 doi:10.20944/preprints202211.0033.v1

*Brief Report*

# Long COVID Classification: Findings from a Clustering Analysis in the Predi-COVID Cohort Study

**Aurélie Fischer[1*], Nolwenn Badier[1], Lu Zhang[2], Abir Elbéji[1], Paul Wilmes[3,4], Pauline Oustric[5], Charles Benoy[6,9], Markus Ollert[7,8], Guy Fagherazzi[1]**

[1]Deep Digital Phenotyping Research Unit. Department of Population Health. Luxembourg Institute of Health, 1 A-B rue Thomas Edison, L-1445 Strassen, LUXEMBOURG
[2]Bioinformatics Platform. Luxembourg Institute of Health, 1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg
[3]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, 7, avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg
[4]Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, 6, Avenue du Swing, L-4367 Belvaux, Luxembourg
[5]Association Après J20 COVID Long France, Lucé, France
[6]Centre Hospitalier Neuro-Psychiatrique, Ettelbruck, Luxembourg
[7]Department of Infection and Immunity. Luxembourg Institute of Health, 29, Rue Henri Koch, L-4354 Esch-sur-Alzette, LUXEMBOURG
[8]Department of Dermatology and Allergy Center, Odense Research Center for Anaphylaxis (ORCA), University of Southern Denmark, Odense, 5000 C, Denmark
[9]Psychiatric Hospital of the University of Basel, Switzerland
* Correspondence: aurelie.fischer@lih.lu

**Abstract:** The increasing number of people living with Long COVID requires the development of more personalized care, as for now limited treatment options and rehabilitation programs adapted to the variety of Long COVID presentations are available. Our objective was to design an easy-to-use Long COVID classification to help stratifying people with Long COVID. Individual characteristics and a detailed set of 62 self-reported persisting symptoms together with quality of life indexes 12 months after initial COVID-19 infection were collected in a cohort of SARS-CoV-2 infected people in Luxembourg. A hierarchical ascendant classification (HAC) was used to identify clusters of people. We identified 3 patterns of Long COVID symptoms with a gradient in disease severity. Cluster-Mild encompassed almost 50% of the study population and was composed of participants with less severe initial infection, fewer comorbidities, and fewer persisting symptoms (mean=2.9). Cluster-Moderate was characterized by a mean of 11 persisting symptoms and a poor sleep and respiratory quality of life. Cluster-Severe was characterized by a higher proportion of women and smokers as in the other clusters, with a higher number of Long COVID symptoms, in particular of vascular, urinary, and skin symptoms. Our study evidenced that Long COVID can be stratified in 3 sub-categories in terms of severity. If replicated in other populations, this simple classification will help clinicians to personalize the care of people with Long COVID.

**Keywords:** Clustering; COVID-19; Long COVID; disease severity

## 1. Introduction

It is now estimated that a mean of 10 to 20% of the people infected by the SARS-CoV-2 experience persisting and fluctuating symptoms more than 12 weeks after the acute infection[1][2]. This syndrome has been called “Long COVID” by patients themselves and has a high impact on the quality of life of the affected people and as a consequence on the whole healthcare system.

 doi:10.20944/preprints202211.0033.v1

Long COVID has been defined by WHO as a condition that occurs 3 months after infection with SARS CoV-2, with symptoms that last at least 2 months and cannot be explained by any other diagnosis[3], but this definition does not account for the substantial intra-group variability in the different presentations of Long COVID.

Many studies described Long COVID in post-hospitalization cohorts[4][5][6] and in population-based studies of less severe forms of COVID-19 with similar results[7][8], the most common reported symptoms being fatigue, shortness of breath, cognitive dysfunction, with usually a major impact on daily life [9][7][8]. Long COVID affects many organs with pulmonary, cardiac, thromboembolic, neurologic, and renal sequelae. However, their distribution and intensity in the general population are largely heterogeneous [10].

A one-size-fits-all care strategy for people with Long COVID is therefore not possible and a better understanding of sub-forms of Long COVID would allow to develop personalized care for people with Long COVID or be integrated as a screening tool for future clinical trials[11]. To date, few studies used clustering analysis to identify and characterize different Long COVID phenotypes[8,12,13].

In this study, we hypothesized that Long COVID can be stratified in different clinically relevant sub-groups. To test this hypothesis, we applied hierarchical clustering to study participants with Long COVID from the Predi-COVID cohort study.

## 2. Materials and Methods

### *STUDY POPULATION*

We used data from the Predi-COVID study, a prospective cohort study of persons with a PCR-confirmed diagnosis of COVID-19 in Luxembourg. The study design and objectives have been published previously[14]. Participants were followed-up at 12 months with a self-reported questionnaire to update their general health status, persisting symptoms, and quality of life. The Predi-COVID study was approved by the National Research Ethics Committee of Luxembourg (study number 202003/07) and by the Luxembourg Ministry of Health as the authorizing body in April 2020.

Individual characteristics, comorbidities, and initial symptoms were collected at inclusion in the Predi-COVID study. Initial COVID-19 disease severity ("Asymptomatic," "Mild illness," and "Moderate/severe illness") has been previously assessed as described elsewhere[15][16].

Persisting symptoms were collected using a list of 62 symptoms[11], further divided into 8 categories: ear/nose/throat symptoms, neurological and ocular symptoms, general symptoms, cardio-respiratory symptoms or diseases, gastrointestinal symptoms, vascular and ganglionic symptoms or diseases, urinary symptoms, and skin symptoms (see Online Supplementary Table 1 for the full list).

Sleep quality was assessed using the Pittsburgh Sleep Quality Index [17]. The respiratory quality of life was assessed with the VQ11 questionnaire (global score and 3 subscores) [18]. Finally, participants were asked whether they could envisage coping with their current health status in the long term (yes/no).

Inclusion criteria for our analysis were: adult participants with a complete 12-month questionnaire and baseline data available and who declared at least one persisting symptom.

### *CLUSTERING AND STATISTICAL ANALYSIS*

The clustering was based on the following features: sociodemographic characteristics, initial classification of COVID-19 disease severity, comorbidities, symptoms at inclusion, and quality of life (See Online Supplementary Table 2 for the full list).

A Hierarchical Ascendant Classification (HAC) was used to construct clusters[8]. The optimal number of clusters has been determined using the “elbow” method calculating the distortion depending on the number of clusters with the objective to maintain clinical interpretability and sufficient cluster size. The cluster stability was assessed with the Jaccard Similarity Index. A simple imputation was done for variables if they had less than

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

5% of missing data (using median for quantitative variables and main modality for categorical variables) and multiple imputations using the mice package from R otherwise. Data were described with numbers and percentages for categorical variables and with mean and standard deviation for numerical variables. We performed all the analysis using the R software[19] and generated the figures using the ggplot2 R package[20].

## 3. Results

### *3.1. POPULATION STUDY CHARACTERISTICS*

We initially included 545 participants between May 2020 and May 2021 with an available follow-up questionnaire 12 months after their primary infection. Participants with incomplete questionnaires were excluded (n=54) as well as participants with an age below 18 (n=1), participants without any information about their study inclusion (n=19) or about their initial COVID-19 severity classification (n=3). Participants who did not experience any symptoms at 12 months were removed (n=180). Finally, 288 participants were considered in the analysis (see Online Supplementary Figure 1).

The majority of the overall study participants were females (59%) and were not hospitalized at the time of COVID-19 (97%). The average age was 43 years (sd=12) and 16% of the participants were smokers. One-third (33 %) of the participants had a moderate/severe form of the initial COVID-19. Sixty percent of the participants experienced poor sleep quality (PSQI total score>5) and 28% had a poor respiratory quality of life (VQ11 global score>22). Few participants had comorbidities prior to COVID-19 diagnosis (14%) and they had an average of 2.38 (sd=0.33) comorbidities. Hypertension was the most frequent one (13%). At the time of inclusion, the most frequent symptoms were fatigue/malaise (47%), fever (34%), cough (33%), cephalea (27%), and rhinorrhea (26%).

On average, participants declared 8 symptoms (sd=8) after 12 months. Most participants had general symptoms (80%), neurological and ocular symptoms (65%), and cardiorespiratory symptoms (55%).

### *3.2. CLUSTERS*

Based on the elbow curve (see Online Supplementary figure 2), we determined the optimal cluster number as 3, which simultaneously allows good cluster stability (Cluster-Mild, Jaccard=0.5707; Cluster-Moderate, Jaccard=0.7556; and Cluster-Severe, Jaccard=0.8297), clinical interpretability, and sufficient cluster size for each cluster. We labeled them according to their distinguishing characteristics. The characteristics of the overall study population and of the 3 clusters are shown in Table 1.

**Table 1:** Participants' characteristics in the overall study population and by cluster.

| | | **Overall Population** *N = 288* | **Cluster – Mild** *N = 139 (48.26%)* | **Cluster – Moderate** *N = 106 (36.81%)* | **Cluster – Severe** *N = 43 (14.93%)* | **P value*** |
|---|---|---|---|---|---|---|
| **Sociodemographic Characteristics and Initial Severity Classification** | Female N(%) | 170 (59%) | 73 (53%) | 66 (62%) | 31 (72%) | 0.053 |
| | Age (Years) | 43 ±12 | 42 ±12 | 43 ±12 | 45 ±14 | 0.360 |
| | Body Mass Index (kg/m2) | 26.4 ±5.5 | 25.8 ±5.1 | 27.0 ±5.8 | 26.7 ±5.7 | 0.224 |
| | Smoker N(%) | 45 (16%) | 16 (12%) | 15 (14%) | 14 (33%) | 0.027 |
| | Moderate/severe illness N(%) | 95 (33%) | 34 (24%) | 41 (39%) | 20 (47%) | 0.015 |
| **Comorbidities** | At least one comorbidity N(%) | 40 (14%) | 12 (8.6%) | 16 (15%) | 12 (28%) | 0.007 |
| | Number of comorbidities Mean(SD) | 2.38 ±0.33 | 2.37 ±0.25 | 2.34 ±0.16 | 2.48 ±0.68 | 0.001 |
| | Hypertension N(%) | 38 (13%) | 14 (10%) | 12 (11%) | 12 (28%) | 0.015 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| | Cardiac diseases N(%) | 11 (3.8%) | 3 (2.2%) | 6 (5.7%) | 2 (4.7%) | 0.311 |
| | Asthma N(%) | 14 (4.9%) | 4 (2.9%) | 8 (7.5%) | 2 (4.7%) | 0.200 |
| | Diabetes N(%) | 13 (4.5%) | 3 (2.2%) | 4 (3.8%) | 6 (14%) | 0.009 |
| **Symptoms at inclusion N(%)** | Fever | 98 (34%) | 45 (32%) | 36 (34%) | 17 (40%) | 0.688 |
| | Cough | 96 (33%) | 41 (29%) | 38 (36%) | 17 (40%) | 0.362 |
| | Cough_sputum | 27 (9.4%) | 11 (7.9%) | 9 (8.5%) | 7 (16%) | 0.279 |
| | Sore throat | 50 (17%) | 17 (12%) | 24 (23%) | 9 (21%) | 0.076 |
| | Rhinorrhea | 76 (26%) | 35 (25%) | 31 (29%) | 10 (23%) | 0.708 |
| | Earache | 22 (7.6%) | 8 (5.8%) | 10 (9.4%) | 4 (9.3%) | 0.490 |
| | Chest_pain | 19 (6.6%) | 4 (2.9%) | 11 (10%) | 4 (9.3%) | 0.036 |
| | Myalgia | 51 (18%) | 11 (7.9%) | 28 (26%) | 12 (28%) | <0.001 |
| | Arthralgia | 25 (8.7%) | 4 (2.9%) | 14 (13%) | 7 (16%) | 0.001 |
| | Fatigue | 136 (47%) | 47 (34%) | 60 (57%) | 29 (67%) | <0.001 |
| | Dyspnea | 33 (11%) | 10 (7.2%) | 16 (15%) | 7 (16%) | 0.067 |
| | Cephalea | 77 (27%) | 27 (19%) | 36 (34%) | 14 (33%) | 0.022 |
| | Abdominal pain | 14 (4.9%) | 4 (2.9%) | 3 (2.8%) | 7 (16%) | 0.004 |
| | Nausea | 13 (4.5%) | 5 (3.6%) | 4 (3.8%) | 4 (9.3%) | 0.289 |
| | Diarrhea | 20 (6.9%) | 5 (3.6%) | 8 (7.5%) | 7 (16%) | 0.019 |
| **Persisting symptoms at 12 months N(%)** | Ear Nose Throat (ENT) symptoms | 110 (38%) | 24 (17%) | 65 (61%) | 21 (49%) | <0.001 |
| | Neurological symptoms | 188 (65%) | 51 (37%) | 101 (95%) | 36 (84%) | <0.001 |
| | General symptoms | 229 (80%) | 80 (58%) | 106 (100%) | 43 (100%) | <0.001 |
| | Cardio-respiratory symptoms | 159 (55%) | 33 (24%) | 87 (82%) | 39 (91%) | <0.001 |
| | Gastrointestinal symptoms | 63 (22%) | 7 (5.0%) | 32 (30%) | 24 (56%) | <0.001 |
| | Vascular symptoms | 76 (26%) | 10 (7.2%) | 29 (27%) | 37 (86%) | <0.001 |
| | Urinary symptoms | 16 (5.6%) | 2 (1.4%) | 0 (0%) | 14 (33%) | <0.001 |
| | Skin symptoms | 66 (23%) | 17 (12%) | 12 (11%) | 37 (86%) | <0.001 |
| **Number of persisting symptoms at 12 months Mean(SD)** | Total number of symptoms | 8 ±8 | 2.89 ±2.15 | 11.5 ±5.7 | 18 ±9 | <0.001 |
| | Number ENT symptoms | 0.70 ±1.11 | 0.25 ±0.63 | 1.12 ±1.24 | 1.09 ±1.44 | 0.079 |
| | Number neurological symptoms | 2.12 ±2.28 | 0.72 ±1.27 | 3.27 ±2.07 | 3.79 ±2.63 | <0.001 |
| | Number general symptoms | 3.02 ±2.86 | 1.19 ±1.48 | 4.04 ±2.30 | 6.44 ±3.13 | <0.001 |
| | Number cardio-respiratory symptoms | 1.36 ±1.72 | 0.42 ±0.92 | 2.02 ±1.65 | 2.81 ±2.11 | 0.002 |
| | Number gastrointestinal symptoms | 0.39 ±0.87 | 0.079 ±0.382 | 0.48 ±0.86 | 1.19 ±1.35 | 0.010 |
| | Number vascular symptoms | 0.39 ±0.75 | 0.09 ±0.33 | 0.41 ±0.73 | 1.35 ±0.95 | 0.356 |
| | Number urinary symptoms | 0.07 ±0.32 | 0.01 ±0.11 | 0.00 ±0.00 | 0.44 ±0.70 | 0.610 |
| | Number skin symptoms | 0.27 ±0.54 | 0.14 ±0.38 | 0.13 ±0.39 | 1.05 ±0.62 | 0.570 |
| **Quality of life N(%)** | Could not envisage coping with symptoms long term | 45 (16%) | 11 (7.9%) | 24 (23%) | 10 (23%) | 0.002 |
| | Poor sleep# | 239 (83%) | 102 (73%) | 99 (93%) | 38 (88%) | <0.001 |
| | Altered respiratory quality of life& at 1 year | 81 (28%) | 8 (5.8%) | 51 (48%) | 22 (51%) | <0.001 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

Sleep quality was assessed using the PSQI questionnaire. A categorical variable was generated using the PSQI score: # poor sleep was defined as PSQI total score > 5. The respiratory quality of life was assessed using the VQ11 questionnaire, initially developed for COPD patients. One global score and 3 sub-scores (functional, psychological and relational) were calculated as described elsewhere and categorical variables were generated. An altered respiratory quality of life& was defined as VQ11 global score > 22. *P-values are determined using the ANOVA Significant Difference test for continuous variables (age and BMI) and the Fisher's exact test for categorical variables.

Cluster - Mild contains 139 participants (48.26%). Compared with the overall study population, the initial disease severity was classified as moderate/severe for only 24% of the members of Cluster - Mild. Individuals in this cluster had a less impacted quality of life than the overall study population: only 7.9% declared that they could not envisage coping with their symptoms in the long term, 40% of them had poor sleep quality and 5.8% had a poor respiratory quality of life. Overall, participants in Cluster - Mild had fewer comorbidities (8.6%). At 12 months, participants declared fewer symptoms overall (mean number=2.89, sd=2.15). The symptoms were mostly grouped in the following categories: general symptoms (58%), neurological and ocular symptoms (37%), and cardio-respiratory symptoms or diseases (24%).

Cluster - Moderate contains 106 participants (36.81%). Compared with the overall study population, members were slightly more frequently female (62%) and presented more frequently a moderate/severe form of the initial illness (39%). Quality of life was more impacted with 23% of Cluster - Moderate declaring that they could not envisage coping with their symptoms in the long term, 78% of them having a poor sleep quality, and 48% having a poor respiratory quality of life. Comorbidities were similar in Cluster - Moderate and in the overall study population but participants declared a higher number of symptoms at 12 months (mean=11.5, sd=5.7). All participants had general symptoms (100%), and a large majority also had neurological and ocular symptoms (95%) and cardio-respiratory symptoms or diseases (82%). Most participants also had ENT symptoms (61%).

Cluster - Severe contains 43 participants (14.93%). Compared with the overall study population, members were a majority of females (72%). Participants were more frequently smokers (33%) and 47% had an initial moderate/severe acute illness. As for Cluster - Moderate, the quality of life in Cluster - Severe was highly impacted with 84% of them having poor sleep quality and 51% having a poor respiratory quality of life. Overall, participants in Cluster - Severe presented more comorbidities at inclusion (28%), hypertension being the most frequent one (28%). At 12 months, participants had a high number of symptoms (mean=18, sd=9). The presentation of symptoms was similar to Cluster - Moderate for general, neurological, and cardio-respiratory symptoms: all participants had general symptoms (100%), 84% had neurological and ocular symptoms or diseases, and 91% had cardio-respiratory symptoms or diseases. What characterizes Cluster - Severe is the high frequencies of vascular, skin, and urinary symptoms (86%, 86%, and 33%, respectively).

The symptom distribution by symptom categories in the 3 clusters is represented in Figure 1 which shows the differences between the clusters.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

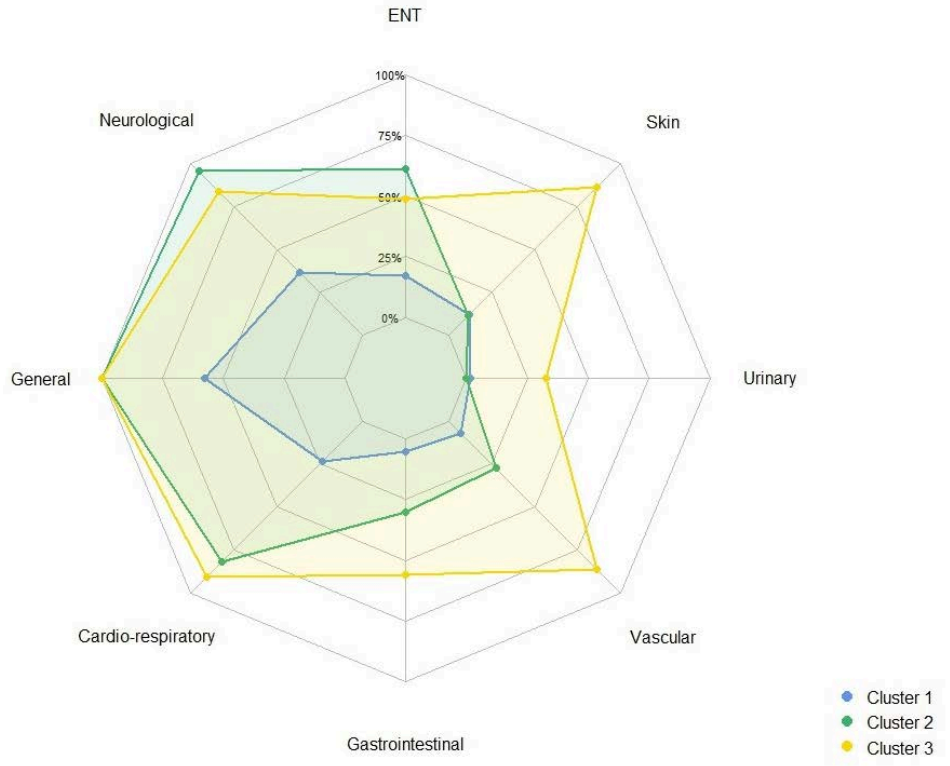



**Figure 1: Distribution of Long COVID symptoms (in %) by symptom categories in the 3 clusters.**

## 4. Discussion

In this study, we identified 3 clusters of Long COVID in people with persisting symptoms 12 months after acute infection with a clear gradient in Long COVID severity. Cluster - Mild represented almost half of the study population and was composed of participants with less severe initial infection, fewer comorbidities and with few persisting symptoms (mean=2.9), mainly in the general, neurological, or cardiorespiratory categories. Individuals in Cluster - Moderate declared a mean of 11.5 persisting symptoms and had a poor quality of sleep and of respiratory quality of life. Cluster - Severe was characterized by a higher proportion of women, of smokers, with a higher number of preexisting comorbidities than in Clusters - Mild and Moderate. Strikingly, participants from Cluster - Severe declared more persisting symptoms in total than those from Cluster - Moderate (mean=18), with a similar pattern of general, neurological, and cardio-respiratory symptoms, but is distinct by higher occurrences of vascular, urinary, and skin symptoms.

Few studies investigated clustering analysis of Long COVID patients. Kenny et al. applied similar clustering methods to a prospective cohort of 233 COVID-19-infected patients with ongoing symptoms at least 4 weeks after acute infection and described also 3 clusters: the larger one constituted by participants with a lower number of persisting symptoms (mean=2) and 2 characterized by a higher number of persisting symptoms (mean=4 and 6) and more functional impairments. As in our study, the distribution of persisting symptoms was different between the 2 most severe clusters, with one cluster grouping cardio-respiratory and general symptoms, and the other one with a predominance of pain-related symptoms. The time and method of symptom evaluation were different as it was done in person during a visit to a clinic and the median time of symptom duration was 18 weeks[13]. Another study identified 3 different clusters among a cohort of 1969 post-hos-

 doi:10.20944/preprints202211.0033.v1

pitalized COVID-19 patients in Spain[12]: one cluster grouped patients with fewer comorbidities and symptoms at the hospital inclusion, less persisting symptoms, and had a preserved quality of life, and the 2 other clusters were constituted of patients with more pre-existing comorbidities, a higher number of symptoms during the acute phase, a higher number of persisting symptoms and higher impact on quality of life (higher level of anxiety and altered sleep quality). One cluster was also characterized by respiratory symptoms (dyspnea at rest 73.4%) and particularly high limitations in daily activities (92.1% for social activities and 93.3% for instrumental daily activities). The overall number of symptoms in each cluster was lower than in our clusters because their clustering included also participants without persisting symptoms.

Another study conducted in the United Kingdom in 2022 also described groups of people with Long COVID. More participants (N=2550) were recruited, via an online survey, with a mean duration of illness of 7.2 months (sd=1.8). The mean age was similar to our participants, as was the greater presence of women and comorbidities. The most common first symptoms (fatigue, headache, chest pain, shortness of breath, and cough), persistent symptoms (fatigue, cognitive dysfunction, chest pain, shortness of breath, headache, and muscle pain), number of symptoms experienced, and organ systems affected were also similar. Participants were asked to report the presence or absence of 35 symptoms, and two groups were identified. The first group (88.8%) had mainly cardiopulmonary, cognitive, and fatigue symptoms and the second group had more multisystem symptoms [8], which is relatively well aligned with our findings.

Reese et al applied an adapted Phenomizer algorithm to classify patients with Long COVID, based on the ICD-10 diagnosis code U09.9 for Post COVID-19 condition, and identified 6 clusters[21]. Although the clustering method was different and based on medical records data, this study also identified 2 “severe” clusters with more pre-existing comorbidities, an increased initial illness, and a wide range of Long COVID symptoms.

The overrepresentation of women in the most severe cluster is consistent with findings from other studies[8,13].

Finally, despite different analysis time points, similar results were found in these different studies which confirm that our findings are relevant despite the fluctuating character of Long COVID.

*STRENGTHS AND LIMITATIONS*

This study has several strengths. First, a large list of 62 symptoms was considered, distributed in 8 categories that cover the complex symptomatology of Long Covid. Participants with different forms of initial illness severity were represented. All participants had a documented initial COVID-19 infection, confirmed by a PCR test and their symptoms were assessed 12 months after acute infection.

This study also has some limitations. The analyses were done on a moderate sample size and, as in any selected study population, results may not be directly extrapolated to all people with Long COVID. External validation in a larger population would be of the highest interest to confirm these results. Information on pre-existing symptoms before COVID-19 infection was missing and symptoms were self-reported which could lead to biases in the estimation of the number of persisting symptoms attributable to COVID-19. However, this may not affect the main message of our findings. The participants in the present study were included before the Omicron wave, thus we can not ensure that our results can be extended to Long COVID following infection by the Omicron variant. Recent studies demonstrated that infection by Omicron variants leads to a 24 to 50% risk reduction of developing Long COVID, however, there were no differences in the distribution of Long COVID symptoms and the risk of neurological and psychiatric sequelae remains the same after infection by Omicron[22][23][24].

## 5. Conclusions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

Our study highlighted 3 clinically relevant subgroups of people with Long COVID of increasing severity, but also with different patterns of symptoms. Such stratification of Long Covid will help healthcare professionals to define precision health strategies, improve the triage of patients, and personalize the care of people with Long COVID.

**Supplementary Materials:** Not applicable.

**Author Contributions:** N.B., A.F., and G.F. wrote the manuscript. All the authors interpreted the data, critically revised the manuscript, and approved the final version. All authors have read and agreed to the published version of the manuscript.

**Funding:** The Predi-COVID study is supported by the Luxembourg National Research Fund (FNR) (Predi-COVID, grant number 14716273), the André Losch Foundation, and the Luxembourg Institute of Health.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the National Research Ethics Committee of Luxembourg (study number 202003/07).

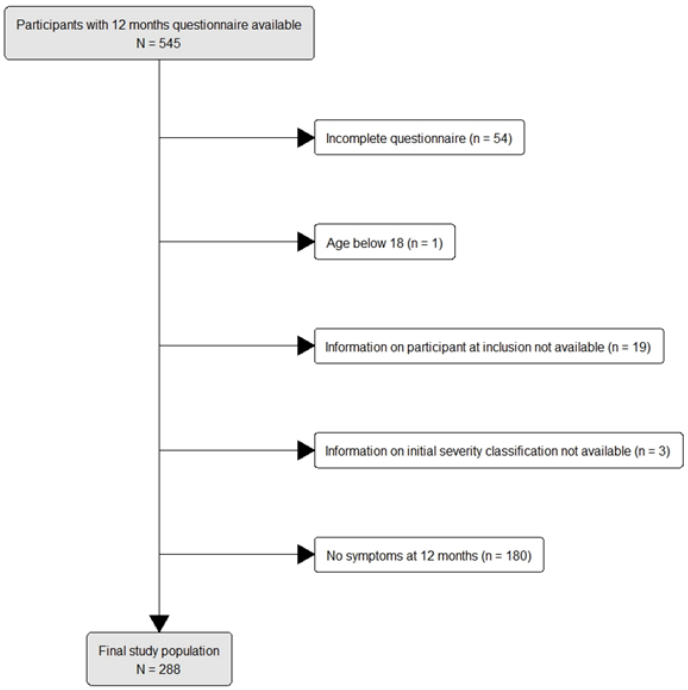
**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are available from the corresponding author upon reasonable request.

**Acknowledgments:** In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

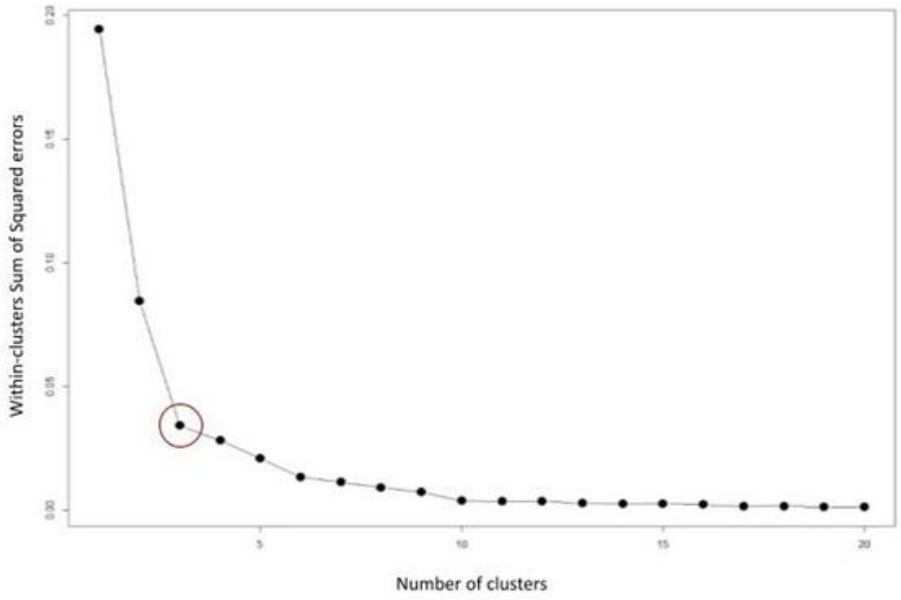
**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A




**Supplementary Figure A1.** Flowchart of participants included in the analyses (N = 288).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1




**Supplementary Figure A2.** Determination of cluster number:.

**Supplementary Table A1: Full list of persisting symptoms considered in the 12-months questionnaire**

| Ear/Nose/Throat Symptoms | Neurological and ocular symptoms | General symptoms | Cardio-respiratory symptoms or diseases | Gastrointestinal symptoms | Vascular and ganglionic symptoms or diseases | Urinary symptoms | Skin symptoms |
|---|---|---|---|---|---|---|---|
| Loss of taste | Tremors | Fatigue | Shortness breath | Nausea | Hypertension | Urinary pain | Skin rashes |
| Loss of smell | Headaches | Irritability | Chest tightness | Vomiting | Hypotension | Urinary infections | Dry skin |
| Runny nose, cold or rhinitis | Migraines | Anxiety | Dry cough | Diarrhea | Adenopathies | Dialysis | Blue fingers |
| Sinus pain | Mental confusion | Depression | Fatty cough | Heartburn | Circulation disorders | | |
| Ear pain | Malaise | Sweating | Tachycardia | Abdominal pain | Hematoma | | |
| Sore throat | Convulsions | Fever | Arrhythmia | | | | |
| | Balance | Loss of appetite | Myocarditis | | | | |
| | Memory | Loss weight | Heart failure | | | | |
| | Fatigue in eyes | Thirst | Burning chest | | | | |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Hallucina-tions | Ants | Chest pain | | | | |
| | Sensitivity to light | Muscle pain (upper limbs) | Wheezing | | | | |
| | Conjunctivi-tis | Muscle pain (lower limbs) | Coughing blood | | | | |
| | | Back pain | | | | | |
| | | Allergy | | | | | |
| | | Loss hair | | | | | |
| | | Difficulty walking | | | | | |

**Supplementary Table A2: Full list of features included in the clustering**

| | |
|---|---|
| **Sociodemographic Characteristics and Initial Severity Classification** | Inclusion at home or at hospital |
| | Gender |
| | Age |
| | BMI |
| | Weight loss in last 6 months |
| | Smoking status |
| | Classification severity initial illness |
| | Blood type |
| **Comorbidities** | Hypertension |
| | Cardiac diseases |
| | Pulmonary diseases |
| | Asthma |
| | Renal diseases |
| | Hepatic diseases |
| | Neurological diseases |
| | Cancer |
| | Hematological diseases |
| | Obesity |
| | Diabetes |
| | Rheumatological diseases |
| | Malnutrition |
| | COPD |
| **Symptoms at inclusion** | Fever |
| | Cough |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0033.v1

| | |
|---|---|
| | Cough_sputum |
| | Cough hemoptysis |
| | Sore throat |
| | Rhinorrhea |
| | Earache |
| | Wheezing |
| | Chest_pain |
| | Myalgia |
| | Arthralgia |
| | Fatigue |
| | Dyspnea |
| | Chest tightness |
| | Cephalea |
| | Confusion |
| | Abdominal pain |
| | Nausea |
| | Diarrhea |
| | Conjunctivitis |
| | Skin rash |
| | Lymphadenopathy |
| | Fall |
| | Hemorrhage |
| **Persisting symptoms at 12 months by categories** | Ear Nose Throat (ENT) symptoms |
| | Neurological symptoms |
| | General symptoms |
| | Cardio-respiratory symptoms |
| | Gastrointestinal symptoms |
| | Vascular symptoms |
| | Urinary symptoms |
| | Skin symptoms |
| **Quality of life** | Could not envisage coping with symptoms long term N(%) |
| | PSQI score |
| | VQ11 global score |
| | VQ11 Functional component score |
| | VQ11 Psychological component score |
| | VQ11 Relational component score |

**References**

1. Ballering AV, van Zon SKR, Olde Hartman TC, Rosmalen JGM, Lifelines Corona Research Initiative. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. Lancet. 2022;400: 452–461.

2. Coronavirus disease (COVID-19): Post COVID-19 condition. [cited 21 Oct 2022]. Available: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-post-covid-19-condition

3. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization; 6 Oct 2021 [cited 29 Sep 2022]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

4. Carfì A, Bernabei R, Landi F. Persistent Symptoms in Patients After Acute COVID-19. JAMA. 2020;324. doi:10.1001/jama.2020.12603

5. Halpin SJ, McIvor C, Whyatt G, Adams A, Harvey O, McLean L, et al. Postdischarge symptoms and rehabilitation needs in survivors of COVID-19 infection: A cross-sectional evaluation. J Med Virol. 2021;93. doi:10.1002/jmv.26368

6. Garrigues E, Janvier P, Kherabi Y, Le Bot A, Hamon A, Gouze H, et al. Post-discharge persistent symptoms and health-related quality of life after hospitalization for COVID-19. J Infect. 2020;81. doi:10.1016/j.jinf.2020.08.029

7. Tenforde MW, Kim SS, Lindsell CJ, Billig RE, Shapiro NI, Files DC, et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network - United States, March-June 2020. MMWR Morb Mortal Wkly Rep. 2020;69. doi:10.15585/mmwr.mm6930e1

8. 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. doi:10.1371/journal.pone.0264331

9. A clinical case definition of post COVID-19 condition by a Delphi consensus, 6 October 2021. World Health Organization; 6 Oct 2021 [cited 2 Aug 2022]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1

10. Amenta EM, Spallone A, Rodriguez-Barradas MC, El Sahly HM, Atmar RL, Kulkarni PA. Postacute COVID-19: An Overview and Approach to Classification. Open Forum Infectious Diseases. 2020;7. doi:10.1093/ofid/ofaa509

11. Fischer A, Zhang L, Elbéji A, Wilmes P, Oustric P, Staub T, et al. Long COVID Symptomatology After 12 Months and Its Impact on Quality of Life According to Initial Coronavirus Disease 2019 Disease Severity. Open Forum Infect Dis. 2022;9: ofac397.

12. Fernández-de-Las-Peñas C, Martín-Guerrero JD, Florencio LL, Navarro-Pardo E, Rodríguez-Jiménez J, Torres-Macho J, et al. Clustering analysis reveals different profiles associating long-term post-COVID symptoms, COVID-19 symptoms at hospital admission and previous medical co-morbidities in previously hospitalized COVID-19 survivors. Infection. 2022. doi:10.1007/s15010-022-01822-x

13. 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. Open Forum Infect Dis. 2022;9: ofac060.

14. Fagherazzi G, Fischer A, Betsou F, Vaillant M, Ernens I, Masi S, et al. Protocol for a prospective, longitudinal cohort of people with COVID-19 and their household members to study factors associated with disease severity: the Predi-COVID study. BMJ Open. 2020;10: e041834.

15. Clinical Spectrum. In: COVID-19 Treatment Guidelines [Internet]. [cited 3 Aug 2022]. Available: https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/

16. Capelle CM, Ciré S, Domingues O, Ernens I, Hedin F, Fischer A, et al. Combinatorial analysis reveals highly coordinated early-stage immune reactions that predict later antiviral immunity in mild COVID-19 patients. Cell Reports Medicine. 2022; 100600.

17. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28. doi:10.1016/0165-1781(89)90047-4

18. Ninot G, Soyez F, Préfaut C. A short questionnaire for the assessment of quality of life in patients with chronic obstructive pulmonary disease: psychometric properties of VQ11. Health Qual Life Outcomes. 2013;11: 179.

19. Website. Available: R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

20. Gómez-Rubio V. ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). Journal of Statistical Software. 2017. doi:10.18637/jss.v077.b02

21. Reese J, Blau H, Bergquist T, Loomba JJ, Callahan T, Laraway B, et al. Generalizable Long COVID Subtypes: Findings from the NIH N3C and RECOVER Programs. medRxiv. 2022. doi:10.1101/2022.05.24.22275398

22. Antonelli M, Pujol JC, Spector TD, Ourselin S, Steves CJ. Risk of long COVID associated with delta versus omicron variants of SARS-CoV-2. The Lancet. Elsevier BV; 2022. pp. 2263–2264.

23. Morioka S, Tsuzuki S, Suzuki M, Terada M, Akashi M, Osanai Y, et al. Post COVID-19 condition of the Omicron variant of SARS-CoV-2. J Infect Chemother. 2022;28: 1546–1551.

24. Taquet M, Sillett R, Zhu L, Mendel J, Camplisson I, Dercon Q, et al. Neurological and psychiatric risk trajectories after SARS-CoV-2 infection: an analysis of 2-year retrospective cohort studies including 1 284 437 patients. Lancet Psychiatry. 2022;9: 815–827.