



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Original Article

## Symptom-based clusters of hospitalized patients with severe acute respiratory illness by SARS-CoV-2 in Brazil



Leticia Martins Raposo<sup>a,1,\*</sup>, Gabriel Ferreira Diaz Abreu<sup>b</sup>, Felipe Borges de Medeiros Cardoso<sup>b</sup>, André Thiago Jonathas Alves<sup>c</sup>, Paulo Tadeu Cardozo Ribeiro Rosa<sup>d</sup>, Flávio Fonseca Nobre<sup>d</sup>

<sup>a</sup> Departamento de Métodos Quantitativos, Centro de Ciências Exatas e Tecnologia, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>b</sup> Escola de Medicina e Cirurgia, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>c</sup> Our Lady of Mercy School, Rio de Janeiro, Brazil

<sup>d</sup> Programa de Engenharia Biomédica, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia (COPPE), Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

## ARTICLE INFO

## Article history:

Received 22 February 2022

Received in revised form 13 April 2022

Accepted 21 April 2022

## Keywords:

COVID-19

Severity

Symptoms

Coronavirus Infections

Cluster Analysis

## ABSTRACT

**Background:** COVID-19 has shown a broad clinical spectrum, ranging from asymptomatic to mild, moderate, and severe infections. Many symptoms have already been identified as typical of COVID-19, but few studies show how they can be useful in identifying clusters of patients with different severity of illness. This interpretation may help to recognize the different profiles of symptoms of COVID-19 expressed in a population at certain time. The aim of this study was to identify symptom-based clusters of hospitalized patients with severe acute respiratory illness by SARS-CoV-2 in Brazil. The clusters were evaluated based on sociodemographic characteristics, admission to the Intensive Care Unit (ICU), use of respiratory support, and outcome.

**Methods:** The Multiple Correspondence Analysis (MCA)-based cluster analysis was applied to symptoms presented before admission. Pearson's chi-square test was used to compare the proportions of symptoms between the clusters and to examine differences in the calculated rates for the following variables: sex, age group, race, Brazilian region, use of respiratory support, admission to the ICU and outcome.

**Results:** Three COVID-19 clusters with distinct symptom profiles were identified by MCA-based cluster analysis. Cluster 1 had the mildest severity profile, with the lowest frequencies for most symptoms investigated. Cluster 2 had a severe respiratory profile, with the highest frequencies of patients with dyspnea, respiratory discomfort and O<sub>2</sub> saturation < 95%. Cluster 2 was also the most prevalent in all Brazilian regions and had the highest percentages of patients who used invasive respiratory support (27.4%) (p-value < 0.001), were admitted to the ICU (42.6%) (p-value < 0.001) and died (39.0%) (p-value < 0.001). Cluster 3 had a prominent profile of gastrointestinal symptoms.

**Conclusions:** The study identified three distinct COVID-19 clusters based on the symptoms presented by patients with severe acute respiratory illness by SARS-CoV-2, but without distinction in their prevalence in the Brazilian regions.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of King Saud Bin Abdulaziz University for Health Sciences.

CC\_BY\_NC\_ND\_4.0

\* Correspondence to: Departamento de Métodos Quantitativos, Centro de Ciências Exatas e Tecnologia, Universidade Federal do Estado do Rio de Janeiro, Av. Pasteur, 458, Rio de Janeiro, Brazil.

E-mail address: [leticia.raposo@uniriotec.br](mailto:leticia.raposo@uniriotec.br) (L.M. Raposo).

<sup>1</sup> ORCID: 0000-0003-0613-5582

## Introduction

The Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) is responsible for the Coronavirus Disease 2019 (COVID-19), which has a broad clinical spectrum, ranging from asymptomatic to mild, moderate, and severe (~ 15%) infections. About 5% of the cases can progress to critical disease with complications such as acute respiratory distress syndrome, thromboembolism, and/or multi-organ failure [1,2].

The most common symptoms of COVID-19 are fever, dry cough, and fatigue. Some patients experience loss of taste or smell, sore throat, headache, muscle or joint pain, nausea or vomiting, diarrhea, among others. Patients with severe COVID-19 may also present shortness of breath, loss of appetite, and confusion [1,3]. The most severe cases often require hospitalization in an Intensive Care Unit (ICU) to ensure assisted respiratory support and other medical treatments [4]. Comorbidities such as asthma, diabetes, cardiovascular and pulmonary diseases, and characteristics such as age, sex, and race have also been associated with the severity of the disease [5–9].

Many symptoms have already been identified as typical of COVID-19, but few studies show how they can be useful in identifying clusters of patients with different severity of illness. This interpretation may help to recognize different profiles of symptoms of COVID-19 expressed in a population at a certain time.

It is estimated that as of May 09, 2022, more than 517 million SARS-CoV-2 infections have been confirmed, resulting in 6,276,846 COVID-19-related deaths worldwide. Among active cases, 0.1% (39,764 of 38,891,845) are in severe or critical condition [10]. In Brazil, to date, 30,564,536 cases have been recorded, with 664,189 deaths. The country was in the third position among the countries in the world with the highest number of registered cases and second place concerning the number of deaths [10].

Brazil has a flu epidemiological surveillance system that has been recording cases of SARS-CoV-2 infections, including patient information such as sociodemographic characteristics, symptoms, comorbidities, hospitalization, evolution, among others. In this study, we sought to assess the symptoms presented in Brazilian patients with severe acute respiratory illness by SARS-CoV-2 and to identify severity profiles according to these epidemiological characteristics.

## Materials and methods

This study is based on the public database SIVEP-Gripe (*Sistema de Informação de Vigilância Epidemiológica da Gripe*—Flu Epidemiological Surveillance System) [11], a system for recording cases of severe acute respiratory illness (SARI) in Brazil, including data from COVID-19, maintained by the Ministry of Health (<https://opendatasus.saude.gov.br/dataset/bd-srag-2021>). The Brazilian Guide for Epidemiological Surveillance considered a case of SARI when a person with flu-like illness has dyspnea/respiratory discomfort, or persistent pressure, or pain in the chest, or O<sub>2</sub> saturation < 95% in room air, or a bluish discoloration (cyanosis) of the lips or face [12]. The original data used in this study can be accessed at <https://zenodo.org/record/5776398#.YbdFcdDMJPY>. The data was downloaded on November 10, 2021.

We included in the study patients aged ≥ 18 years, classified as a case of SARI, with a positive RT-PCR test for SARS-CoV-2, who were admitted to the hospital and with the known outcome as hospital discharge (cure) or death by COVID-19. A stratified random sample proportional to Brazilian states, with an approximate size of 10,000, was considered in the analysis. The date of COVID-19 diagnosis spans the time interval from February, 1 to October 31, 2021.

The variables included in the study were sociodemographic characteristics – Brazilian region (North; Northeast; Center-West; Southeast and South), sex (male and female), age group (18–59 years old; ≥ 60 years old), and race (white and non-white); signs and symptoms (fever, cough, sore throat, dyspnea, respiratory discomfort, O<sub>2</sub> saturation < 95%, diarrhea, vomiting, abdominal pain, fatigue, and loss of smell or taste); hospitalization (ICU admission, use of respiratory support—invasive, non-invasive and non-use); and outcome (cure or death by SARI). Missing values for symptoms were

considered as the absence of these characteristics, and missing values for ICU admissions were considered as non-ICU admissions [13].

A descriptive analysis was performed. Absolute and percentage frequencies were calculated to summarize the main characteristics of the participants in the study.

Multiple correspondence analysis (MCA) was used to explore the joint association between signs and symptoms. MCA enables the identification of associations among variables when considered simultaneously. This analysis allows a geometrical representation of the results by locating each variable category as a point on the Cartesian axes, according to the frequency of the modalities, so that, closer points indicate that the subjects share the corresponding categories, which finally leads to categories being associated. MCA was performed using the R package FactoMineR [14], and the factoextra package [15] was used to extract the MCA results.

The hierarchical clustering on principal components from MCA was used to identify groups of patients with similar profiles according to the symptoms. The Euclidean distance with Ward's agglomeration method was used. This step was performed by the HCPC function from the FactoMineR package, which uses the V test to compare the proportion of the categories in a cluster to the proportion of the categories in the population. The test is based on the hypergeometric distribution [16]. The algorithm determined the optimum number of clusters.

Pearson's chi-square test was used to compare the proportions of symptoms between the clusters and to examine differences in the rates of sex, age group, race, Brazilian region, use of respiratory support, admission to the ICU and outcome based on the clusters.

All analyzes were performed in R version 4.1.0 software [17]. A  $p < 0.05$  was considered statistically significant.

## Results

A total of 10,011 hospitalized Brazilian patients with SARI by SARS-CoV-2 confirmed by RT-PCR were included in the study. Most hospitalized patients were from the Southeast region (54.9%), adults (56.8%), male (56.4%), and white (58.4%). Among the symptoms, the most common were dyspnea (75.6%), O<sub>2</sub> saturation < 95% (72.9%), cough (70.6%), respiratory discomfort (59.9%) and fever (55.6%). Over 60% of patients used non-invasive respiratory support, 38.2% were admitted to ICU and 65.2% were cured (Table 1).

The first two dimensions captured 32.9% of the total variability (18.7% and 14.2% by the first and second dimensions, respectively). Five dimensions captured 60% of the total variability. Fatigue and respiratory discomfort were the most correlated with dimension 1. O<sub>2</sub> saturation < 95% and dyspnea were the most correlated with dimension 2, with respiratory discomfort also well correlated. Cough and fever were more correlated to dimension 3, fatigue, vomiting and loss of smell and taste to dimension 4, and sore throat to dimension 5.

Fig. 1 shows the categories of variables on the MCA map. The first component opposes the categories that represent the presence of symptoms (in the first and fourth quadrants) with the categories of the absence of symptoms (in the second and third quadrants). The second component distinguishes the categories that indicate the presence of abdominal pain, vomiting, diarrhea, loss of smell or taste, sore throat, fever, and cough; with the categories that indicate the presence of fatigue, dyspnea, O<sub>2</sub> saturation < 95%, and respiratory discomfort.

We observed proximity between abdominal pain, vomiting, diarrhea, and loss of smell or taste in the first quadrant, which are well represented in the factor map for both dimensions. Fever, cough, sore throat and fatigue were located close together in the

**Table 1**  
Epidemiological characteristics of the studied sample of Brazilian patients hospitalized with SARI by SARS-CoV-2 confirmed by RT-PCR.

Characteristic	N = 10,011 (%)
Brazilian region	
North	345 (3.4%)
Northeast	1054 (10.5%)
Center-West	913 (9.1%)
Southeast	5499 (54.9%)
South	2200 (22.0%)
Age group (years old)	
18 – 59	5684 (56.8%)
≥ 60	4327 (43.2%)
Sex	
Female	4363 (43.6%)
Male	5648 (56.4%)
Race	
White	5842 (58.4%)
Non-white	4169 (41.6%)
Symptoms	
Fever	5563 (55.6%)
Cough	7066 (70.6%)
Sore throat	1784 (17.8%)
Dyspnea	7564 (75.6%)
Respiratory discomfort	5998 (59.9%)
O2 saturation < 95%	7303 (72.9%)
Diarrhea	1377 (13.8%)
Vomiting	804 (8.0%)
Abdominal pain	597 (6.0%)
Fatigue	3059 (30.6%)
Loss of smell or taste	1331 (13.3%)
Respiratory support	
None	1501 (15.0%)
Non-invasive	6156 (61.5%)
Invasive	2354 (23.5%)
ICU	3828 (38.2%)
Outcome	
Cure	6531 (65.2%)
Death	3480 (34.8%)

right quadrant, with the first three categories in the first quadrant and the last one in the fourth quadrant. Respiratory distress, O2 saturation < 95%, and dyspnea were located close together in the fourth quadrant (Fig. 1).

Table 2 shows the three clusters obtained after MCA using Euclidean distance with Ward's agglomeration method. Cluster 1 has 2541 observations, Cluster 2 has 5849 observations and Cluster 3 1621 observations. The results are described according to their within-cluster membership i.e. the percentage of individuals within a cluster (class) who present the category (modality) (Mod.Cla), and by their across-cluster membership i.e. the percentage of all individuals belonging to a class (cluster) that selected the respective modality (category) (Cla.Mod). Only positive categories (presence of symptom) were included in Table 2. There were statistically significant differences between the percentages of the three clusters and in relation to the global proportion for all symptoms ( $p < 0.001$ ), except for cough, which showed no difference between global and Cluster 2 frequencies.

Cluster 1 showed the lowest percentages of individuals with a certain symptom when compared to the overall percentage ( $p < 0.001$ ). Comparing to the other clusters, Cluster 1 had the lowest percentages of individuals with fever (49.9%), cough (62.1%), sore throat (11.2%), fatigue (15.6%), dyspnea (31.4%), respiratory discomfort (9.3%), and O2 saturation < 95% (23.3%).

Cluster 2 had the largest number of patients and the highest percentages for respiratory symptoms such as dyspnea (92.5%), respiratory discomfort (78.1%), and O2 saturation < 95% (91.8%). Most

patients with these symptoms are in Cluster 2. On the other hand, it is also the cluster with the lowest percentages of gastrointestinal symptoms, such as abdominal pain (0.3%), vomiting (0.1%), diarrhea (2.7%), and loss of smell or taste (9.2%).

Cluster 3 had the highest percentages of the classical symptoms of a respiratory infection, such as fever (74.3%), cough (84.5%), and sore throat (34.5%). Also, high frequencies of gastrointestinal symptoms, such as abdominal pain (32.2%), vomiting (44.2%), diarrhea (62.1%), and loss of smell or taste (33.4%). Fatigue was greater in this group, being observed in 54.8% of patients. Respiratory symptoms were also high, but not as high as those in Cluster 2. About 83% had dyspnea, 73.6% respiratory discomfort and 82.7% had O2 saturation < 95%.

Table 3 shows the sociodemographic and hospitalization characteristics of patients in each cluster. It is possible to observe a statistically significant difference between the frequencies of the three clusters regarding the variables: age group, gender, respiratory support, ICU and outcome.

Cluster 3 had a higher proportion of patients aged 18–59 years (64.0%) when compared to clusters 1 and 2 ( $p$ -value < 0.001). For gender, Cluster 3 had a higher proportion of females (47.4%) when compared to the others ( $p$ -value < 0.001). Cluster 1 had a higher proportion of patients who did not need respiratory support (29.4%), while clusters 2 and 3 had higher proportions of non-invasive support (63.8% and 64.6%, respectively). Cluster 2 had the highest percentages of patients who needed invasive respiratory support (27.4%) ( $p$ -value < 0.001), were admitted to the ICU (42.6%) ( $p$ -value < 0.001) and died (39.0%) ( $p$ -value < 0.001).

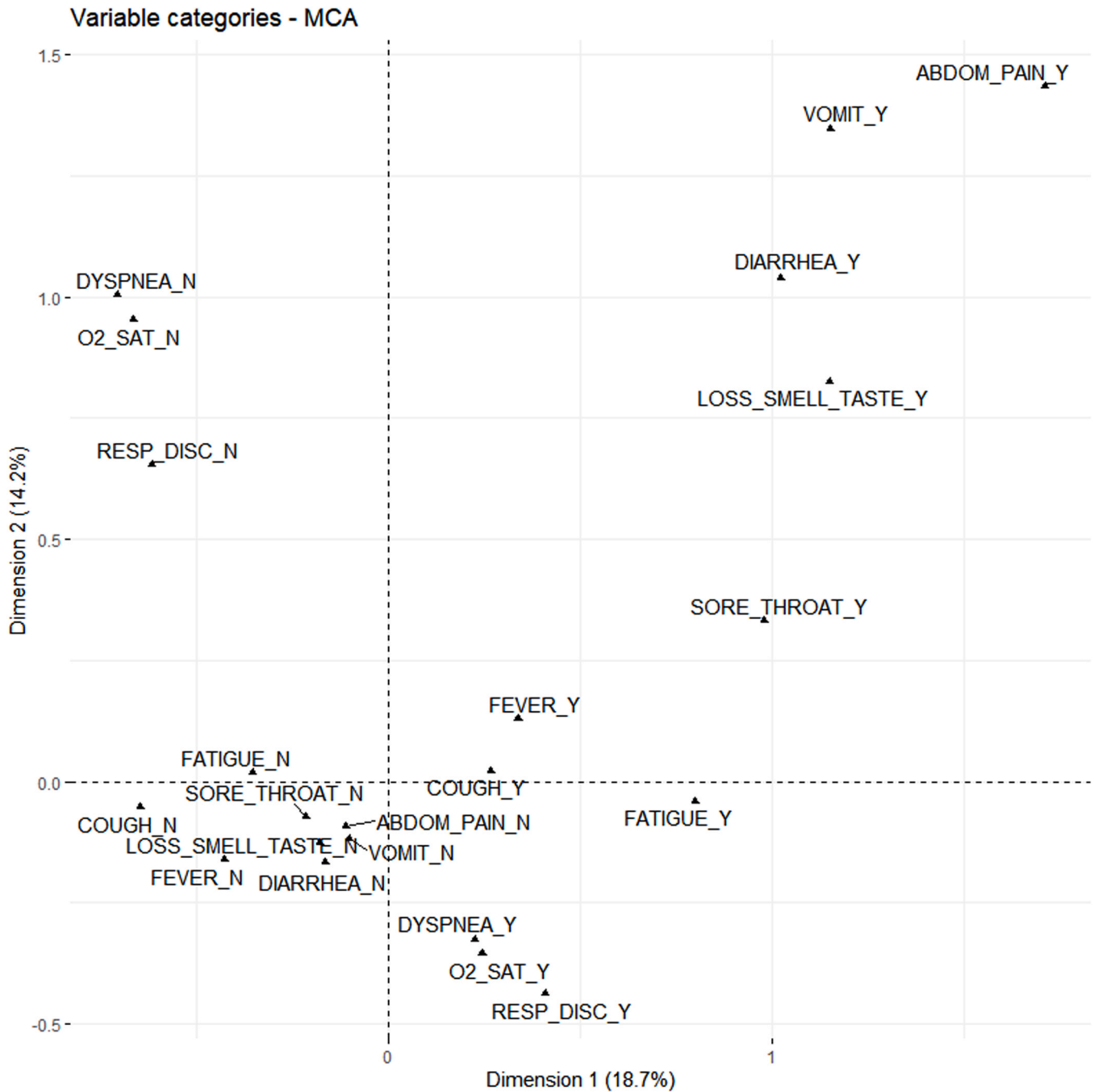
Fig. 2 shows the distribution of clusters for each Brazilian region. It is possible to observe that there were no differences between the frequencies of each cluster between the five units ( $p$ -values < 0.001). Cluster 2 was the most prevalent in all regions, with frequencies ranging from 57.7% in the Southeast region to 59.5% in the South region.

## Discussion

This study aimed to identify the profile of symptoms in the Brazilian COVID-19 cases using MCA combined with cluster analysis. We also evaluated how these profiles differed in terms of socio-demographic characteristics, ICU admission, use of respiratory support and outcome.

Three different clusters were identified according to the prevalence of symptoms. Cluster 1 had the lowest severity profile, with the lowest frequencies for most symptoms investigated. Cluster 2 showed a severe respiratory profile, with the highest frequency of patients with dyspnea, respiratory discomfort, and oxygenation. Cluster 3 presented a profile with a certain severity, which involved a strong profile of gastrointestinal symptoms. Such characteristics were not present in the other two clusters.

Similar to the present study, Han et al. also identified three clusters, labeled as “severe”, “classical” and “atypical” [18]. These clusters were based on the symptoms of 1035 COVID-19 patients from Wuhan, China. Unlike our study, they used laboratory data. Comparing the results, the “severe” cluster in the study by Han and colleagues had high frequencies of fever, fatigue, sore throat, dyspnea, diarrhea, abdominal pain, and vomiting. This profile is similar to our Cluster 3, which also presented high frequencies for these symptoms. The profile of Cluster 2 in our study is like the “classical” cluster, which presented intermediate frequencies compared to those of the “severe” and “atypical” clusters. The “atypical” cluster presented the lowest levels of all indicators among the three



**Fig. 1.** Multiple correspondence analysis map (projections on the first two dimensions) for the categories included in the analysis. The variables considered in this analysis were abdominal pain, cough, diarrhea, dyspnea, fatigue, fever, loss of smell or taste, O2 saturation < 95%, respiratory discomfort, sore throat, and vomiting. The letters “N” and “Y” represent, respectively, the negative (no) and positive (yes) categories.

clusters, being similar to our Cluster 1, being also the cluster with the lowest frequencies for most symptoms investigated. However, in the study conducted by Han et al., the severity of the clusters was also determined, in addition to the symptoms, by considering the presence of syndromes and laboratory findings. This study, on the other hand, only compares cluster profiles by symptoms.

Other studies have also identified groups of symptoms in COVID-19, although using other approaches, such as exploratory factor

analysis (EFA). Dixon et al. applied EFA to identify major groups of symptoms associated with SARS-CoV-2 infection [19]. The first group of symptoms comprised loss of smell, loss of taste, and fever. The second group: shortness of breath, cough, and chest pain. The third group: fatigue and myalgia. The fourth group: vomiting and diarrhea. Finally, the fifth group of symptoms comprised runny nose and sore throat. Observing the factorial map obtained with the MCA, the present study also observed proximity between the symptoms of

**Table 2**

Percentage of subjects within the cluster who present the symptom (Mod.Cla) and percentage of all subjects who present the symptom distributed across the clusters (Cla.Mod). Only items with a p-value less than 5% were included, as this shows that one category is significantly linked to the other categories.

	Cluster 1 (n = 2541)		Cluster 2 (n = 5849)		Cluster 3 (n = 1621)		Global (n = 10,011)
	Cla.Mod	Mod.Cla	Cla.Mod	Mod.Cla	Cla.Mod	Mod.Cla	
Fever	22.8	49.9	55.6	52.8	21.7	74.3	55.6
Cough	22.3	62.1	58.3	70.4 <sup>ns</sup>	19.4	84.5	70.6
Sore throat	16.0	11.2	52.6	16.1	31.4	34.5	17.8
Abdominal pain	9.9	2.3	2.7	0.3	87.4	32.2	6.0
Vomiting	9.8	3.1	1.0	0.1	89.2	44.2	8.0
Diarrhea	15.5	8.4	11.3	2.7	73.1	62.1	13.8
Loss of smell or taste	19.0	10.0	40.3	9.2	40.6	33.4	13.3
Fatigue	13.0	15.6	44.1	26.8	29.0	54.8	30.6
Dyspnea	10.6	31.4	71.6	92.5	17.9	83.4	75.6
Respiratory discomfort	3.9	9.3	76.2	78.1	19.9	73.6	59.9
O2 saturation < 95%	8.1	23.3	73.5	91.8	18.4	82.7	73.0

ns: there was no significant difference regarding the global percentage.

**Table 3**

Characteristics of the three COVID-19 clusters identified by hierarchical clustering on principal components from multiple correspondence analysis.

Characteristic	Cluster 1 N = 2541 (%)	Cluster 2 N = 5849 (%)	Cluster 3 N = 1621 (%)	p-value <sup>1</sup>
Brazilian region				0.327
N-NE	366 (14.4)	825 (14.1)	208 (12.8)	
CW-S-SE	2175 (85.6)	5024 (85.9)	1413 (87.2)	
Age group (years old)				< 0.001
18 – 59	1392 (54.8)	3255 (55.7)	1037 (64.0)	
≥ 60	1149 (45.2)	2594 (44.3)	584 (36.0)	
Sex				< 0.001
Female	1125 (44.3)	2469 (42.2)	769 (47.4)	
Male	1416 (55.7)	3380 (57.8)	852 (52.6)	
Race				0.601
Non-white	1039 (40.9)	2459 (42.0)	671 (41.4)	
White	1502 (59.1)	3390 (58.0)	950 (58.6)	
Respiratory support				< 0.001
None	746 (29.4)	515 (8.8)	240 (14.8)	
Non-invasive	1379 (54.3)	3730 (63.8)	1047 (64.6)	
Invasive	416 (16.4)	1604 (27.4)	334 (20.6)	
ICU				< 0.001
No	1729 (68.0)	3356 (57.4)	1098 (67.7)	
Yes	812 (32.0)	2493 (42.6)	523 (32.3)	
Outcome				< 0.001
Cure	1836 (72.3)	3566 (61.0)	1129 (69.6)	
Death	705 (27.7)	2283 (39.0)	492 (30.4)	

Categorical variables are expressed as counts and percentages (%) and the Pearson's Chi-squared test was applied.

diarrhea and vomiting. However, the loss of smell and taste was closer to the symptoms previously mentioned than with fever, different from what was found in [19].

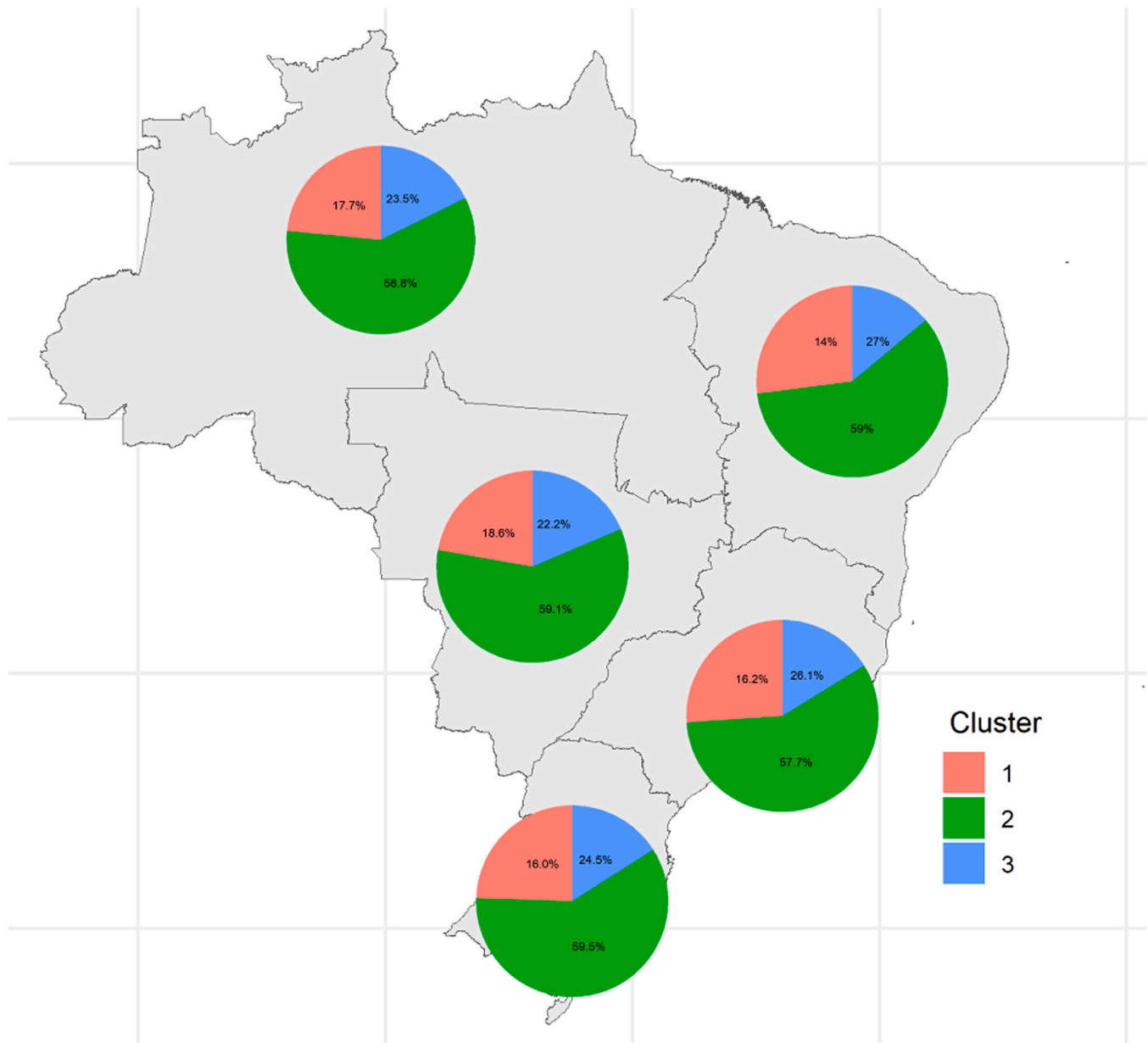
Luo et al. also used factor analysis to identify symptom clusters in COVID-19 patients [20]. Five factors were also identified, as in Dixon et al. The first group of symptoms comprised fever, sore throat, diarrhea, palpitation and chest tightness. The second group: dizziness and headache. The third group: cough, expectoration, and wheezing. The fourth group: dry mouth and bitter taste in the mouth. The fifth group of symptoms comprised poor appetite and fatigue. Many symptoms assessed by Luo and colleagues were not present in our database, but we also observed a correlation between fever and sore throat.

Sudre et al. identified six distinct groups of COVID-19 from symptom data collected in a regular recording app, where users could daily record their health and any potential new symptoms of COVID-19. They applied an unsupervised time series clustering over symptoms. The identified groups were classified into “Flu-like” with no fever; “Flu-like” with fever; gastrointestinal; severe level one, fatigue; severe level two, confusion; severe level three, abdominal and respiratory symptoms [21,22]. We observed that Cluster 3 identified in the present study is similar to the Severe level three, abdominal and respiratory cluster, with high percentages for symptoms of these two classes.

When the symptoms clusters were evaluated, Cluster 2 was shown to be the most severe. It had the highest percentages of ICU admission, use of invasive respiratory support, and death. It was the one with the highest frequencies of dyspnea, respiratory discomfort, and O2 saturation < 95%. This shows that these symptoms make the patient's profile more critical, setting alerts for monitoring the disease. Clusters 1 and 3 did not show significant differences regarding the proportions of ICU admission and death. Cluster 3 had higher percentages of use of respiratory support when compared to Cluster 1. Thus, we can classify Cluster 3 as moderate severity and Cluster 1 with a mild profile.

When the clusters were evaluated, based on sociodemographic characteristics, no differences were observed between them for race and Brazilian regions. In fact, for all Brazilian regions, cluster 2 was the most prevalent, followed by clusters 1 and 3. Clusters 1 and 2 had higher percentages of males and elderly (>= 60 years).

Our study had some limitations. The use of secondary data does not guarantee that all patients who presented SARI by SARS-CoV-2 in different Brazilian regions were notified by the surveillance system, since not everyone was tested and/or hospitalized. To ease this situation, we only used data from 2021, when access to RT-PCR exams was greater. It is also noteworthy that the cases evaluated were only patients hospitalized with SARI by SARS-CoV-2. Thus, the profiles identified here refer to cases with some degree of severity, not considering asymptomatic or symptomatic cases of people infected with SARS-CoV-2, and who did not require hospitalization. Finally, although the vaccination campaign against COVID-19 started in 2021 in Brazil, in this study we did not assess differences in symptom profiles between vaccinated and unvaccinated hospitalized patients. The incompleteness of vaccination data and the use of vaccines (Pfizer – BioNTech, Oxford – AstraZeneca, CoronaVac, and Janssen) with different schedules and combinations require deeper analysis that will be the topic of further studies. It is also important to



**Fig. 2.** Distribution of clusters for each of the five Brazilian regions. Cluster 2 was the most prevalent in all regions.

highlight that in the database used there was no information about which strain of SARS-CoV-2 the patient was infected with.

### Conclusions

In our study, Brazilian hospitalized patients with SARI by SARS-CoV-2 were divided into three clusters with different symptom profiles using MCA-based cluster analysis. We identified three symptom-based clusters: milder cases, moderate cases with gastrointestinal symptoms and more severe cases with respiratory symptoms. It was not possible to observe differences in the proportion of clusters between the five Brazilian regions, with the most severe cluster being the most prevalent in all regions.

### Funding

G. Abreu and F. Cardoso received scientific initiation grants from Universidade Federal do Estado do Rio de Janeiro and L. Raposo received a grant from Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

### Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- [1] WHO. Coronavirus disease (COVID-19) 2020. (<https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19>) (Accessed 26 October, 2020).
- [2] Organization WH. Clinical management of COVID-19: interim guidance 2020. World Health Organization; 27 2020.
- [3] Pascarella G, Strumia A, Piliago C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: a comprehensive review. *J Intern Med* 2020;288:192–206. <https://doi.org/10.1111/joim.13091>
- [4] García LF. Immune response, inflammation, and the clinical spectrum of COVID-19. *Front Immunol* 2020;11:1441. <https://doi.org/10.3389/fimmu.2020.01441>
- [5] Menezes Soares R, de C, Mattos LR, Raposo LM. Risk factors for hospitalization and mortality due to COVID-19 in Espírito Santo State, Brazil. *Am J Trop Med Hyg* 2020;103:1184–90. <https://doi.org/10.4269/ajtmh.20-0483>
- [6] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA J Am Med Assoc* 2020;323:1239–42. <https://doi.org/10.1001/jama.2020.2648>
- [7] Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in coronavirus disease 2019 patients: a systematic review and meta-analysis. *Int J Infect Dis* 2020;94:91–5. <https://doi.org/10.1016/j.ijid.2020.03.017>
- [8] Wolff D, Nee S, Hickey NS, Marscholke M. Risk factors for Covid-19 severity and fatality: a structured literature review. *Infection* 2020;1:1. <https://doi.org/10.1007/s15010-020-01509-1>
- [9] Kaeuffer C, Hyaric Cle, Fabacher T, Mootien J, Dervieux B, Ruch Y, et al. Clinical characteristics and risk factors associated with severe COVID-19: prospective analysis of 1,045 hospitalised cases in North-Eastern France, March 2020. *Eurosurveillance* 2020;25:2000895 <https://doi.org/10.2807/1560-7917.ES.2020.25.48.2000895>
- [10] Worldmeter. Covid-19 coronavirus pandemic 2022. (<https://www.worldometers.info/coronavirus/>) (Accessed 09 May, 2022).
- [11] Ministério da Saúde. SRAG 2021 - Banco de Dados de Síndrome Respiratória Aguda Grave - incluindo dados da COVID-19 2021. (<https://opendatasus.saude.gov.br/dataset/bd-srag-2021>) (Accessed 17 November, 2021).
- [12] Ministério da Saúde. Guia de vigilância epidemiológica - Emergência de saúde pública de importância nacional pela doença pelo Coronavírus 2019. Brasília: 2020.
- [13] Baqui P, Bica I, Marra V, Ercole A, van der Schaar M. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *Lancet Glob Heal* 2020;8:e1018–26. [https://doi.org/10.1016/S2214-109X\(20\)30285-0](https://doi.org/10.1016/S2214-109X(20)30285-0)
- [14] Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. *J Stat Softw* 2008. <https://doi.org/10.18637/jss.v025.i01>
- [15] Kassambara A. factoextra: Extract and Visualize the Results of Multivariate Data Analyses 2016, pp. 1–74.
- [16] Husson F, Lê S., Pagès J. Exploratory multivariate analysis by example using R. 2010. <https://doi.org/10.1201/b10345>.
- [17] R Development Core Team. R Software. *R A Lang Environ Stat Comput* 2021.
- [18] Han L, Shen P, Yan J, Huang Y, Ba X, Lin W, et al. Exploring the clinical characteristics of COVID-19 clusters identified using factor analysis of mixed data-based cluster analysis. *Front Med* 2021;8:1122. <https://doi.org/10.3389/FMED.2021.644724/BIBTEX>
- [19] Dixon BE, Wools-Kaloustian K, Fadel WF, Duszynski TJ, Yiannoutsos C, Halverson PK, et al. Symptoms and symptom clusters associated with SARS-CoV-2 infection in community-based populations: results from a statewide epidemiological study. *MedRxiv* 2020. <https://doi.org/10.1101/2020.10.11.20210922>
- [20] Luo Y, Wu J, Lu J, Xu X, Long W, Yan G, et al. Investigation of COVID-19-related symptoms based on factor analysis. *Ann Palliat Med* 2020;9. <https://doi.org/10.21037/APM-20-1113>
- [21] Sudre CH, Lee KA, Lochlainn MN, Varsavsky T, Murray B, Graham MS, et al. Symptom clusters in COVID-19: a potential clinical prediction tool from the COVID symptom study app. *Sci Adv* 2021;7:4177. <https://doi.org/10.1126/SCIADV.ABD4177>
- [22] Wise J. Covid-19: study reveals six clusters of symptoms that could be used as a clinical prediction tool. *BMJ* 2020;370:m2911. <https://doi.org/10.1136/BMJ.M2911>