XI CONGRESSO NAZIONALE – SISMEC *Bari, 15-18th September*

CHAID analysis and Logistic Regression for identification of predictive scores in Post-Covid-19 Syndrome

Maria De Martino[1], Maddalena Peghin[2], Alvisa Palese[3], Carlo Tascini[2] , Miriam Isola[1]

[1]Division of Medical Statistic, Department of Medicine (DAME), University of Udine, 33100 Udine, Italy [2]Infectious Diseases Division, Department of Medicine, University of Udine and Azienda Sanitaria Universitaria Friuli Centrale (ASUFC), 33100, Udine, Italy [3]Department of Medical Sciences, University of Udine, Udine, Italy

Introduction

From the beginning of the SARS-CoV-*2* pandemic, descriptions of the symptomology and underlying mechanisms have been focused on COVID-19 and short-term outcomes. However, the so-called "post-COVID-19 syndrome", describing the experience of persistent symptoms after recovering from the initial acute COVID-19, has recently attracted attention. The aim is to compare two scores created for predicting the presence of symptoms after six months from recovery. The first score is created using a model of multivariable logistic regression, while for the second one the chi-squared automatic interaction detection (CHAID) analysis method is used. This last algorithm offers an alternative method from the classical one, being able to create scores considering interactions between variables.

Materials & Methods

Predictor	Post-COVID-19 syndrome	
	Yes (N=241)	No (N=358)
Female gender	146 (60.6)	174 (48.6)
Age		
18-40	50 (20.8)	91 (25.4)
41-60	109 (45.2)	138 (38.6)
>60	82 (34.0)	129 (36.0)
Number of symptoms		
0-1	25 (10.4)	156 (43.6)
2	35 (14.5)	78 (21.8)
3	49 (20.3)	51 (14.2)
4	53 (22.0)	36 (10.1)
≥5	79 (32.8)	37 (10.3)
Management		
Outpatients	157 (65.1)	285 (79.6)
Ward	70 (29.1)	64 (17.9)
ICU	14 (5.8)	9 (2.5)

Table 1. *Prevalence Post-COVID-19 syndrome.*

Population: 599 patients with a diagnosis of COVID-19 from 1 March to 30 May 2020

Aim: creation of a score to predict presence of symptoms after six months from recovery

Outcome: presence of symptoms after six months from recovery

Candidate predictors: gender, group of age, number of symptoms during acute onset disease, management of the patient

Statistical methods:

Score 1: Multivariable logistic regression

Score 2: CHAID

Performance evaluation: AUC (area under the curve)

Results

Multivariable logistic regression model

1. Variables resulting significantly associated to the Post-COVID-19 syndrome in the univariable analysis and clinically significant variables were included in a multivariable model to identify independent predictors (Table 2).
2. Based on results of the multivariable model, each significant predictor was weighted according to the resulting odds ratio (OR).
3. The score of each patients was determined summing the weights up to a total. The final score ranged from 0 to 7.

Predictors	Univariable Analysis			Multivariable Analysis			Score weight
	OR	95% CI	p	OR	95% CI	p	
Female gender	1.62	(1.17,2.26)	0.004	1.55	(1.06,2.27)	0.025	**1**
Age							
18-40	ref			ref			
41-60	1.44	(0.94,2.20)	0.096	0.96	(0.59,1.57)	0.887	/
>60	1.16	(0.74,1.80)	0.518	1.01	(0.60,1.71)	0.963	/
Number of symptoms							
0-1	ref			ref			
2	2.8	(1.56,5.00)	0.001	2.41	(1.32,4.39)	0.004	**1**
3	5.99	(3.37,10.67)	<0.001	5.79	(3.23,10.38)	<0.001	**2**
4	9.19	(5.05,16.70)	<0.001	8.04	(4.34,14.93)	<0.001	**3**
≥5	13.32	(7.50,23.68)	<0.001	11.85	(6.58,21.36)	<0.001	**4**
Management							
Outpatients	ref			ref			
Ward	1.98	(1.34,2.93)	0.001	1.91	(1.21,3.03)	0.006	**1**
ICU	2.82	(1.19,6.67)	0.018	3.25	(1.22,8.65)	0.018	**2**

Table 2. *Univariable and multivariable logistic regression.*

1. The score obtained from the multivariable logistic regression reported an AUC of 0.762 (95% CI 0.725-0.800).
2. The score obtained from the CHAID analysis reported an AUC of 0.741 (95% CI 0.703-0.779).
3. The two areas resulted to be significantly different (p=0.014), which is probably due to the size of the sample

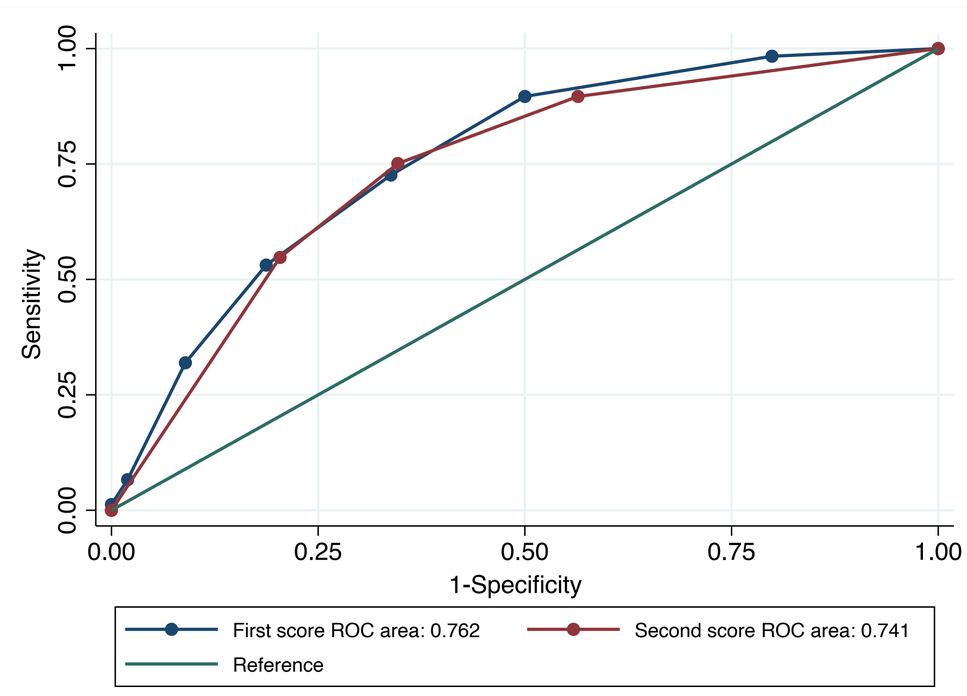
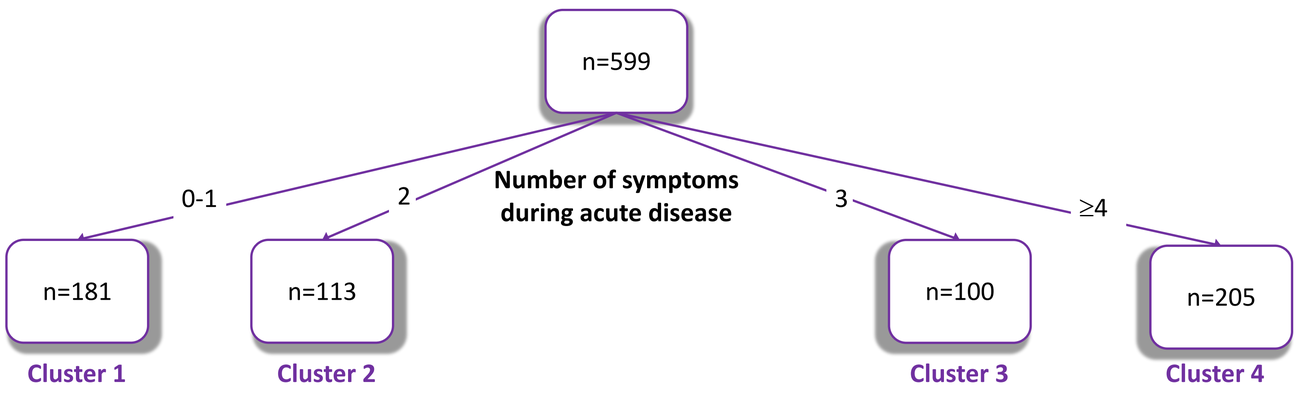


Figure 1. *ROC curves referred to the two methods.*

CHAID analysis

1. The score was created using the CHAID algorithm considering the four candidate predictors.
2. The analysis identified 4 clusters according to number of symptoms during acute onset disease.



Conclusions

CHAID procedure could represent an optimal method to create predictive scores, even if it has some limitations, for example the variables considered in the analysis have to be categorical.
In this peculiar population, CHAID analysis and logistic model showed a similar predictive performance, easy use in clinical practice should guide the final score selection.