters follow over a number of time cut-off points while they persist. This can enable health professionals to better suspect and identify a Long COVID condition in clinical appointments by symptoms and cluster evolution at different moments

in time. Learning about cluster trends might also help health systems to improve their delivery of care to Long COVID patients [63].

The clusters defined in our study are justified for two different reasons. Firstly, a mathematical validation to choose the clustering hyperparameters was performed: The number of clusters (from 2 to 8) and degree of fuzziness (from 1.1 to 1.8, per 0.1) was validated were chosen through by validation indices calculated 100 times in order to account for the random nature of the clustering initialisation. In addition, the most determinant conditions on each cluster were selected through the OE and the exclusivity. Secondly, the mechanisms by which long COVID-19 manifests are multiple, complex, and often overlap. The clusters obtained, such as the multisystemic one, are conditioned by various pathophysiological mechanisms, including Mast Cell Activation Syndrome), Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, and Postural Orthostatic Tachycardia Syndrome. These are justified in the different clusters observed in this paper [63]. In our data, the most prevalent clusters observed were Multisystemic and Multisystemic-predominantly disautonomic. We noted that these clusters stabilised over time with either the second becoming part of the former or the former becoming part of the Heterogeneous group. Furthermore, the transitions over time of clusters might suggest a tendency towards unspecificity or heterogeneity of symptoms that could point to an improvement in symptoms or greater adaptation of people to the symptoms after a long period of experiencing them. Kenny et al. report that the most heterogeneous of the three clusters they found is the one that includes the most people and suggest that this heterogeneity may be a sign of recovery [26]. Contrary to our results, Whitaker et al. [64] identify two stable clusters over time, one of which includes fatigue, shortness of breath and chest pain or tightness and the other with a high prevalence of smell and taste disturbances [64]. Cluster changes over time underscore Long COVID's multisystemic nature. Data analysed using cluster methodology indicate that there is no specific timeline for recovery from long COVID, as it appears to depend on individual risk factors, including psychological factors, and the severity and spectrum of symptoms experienced. Some studies indicate that the total time to complete symptom resolution reported in the literature for patients with long COVID is highly variable, with the average time to symptom resolution being 4 months in non-hospitalized patients and 9 months in those with more serious cases [29, 65, 66].

The menstrual cluster and menstrual symptoms increased across the three cut-off points probably

because over time there are more cycles to assess the disturbance. Most of the reviewed studies on persistent COVID that feature clusters do not include symptoms relating to the menstrual cycle [21, 26, 44, 61, 67–69]. Those that did consider them found changes in the volume and duration of the cycle; some saw them as part of a heterogeneous group of genitourinary symptoms, where 62.5% of respondents reported disorders, while others included them in a group of gynaecological disorders which remained stable over time [2, 70, 71]. We included menstrual symptoms in our study at the request of the group of people affected and because the rest of the research team was concerned that this information was often downplayed in the medical setting. It also speaks to the need to make menstrual health visible and relevant to women's health research as a public health issue and also as a matter of human rights [72].

Several studies examine the evolution and transitions over time of clusters, yet there are no common clusters across studies [2, 21, 26, 44, 64, 68, 73, 74]. Between-study differences are due to the varying symptom classification, the analysis techniques used, and the number of people included in each study that shape the symptom clusters identified. These differences are also a result of the time at which symptoms are identified in relation to the initial disease [2, 18, 23, 37, 51–54]. This heterogeneity hampers comparison between studies.

Thus, the evolution and transitions of long COVID-19 symptom clusters over time are complex and variable, with different trajectories and phenotypes being identified. Further research is needed to better understand the long-term implications of these symptoms and to guide monitoring and treatment strategies for individuals with long COVID-19.

Strengths and limitations

The study's strengths include the fact that it is co-created and stems from a commitment made to the people in the Long COVID-19 group in Catalonia. The analyses have been differentiated by sex, whereas few studies have stratified persistent COVID results by sex [75]. Moreover, this is a longitudinal study that involves cluster analysis. The inclusion of menstrual symptoms is not described in many publications on persistent COVID and is one of this study's strengths.

Compared with hierarchical clustering, fuzzy c-means cluster analysis is less susceptible to outliers in the data, choice of distance measure and the inclusion of inappropriate or irrelevant variables [76]. Nevertheless, some disadvantages of the method are that there may be different solutions for each set of seed points and there is

no guarantee of optimal clustering [77]. To minimise this shortcoming, we carried out 100 cluster realisations with different seed points to use the average result of all of them. In addition, although the method is not efficient when a large number of potential cluster solutions are to be considered, this was not the case of our study [76].

However, this study is not without limitations. Not least of them is the likelihood of recall bias since recruitment began in December 2020 and we also included individuals already infected in the first wave and therefore with retrospective data in this subgroup. The fact that this is a self-reported survey may be a limitation for some, although we think it values the experience of the affected person as a source of knowledge in addition to how a professional might subjectively assess an affected person's narrative.

The individuals included in the study were part of the social networks of activists, close people or contacts of contacts. We are aware that we have not been able to access all people with long COVID and that can introduce a selection bias. At the time of data collection this was a possible and feasible way. Two reasons account for this: 1) the limited number of face-to-face meetings due to outbreak restrictions 2) the limitation due to the physical conditions of the participants. Our sampling was performed by convenience and snowball sampling, with the advantages and disadvantages of this sampling strategy.

The inclusion of people with an end date of symptoms in the main analysis could lead to a bias, but two things might be of consideration. On one hand, these people had more than 3 months of symptom evolution so, they were labelled as Long COVID. On the other hand, as there is no definition for "recovery" (relapses being a common evolution of the condition), we consider it was better to include them and follow them up in the second phase of the study to see if they relapsed or not.

At the beginning of the pandemics, the lack of tests for non-hospitalised patients made it hard to confirm a SARS-COV-2 infection. Although the inclusion of people who never tested positive for SARS-COV-2 could be seen as a limitation, we see it as a matter of justice to people affected who had no access to the test.

The gender imbalance can introduce biases and limit the generalizability of the study findings, as the experience of men with Long COVID may not be accurately reflected due to the lower number of men.

We are aware that the selection of sex, age and systems as variables and no other variables such as comorbidities or disease severity provides one perspective of understanding Long COVID from multiple perspectives existing, such as the quality and relevance of the results are highly dependent on the input variables chosen by the analysis.

The respondents were probably not representative of people with persistent COVID as most of them were members of the Long COVID-19 group in Catalonia, albeit the description of the characteristics of this group is also one of our study's strengths. There may thus be a selection bias in the fact that many of the participants were recruited by the Long COVID-19 group in Catalonia and were more willing to participate in a study about their condition. So, replication of the study using different datasets and populations could be necessary to assess the generalizability of the results.

Not having a control group of non-infected participants could alter the validation of the finding.

Vaccination status and reinfection were not considered in our questionnaire. Recruitment started before the announcement of the vaccination programme (which started on 27th December 2020) in Spain. Vaccinated status and reinfection might be confounding factors when assessing the frequency of symptoms in those who reported symptom onset in 2021 [78, 79].

Conclusions

People with persistent COVID in our cohort reported general and neurological symptoms as the most frequent initial symptoms followed by respiratory symptoms in both women and men. Over time, neurocognitive symptoms displaced respiratory symptoms in women, while respiratory symptoms remained the third most frequent symptom group in men. The greatest differences between sex were found in dermatological and olfactory symptoms which were more frequent in women at all time cut-off points. In cluster analysis, evolution towards a more heterogeneous cluster over time might suggest stabilisation of the disease or adaptation to the symptoms. Heterogeneity of symptoms may render the clinical picture vague and indeterminate. This, coupled with potential gender bias, restricted access to diagnostic testing during the first wave and the change in current Spanish protocols for screening for SARS-COV-2 infection, may interfere with and hinder recognition of and care for people with persistent symptoms.

Abbreviations

| | |
|-----------------|--|
| SARS-COV-2 | Severe Acute Respiratory Syndrome Coronavirus 2 |
| COVID-19 | coronavirus disease |
| ICU | Intensive Care Unit |
| PCR | Polymerase Chain Reaction |
| RAT | Rapid Antigen Test |
| WHO | World Health Organisation |
| EHR | Emergency Health Room |
| IDIAP Jordi Gol | Fundació de Recerca en Atenció Primària de Salut Jordi Gol |
| REDCap | Research Electronic Data Capture |
| RENAVE | Red Nacional de Vigilancia Epidemiológica |
| PCAMix | Principal Component Analysis of Mixed Data |
| BMI | Body Mass Index |
| ENT | Ear, nose, throat |
| ACE-2 | Angiotensin Converting Enzyme 2 |
| TMPRSS-2 | Transmembrane protease serine 2 |
| IL-6 | Interleukin 6 |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-023-08954-x>.

Additional file 1: Figure S1. How symptoms evolution graphics were constructed.

Additional file 2: Figure S2. Graphs representing T-Trend of each group of symptoms. A= Menstrual; B= Olfactory; C= Cardiologic; D= Dermatologic; E= Digestive; F= Disautonomic; G= Sexual; H= General; I= Gynaecological; J= Neurocognitive; K= Neurologic; L= Ophthalmologic; M= Taste and Smell; N= Ear, Nose and Throat; O= Respiratory; P= Rheumatic; Q=Urologic; R=Upper Respiratory Ways.

Additional file 3: Figure S3. Visualization of the records on the first two PCAMix dimensions, coloured by cluster (A), and visualization of the squared loadings (magnitude and direction of the coefficients for the original variables) (B).

Additional file 4: Table S1. Symptoms classification by system.

Additional file 5: Table S1. Symptoms groups overtime by sex with t-Trend.

Additional file 6: Table S2. Characteristics of end-date of symptoms cohort.

Additional file 7: Table S3. Symptoms by sex at baseline.

Additional file 8: Table S4. Symptoms by sex at 22-60 days.

Additional file 9: Table S5. Symptoms by sex at ≥ 3 months.

Additional file 10: Table S6. Symptoms by system by sex overtime.

Additional file 11: Table S7. Symptoms by system by age.

Additional file 12: Table S8. Symptoms by system by wave over time.

Additional file 13: Table S9. Symptoms by system by PCR or RAT result.

Acknowledgements

We would like to thank people affected with Long COVID in Catalonia for their participation in this project and their tenacity and also all the health professionals who collaborated in recruiting patients. Special acknowledgement goes to the Health Department in Catalonia for the initial funding of this study.

Authors' contributions

GT, DP, CJA, VR, CV, AB and LMP participated in the design of the study. TL contributed to the data analysis. LC performed the cluster analysis, its interpretation and Figs. 5, 6, 7. GT performed the main analysis, wrote the draft of the main manuscript text and prepared Figs. 1, 2, 3, 4, Table 1 and the supplementary data. All the authors participated in the critical review of the manuscript and approved the final version.

Funding

Funding was obtained from the Health Department in Catalonia and the project also received a research grant from the Carlos III Institute of Health, Ministry of Economy and Competitiveness (Spain), awarded on the call for the creation of Health Outcomes-Oriented Cooperative Research Networks (RICORS), with reference RD21/0016/0029, co-funded with European Union – NextGenerationEU funds. The study's funders had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Availability of data and materials

In accordance with current European and national law, the data used in this study are only available for the researchers participating in this project. Thus, we are not allowed to distribute the data or make them publicly available to other parties. The original REDCap questionnaire will be available under request. For further information, contact the corresponding author.

Declarations

Ethics approval and consent to participate

This study follows all national and international regulations in the Declaration of Helsinki and Principles of Good Research Practice and was approved by the Clinical Research Ethics Committee of IDIAPJGol (20/165-PCV) on 1st October 2020. Anonymity and confidentiality of data were always ensured by the REDCap platform pursuant to Spain's Data Protection and Digital Rights Safeguards Act 3/2018.

The ethics committees of the Institut Universitari d'Investigació en Atenció Primària Jordi Gol i Gurina (IDIAPJGol) (code 20/165-PCV) approved the study protocol. All participants recruited in the study were fully informed about the study protocol and signed informed consent forms to participate. They consented to the use of their personal data for research and agreed to the applicable regulations, privacy policies and terms of use. Participant data was anonymised using a numerical order-based coding system and securely stored in a database.

The study's participants were directly involved in the design and analysis of the reported data. The corresponding author (DP) had full access to all data, while TL and LCRB had access to the raw data.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Centre d'Atenció Primària Les Indianes, Gerència Territorial de Barcelona, Institut Català de la Salut, Barcelona, Spain. ²Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain. ³Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. ⁴Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS) (RD21/0016/0029) Instituto de Salud Carlos III, Madrid, Spain. ⁵Unitat de Suport a la Recerca Metropolitana Nord, Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Mataró, Spain. ⁶Grup de Recerca en Impacte de les Malalties Cròniques i les seves Trajectòries (GRIMTra), (2021 SGR 01537), Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAPJGol), 08008 Barcelona, Spain. ⁷Direcció d'Atenció Primària Metropolitana Nord Institut Català de Salut, Barcelona, Spain. ⁸Universitat Politècnica de Catalunya BarcelonaTech (UPC), Barcelona, Spain. ⁹Members of the Col·lectiu d'Afectades i Afectats Persistents per COVID-19 a Catalunya, Barcelona, Spain. ¹⁰Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Spain.

Received: 11 April 2023 Accepted: 27 December 2023

Published online: 15 January 2024

References

1. Ngozi AC. El peligro de la historia única. Penguin Random House; 2018.
2. Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* [Internet]. 2021;38:101019. Available from: <https://doi.org/10.1016/j.eclinm.2021.101019>.
3. Carfi A, Bernabei R, Landi F; Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent Symptoms in Patients After Acute COVID-19. *JAMA*. 2020;324(6):603–5. <https://doi.org/10.1001/jama.2020.12603>.
4. Callard F, Perego E. How and why patients made Long Covid. *Social Science & Medicine* [Internet]. 2021;268:113426. Available from: <https://doi.org/10.1016/j.socscimed.2020.113426>.
5. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. 2022;22(4):e102–7.

6. Lund LC, Hallas J, Nielsen H, Koch A, Mogensen SH, Brun NC, et al. Post-acute effects of SARS-CoV-2 infection in individuals not requiring hospital admission: a Danish population-based cohort study. *Lancet Infect Dis*. 2021;21(10):1373–82.
7. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics [Internet]. [cited 2022 Jul 15]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/7july2022>
8. Chen C, Hauptert SR, Zimmermann L, Shi X, Fritsche LG, Mukherjee B. Global Prevalence of Post COVID-19 Condition or Long COVID: A Meta-Analysis and Systematic Review 2 Running Title: Post COVID-19 Condition Meta-Analysis 3 4. *J Infect Dis*. 2022;(jiac136):1–32. Available from: https://academic.oup.com/journals/pages/open_access/funder_policies/chorus/standard_publication_model
9. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. *Sci Rep*. 11(1):16144. <https://doi.org/10.1038/s41598-021-95565-8>.
10. Brown DA, Brien O. Conceptualising Long COVID as an episodic health condition commentary handling editor Seye Abimbola. *BMJ Glob Heal*. 2021;6:7004. Available from: <http://gh.bmj.com/>
11. Díez Antón JM, Blanco J, Bassat Q, Sarukhan A, Campins M, Guerri R, et al. Post-acute COVID syndrome (PACS): definition, impact and management a report of the multidisciplinary collaborative Group for the Scientific Monitoring of COVID-19 (GCMSC) members of the GCMSC group: Gema M Lledó (invited contributor), Jacobo Sellares (invited contributor), Carlos Brotons, Mireia Sans. 2021 Available from: <http://hdl.handle.net/2445/178471>
12. Tejerina F, Catalan P, Rodriguez-Grande C, Adan J, Rodriguez-Gonzalez C, Muñoz P, et al. Post-COVID-19 syndrome. SARS-CoV-2 RNA detection in plasma, stool, and urine in patients with persistent symptoms after COVID-19. *BMC Infect Dis*. 2022;22(1):1–8. Available from: <https://bmcfinfectedis.biomedcentral.com/articles/10.1186/s12879-022-07153-4>
13. Phetsouphanh C, Darley DR, Wilson DB, Howe A, Munier CML, Patel SK, et al. Immunological dysfunction persists for 8 months following initial mild-to-moderate SARS-CoV-2 infection. *Nat Immunol*. 2022;23(2):210–6.
14. Romero-Duarte Á, Rivera-Izquierdo M, Guerrero-Fernández de Alba I, Pérez-Contreras M, Fernández-Martínez NF, Ruiz-Montero R, et al. Seque-lae, persistent symptomatology and outcomes after COVID-19 hospitalization: the ANCOHVID multicentre 6-month follow-up study. *BMC Med*. 2021 19(1):1–13. Available from: <https://bmcmecicine.biomedcentral.com/articles/10.1186/s12916-021-02003-7>
15. Wang EY, Mao T, Klein J, Dai Y, Huck JD, Jaycox JR, et al. Diverse functional autoantibodies in patients with COVID-19. *Nature*. 2021;595:283–8. <https://doi.org/10.1038/s41586-021-03631-y>.
16. Kit Yeoh Y, Zuo T, Chung-Yan Lui G, Zhang F, Liu Q, Li AY, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*. 2021;70(4):698–706. <https://doi.org/10.1136/gutjnl-2020-323020>.
17. Tang Y, Liu J, Zhang D, Xu Z, Ji J, Wen C. Cytokine storm in COVID-19: the current evidence and treatment strategies. *Front Immunol*. 2020;11:1–13.
18. Proal AD, VanElzakker MB. Long COVID or post-acute sequelae of COVID-19 (PASC): an overview of biological factors that may contribute to persistent Symptoms. *Front Microbiol*. 2021;23:1494.
19. Klein J, Wood J, Jaycox J, Lu P, Dhodapkar RM, Gehlhausen JR, et al. Distinguishing features of Long COVID identified through immune profiling. *medRxiv*. 2022;623(7985):139–48. Available from: <https://www.medrxiv.org/content/10.1101/2022.08.09.22278592v1>
20. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med*. 27(4):626–31. <https://doi.org/10.1038/s41591-021-01292-y>.
21. Ziauddeen N, Gurdasani D, O'hara ME, Hastie C, Roderick P, Yao G, et al. Characteristics and impact of Long Covid: Findings from an online survey. *PLoS one*. 2022;17(3):e0264331. <https://doi.org/10.1371/journal.pone.0264331>.
22. Subramanian A, Nirantharakumar K, Hughes S, Myles P, Williams T, Gokhale KM, et al. Symptoms and risk factors for long COVID in non-hospitalized adults. *Nat Med*. 2022;28(8):1706–14. Available from: <https://www.nature.com/articles/s41591-022-01909-w>
23. Maglietta G, Diodati F, Puntoni M, Lazzarelli S, Marcomini B, Patrizi L, et al. Clinical medicine prognostic factors for post-COVID-19 syndrome: a systematic review and Meta-analysis. *J Clin Med*. 2022;11(6):1541.
24. Smith PB, Easterbrook MJ, Al-Selim H, Miu V, Lun C, Koc Y, et al. Sex Differences in Self-Construct and in Depressive Symptoms: Predictors of Cross-National Variation. *J Cross Cult Psychol*. 51(7-8):616–35. <https://doi.org/10.1177/0022022120939655>.
25. Estiri H, Strasser ZH, Brat GA, Semenov YR, Patel CJ, Murphy SN. Evolving phenotypes of non-hospitalized patients that indicate long COVID. *BMC Med*. 2021;19(1):1–10. Available from: <https://bmcmecicine.biomedcentral.com/articles/10.1186/s12916-021-02115-0>
26. Kenny G, Mccann K, O'brien C, Savinelli S, Tinago W, Yousif O, et al. Identification of Distinct Long COVID Clinical Phenotypes Through Cluster Analysis of Self-Reported Symptoms.
27. Canas LS, Molteni E, Deng J, Sudre CH, Murray B, Kerfoot E, et al. Profiling post-COVID syndrome across different variants of SARS-CoV-2. *medRxiv*. 2022:2022–07. Available from: <https://www.medrxiv.org/content/10.1101/2022.07.28.22278159v1>
28. Jaques-Aviñó C, Pons-Vigués M, Elsie Mcghe J, Rodríguez-Giralt I, Medina-Perucha L, Mahtani-Chugani V, et al. Participación pública en los proyectos de investigación: formas de crear conocimiento colectivo en salud. *Gac Sanit*. 2020;34(2):200–3. Available from: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112020000200018&lng=es&nrm=iso&tling=es
29. Servier C, Porcher R, Pane I, Ravaud P, Tran V-T. Trajectories of the evolution of post-COVID-19 condition, up to two years after symptoms onset. *Int J Infect Dis*. 2023;133:67–74. <https://doi.org/10.1016/j.ijid.2023.05.007>.
30. Wynberg E, Van Willigen HDG, Dijkstra M, Boyd A, Kootstra NA, Van Den Aardweg JG, et al. Evolution of coronavirus disease 2019 (COVID-19) Symptoms during the first 12 months after illness onset. *Clin Infect Dis*. 2022;75(1):E482–90.
31. Col·lectiu d'afectats i afectats persistents per la COVID-19 a Catalunya. Home [Internet] [cited 2022 Sep 19]. Available from: <https://sites.google.com/xtec.cat/afectatspersistentscovid19/home>
32. Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. *BMJ*. 2020;370 Available from: <https://www.bmj.com/content/370/bmj.m3026>
33. Assaf G, Davis H, McCorkell L, Akrami A, Wei H, Brooke O, et al. How does COVID-19 recovery actually look like? [Internet]. 2020. Available from: <https://patientresearchcovid19.com/research/report-1/>
34. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'neal L, et al. The REDCap consortium: building an International Community of Software Platform Partners Graphical Abstarct HHS public access. *J Biomed Inf*. 2019;95:103208.
35. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81. <https://doi.org/10.1016/j.jbi.2008.08.010>.
36. Ministerio de Sanidad. Contenido. Inf RENAVE n°113 [Internet]. 2022 [cited 2022 Jul 15]; Available from: https://www.mscbes.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/COVID19_Estrategia_vigilancia_y_control_e_indica
37. Chavent M, Kuentz-Simonet V, Labenne A, Saracco J. Multivariate Analysis of Mixed Data: The R Package PCAmixdata. 2014 Oct 23 [cited 2022 Sep 19]; Available from: <https://cran.r-project.org/package=PCAmixdata>
38. Karlis D, Saporta G, Spinakis A. A Simple Rule for the Selection of Principal Components. *Commun Stat-Theory Methods*. 2006;32(3):643–66. Available from: <https://www.tandfonline.com/doi/abs/10.1081/STA-120018556>
39. Bezdek JC, Ehrlich R, Full W. FCM: the fuzzy c-means clustering algorithm. *Comput Geosci*. 1984;10(2–3):191–203.
40. Tazzeo C, Rizzuto D, Calderón-Larrañaga A, Roso-Llorach A, Marengoni A, Welmer A-K, et al. Multimorbidity patterns and risk of frailty in older community-dwelling adults: a population-based cohort study. *Age Ageing*. 2021;50:2183–91. <https://doi.org/10.1093/ageing/afab138>.

41. Violán C, Fernández-Bertolín S, Guisado-Clavero M, Foguet-Boreu Q, Valderas JM, Manzano JV, et al. Five-year trajectories of multimorbidity patterns in an elderly Mediterranean population using hidden Markov models. *Sci Rep*. 10(1):168799. <https://doi.org/10.1038/s41598-020-73231-9>.
42. Guisado-Clavero M, Roso-Llorach A, López-Jimenez T, Pons-Vigués M, Foguet-Boreu Q, Muñoz MA, et al. Multimorbidity patterns in the elderly: a prospective cohort study with cluster analysis. *BMC Geriatr*. 2018;18(1):1–11. Available from: <https://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-018-0705-7>
43. Violán C, Foguet-Boreu Q, Fernández-Bertolín S, Guisado-Clavero M, Cabrera-Bean M, Formiga F, et al. Soft clustering using real-world data for the identification of multimorbidity patterns in an elderly population: cross-sectional study in a Mediterranean population. Available from: <http://bmjopen.bmj.com/>
44. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med*. 2021;27(4):626–31.
45. Blomberg B, Greve-Isdahl Mohn K, Albert Brokstad K, Zhou F, Waag Linchansen D, Hansen B-A, et al. Long COVID in a prospective cohort of home-isolated patients. *Nat Med*. 27(9):1607–13. <https://doi.org/10.1038/s41591-021-01433-3>.
46. Hastie CE, Lowe DJ, McAuley A, Winter AJ, Mills NL, Black C, et al. Outcomes among confirmed cases and a matched comparison group in the Long-COVID in Scotland study. *Nat Commun*. 2022;13(1):1–9. Available from: <https://www.nature.com/articles/s41467-022-33415-5>
47. Mateu L, Tebe C, Loste C, Santos JR, Lladós G, López C, et al. Determinants of the onset and prognosis of the post-COVID-19 condition: a 2-year prospective observational cohort study. *Lancet Reg Heal–Eur*. 2023;33:100724. Available from: www.gisaid.org
48. Rodríguez Ledo P, Armenteros del Olmo L, Rodríguez Rodríguez E, Gómez AF. Descripción de los 201 síntomas de la afectación multiorgánica producida en los pacientes afectados por la COVID-19 persistente. *Med Gen y Fam*. 2021;10(2):60–8.
49. Fernández-De-las-peñas C, Martín-Guerrero JD, Pellicer-Valero ÓJ, Navarro-Pardo E, Gómez-Mayordomo V, Cuadrado ML, 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):1–10.
50. Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, Brinton RD, Carrero JJ, DeMeo DL, et al. Sex and gender: modifiers of health, disease, and medicine. *Lancet*. 2020;396(10250):565–82. Available from: <http://www.thelancet.com/article/S0140673620315610/fulltext>
51. Barsky AJ, Peekna HM, Borus JF. Somatic symptom reporting in women and men. *J Gen Intern Med*. 2001;16(4):266–75. <https://doi.org/10.1046/j.1525-1497.2001.00229.x>.
52. Scholz U, Bierbauer W, Lüscher J. Social Stigma, Mental Health, Stress, and Health-Related Quality of Life in People with Long COVID. *Int J Environ Res Public Health*. 2023;20(5):3927.
53. Pantelic M, Ziauddeen N, Boyes M, O'Hara ME, Hastie C, Alwan NA. Long Covid stigma: estimating burden and validating scale in a UK-based sample. *medRxiv*. 2022;17(11):e0277317. Available from: <https://www.medrxiv.org/content/10.1101/2022.05.26.22275585v1>
54. Damant RW, Rourke L, Cui Y, Lam GY, Smith MP, Fuhr DP, et al. Reliability and validity of the post COVID-19 condition stigma questionnaire: a prospective cohort study. *eClinicalMedicine*. 2023;55:101755. <https://doi.org/10.1016/j.eclinm.2022.101755>.
55. Pelletier R, Choi J, Winters N, Eisenberg MJ, Bacon SL, Cox J, et al. Sex differences in clinical outcomes after premature acute coronary syndrome. *Can J Cardiol*. 2016;32(12):1447–53. Available from: <http://www.onlinecjc.ca/article/S0828282X16301441/fulltext>
56. Goërtz YMJ, Van Herck M, Delbressin JM, Vaes AW, Meys R, Machado FVC, et al. Persistent symptoms 3 months after a SARS-CoV-2 infection: the post-COVID-19 syndrome? *ERJ Open Res*. 2020;6(4):00542–2020.
57. Ballering AV, van Zon SK, Olde Hartman TC, Rosmalen JG. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. *Lancet*. 2022;400(10350):452. Available from: <https://www.pmc/articles/PMC9352274/>
58. Sigfrid L, Cevik M, Jesudason E, Lim WS, Rello J, Amuasi J, et al. What is the recovery rate and risk of long-term consequences following a diagnosis of COVID-19? - a harmonised, global longitudinal observational study protocol. *BMJ Open*. 2021;11(3):e043887.
59. Guo P, Benito Ballesteros A, Yeung SP, Liu R, Saha A, Curtis L, et al. COV-COG 1: factors predicting physical, neurological and cognitive Symptoms in Long COVID in a community sample. A first publication from the COVID and cognition study. *Front Aging Neurosci*. 2022;14:1–24.
60. Kye Jyn Tan B, Han R, Zhao JJ, Kye Wen Tan N, Sen Hui Quah E, Jing-Wen Tan C, et al. Prognosis and persistence of smell and taste dysfunction in patients with covid-19: meta-analysis with parametric cure modelling of recovery curves. *Bmj*. 378 Available from: <http://www.bmj.com/>
61. Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. *BMJ Glob Heal*. 2021;6(9):e005427. Available from: <https://gh.bmj.com/content/6/9/e005427>
62. Caspersen IH, Magnus P, Trogstad L. Excess risk and clusters of symptoms after COVID-19 in a large Norwegian cohort. *Eur J Epidemiol*. 2022;37(5):539–48.
63. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol*. 2023;21:133–46.
64. Whitaker M, Elliott J, Chadeau-Hyam M, Riley S, Darzi A, Cooke G, et al. Persistent COVID-19 symptoms in a community study of 606,434 people in England. *Nat Commun*. 2022;13(1):1957.
65. Perumal R, Shunmugam L, Naidoo K, Abdool Karim SS, Wilkins D, Garzino-Demo A, et al. Long COVID: a review and proposed visualization of the complexity of long COVID. *Front Immunol*. 2023;14:1–18.
66. Gottlieb M, Spatz ES, Yu H, Wisk LE, Elmore JG, Gentile NL, et al. Long COVID Clinical Phenotypes up to 6 Months After Infection Identified by Latent Class Analysis of Self-Reported Symptoms. In: *Open Forum Infectious Diseases*. Oxford University Press. p. ofad277. <https://doi.org/10.1093/ofid/ofad277>.
67. Whitaker M, Elliott J, Chadeau-Hyam M, Riley S, Darzi A, Cooke G, et al. Persistent COVID-19 symptoms in a community study of 606,434 people in England. *Nat Commun*. 13(1):1957. <https://doi.org/10.1038/s41467-022-29521-z>.
68. Caspersen IH, Trogstad L. Excess risk and clusters of symptoms after COVID-19 in a large Norwegian cohort. *Eur J Epidemiol*. 2022;37(5):539–48. <https://doi.org/10.1101/2021.10.15.21265038>.
69. Huang Y, Pinto MD, Borelli JL, Mehrabadi MA, Abrihim H, Dutt N, et al. COVID Symptoms, Symptom Clusters, and Predictors for Becoming a Long-Hauler: Looking for Clarity in the Haze of the Pandemic. 2021;31(8):1390–8. <https://doi.org/10.1101/2021.03.03.21252086>.
70. MAA A-N, Al-Alwany RR, Al-Rshoud FM, Abu-Farha RK, Zawiah M. Menstrual changes following COVID-19 infection: A cross-sectional study from Jordan and Iraq. *PLoS One*. 2022;17(6):e0270537. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270537>
71. Tran V-T, Porcher R, Pane I, Ravaud P. Course of post COVID-19 disease symptoms over time in the ComPaRe long COVID prospective e-cohort. *Nat Commun*. 13(1):1812. <https://doi.org/10.1038/s41467-022-29513-z>.
72. Babbar K, Martin J, Ruiz J, Parry AA, Sommer M. Menstrual health is a public health and human rights issue. *Lancet Public Heal*. 2022;7(1):e10–1. Available from: <http://www.thelancet.com/article/S2468266721002127/fulltext>
73. Yelin D, Margalit I, Nehme M, Bordas-Martínez J, Pistelli F, Yahav D, et al. Clinical medicine patterns of Long COVID Symptoms: a multi-center cross sectional study. *J Clin Med*. 2022;11:898. <https://doi.org/10.3390/jcm11040898>.
74. Huang Y, Pinto MD, Borelli JL, Mehrabadi MA, Abrihim H, Dutt N, et al. COVID Symptoms, Symptom Clusters, and Predictors for Becoming a Long-Hauler: Looking for Clarity in the Haze of the Pandemic. *medRxiv Prepr Serv Heal Sci [Internet]*. 2021 Mar 5 [cited 2022 Jul 18]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33688670>.
75. Sylvester SV, Rusu R, Chan B, Bellows M, O'keefe C, Nicholson S. Sex differences in sequelae from COVID-19 infection and in long COVID syndrome: a review. *Curr Med Res Opin*. 2022;38(8):1391–9. Available from: <https://www.tandfonline.com/action/journalInformation?journalCode=icmo20>
76. Badsha MB, Mollah MNH, Jahan N, Kurata H. Robust complementary hierarchical clustering for gene expression data analysis by β -divergence. *J Biosci Bioeng*. 2013;116(3):397–407.
77. Fuzzy Clustering [Internet]. [cited 2023 Feb 6]. Available from: <https://reference.wolfram.com/legacy/applications/fuzzylogic/Manual/12.html>
78. Nehme M, Vetter P, Chappuis F, Kaiser L, Guessous I, Team for the CS, et al. Prevalence of Post-COVID Condition 12 Weeks After Omicron Infection Compared With Negative Controls and Association With Vaccination

Status. *Clin Infect Dis*. 2022;76(9):1567–75. Available from: <https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciac947/6905455>

79. Azzolini E, Levi R, Sarti R, Pozzi C, Mollura M, Mantovani A, et al. Association Between BNT162b2 Vaccination and Long COVID After Infections Not Requiring Hospitalization in Health Care Workers. *JAMA*. 2022;328(7):676. Available from: [/pmc/articles/PMC9250078/](https://pubmed.ncbi.nlm.nih.gov/39250078/)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

