Preprints.org

Article

Long COVID's Hidden Complexity: Machine Learning Reveals Why Personalized Care Remains Essential

Eleonora Fresi , Elisabetta Pagani [*] , Federica Pezzetti , Cristina Montomoli , Cristina Monti , Monia Betti , Annalisa De Silvestri , Orlando Sagliocco , Valentina Zuccaro , Raffaele Bruno , Catherine Klersy

Article

Long COVID's Hidden Complexity: Machine Learning Reveals Why Personalized Care Remains Essential

Eleonora Fresi [1,†]**, Elisabetta Pagani** [2,*,†]**, Federica Pezzetti** [3]**, Cristina Montomoli** [4]**, Cristina Monti** [4]**, Monia Betti** [5]**, Annalisa De Silvestri** [1]**, Orlando Sagliocco** [6]**, Valentina Zuccaro** [2]**, Raffaele Bruno** [2,7] **and Catherine Klersy** [1]

[1] Biostatistics & Clinical Trial Center, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

[2] Infectious Diseases Unit, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

[3] Operation Management Next Generation EU, ASST Cremona, Cremona, Italy

[4] Biostatistics and Clinical Epidemiology Unit, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Italy

[5] Pneumology Unit, ASST Cremona, Cremona, Italy

[6] Intensive Care Unit Bolognini Hospital, ASST Bergamo Est

[7] Dipartimento di Scienze cliniche chirurgiche diagnostiche e pediatriche, Università di Pavia

[*] Correspondence: e.pagani@smatteo.pv.it

[†] These authors contributed equally to this work.

Abstract: Background Long Covid can develop in individuals who have had Covid-19, regardless of the severity of their initial infection or of the treatment they received. Several studies examined the prevalence and manifestation of symptoms phenotypes to comprehend the pathophysiological mechanisms associated with these. Numerous articles outlined specific approaches for a multidisciplinary management and treatment of these patients, focusing primarily on those with mild acute illness. The various management models implemented focused on a patient-centred approach, where the specialists were positioned around the patient. On the other hand, the created pathways do not consider the possibility of symptom clusters when determining how to define diagnostic algorithms. **Methods** This is a retrospective longitudinal study that took place at the "Fondazione IRCCS Policlinico San Matteo", Pavia, Italy (SMATTEO) and at the "Ospedale di Cremona", ASST Cremona, Italy (CREMONA). Information was retrieved from the administrative datawarehouse and from two dedicated registries. We included patients discharged with a diagnosis of severe Covid, systematically invited for a 3-month follow-up visit. Unsupervised machine learning was used to identify potential patient phenotypes. **Results** Three-hundred and eighty-two patients were included in these analyses. About one-third of patients were older than 65 years; a quarter were female; more than 80% of patients had multi-morbidities. Diagnoses related to the circulatory system were the most frequent, comprising 46% of cases, followed by endocrinopathies at 20%. PCA (principal component analysis) had no clustering tendency, which was comparable to the PCA plot of a random dataset. The unsupervised machine learning approach confirms these findings. Indeed, while dendrograms for the hierarchical clustering approach may visually indicate some clusters, this is not the case for the PAM method. Notably, most patients concentrated in one cluster. **Conclusion** The extreme heterogeneity of patients affected by post-acute sequelae of Sars-Cov2 infection (PASC) has not allowed the identification of specific symptom clusters with the most recent statistical techniques, thus preventing the generation of common diagnostic-therapeutic pathways.

Keywords: PASC; cluster; phenotype; machine learning; severe COVID

1. Introduction

The British National Institute for Health and Care Excellence (NICE) describes long COVID or post-COVID syndrome (PCS) as the continuation of signs and symptoms that were present during or arised after a Covid-19 infection and persist for more than twelve weeks, with no other explanation for their persistence [1]. In contrast, the United States National Institutes of Health (NIH) refers to Long Covid as sequelae that persist beyond four weeks from the onset of the initial infection, as the definition provided by the Centers for Disease Control and Prevention (CDC) [2]. Nowadays, post-acute sequelae of Sars-Cov2 infection syndrome (PASC) is defined as symptoms that persist, relapse or arise 30 or more days after a Sars-COv2 infection [3].

Many studies examined the residual symptoms reported after contracting SARS-CoV-2, including also the incidence, risk factors, treatment, and management of Long Covid [4,5]. Considering all these factors, it is evident that this virus can potentially result in health lasting consequences [6]. Long Covid can impact individuals who experienced mild symptoms during their initial illness, as well as those who battled against more severe forms of the infection [7,8]. Long Covid can develop in any individual who has had Covid-19, regardless of the severity of the initial infection or of the treatment they received. This includes patients treated in hospital wards or intensive care units, requiring oxygen therapy, continuous positive airway pressure, or invasive ventilation and those not hospitalized [9].

Several studies examined the prevalence and manifestation of different symptom phenotypes to comprehend the underlying pathophysiological mechanisms associated with these symptoms [10]. However, these studies did not differentiate between patients based on the severity of their acute illness [11,12]. They identified various phenotypes in diverse populations of Covid patients, including both those who were hospitalized and those who were not [13]. These studies are focused on deciphering the pathophysiological mechanisms underlying PASC.

Numerous articles outlined specific approaches for a multidisciplinary management and treatment of these patients, focusing primarily on those with mild acute illness [14]. The various management models implemented focused on a patient-centred approach, where the specialists involved were positioned around the patient [15]. The created pathways did not consider the possibility of symptom clusters when determining how to define diagnostic algorithms.

Thus, due to the complexity of the issue, a comprehensive and universally accepted definition is challenging.

Our research seeks to determine how common are lingering symptoms, three months after patients have been released from the hospital following a severe case of COVID-19. Lombardy was a region with a high rate of COVID-19 infections during the initial phases of the pandemic. This research aims also to establish standardized treatment pathways for these patients by identifying common symptom clusters. Clusters may help distinguish unique phenotypes of individuals experiencing PASC syndrome.

2. Material and Methods

2.1. Study Design

This is a retrospective longitudinal study part of a larger research project funded by Fondazione CARIPLO, the "Chronic diseases management after the CoViD-19 epidemic trigger. Capturing data, generating evidence, suggesting actions for health protection. The CHANCE Project" (cod. 2020-4238). This sub-project took place at the "Fondazione IRCCS Policlinico San Matteo", Pavia, Italy (SMATTEO) and at the "Ospedale di Cremona", ASST Cremona, Italy (CREMONA); the study was approved by the ethical committee of Pavia (26 July 2022, protocol number 0036061/22) as well as by the ethical committee of Val Padana (30 September 2022, protocol number 34131).

2.2. Data Source

Discharge data on hospitalization were retrieved from the administrative databases of both hospitals and follow-up data were derived from dedicated clinical COVID registries maintained at both hospitals. Patients with multimorbidities were identified through the ICD9-CM discharge diagnoses up to the 6th.

2.3. Study Population

Individuals with residual symptoms correlated with PASC were enrolled during the outpatient follow-up visit at 3 months after discharge from the two medical facilities. Subjects discharged between March 2020 and March 2022 with a diagnosis of severe COVID-19 were eligible for the study. Specifically, subjects who required either CPAP (Continuous Positive Airway Pressure) or Endotracheal Intubation and exhibited residual symptoms at the 3- to 4-month visit were included. The following discharge ICD9-CM diagnoses were considered: codes 48041, 51891, 9604, 311, 9670, 9671, 9672, and 9390. During the outpatient visit, information about the presence of the residual symptoms were collected, listed in Supplementary Table 1; for the purpose of the analysis, we partially aggregated these symptoms in macro-categories, as identified in the Table 1.

Table 1. Subjects characteristics and disease main categories of the study population. Overall and by participating centre.

	ICD9-CM chapter	Overall N = 382[1]	San Matteo Hosp N = 242[1]	Cremona Hosp N = 140[1]
Sex (F)	-	102 (26.7%)	*72 (29.8%)*	*30 (21.4%)*
Age > 65	-	136 (35.7%)	*84 (34.7%)*	*52 (37.4%)*
Endotracheal intubation	-	97 (25.4%)	*51 (21.1%)*	*46 (32.9%)*
Multimorbidities	-	324 (84.8%)	*200 (82.6%)*	*124 (88.6%)*
Circulatory	7	176 (46.1%)	*126 (52.1%)*	*50 (35.7%)*
Endocrin	3	76 (19.9%)	*66 (27.3%)*	*10 (7.1%)*
Genitourinary	10	34 (8.9%)	*24 (9.9%)*	*10 (7.1%)*
Neurological	6	25 (6.5%)	*21 (8.7%)*	*4 (2.9%)*
Gastroenterological	9	13 (3.4%)	*11 (4.5%)*	*2 (1.4%)*
Cancer	2	12 (3.1%)	*8 (3.3%)*	*4 (2.9%)*
Haematological	4	10 (2.6%)	*9 (3.7%)*	*1 (0.7%)*
Dermatological	12	8 (2.1%)	*5 (2.1%)*	*3 (2.1%)*
Trauma	17	6 (1.6%)	*4 (1.7%)*	*2 (1.4%)*
Mental	5	5 (1.3%)	*4 (1.7%)*	*1 (0.7%)*

Musculoskeletal	13	*4 (1.0%)*	*4 (1.7%)*	*0 (0.0%)*
Other	18	*157 (41.1%)*	*28 (11.6%)*	*129 (92.1%)*
Symptoms	16	*113 (29.6%)*	*7 (2.9%)*	*106 (75.7%)*

[1] n (%).

2.4. Data Analysis

All analyses were performed using the R (v. 4.3.1) software [16]. We used the Fisher exact test to compare patient characteristics between the two hospitals. The prevalence of each category of symptoms was computed together with its exact binomial 95% confidence interval (95% CI). In order to elicit potential different aggregations of patients, we plotted the entire case series over the first two components of a principal component analysis (PCA). For comparison, PCA was also run on the random dataset. More formally, we ~~then~~ applied a series of unsupervised machine learning techniques, such as hierarchical clustering (either agglomerative or divisive) and partition around medoids. These techniques attempt to find subgroups of patients that share common characteristics and differ from the other subgroups. To rank the performance of such methods, we calculated the following indices that measure the separation between potential clusters (the higher the better): the average silhouette width, with a value >0.5 generally considered as an acceptable performance, the separation index (range 0-1) and the cophenetic correlation coefficient (range |0-1|). To further discriminate between the three methods, we computed the entropy, where lower values indicated lower heterogeneity within clusters. The results of the clustering processes were reported graphically as dendrograms or cluster plots. Patients in which more than 50% of the selected variables were missing, did not enter the machine learning approach. A detailed description of these analyses is reported in the supplementary material.

3. Results

3.1. Patient

In our study we included 382 patients discharged with a diagnosis of severe COVID-19 and a 3-month follow-up visit. Their demographic and clinical characteristics are shown in Table 1: about one-third of patients were older than 65 years, and a quarter were female; 25% of this case series had had endotracheal intubation during their hospitalization; more than 80% of patients had multimorbidities. Diagnoses related to the circulatory system were notably the most frequent, including 46% of cases, followed by those within the endocrine ICD9-CM chapter at 20%. All other diagnoses had a prevalence below 10%. Notably, 40% and 30% of diagnoses were unspecified and symptomatic, respectively. Table 2 reports the prevalence of symptoms at follow-up. About 70% of patients (N=253) attending the outpatient clinic had residual symptoms, with 40% of them with 2 or more symptoms. Dyspnea prevalence was largely the highest with 60% of patients affected. Fatigue (40%) and neuropsychological symptoms (30%) were other frequent symptoms.

Table 2. Prevalence of Symptoms at follow-up (95%CI) overall and by participating centre.

Symptom	All (N=382)		San Matteo Hosp (N=242)		Cremona Hosp (N=140)	
	N	% (95%CI)	N	% (95%CI)	N	% (95%CI)

Residual Symptoms	253	67.8 (62.8, 72.5)	148	63.5 (56.9, 69.6)	105	75.0 (66.8, 81.8)
Multiple Symptoms						
1	107	28.0 (23.6, 32.9)	71	29.3 (23.8, 35.6)	36	25.7 (18.9, 33.9)
2	77	20.2 (16.3, 24.6)	46	19.0 (14.4, 24.6)	31	22.1 (15.8, 30.1)
3+	74	19.4 (15.6, 23.8)	36	14.9 (10.8, 20.1)	38	27.1 (20.1, 35.4)
Dyspnea	170	60.9 (54.9, 66.6)	100	68.5 (60.2, 75.8)	70	52.6 (43.8, 61.3)
Fatigue	109	39.8 (34.0, 45.9)	64	45.7 (37.3, 54.3)	45	33.6 (25.8, 42.3)
Neuro-psychological symptoms	69	30.4 (24.6, 36.9)	33	35.9 (26.3, 46.6)	36	26.7 (19.6, 35.1)
Rheumatologic symptoms	47	21.1 (16.0, 27.1)	21	23.6 (15.5, 34.0)	26	19.4 (13.3, 27.3)
Cardiovascular symptoms	47	17.2 (13.0, 22.3)	28	20.3 (14.1, 28.2)	19	14.1 (8.9, 21.4)
Otorhinolaryngological symptoms	28	10.3 (7.1, 14.7)	20	14.5 (9.3, 21.7)	8	6.0% (2.8, 11.8)
Dermatologic symptoms	22	9.8 (6.4, 14.6)	6	6.7 (2.7, 14.5)	16	11.9 (7.1, 18.8)
Cough	18	6.6 (4.1, 10.5)	7	5.1% (2.3, 10.7)	11	8.1 (4.3, 14.4)
Gastrointestinal disorders	19	6.9 (4.3, 10.8)	16	11.4 (6.9, 18.2)	3	2.2 (0.6, 6.9)
Headache	11	4.9 (2.6, 8.9)	9	10.1 (5.0, 18.8)	2	1.5 (0.3, 5.8)

3.2. Unsupervised Machine Learning for the Identification of Patient Aggregation

These analyses included 234 patients with symptoms and sufficient available data.

As shown in Figure 1 (A), there was no clustering tendency at PCA, comparable to the PCA plot of a random dataset (B). The unsupervised machine learning approach confirms these findings. Indeed, while dendrograms for the hierarchical clustering approach may visually indicate some clusters (Figure 2, panels A and B), this is not the case for the PAM method (Figure 2, panel C). Notably, most patients concentrated in one cluster (181 out of 234). Moreover, the internal validation indices did not support the validity of patient aggregation, as evidenced by inferior values (Table 3). Specifically, the average silhouette values were significantly below the acceptable threshold of 0.5 across all cases, indicating that the clusters might overlap or were not well-defined. Similarly, the separation index was close to 0, confirming the lack of separation between the hypothetical clusters. Importantly, none of the three approaches demonstrated superiority over the others.

Table 3. Internal stability indexes for hierarchical (Agglomerative and Divisive) and PAM clustering of patients.

Method	Average Silhouette	Separation Index (SI)	Cophenetic correlation coefficient	Entropy
Agglomerative Clustering	0.31	0.05	0.61	1.10
Divisive Clustering	0.31	0.03	0.74	0.74
PAM Clustering	0.18	0.01	-	1.27

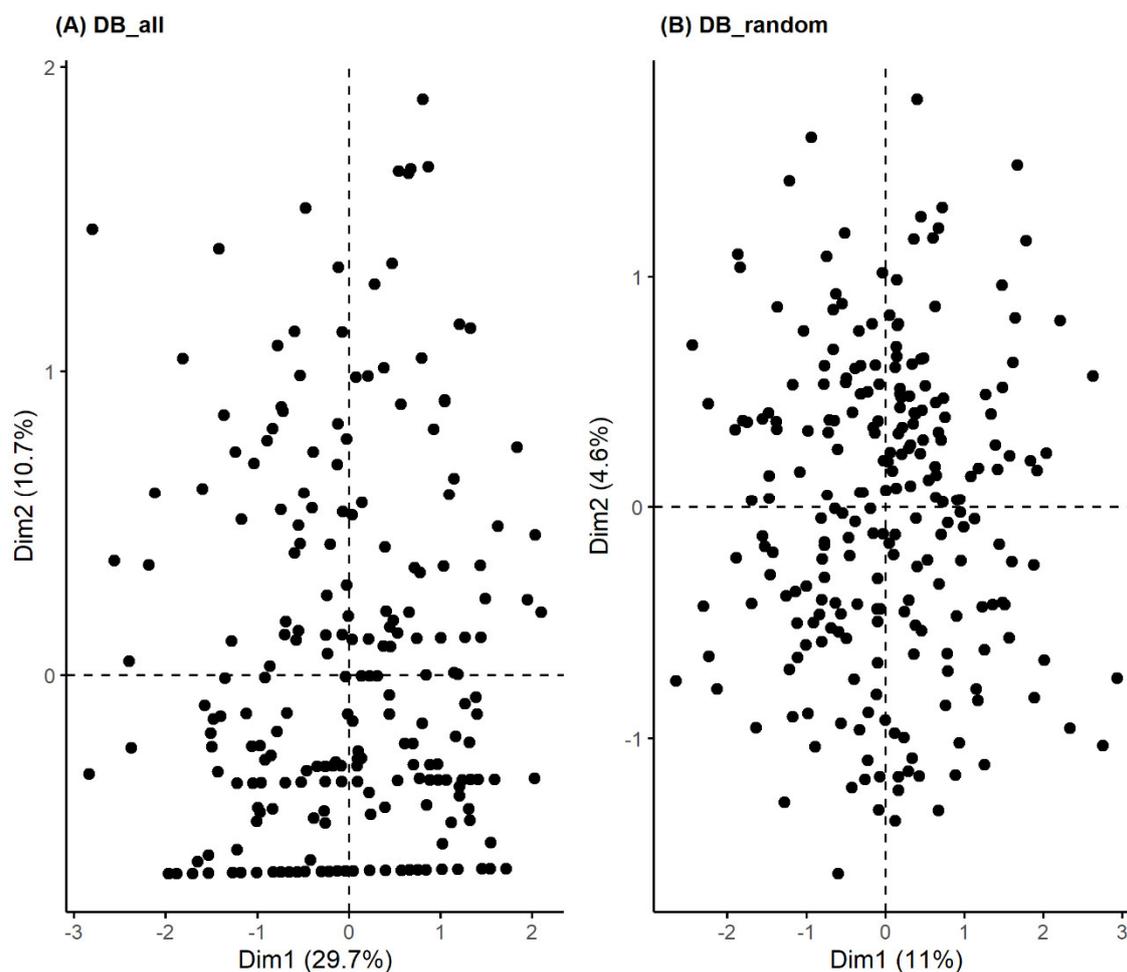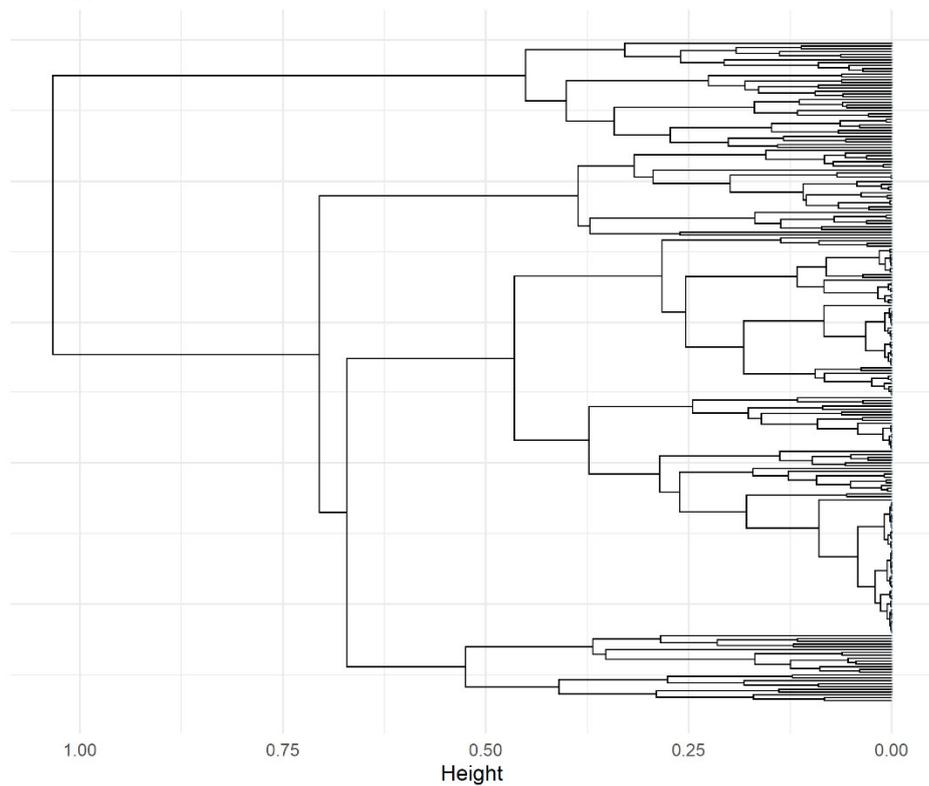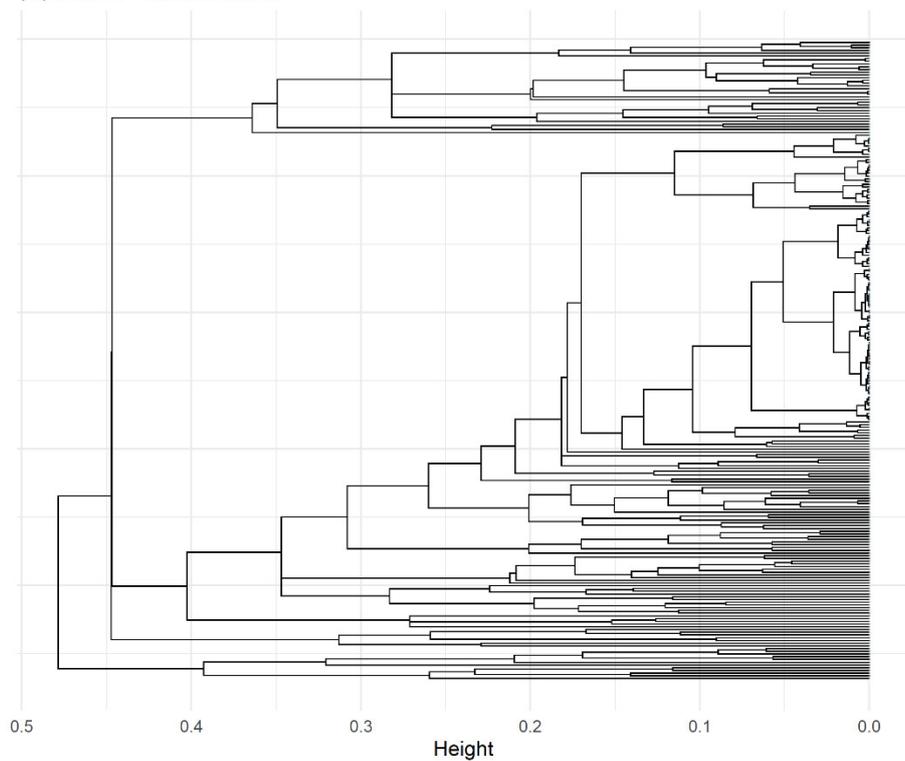(A) DB_all   (B) DB_random

Figure 1. PCA plot over the first two principal components for Patient data (A) and a random generated dataset (B).

(A) Agglomerative Hierarchical
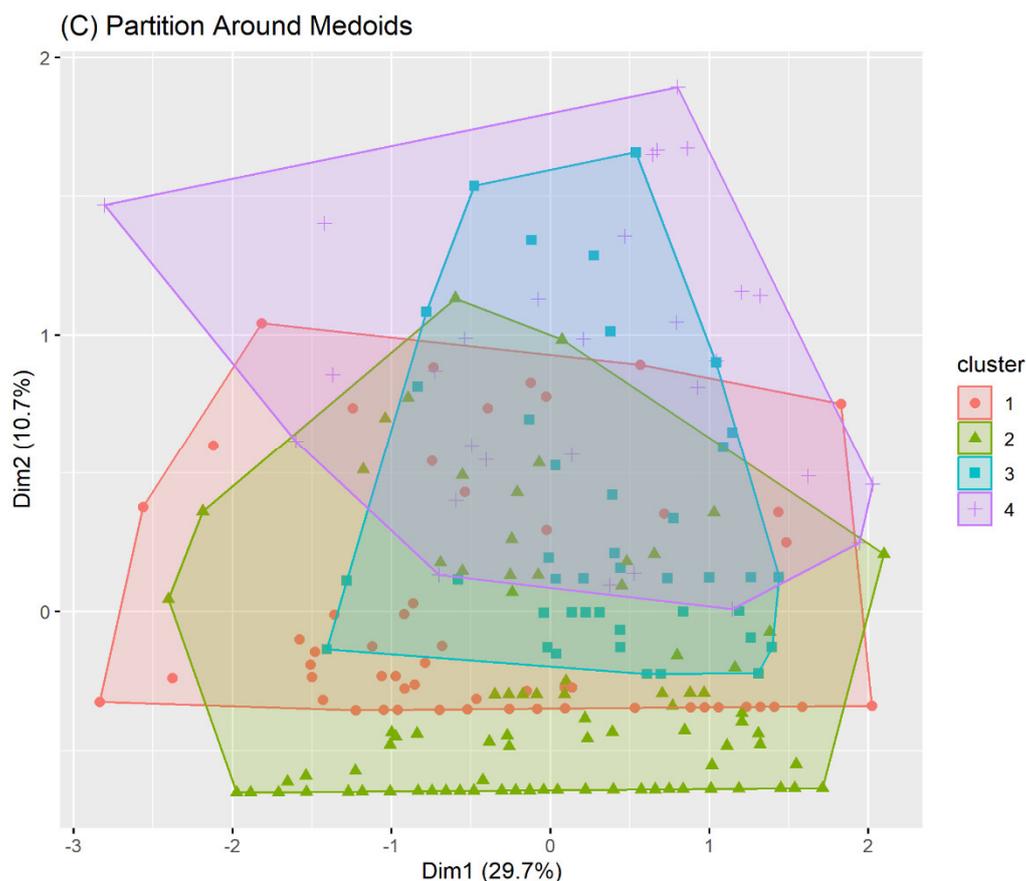

(B) Divisive Hierarchical

Figure 2. (A) Agglomerative hierarchical, (B) Divisive hierarchical and (C) PAM clustering for patients.

4. Discussion and Conclusions

Discussion

Patients included in this study were recruited from two major Centers in Lombardy (IRCCS Policlinico San Matteo Foundation and Cremona Hospital), areas with a high incidence of COVID-19 during the first two waves of the pandemic. The study aimed to evaluate the prevalence of post-acute sequelae of SARS-CoV-2 infection (PASC) symptoms in patients discharged after a severe COVID-19. It sought to identify symptom-based patient clusters to facilitate structured management pathways. Utilizing a retrospective longitudinal design within a larger project funded by Fondazione CARIPLO, the researchers analyzed hospital discharge and follow-up data from COVID registries. A total of 382 patients with severe COVID-19 were included. At the 3-month follow-up, 70% of patients exhibited residual symptoms, predominantly dyspnea, fatigue, and neuropsychological issues.

The results of our investigation, in particular the application of an unsupervised machine learning approach, indicate that there was no discernible clustering of patients, thus precluding the identification of specific phenotypes among individuals, systematically assessed three months after discharge with a diagnosis of severe COVID-19 and residual symptoms of PACS.

Additionally, we observed a limited number of patients who required continuous positive airway pressure (CPAP) or endotracheal intubation during their hospitalization. This observation can be attributed to the fact that patients requiring such interventions were less likely to be discharged alive and, consequently, were unable to participate in the three-month follow-up visit.

Our analysis revealed that the study population was characterized by a remarkably high prevalence of multimorbidities (84.4%), with circulatory and endocrine diseases being the most commonly observed comorbid conditions. This highlights the complexity of managing post-COVID-

19 patients, especially those with pre-existing health conditions, and underscores the importance of comprehensive and tailored medical care to address their diverse needs.

The most common symptoms reported in clusters of PACS patients vary but generally include a range of physical, cognitive, and psychiatric manifestations. Fatigue emerges as a predominant symptom across multiple studies, often accompanied by dyspnea (shortness of breath) and cognitive impairments such as forgetfulness and memory impairment. For instance, one study identified clusters including fatigue alone and combinations of fatigue with other symptoms like dyspnea, chest pain, and cognitive disturbances [17]. Similarly, another study highlighted fatigue, dyspnea, and myalgia as the most common symptoms, with women reporting more symptoms than men [18]. Psychiatric symptoms, including anxiety and depression, are also frequently reported among Long COVID patients. A systematic review found sleep disturbances, depression, post-traumatic stress symptoms, anxiety, and cognitive impairments to be common psychiatric manifestations [19]. Moreover, the risk factors for developing psychiatric symptoms include being female and having a previous psychiatric diagnosis [20]. The heterogeneity of PASC symptoms is further evidenced by the identification of symptom clusters such as gastrointestinal, musculoskeletal, neurocognitive and cardiopulmonary in one study, with neurocognitive symptoms being associated with increased odds of depression and anxiety [21]. Another study proposed three phenotypes of PASC based on symptom severity, with the severe phenotype characterized by fatigue, cognitive impairment, and depression [22]. Research also indicates that the symptomatology of PASC can evolve over time, with variations in symptom clusters observed across different waves of the pandemic and about SARS-CoV-2 variants [23]. Additionally, the presence of symptoms like joint pain, chest discomfort, and hair loss points to the multisystemic nature of PASC [24,25]. In summary, Long COVID presents with a wide array of symptoms, predominantly fatigue, dyspnea, cognitive impairment, and psychiatric symptoms, with significant variability in symptom clusters among patients [26]. In order to highlight the natural history of Long COVID, by mean of an unsupervised machine learning method that used semantic similarity of phenotype data to stratify long COVID patients, a study identified six clusters of PASC patients, which differed with respect to pre-existing comorbidities and with severity of acute COVID disease [27].

In our study, the application of a machine learning method in order to analyze the population of patients hospitalized for severe COVID disease and who developed PASC, confirmed the high heterogeneity of symptoms. However, this heterogeneity does not allow for the identification of common treatment pathways, confirming the need to create diagnosis and treatment pathways focused on every single patient.

Our study population was limited in size and this aspect might hamper the identification of clearly separated clusters. However, the substantial homogeneity of the cohort, with all patients having been discharged after a severe Covid-19 infection, might justify the lack of distinct phenotypes.

5. Conclusion

The extreme heterogeneity of patients affected by PASC has not allowed the identification of specific symptom clusters with the most recent statistical techniques. The characteristics of the different cohorts of patients enrolled in previous studies may have been drivers for the emergence of cohort effects that make the results not generalisable.

In our experience, enrolling a large cohort of consecutive patients with severe acute COVID-19 does not allow to define specific clusters of residual symptoms at a 3-month follow-up that can generate differentiated diagnostic-therapeutic pathways.

References

1. National Institute for Health and Care Excellence (NICE). COVID-19 rapid guideline: managing the long-term effects of COVID-19. 2020 Dec 18.

2. Center for Disease Control and Prevention. COVID-19 Post-COVID Condition: Information for Healthcare Providers. 2022. vols. 1-16.

3. Thaweethai T, Jolley SE, Karlson EW, Levitan EB, Levy B, McComsey GA et al. Development of a Definition of Postacute Sequelae of SARS-CoV-2 Infection JAMA. 2023 Jun 13; 329(22): 1934–1946. doi: 10.1001/jama.2024.2984. PMID: 37278994; PMCID: PMC10214179.

4. Crook H, Raza S, Nowell J, Young M, Edison P. Long Covid-mechanisms, risk factors, and management. BMJ. 2021 Jul; 26: 374: n1648. doi: 10.1136/bmj.n1944. PMID: 34312178.

5. Astin R, Banerjee A, Baker MR, Dani M, Ford E, Hull JH et al. Long COVID: mechanisms, risk factors and recovery. Exp Physiol. 2023 Jan; 108(1): 12-27. doi: 10.1113/EP090802. Epub 2022 Nov 22. PMID: 36412084; PMCID: PMC10103775.

6. Najafi MB, Javanmard SH. Post-COVID-19 syndrome mechanisms, prevention and management. Int J Prev Med 2023; 14: 59. doi: 10.4103/ijpvm.ijpvm_508_21. PMID: 37351054; PMCID: PMC10284243.

7. van Kessel SAM, Olde Hartman TC, Lucassen PLBJ, van Jaarsveld CHM. Post-acute and long-COVID-19 symptoms in patients with mild diseases: a systematic review. Family Practice, 2022, 159–167. doi: 10.1093/fampra/cmab076. PMID: 34268556; PMCID: PMC8414057.

8. Fernández-de-Las-Peñas C, Palacios-Ceña D, Gómez-Mayordomo V, Florencio LL, Cuadrado ML, Plaza-Manzano G, Navarro-Santana M et al. Prevalence of post-COVID-19 symptoms in hospitalized and non-hospitalized COVID-19 survivors: a systematic review and meta-analysis. Eur J Intern Med. 2021; 92:55–70. doi: 10.1016/j.ejim.2021.06.009. Epub 2021 Jun 16. PMID: 34167876; PMCID: PMC8206636.

9. Davis H.E., McCorkell L., Vogel J.M., J. Topol E. Long COVID: major findings, mechanisms and recommendations Nat Rev Microbiol 2023, 21(3):133-146. doi: 10.1038/s41579-022-00846-2. PMID: 37069455; PMCID: PMC10408714.

10. 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 2023 Feb;51(1):61-69. doi: 10.1007/s15010-022-01822-x. Epub 2022 Apr 22. PMID: 35451721; PMCID: PMC9028890.

11. Kisiel MA, Lee S, Malmquist S, Rykatkin O, Holgert S, Janols H et al. Clustering Analysis Identified Three Long COVID Phenotypes and Their Association with General Health Status and Working Ability. Journal of Clinical Medicine. 2023;12(11):3617. doi: 10.3390/jcm12113617. PMID: 37297812; PMCID: PMC10253616.

12. 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. Nature Medicine, august 2022; vol 28: 1706–1714. doi: 10.1038/s41591-022-01909-w. Epub 2022 Jul 25. PMID: 35879616; PMCID: PMC9388369.

13. Seeßle J, Waterboer T, Hippchen T, Simon J, Kirchner M, Lim A et al. Persistent Symptoms in Adult Patients 1 Year After Coronavirus Disease 2019 (COVID-19): A Prospective Cohort Study. Clinical Infectious Diseases 2021; 20 (20):1–8. doi: 10.1093/cid/ciab611. PMID: 34223884; PMCID: PMC8394862.

14. Greenhalgh T, Sivan M, Delaney B, Evans R, Milne R. Long covid—an update for primary care. BMJ 2022;378:e072117

15. Sisó-Almirall A, Brito-Zerón P, Conangla Ferrín L, Kostov B, Moragas Moreno A, Mestres J et al. Long Covid-19: Proposed Primary Care Clinical Guidelines for Diagnosis and Disease Management. Int. J. Environ. Res. Public Health 2021,18, 4350. doi: 10.1136/bmj-2022-072117. PMID: 36137612.

16. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. 2023.

17. Tsuchida T, Yoshimura N, Ishizuka K, Katayama K, Inoue Y, Hirose M et al. Five cluster classifications of long COVID and their background factors: A cross-sectional study in Japan. Clin Exp Med. 2023 Nov;23(7):3663-3670. doi: 10.1007/s10238-023-01057-6. Epub 2023 Apr 7. PMID: 37027067; PMCID: PMC10081305.

11

18.   Bai F, Tomasoni D, Falcinella C, Barbanotti D, Castoldi R, Mulè G et al. Female gender is associated with long COVID syndrome: a prospective cohort study. Clin Microbiol Infect. 2022 Apr;28(4):611.e9-611.e16. doi: 10.1016/j.cmi.2021.11.002. Epub 2021 Nov 9. PMID: 34763058; PMCID: PMC8575536.

19.   Marchi M, Grenzi P, Serafini V, Capoccia F, Rossi F, Marrino P et al. Psychiatric symptoms in Long-COVID patients: a systematic review. Front Psychiatry. 2023 Jun 21;14:1138389. doi: 10.3389/fpsyt.2023.1138389. PMID: 37415689; PMCID: PMC10320160.

20.   Zakia H, Pradana K, Iskandar S. Risk factors for psychiatric symptoms in patients with long COVID: A systematic review. PLoS One. 2023 Apr 7;18(4):e0284075. doi: 10.1371/journal.pone.0284075. PMID: 37027455; PMCID: PMC10081737.

21.   Goldhaber NH, Kohn JN, Ogan WS, Sitapati A, Longhurst CA, Wang A et al. Deep Dive into the Long Haul: Analysis of Symptom Clusters and Risk Factors for Post-Acute Sequelae of COVID-19 to Inform Clinical Care. Int J Environ Res Public Health. 2022 Dec 15;19(24):16841. doi: 10.3390/ijerph192416841. PMID: 36554723; PMCID: PMC9778884.

22.   Kisiel MA, Lee S, Malmquist S, Rykatkin O, Holgert S, Janols H et al. Clustering Analysis Identified Three Long COVID Phenotypes and Their Association with General Health Status and Working Ability. J Clin Med. 2023 May 23;12(11):3617. doi: 10.3390/jcm12113617. PMID: 37297812; PMCID: PMC10253616.

23.   Perlis RH, Santillana M, Ognyanova K, Safarpour A, Lunz Trujillo K, Simonson MD et al. Prevalence and Correlates of Long COVID Symptoms Among US Adults. JAMA Netw Open. 2022 Oct 3;5(10):e2238804. doi: 10.1001/jamanetworkopen.2022.38804. PMID: 36301542; PMCID: PMC9614581.

24.   Chudzik M, Babicki M, Kapusta J, Kałuzińska-Kołat Ż, Kołat D, Jankowski P et al. Long-COVID Clinical Features and Risk Factors: A Retrospective Analysis of Patients from the STOP-COVID Registry of the PoLoCOV Study. Viruses. 2022 Aug 11;14(8):1755. doi: 10.3390/v14081755. PMID: 36016376; PMCID: PMC9415629.

25.   Szabo S, Zayachkivska O, Hussain A, Muller V. What is really 'Long COVID'? Inflammopharmacology. 2023 Apr;31(2):551-557. doi: 10.1007/s10787-023-01194-0. Epub 2023 Mar 25. PMID: 36964860; PMCID: PMC10039447.

26.   Ziauddeen N, Gurdasani D, O'Hara ME, Hastie C, Roderick P et al. Characteristics and impact of Long Covid: Findings from an online survey. PLoS One. 2022 Mar 8;17(3):e0264331. doi: 10.1371/journal.pone.0264331. PMID: 35259179; PMCID: PMC8903286.

27.   Reese JT, Blau H, Casiraghi E, Bergquist T, Loomba JJ, Callahan TJ et al. Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes. EBioMedicine. 2023 Jan; 87:104413. doi: 10.1016/j.ebiom.2022.104413. Epub 2022 Dec 21. PMID: 36563487; PMCID: PMC9769411.