

## Latent class analysis of COVID-19 symptoms in Brazil: results of the PNAD-COVID19 survey

Análises de classes latentes dos sintomas relacionados à COVID-19 no Brasil: resultados da PNAD-COVID19

Análisis de clases latentes de síntomas relacionados con la COVID-19 en Brasil: resultados de la PNAD-COVID19

Rafael da Silveira Moreira <sup>1,2</sup>

doi: 10.1590/0102-311X00238420

### Abstract

*The lack of mass testing for COVID-19 diagnosis creates the need to determine the magnitude of the disease based on its clinical symptoms. The study aimed to analyze the profile of COVID-19 symptoms and related aspects in Brazil. The author analyzed the sample of participants from the Brazilian National Household Sample Survey (PNAD-COVID19) conducted in May 2020. Latent class analysis (LCA) was performed with sociodemographic covariables and 11 symptoms reported by 346,181 individuals. Rao-Scott test and standardized residual analysis were used to measure the association with the pattern of health services use. Spatial scan analysis was performed to identify areas at risk of COVID-19 cases. LCA showed six classes of symptoms based on the pattern of answers by participants: (1) all the symptoms; (2) high prevalence of symptoms; (3) predominance of fever; (4) predominance of cough/sore throat; (5) mild symptoms with predominance of headache; and (6) absence of symptoms. Female sex, brown skin color, the North and Northeast regions of Brazil, and all three older age brackets showed stronger association with the class with all the symptoms (class 1). Most use of health services was also by this group of individuals, but with different profiles of use. Spatial analysis showed juxtaposition of this class with areas at greater risk of COVID-19. These findings underline the importance of investigating symptoms for the epidemiological identification of possible cases in a scenario with low population testing rates.*

COVID-19; Latent Class Analysis; Pandemics; Epidemiological Monitoring

### Correspondence

R. S. Moreira

Instituto Aggeu Magalhães, Fundação Oswaldo Cruz.  
Av. Professor Moraes Rego s/n, Cidade Universitária, Recife, PE  
50670-420, Brasil.  
rafael.moreira@fiocruz.br

<sup>1</sup> Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Brasil.

<sup>2</sup> Faculdade de Medicina, Universidade Federal de Pernambuco, Recife, Brasil.



## Introduction

The pandemic caused by the novel coronavirus and called COVID-19 (corona virus disease, identified in 2019), first reported in December 2019 in Wuhan, China <sup>1,2</sup>, emerged in association with severe forms of pneumonia and high transmissibility. Patients generally present shortness of breath, dry cough, fever, headache, and dyspnea. Fatal cases usually involve progressive respiratory failure with severe lung damage <sup>3,4,5</sup>.

The first case in Brazil was identified on February 26, 2020, and the first death was reported on March 17, 2020. By August 5, 2020, just over five months since the first case, there had been a total of 2,859,073 cases and 97,256 deaths (Brazilian Ministry of Health. <https://covid.saude.gov.br/>, accessed on 05/Aug/2020). With models showing exponential progression, containment measures include testing more cases, enhanced hygiene, and social isolation <sup>6</sup>. Although there are cases of asymptomatic infection <sup>7,8,9</sup>, the demand for health services is generally accompanied by complaints of symptoms associated with the disease. The mean incubation period has been reported at 5.2 days <sup>10</sup>.

Meanwhile, the circulation of asymptomatic cases can mean an increase in the infection rate. Considering the time span from onset of the disease to death, or some 6 to 41 days, with a median of 14 days <sup>11</sup>, early identification of suspected cases is a crucial window of opportunity for better management of the epidemic, both pharmacological and nonpharmacological.

Countries that have achieved success in controlling the epidemic have shown high population testing rates, follow-up of cases and contacts, and greater political and social engagement in the maintenance of social distancing measures. Mass testing also allows identifies the infections that are still at the subclinical level <sup>10,12,13</sup>. Low and middle-income countries with difficulties in the implementation of mass testing must prioritize screening based on definition of clinical cases or presumptive diagnosis <sup>14</sup>.

Brazil presents a much lower testing rate than expected for adequate control of the epidemic, relaxing of social distancing measures due to economic pressures, and instabilities in the coordination of public health measures. The epidemic has thus surged at different paces on various regional geographic scales. This inequality mirrors Brazil's social iniquity, and the pandemic struck a country already debilitated by low economic growth and a public sector (science, education, and health) weakened by cutbacks in public policy investments, resulting from the neoliberal fiscal austerity model, a reality shared by other countries of Latin America <sup>15</sup>.

In the absence of mass testing, knowledge of notified cases generally occurs when there is a search by health services, when these have the means for confirmation of suspected cases. Another form of notification occurs with the confirmation of underlying cause of death. In both situations, notification is limited to patients that appeared for treatment (and obtained access to it), characterizing a scenario prone to the collapse of health systems for failing to display a strategy for active epidemiological surveillance of suspected cases.

The Brazilian Institute of Geography and Statistics (IBGE) has systematically conducted a specific *Brazilian National Household Sample Survey* (PNAD) for COVID-19, called PNAD-COVID19 <sup>16</sup>. Participants answer questions on self-reported symptoms related to COVID-19, the profile of health services use, and their work market status. Given the low mass testing in Brazil, it is essential to determine the pattern of COVID-19 symptoms. The way this pattern appears in the population, plus scientific knowledge on associated factors, can enhance the support for management focused on patients with higher likelihood of positive confirmation of the disease. This should also foster rational and epidemiologically directed use of ICU beds and mechanical ventilators <sup>17</sup>, besides allowing better detection, diagnosis, and timely monitoring of suspected cases. Within this scope, the study aimed to elucidate the pattern of COVID-19 symptoms and the main associated socio-spatial factors.

## Methodology

A cross-sectional analysis was performed with data from PNAD-COVID19 in May 2020. In this sample, questions were collected on COVID-19 symptoms and the profile of health services use related to these symptoms and the repercussions on the work market. Details on the sampling plan can be found on the IBGE <sup>16</sup>. The PNAD COVID19 survey was based on the sample of households from the Permanent PNAD, first quarter of 2019. A linkage was performed for integration with other databases in order to obtain telephone numbers for each household. This allowed identifying a sample with at least one available telephone in each of 193,662 households. These data represent 92% of the basic sample, distributed across sets of some 48 thousand households per week. This sample is fixed, in the sense that the households interviewed in the first month of data collection remained in the sample in the subsequent months until the end of the survey. All the residents in the selected households were invited to answer the survey, conducted by computer-assisted telephone interviewing (CATI).

The study considered data on symptoms, sociodemographic profile, and use of health services. The questions on symptoms referred to the presence in the reference week (week anterior to the interview) of 11 symptoms: fever, cough, sore throat, shortness of breath, headache, chest pain, nausea, stuffy or runny nose, fatigue, sore eyes, and loss of sense of smell or taste.

In relation to the profile of health services use, the survey asked whether household members had gone to health services and which measures had been taken (stayed at home, called a healthcare worker, self-medication, prescribed medication, or received a visit from a health worker from the SUS or private services). The following healthcare services were listed: basic healthcare unit (UBS in Portuguese) or family health team, urgent care unit (UPA in Portuguese), hospital of the Brazilian Unified National Health System (SUS), private physician's office, private urgent care service, or private hospital. Finally, the survey asked about hospitalization for COVID-19.

Sociodemographic data were sex, skin color, schooling, age, and major geographic region. In order to spatially contextualize COVID-19 cases with symptoms, the survey used the number of accumulated confirmed cases as of June 2020 per 100,000 inhabitants according to Health Regions and states of the country. The basis was the number of accumulated cases as of May 31, 2020, provided by the Ministry of Health (<https://covid.saude.gov.br/>, accessed on 05/Aug/2020).

Latent class analysis (LCA) with covariables was used to establish the profile of COVID-19 symptoms (dependent variable). This statistical procedure aims to group individuals according to similar patterns of answers, modeled with covariables. This forms classes with greater intraclass homogeneity and greater interclass heterogeneity. Latent classes were generated from the 11 symptoms self-reported by participants in the PNAD-COVID19 survey. However, to guarantee local independence between the indicator variables, it was necessary to perform three groupings. These were: (1) cough with sore throat; (2) shortness of breath with chest pain; and (3) nausea with fatigue. We thus ended up with 8 indicator variables rather than the initial 11 variables.

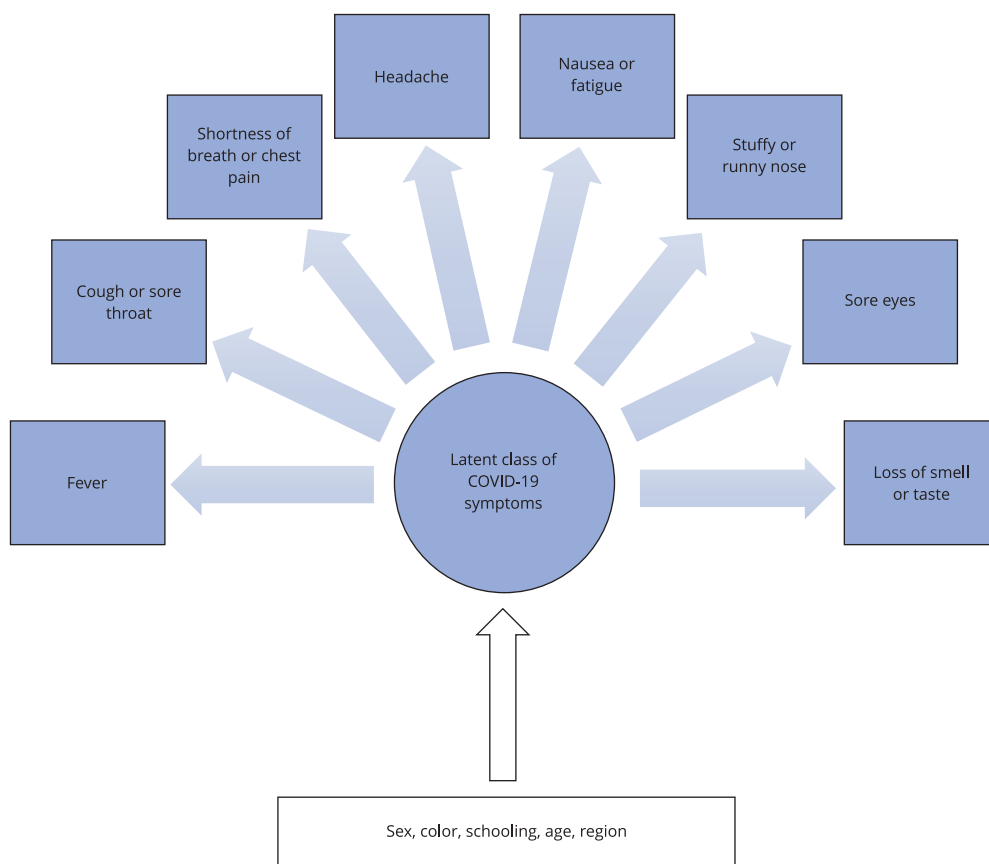
In the construction of this latent variable, models with different numbers of latent classes (categories) were created and tested until finding the ideal model to describe this variable. The following criteria were observed in this choice: Akaike information criterion (AIC), Bayesian information criterion (BIC), and adjusted Bayesian information criterion (adjusted BIC), always observing the lowest values when comparing the current model with the previous one. The highest entropy value was also considered <sup>18</sup>. Besides these criteria, two other statistical tests were performed (Vuong-Lo-Mendell-Rubin and Lo-Mendell-Rubin) to verify whether the chosen number of classes was better in terms of the model's fit when compared to the number of classes in the previous model.

To test the association between latent classes and independent socioeconomic variables, we estimated simple and multiple multinomial logistic regression models, with odds ratio (OR) as the measure of effect and its respective 95% confidence intervals (95%CI). The reference category was the latent class without symptoms. Estimation in the LCA models with sociodemographic covariables used the Mplus 6.12 (<https://www.statmodel.com/>) package, considering the characteristics of the complex sample design. Level of significance was set at 5%. Figure 1 illustrates the model representing the latent class analysis with covariables.

Analysis of the association with the profile of health services use was performed with Rao-Scott tests <sup>19</sup>, an adaptation of the chi-square test to complex samples. Standardized residual analysis (SRA)

**Figure 1**

Latent class analysis (LCA) model with covariables.



was performed to identify excess observations in each category of the dependent variable with the categories of the independent variables. Standardized residuals were considered significant when they were greater than 1.96 in a one-tailed test with 2.5% significance. SRA consists of examining the residuals (differences between observed and expected values in contingency tables) in standardized form, that is, expressed as units of standard deviation. In this sense, in the distribution of probabilities of occurrence, standardized residuals higher than 1.96 or lower than -1.96 have low odds of occurrence ( $\pm 2,5\%$ )<sup>20,21</sup>.

In complex samples, the standardized residuals can be quite large, since the standard errors produced without Rao-Scott correction tend to be smaller than the standard errors with the proper correction. Thus, the SPSS for Windows version 20 (<https://www.ibm.com/>) offers the option of analysis of crossed tables in complex samples. With this option, both Rao-Scott correction of the chi-square test and calculation of standardized residuals are adjusted for the complex sampling plan.

Non-incorporation of variables in the profile of health services use in the estimation of latent classes (LCA with covariables) can produce biases due to measurement errors. Still, according to Wang & Wang<sup>22</sup>, the use of LCA (without covariables) mixed with other databases, with the aid of other packages like SPSS, can be executed when the most likely model for class belonging is adequate, that is, when entropy is high (greater than 0.80). In our case, entropy with six classes was 0.928. So,

even assuming a possible limitation due to measurement bias, high entropy allows an analysis of classes via other alternatives besides the model with covariables. Another important issue was that due to the high number of individuals that did not seek health services, especially in the absence of symptoms, we preferred to conduct only a bivariate analysis with contingency tables, since the inclusion of these variables in a multiple multinomial logistic regression model would violate both the model's parsimony and the absence of collinearity between these independent variables.

All analyses included the complex sample's parameters. Since this was a large sample, sampling design should be considered in the analysis of statistical significance in the tests and resulting estimates.

Spatial analysis included scan test with confirmed COVID-19 cases. The maximum circle allowed by the program is 50% of the population at risk. However, this high proportion would include half of the Brazilian population, hindering the identification of areas at lower risk. Thus, the size of the spatial window was based on the value indicated on the results of the Gini index in SaTScan (<https://www.satscan.org/>), provided after a first standard spatial scan (50%), generating an ideal value of 2% of the population at risk. When the analysis finds a circle with relative risk greater than the risk outside the circle, Monte Carlo type replicates are processed to calculate the p-value, considering 5% significance<sup>23</sup>. Thus, a thematic map was built, presenting the number of cases per 100,000 inhabitants according to quintile, proportional distribution of latent classes, and areas with relative risk with statistical significance.

Since the data are publicly available on the Internet and with no form of individual identification, the study did not require submission to the Institutional Review Board for research in human subjects.

## Results

Figure 2 illustrates the unconditional distribution of latent classes and the conditional distribution of affirmative answers to each of the eight symptoms according to latent class analysis classification. The model that displayed the best fit was with six classes. They were classified as: (1) all the symptoms; (2) high prevalence of symptoms; (3) predominance of fever; (4) predominance of cough/sore throat; (5) mild symptoms with predominance of headache; and (6) absence of symptoms.

**Figure 2**

Distribution of unconditional probabilities of latent classes and conditional probabilities of eight symptoms related to COVID-19 according to six latent classes. Brazil, 2020.

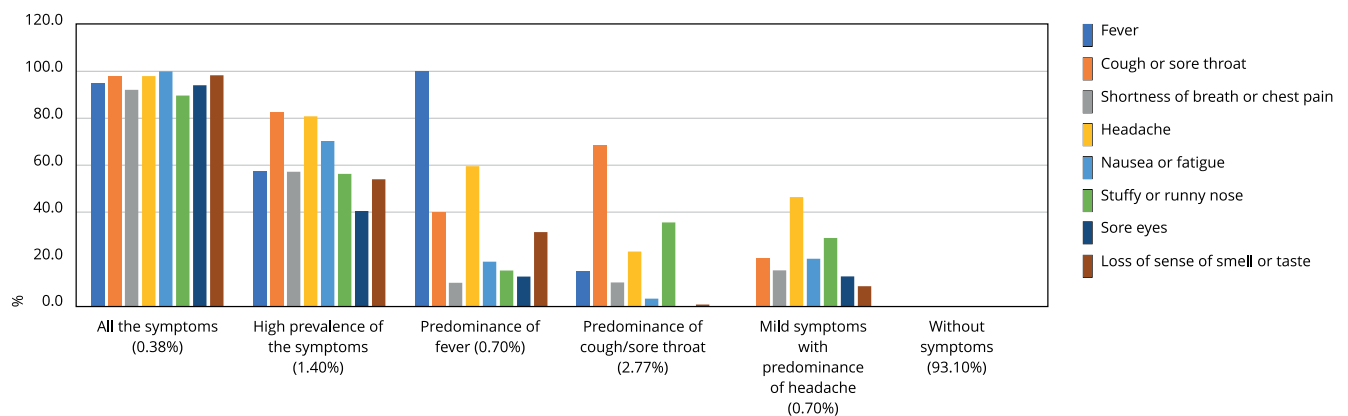


Table 1 presents the results of the LCA multiple multinomial logistic regression model with sociodemographic covariables. Females showed higher odds of being classified in the classes with more symptoms. Females showed 31% higher odds of presenting all the symptoms (class 1). Brown color showed 34% higher odds of presenting all the symptoms, while black color presented 55% higher odds of predominance of fever, when compared to white color. Asian-descendant and indigenous individual presented 80% higher odds of high prevalence of symptoms (class 2).

Categories with more schooling showed higher odds of having all the symptoms (class 1), high prevalence of symptoms (class 2), and predominance of cough/sore throat (class 4). The same was true for all quartiles of more advanced age in all the classes, except for the class of mild symptoms (class 5). Compared to the Southeast region, the North and Northeast of Brazil presented high odds of belong-

**Table 1**

Multiple multinomial logistic regression model with estimates of odds ratios (OR) and 95% confidence intervals (95%CI) for latent class analysis with sociodemographic covariables, Brazil, 2020.

	All the symptoms		High prevalence of the symptoms		Predominance of fever		Predominance of cough/sore throat		Mild symptoms with predominance of headache	
	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI	OR	95%CI
Sex										
Male	1.00		1.00		1.00		1.00		1.00	
Female	1.31	1.13-1.53	1.10	0.97-1.25	2.13	1.80-2.51	2.11	1.91-2.35	0.75	0.61-0.93
Color										
White	1.00		1.00		1.00		1.00		1.00	
Black	1.19	0.92-1.53	1.10	0.91-1.34	1.55	1.24-1.93	1.15	0.98-1.34	1.06	0.80-1.39
Asian-descendant and Indigenous	1.43	0.95-2.15	1.80	1.08-3.01	1.02	0.50-2.05	1.33	0.98-1.81	1.47	0.94-2.28
Brown	1.34	1.13-1.58	1.16	1.02-1.31	1.20	1.04-1.39	1.06	0.97-1.15	1.03	0.91-1.18
Schooling										
None	1.00		1.00		1.00		1.00		1.00	
Incomplete primary	2.01	1.41-2.86	1.32	0.99-1.76	1.26	0.84-1.87	1.97	1.40-2.87	0.89	0.67-1.18
Primary and secondary	3.22	2.28-4.54	1.84	1.33-2.55	1.36	0.66-2.79	2.14	1.45-3.15	0.84	0.60-1.18
University and graduate	3.18	2.23-4.52	1.24	0.85-1.81	1.65	0.82-3.30	3.04	2.01-4.61	1.27	0.71-2.26
Age										
Quartile 1 (1 years and under)	1.00		1.00		1.00		1.00		1.00	
Quartile 2 (19-36 years)	2.88	2.27-3.66	1.55	1.28-1.87	7.07	2.52-19.86	5.30	3.73-7.54	0.92	0.65-1.29
Quartile 3 (37-53 years)	3.64	2.90-4.57	1.73	1.38-2.17	9.64	3.34-27.84	6.01	4.12-8.78	0.71	0.43-1.16
Quartile 4 (53 years and older)	2.23	1.71-2.91	1.17	0.95-1.44	9.63	2.33-39.73	3.63	2.27-5.81	1.25	0.65-2.38
Region										
Southeast	1.00		1.00		1.00		1.00		1.00	
North	8.35	5.21-13.39	9.25	5.79-14.78	1.27	0.69-2.36	1.04	0.86-1.25	0.54	0.27-1.06
Northeast	2.30	1.69-3.14	5.51	3.68-8.25	0.83	0.45-1.53	1.09	0.95-1.24	0.63	0.43-0.92
South	0.27	0.13-0.57	0.29	0.13-0.64	1.11	0.73-1.69	1.14	1.01-1.29	0.90	0.73-1.11
Central	0.34	0.15-0.76	0.73	0.48-1.11	0.69	0.54-0.87	0.82	0.69-0.97	0.51	0.38-0.68

ing to the classes with more symptoms (classes 1 and 2). The South of Brazil showed 14% higher odds of belonging to the class with predominance of cough/sore throat (class 4).

Table 2 presents the distribution of the relative frequency of latent classes according to the profile of health services use. Health services use was associated with the prevalence of the latent classes of all the symptoms (class 1), high prevalence of symptoms (class 2), and predominance of fever (class 3). Except for classes 5 (mild symptoms with predominance of headache) and 6 (absence of symptoms), all the individuals classified in the other classes stayed at home. However, the acts of calling a healthcare worker, taking medication (prescribed or self-medicated), and receiving a visit from a private healthcare worker were associated with the classes with the most severe symptoms. Receiving a visit from a healthcare worker from the SUS was associated with both having all the symptoms and mild symptoms with predominance of headache. The most common healthcare service for individuals with all the symptoms was the UPA under the SUS. For individuals with higher prevalence of cough and sore throat, the UBS was the most frequently used service. Hospitals, both public and private, were not the priority services for COVID-19 care, nor were private physicians' offices. Private urgent care services were used by individuals with high prevalence of symptoms. There was an excess of individuals in class 1, with all the symptoms, who were hospitalized or attempted hospital admission but were unsuccessful.

**Table 2**

Percent distributions of latent classes of COVID-19 symptoms according to profile of health services use. Brazil, 2020.

Variables	All the symptoms (n = 1,315; 0.38%)	High prevalence of symptoms (n = 4,204; 1.21%)	Predominance of fever (n = 2,391; 0.70%)	Predominance of cough/sore throat (n = 9,598; 2.77%)	Mild symptoms with predominance of headache (n = 6,361; 1.84%)	Without symptoms (n = 322,310; 93.10%)	Total (N = 346,179; 100.00%)
<b>Went to healthcare service</b>							
Yes	47.00 **	31.90 **	28.80 **	13.30	12.70	8.00	15.50
No	53.00	68.10	71.20	86.70 **	87.30 **	92.00 **	84.50
<b>Measure taken</b>							
Stayed at home *							
Yes	94.30 **	91.00 **	92.00 **	84.80 **	83.10	80.20	83.90
No	5.70	9.00	8.00	15.20	16.90	19.80 **	16.10
Called a healthcare worker *							
Yes	23.00 **	16.30 **	11.30 **	5.10	5.50	3.10	6.10
No	77.00	83.70	88.70	94.90 **	94.50	96.90 **	93.90
Self-medication *							
Yes	68.70 **	72.10 **	78.20 **	55.70	55.20	54.10	58.00
No	31.30	27.90	21.80	44.30 **	44.80 **	45.90 **	42.00
Medication prescribed by physician *							
Yes	27.10 **	19.60 **	11.40	12.00	18.00 **	13.70	14.70
No	72.90	80.40	88.60 **	88.00 **	82.00	86.30 **	85.30
Received visit from a healthcare worker from the SUS *							
Yes	6.00 **	4.00	4.40	2.80	3.80 **	2.40	3.10
No	94.00	96.00	95.60	97.20	96.20	97.60 **	96.90

(continues)

Table 2 (continued)

Variables	All the symptoms (n = 1,315; 0.38%)	High prevalence of symptoms (n = 4,204; 1.21%)	Predominance of fever (n = 2,391; 0.70%)	Predominance of cough/sore throat (n = 9,598; 2.77%)	Mild symptoms with predominance of headache (n = 6,361; 1.84%)	Without symptoms (n = 322,310; 93.10%)	Total (N = 346,179; 100.00%)
Received visit from a private healthcare worker *							
Yes	4.10 **	2.00 **	1.90 **	0.80	0.60	0.60	0.90
No	95.90	98.00	98.10	99.20	99.40 **	99.40 **	99.10
<b>Healthcare service</b>							
UBS/Family health team *							
Yes	47.50	48.20	46.90	54.20 **	48.90	41.60	48.20
No	52.50	51.80	53.10	45.80	51.10	58.40 **	51.80
UPA *							
Yes	34.60 **	28.30 **	26.00	18.50	15.90	18.00	22.90
No	65.40	71.70	74.00	81.50 **	84.10 **	82.00 **	77.10
Hospital in the SUS *							
Yes	24.10	20.60	21.40	14.80	18.50	20.40	19.50
No	75.90	79.40	78.60	85.20 **	81.50	79.60	80.50
Private physician's office							
Yes	8.30	7.90	7.80	9.20	10.50	12.40	9.40
No	91.70	92.10	92.20	90.80	89.50	87.60	90.60
Private urgent care service *							
Yes	4.50	6.10 **	4.10	1.90	2.20	3.00	3.70
No	95.50	93.90	95.90	98.10 **	97.80 **	97.00	96.30
Private hospital							
Yes	14.30	13.10	10.60	10.00	11.80	11.30	11.80
No	85.70	86.90	89.40	90.00	88.20	88.70	88.20
Hospitalized *							
Yes	23.30 **	13.20	8.20	7.80	3.80	9.40	10.90
No	70.60	85.20	90.20	91.40 **	95.00 **	90.20	87.30
Unable to obtain admission	6.10 **	1.70	1.60	0.80	1.20	0.50	1.80

SUS: Brazilian Unified National Health System; UBS: basic unit service; UPA: urgent care unit.

\* Rao-Scot test with p-value < 0.05;

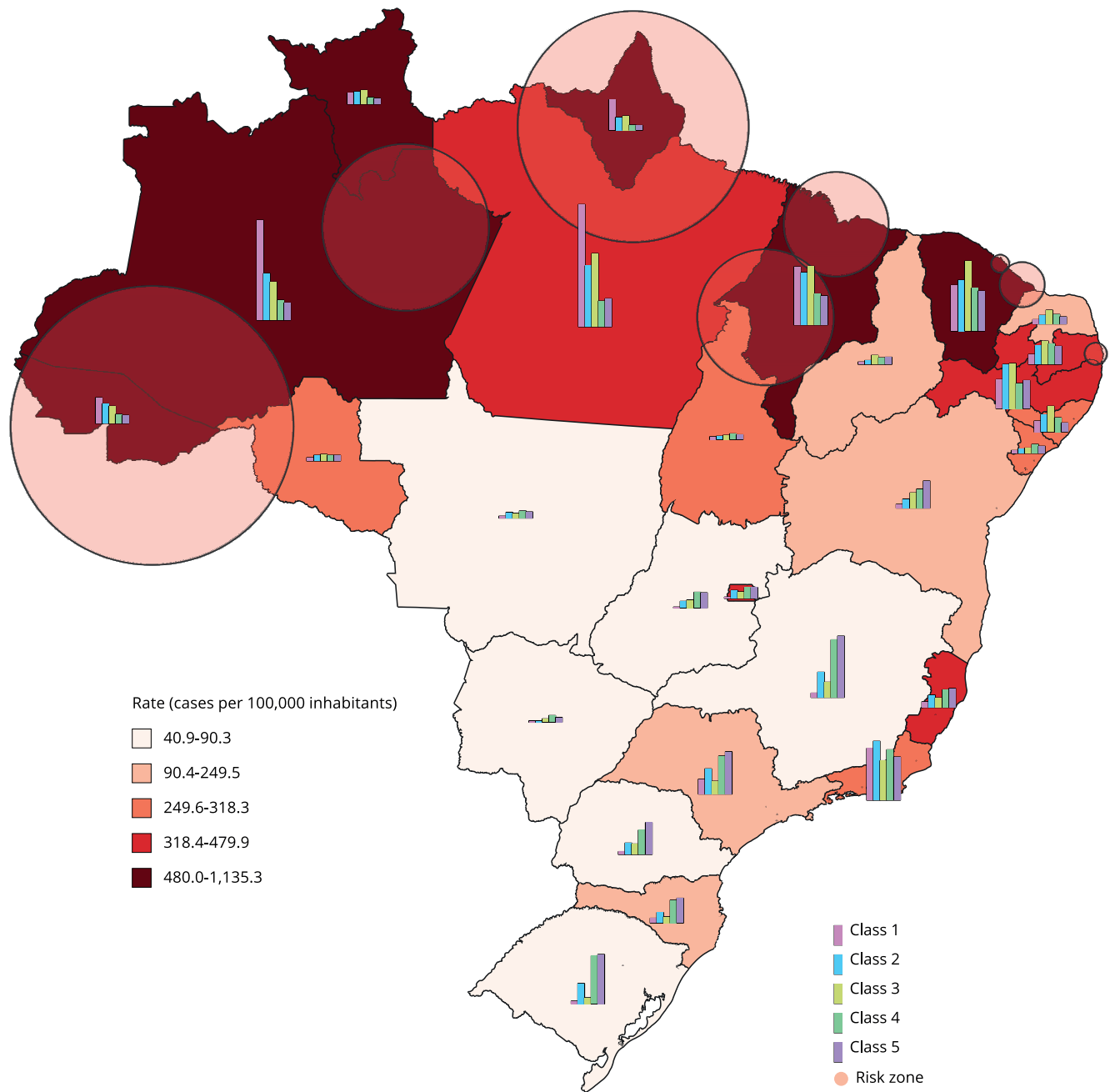
\*\* Standardized residual > 1.96.

As for spatial distribution of the quintiles of confirmed cases per 100,000 inhabitants and circular areas with greater relative risk of the disease, we observed a juxtaposition of a large share of the regions/areas with higher proportions of individuals classified as having all the symptoms (class 1). The class of individuals without symptoms (class 6) was not illustrated spatially in order to facilitate viewing the composition of the other classes, whose order of magnitude is quite diferente (Figure 3).



**Figure 3**

Spatial distribution of latent classes of symptoms (bars), COVID-19 case rates (per 100,000 inhabitants in quintiles), and areas at spatial risk in the COVID-19 case rate (circles). Brazil, 2020.



## Discussion

LCA allowed understanding the way COVID-19 symptoms are grouped among individuals with similar prodromic patterns. It is thus a better alternative than the imposition of a cutoff point that only classifies whether the individual had one or more symptoms. There can often be individuals with the same number of symptoms but of a sharply distinct nature. Considering the epidemiology and pathogenesis of COVID-19, the most common symptoms at the onset of the disease are fever, cough, and fatigue, while other symptoms include the production of sputum, headache, hemoptysis, diarrhea, dyspnea, and lymphopenia <sup>24</sup>. Thus, other symptoms like diarrhea and hemoptysis could be added to the investigation of symptoms in order to capture more suspected cases.

The demographic profile associated with the class of all the symptoms (classe 1) consisted mainly of brown women from the North and Northeast and all three older age brackets. The profile was thus not limited to the older quartile (53 years and older), emphasizing the groups of young adults (19-36 years) and middle-aged adults (37-53 years) with higher odds of belonging to all the classes, except mild symptoms (class 5). Having any level of schooling was associated with classes 1, 2, and 4. This may suggest greater understanding and capacity for voicing the magnitude of self-reported symptoms.

The search for health services was associated with three classes of symptoms, ranging from all the symptoms to predominance of fever. On the one hand, this finding shows the concern for seeking medical care in the presence of these symptoms in general, even if it is a fever. This may reflect the awareness-raising campaigns on the most common COVID-19 symptoms. On the other hand, it may generate higher probability of infection among those that are not confirmed, besides generating greater overload on health services. In such cases, teleconsultation initiatives have emerged as a possibility for follow-up of patients classified with less severe symptoms.

While telehealth was already growing in recent years in Brazil, the pandemic accelerated the field's legislation and regulation, given this strategy's potential for dealing with COVID-19. The possibilities for remote care and treatment, screening, triage, prevention, monitoring, detection, and surveillance shape a scenario of strengthening telehealth in the healthcare services' practice in the territory. The applications range from teleconsultations to tediagnosis, telemonitoring, teleregulation, tele-education, and second opinions. However, the full consolidation of telehealth requires investments in organizational models, systems, services, human resources, and infrastructure, including allocation of more budget funds and time for implementation <sup>25</sup>. Considering Brazil's historical social inequality and the pandemic's rapid growth, the country is still falling far short of universal coverage for these services.

The fact that the UPA of the SUS were the service most used by individuals with highest prevalence of all the symptoms makes the UPA the portal of entry for most individuals with multiple symptoms. This finding suggests this point of care in the system as the priority for testing, and not only the hospitals or UBS. However, Medina et al. <sup>26</sup> emphasize the key role of primary healthcare (PHC) to deal with the pandemic. They propose action by PHC along four main lines: (1) health surveillance in the territories; (2) care for patients with COVID-19; (3) social support for the most vulnerable; and (4) continuity of routine activities in PHC.

In relation to health surveillance, the typical capillarity of PHC, oriented by the Family Health Strategy (ESF), provides for the detection, notification, and follow-up of cases. However, the role of community leadership and territorial proximity of the community health agents favor engagement in measures of social isolation and local empowerment, dissemination of accurate information, and support for educational activities. An important part of care for COVID-19 patients is the organization of different patient flows depending on the severity of the disease, expressed by the symptoms and a history of contacts with infected persons. In this line of care, triage, teleconsultation, monitoring of symptoms, and linkage with other points in the network of care are essential for the continuity of healthcare. Social support for vulnerable groups, especially elderly individuals and those with comorbidities, should be articulated with the other mechanisms of social protection. This line of action features shelters in hotels, schools, home support, and other initiatives in community solidarity. Finally, other routine activities proper to PHC cannot be neglected, since they aim to keep the population protected through continuity of the vaccination calendar, monitoring of the main health problems,

and follow-up of individuals whose diseases are targets of intervention and resolution in this line <sup>26</sup>. In this sense, LCA of the information collected systematically by the PNAD-COVID19 survey proved to be a powerful tool to support the four lines of action in PHC, to the extent that they allow surveillance, screening, and monitoring for adequate linkage to the other points in the network.

However, there was an excess contingent of individuals classified with all the symptoms (class 1) who sought hospitalization but were unsuccessful. This calls attention to the collapse of health systems that were unable in many cases to meet the demand for hospitalization of more severe cases.

A recent longitudinal study in London, United Kingdom, using the app called *COVID Symptom Study Smartphone* detected and validated six distinct groups/clusters of COVID-19 symptoms that could require different levels of medical care <sup>27</sup>. Fourteen questions were used on symptoms reported by study participants. Clusters 1 and 2 represented forms of COVID-19 with 1.5% and 4.4%, respectively, requiring respiratory support. These clusters predominantly presented upper respiratory tract symptoms and were distinguished from each other by pain in cluster 2 compared to cluster 1 and slightly increased reports of fever in cluster 2. Group 3 showed stronger isolated gastrointestinal symptoms (diarrhea, missed meals) and a relative reduction in the need for respiratory support (3.7%). Clusters 4, 5, and 6 included participants that reported more severe COVID-19 symptoms, with 8.6%, 9.9%, and 19.8% of individuals who required respiratory support, respectively. These three groups had distinct presentations, with cluster 4 marked by the early presence of severe fatigue and cluster 5 by persistent pain and cough. Individuals in cluster 5 also reported confusion and severe fatigue. Finally, individuals in cluster 6 reported more severe symptoms such as difficulty breathing, including early onset of shortness of breath accompanied by chest pain.

In the current study, although most individuals were grouped in the class without symptoms, specific knowledge of the characteristics associated with the other latent classes is essential for identification of groups at increased risk. It also allows understanding the pattern of health services use in order to provide the best and timeliest diagnostic and patient care strategy. In this sense, seeking to apply epidemiological reasoning to the diagnostic criteria for COVID-19, Tan <sup>28</sup> identifies as the best detection method the one conducted in three stages: suspected case, diagnosed clinical case, and definitively diagnosed case. The CDC defines COVID-19 patients as those with cough, shortness of breath or difficulty breathing, fever, chills, muscle pain, sore throat, and loss of taste or smell as symptoms, besides close contact with confirmed COVID-19 patients <sup>29</sup>.

In fact, spatial analysis showed the juxtaposition of regions with higher case rates and circular areas at high risk with higher frequency of multiple symptoms (class 1). These findings provide an approximation between the symptoms and confirmed cases, corroborating this association. However, this approximation can only be done at the ecological level, considering the cluster effect. Thus, we cannot attribute a COVID-19 diagnosis to individuals classified with all the symptoms (class 1) and must thus avoid the ecological fallacy.

In a brief synthesis, considering the pandemic hecatomb and the absence of a vaccine or scientifically proven pharmacological treatment, the person-space-time epidemiological triad should be revisited in the same investigative spirit of John Snow in confronting the cholera epidemic in London in the 19th century <sup>30</sup>. In this sense, the recording of different symptom profiles should become a permanent activity in epidemiological surveillance systems. This study's results showed similarity between the factors associated with COVID-19 cases and latent classes of symptoms. The difference lies in the operational ease based on self-report of symptoms recorded by the PNAD-COVID19 survey in relation to the still-distant scenario of mass testing.

## Additional information

ORCID: Rafael da Silveira Moreira (0000-0003-0079-2901).

## Acknowledgments

The authors wish to thank André Luiz Sá de Oliveira of the Center for Statistics and Geoprocessing of the Aggeu Magalhães Institute, Oswaldo Cruz Foundation (Pernambuco State, Brazil), for the map's excellent vector editing.

## References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; 382:727-33.
2. World Health Organization. Novel coronavirus – China. <http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/> (accessed on 05/Aug/2020).
3. Tsang KW, Ho PL, Ooi GC, Yee WK, Wang T, Chan-Yeung M, et al. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 2003; 348:1975-83.
4. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 2003; 348:1967-76.
5. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020; 8:475-81.
6. Jefferson T, Mar CBD, Dooley L, Ferroni E, Al-Ansary LA, Bawazeer GA, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database Syst Rev* 2011; (7):CD006207.
7. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *N Engl J Med* 2020; 382:970-1.
8. Yu P, Zhu J, Zhang Z, Han Y. A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. *J Infect Dis* 2020; 221:1757-61.
9. Li C, Ji F, Wang L, Wang L, Hao J, Dai M, et al. Asymptomatic and human-to-human transmission of SARS-CoV-2 in a 2-family cluster, Xuzhou, China. *Emerg Infect Dis* 2020; 26:1626-8.
10. Li X, Guan P, Wu X, Wang L, Zhou Y, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; 382:1199-207.
11. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *J Med Virol* 2020; 92:441-7.
12. To KKW, Tsang OT, Yip CC, Chan K, Wu T, Chan JM, et al. Consistent detection of 2019 novel coronavirus in saliva. *Clin Infect Dis* 2020; 71:841-3.
13. Kim YJ, Sung H, Ki CS, Hur M. COVID-19 Testing in South Korea: current status and the need for faster diagnostics. *Ann Lab Med* 2020; 40:349-50.
14. Hopman J, Allegranzi B, Mehtar S. Managing COVID-19 in low- and middle-income countries. *JAMA* 2020; 323:1549-50.

15. Lima NVT, Buss PM, Paes-Sousa R. A pandemia de COVID-19: uma crise sanitária e humanitária. *Cad Saúde Pública* 2020; 36:e00177020.
16. Instituto Brasileiro de Geografia e Estatística. PNAD-COVID19: plano amostral e ponderação. <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101727.pdf> (accessed on 05/Aug/2020).
17. Moreira RS. COVID-19: unidades de terapia intensiva, ventiladores mecânicos e perfis latentes de mortalidade associados à letalidade no Brasil. *Cad Saúde Pública* 2020; 36:e00080020
18. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Equ Modeling* 2007; 14:535-69.
19. Rao JNK, Scott AJ. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Ann Stat* 1984; 12:46-60.
20. Pereira JCR. Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais. São Paulo: Edusp; 2004.
21. Haberman SJ. Analysis of qualitative data: introductory topics. New York: Academic Press; 1978.
22. Wang J, Wang X. Structural equation modeling: applications using Mplus. Chichester: John Wiley & Sons; 2012.
23. Kulldorff MA. Spatial scan statistic. *Commun Stat Theory Methods* 1997; 26:1481-96.
24. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* 2020; 109:102433.
25. Caetano R, Silva AB, Guedes ACCM, Paiva CCN, Ribeiro GR, Santos DL, et al. Desafios e oportunidades para telessaúde em tempos da pandemia pela COVID-19: uma reflexão sobre os espaços e iniciativas no contexto brasileiro. *Cad Saúde Pública* 2020; 36:e00088920.
26. Medina MG, Giovanella L, Bousquat A, Mendonça MHM, Aquino R. Atenção primária à saúde em tempos de COVID-19: o que fazer? *Cad Saúde Pública* 2020; 36:e00149720.
27. Sudre C, Lee K, Lochlainn M, et al. Symptom clusters in covid19: a potential clinical prediction tool from the COVID Symptom study app. *medRxiv* 2020; 6 jun. <https://www.medrxiv.org/content/10.1101/2020.06.12.20129056v1>.
28. Tan HZ. An epidemiologic thinking on the diagnosis criteria of COVID-19. *Zhonghua Liu Xing Bing Xue Za Zhi* 2020; 41:998-9.
29. Centers for Disease Control and Prevention. Novel coronavirus, (2019). <https://www.cdc.gov/coronavirus/2019-ncov/about/transmission.html> (accessed on 05/Aug/2020).
30. Cameron D, Jones IG. John Snow, the broad street pump and modern epidemiology. *Int J Epidemiol* 1983; 12:393-6.

## Resumo

*A ausência de testagens em massa para o diagnóstico da COVID-19 gera a necessidade de conhecer a dimensão da doença por meio da sua sintomatologia clínica. O objetivo foi investigar o perfil de sintomas relacionados à COVID-19 e aspectos relacionados. Foi analisada a amostra de participantes da Pesquisa Nacional por Amostra de Domicílios (PNAD-COVID19) realizada em maio de 2020. Foi realizada análise de classes latentes (ACL) com covariáveis sociodemográficas sobre 11 sintomas relatados por 346.181 indivíduos. Foram utilizados testes de Rao-Scott e análise de resíduos padronizados para mensurar a associação com o padrão de utilização dos serviços de saúde. Análise espacial de varredura foi realizada para identificar as áreas de risco para os casos de COVID-19. A ACL mostrou seis classes de sintomatologia, segundo o padrão de respostas dos indivíduos analisados: (1) todos os sintomas; (2) prevalência alta dos sintomas; (3) predominância de febre; (4) predominância de tosse/dor de garganta; (5) leves sintomas com predominância de dor de cabeça e (6) ausência de sintomas. Pessoas do sexo feminino, cor parda, provenientes das regiões Norte e Nordeste e em todas as três faixas etárias mais velhas apresentaram maior associação com a classe com todos os sintomas (classe 1). A maioria da procura por serviços também foi realizada por esse grupo de indivíduos, porém com distintos perfis de uso. A análise espacial mostrou sobreposição dessa classe com áreas de maior risco de casos de COVID-19. Os achados sustentam a importância da investigação dos sintomas, servindo para a identificação epidemiológica de possíveis casos em um cenário com baixa taxa de testagem populacional.*

*COVID-19; Análise de Classes Latentes; Pandemias; Monitoramento Epidemiológico*

## Resumen

*La ausencia de pruebas masivas para el diagnóstico de la COVID-19 genera la necesidad de conocer la dimensión de la enfermedad mediante su sintomatología clínica. El objetivo fue investigar el perfil de síntomas relacionados con la COVID-19 y sus aspectos relacionados. Se analizó la muestra de participantes de la Encuesta Nacional por Muestra de Domicilios (PNAD-COVID 19), realizada en mayo de 2020. Se efectuó el análisis de clases latentes (ACL) con covariables sociodemográficas sobre 11 síntomas mencionados por 346.181 individuos. Se utilizaron pruebas de Rao-Scott y análisis de residuos estandarizados para medir la asociación con el patrón de utilización de los servicios de salud. Se efectuó un análisis espacial de barrido para identificar las áreas de riesgo con casos de COVID-19. La ACL mostró seis clases de sintomatología, según el patrón de respuestas de los individuos analizados: (1) todos los síntomas; (2) prevalencia alta de los síntomas; (3) predominancia de fiebre; (4) predominancia de tos/dolor de garganta; (5) leves síntomas con predominancia de dolor de cabeza y; (6) ausencia de síntomas. Personas del sexo femenino, mestizos/mulatos, provenientes de las regiones Norte y Nordeste y en las tres franjas etarias más viejas presentaron una mayor asociación con la clase de todos los síntomas (clase 1). La mayoría de búsquedas de servicios también se realizó por parte de este grupo de individuos, aunque con distintos perfiles de uso. El análisis espacial mostró sobreposición de esta clase con áreas de mayor riesgo de casos de COVID-19. Estos hallazgos sostienen la importancia de la investigación de los síntomas, sirviendo para la identificación epidemiológica de posibles casos en un escenario con baja tasa de pruebas entre la población.*

*COVID-19; Análisis de Clases Latentes; Pandemias; Monitoreo Epidemiológico*

Submitted on 10/Aug/2020

Final version resubmitted on 06/Oct/2020

Approved on 09/Oct/2020