

1  
2  
3  
4 What most influences severity and death of COVID-19 patients in Brazil? Is it clinical, social, or  
5 demographic factors? An observational study.  
6

7 Ana Carolina C. N. Mafra<sup>1\*</sup>, Régis Rodrigues Vieira<sup>1</sup>, Camila N. Monteiro<sup>1</sup>, Denise F. B.  
8 Cavalcante<sup>1</sup>, João L. Miraglia<sup>1</sup>, Daiana Bonfim<sup>1</sup>, Danielle C. Palacio<sup>1</sup>, Alessandra C. F. Martins<sup>1</sup>,  
9 Letícia Yamawaka de Almeida<sup>1</sup>, João Peres Neto<sup>1</sup>

10

11

12

13 <sup>1</sup> Hospital Israelita Albert Einstein, São Paulo, São Paulo, Brazil.

14

15

16 \* Corresponding author

17 Email: [ana.mafra@einstein.br](mailto:ana.mafra@einstein.br) (ACCNM)

18

19

20

21

22

23

24

## 25 **Abstract**

26 **Objective:** This study aimed to assess the space distribution and factors associated with the risk  
27 of severe acute respiratory syndrome (SARS) and death in COVID-19 patients, based on routine  
28 register data; and to develop and validate a predictive model of the risk of death from COVID-19.

29 **Methods:** A cross-sectional, epidemiological study of positive SARS-CoV-2 cases, reported in  
30 the south region of the city of São Paulo, SP, Brazil, from March 2020 to February 2021. Data  
31 were obtained from the official reporting databases of the Brazilian Ministry of Health for  
32 influenza-like illness (ILI) (esus-VE, in Portuguese) and for patients hospitalized for SARS  
33 (SIVEP-Gripe). The space distribution of cases is described by 2D kernel density. To assess  
34 potential factors associated with the outcomes of interest, generalized linear and additive logistic  
35 models were adjusted. To evaluate the discriminatory power of each variable studied as well as  
36 the final model, C-statistic was used (area under the receiver operating characteristics curve).  
37 Moreover, a predictive model for risk of death was developed and validated with accuracy  
38 measurements in the development, internal and temporal (March and April 2021) validation  
39 samples.

40 **Results:** A total of 16,061 patients with confirmed COVID-19 were enrolled. Morbidities  
41 associated with a higher risk of SARS were obesity (OR=25.32) and immunodepression  
42 (OR=12.15). Morbidities associated with a higher risk of death were renal disease (OR=11.8) and  
43 obesity (OR=8.49), and clinical and demographic data were more important than the territory *per*  
44 *se*. Based on the data, a calculator was developed to predict the risk of death from COVID-19,  
45 with 92.2% accuracy in the development sample, 92.3% in the internal validation sample, and  
46 80.0% in the temporal validation sample.

47 **Conclusions:** The risk factors for SARS and death in COVID-19 patients seeking health care, in  
48 order of relevance, were age, comorbidities, and socioeconomic factors, considering each  
49 discriminatory power.

50  
51  
52  
53

## 54 **Introduction**

55           In December 2019, the first cases of SARS-CoV-2 infection were described in Wuhan,  
56 China, and 16 months later COVID-19 has affected approximately 140 million people and resulted  
57 in more than 3 million deaths worldwide [1].

58           The American continent is currently the region most affected by the COVID-19 pandemic,  
59 and Latin America has one of the highest COVID-19 mortality rates in the world, driven by the  
60 inequality within its countries. Brazil has more than 200 million inhabitants and is a country  
61 marked by social inequality. Up to April 2021, SARS-CoV-2 has infected more than 14 million  
62 people and caused over 382 thousand deaths. São Paulo is the largest city in the country, with 12  
63 million inhabitants, and more than 980 thousand have been infected and 26 thousand have died  
64 [1,2].

65           The country has a mixed public and private health system, with one of the largest public  
66 health care programs in the world; in 2020, Brazil's Unified Health Care System (SUS) completed  
67 30 years of existence. Approximately 75% of the Brazilian population depends exclusively on the  
68 public health care system. Despite the advances following expansion of health services,  
69 particularly in Primary Health Care (PHC), SUS has suffered with chronic underfunding, resource  
70 scarcity and deficient delivery of services [3].

71           In the last five years, Brazil has undergone structural changes leading to more severe  
72 inequalities, which have strongly impacted the response to the pandemic. The following changes  
73 stand out: the spending ceiling that has frozen health care financing starting in 2016 through the  
74 next 20 years; changes in the National Primary Care Policy, in 2017, leading to uncertainty with  
75 flexibilization of the need for community health workers, and weakening of the Family Health  
76 Strategy, one of the pillars of PHC in the country. Additionally, in 2019 the program "Previne

77 Brasil” [Prevent Brazil] was launched and changed the PHC funding scenario, increasing doubts  
78 about the ability of SUS to keep its universal nature, at the risk of reducing PHC to a selective  
79 model, based on focal targets [4].

80 The risk factors strongly associated with the severity of COVID-19 are known, particularly  
81 advanced age, cancer, chronic renal disease, chronic obstructive pulmonary disease,  
82 cardiovascular disease, prior transplantation, obesity, pregnancy, sickle cell anemia, smoking and  
83 diabetes mellitus [5-8]. However, there is no specific data on risk factors associated with high  
84 mortality rates in territories of low-income and household overcrowding, such as Brazilian slums  
85 [9].

86 The combination of the SARS-CoV-2 epidemic with chronic non-communicable diseases  
87 (NCD) in a social context of poverty and inequality has disproportionately affected the so-called  
88 slums, i.e. densely populated and poverty-stricken areas [10].

89 The territorial distribution of COVID-19 in Brazil started in large, more populous urban  
90 centers, where trade relations are more intense, and flowed along the richest transport routes in the  
91 country, shaping virus spread, and ultimately reaching poorer regions [11]. São Paulo is one of the  
92 cities with the highest incidence of cases, mostly driven by its high population density. However,  
93 there is no specific data on how virus transmission occurred within the most vulnerable areas of  
94 the city [12], such as the districts of Vila Andrade and Campo Limpo, regions stricken by great  
95 social inequality and extreme vulnerability.

96 Despite the known social inequalities among countries and even regions within the same  
97 country, and the need for differentiated responses in highly vulnerable and resource-scarce  
98 settings, little is known about the specific realities of virus spread and the risk factors present in  
99 slums. The data extrapolated is from regions with vastly different conditions, which may not

100 correspond to the real world [11]. Hence, this study aimed to assess the space distribution and  
101 factors associated with the risk of SARS and death from COVID-19, based on routinely collected  
102 register data, as well as to develop and validate a predictive model of risk of death from COVID-  
103 19.

104

## 105 **Methods**

106 This was a cross-sectional, epidemiological study that included positive COVID-19 cases  
107 reported in the south region of the city of São Paulo, SP, Brazil, between March 2020 and February  
108 2021. We conduct the work following the STROBE checklist (S2 Text).

109 The study territory includes two administrative districts, Campo Limpo and Vila Andrade,  
110 which together have an estimated population of approximately 400 thousand people (Fig 1A) [13].  
111 In this region, social vulnerability can be measured by the São Paulo Social Vulnerability Index  
112 (IPVS, in Portuguese) [14], which ranges from 1 for extremely low vulnerability (navy areas) to 6  
113 for very high vulnerability (purple areas) (Fig 1B). The GeoSES is another socioeconomic  
114 indicator [15] that can express the region's vulnerability. It is a composite index that summarizes  
115 the Brazilian socioeconomic context, and is composed by seven dimensions: education, mobility,  
116 poverty, wealth, income, segregation and deprivation of resources and services. The index can  
117 range from -1 (worst socioeconomic context) to 1 (best socioeconomic context). The districts of  
118 Vila Andrade and Campo Limpo have average values of -0.03 and -0.36 respectively (Fig 1C).

119

120 **Fig 1. Study territory.** A: Satellite image and photographs of the region. B: Distribution of the  
121 São Paulo Social Vulnerability Index. C: Distribution of the GeoSES Socioeconomic Index.

122

123           The distribution of the health care facilities, private and public, accessed by the study  
124 population can be visualized in a map available in S1 Fig. Health care facilities were categorized  
125 as private or public according to the National Registry of Health Care Facilities (CNES, in  
126 Portuguese), and was considered in the analyses as an indicator of health care access, since public  
127 health services are universally accessible, while only those with health insurance or high income  
128 have access to private services.

129           The data reviewed come from official reporting databases of the Brazilian Ministry of  
130 Health, including two systems: one for influenza-like illness (esus-VE), with milder cases of the  
131 disease reported by any health care facility, and another (SIVEP-Gripe) for patients hospitalized  
132 with severe acute respiratory syndrome (SARS). The residence location of individuals included in  
133 the study was geocoded by GISA/CEINFO - SMS, and made available by the Health Surveillance  
134 Unit of Campo Limpo. The study database was anonymized, with all identifiable patient  
135 information removed. This study was approved by the Institutional Review Boards of Hospital  
136 Israelita Albert Einstein and the São Paulo Health Department, protocol numbers 4.462.994 and  
137 4.648.956, respectively.

138           The study population was defined as all COVID-19 cases confirmed by PCR. The outcome  
139 variables were progression to SARS and death, categorized as: influenza-like illness (ILI)/severe  
140 acute respiratory syndrome (SARS) and discharge/death, respectively.

141           The independent variables were sociodemographic characteristics, comorbidities, and  
142 symptoms. The following inclusion criteria were used: to be a resident of the study region, aged  
143 18 years and older, with laboratory confirmation of COVID-19 by PCR. When the same patient  
144 had multiple records, and ILI had been reported previously to SARS, only the SARS record was  
145 considered (Fig 2).

146 **Fig 2. Study data inclusion and exclusion flow.**

147

148 Results were presented as absolute and relative frequencies, overall and by outcome (death  
149 or SARS). For simple comparison between outcome groups, the chi-square test was used for  
150 categorical variables and Student's t-test was used for numerical variables.

151 The spatial distribution of cases was described by a 2D kernel density.

152 To assess potential factors associated with the outcomes of interest, generalized linear and  
153 additive logistic models were adjusted, and the latter were used for inclusion of the spatial location  
154 of subjects' place of residence (latitude and longitude). First, simple models were adjusted, and  
155 variables with a p value  $\leq 0.2$ , and with less than 10% of missing values were considered for the  
156 multiple models. No data imputation was used, so only complete observations were used in the  
157 analyses. Symptoms were not included in the model, so it would only evaluate conditions that  
158 existed prior to COVID-19. The selection of variables for the final model was performed by  
159 backwards stepwise elimination, with a p value  $< 0.05$  as the criteria to inclusion in the model. In  
160 the final model, the quality of adjustment was verified by the magnitude of standard errors, the  
161 influence of inclusion or exclusion of variables in the estimated odds ratio, and the variance  
162 inflation factor, which was not greater than 1.2.

163 To assess the discriminatory power of each variable evaluated as well as the final model,  
164 the C-statistic was used (area under the receiver operating characteristics curve).

165 Since the multiple model presented a good discriminatory power, a predictive model for  
166 death was developed, with the main clinical and demographic characteristics of patients arriving  
167 at health care facilities for COVID-19 evaluation. In this analysis, the symptoms were also  
168 considered, and we followed the same method of variable selection presented above. The data was  
169 randomly split in two sets with a 2:1 ratio for training and testing, respectively. The predictive



170 performance of the model was assessed by accuracy measurements, discriminatory power and the  
171 Hosmer-Lemeshow test [16]. The temporal validation used data from patients over 18 years of  
172 age, with a positive COVID-19 test and who had symptoms onset between March and April of  
173 2021.

174 All analyses were performed with the R software, version 3.6.3 [17], with the mgcv [18],  
175 ggplot2 [19] and viridis [20] packages. A 5% significance level was adopted.

176

## 177 Results

178 The study investigated a population of 16,061 patients with confirmed COVID-19, between  
179 March 2020 and February 2021, with 1,925 (11.98%) cases of SARS and 375 (2.37%) deaths.

180 The age of included individuals ranged from 18 to 103 years, with a mean of 42.1 years  
181 (standard deviation 14.9) and a median of 40 years.

182 The baseline characteristics of study participants, stratified by outcome, and the C-statistic  
183 for each variable and outcome can be found at Table 1.

184 Age presented the greatest C-statistic for the prediction of death and SARS, followed by  
185 the presence of symptoms, such as dyspnea, fever and cough, and the presence of comorbidities,  
186 like heart disease and diabetes.

187

188 **Table 1. Baseline characteristics of study participants, overall and stratified by study**  
189 **outcome, and C-statistics for each variable and outcome. N=16,031 confirmed for COVID-**  
190 **19.**

	Overall	Missing rate (%)	SARS (yes)	C SARS	Death (yes)	C Death
N	16,061		1,925		375	

**Profile**

Mean age (SD)	42.10 (14.94)	0.0	56.93 (16.77)**	0.78	68.08 (14.48)**	0.90
<b>Age Range</b>		0.0		0.75		0.86
18 to 40 years	7892 (49.1)		337 (17.5)**		18 (4.8)**	
40 to 60 years	6022 (37.5)		736 (38.2)		82 (21.9)	
60 to 80 years	1830 (11.4)		653 (33.9)		184 (49.1)	
80 years and older	317 (2.0)		199 (10.3)		91 (24.3)	
<b>Sex</b>		0.0		0.56		0.57
Female	8804 (54.8)		852 (44.3)**		153 (40.8)**	
Male	7230 (45.0)		1073 (55.7)		222 (59.2)	
Undefined	27 (0.2)		0 (0.0)		0 (0.0)	
<b>Race/skin color</b>		24.9		0.54		0.53
White	5925 (49.1)		675 (51.4)**		144 (49.3)	
Black	663 (5.5)		63 (4.8)		21 (7.2)	
Yellow	901 (7.5)		35 (2.7)		10 (3.4)	
Brown	4563 (37.8)		539 (41.1)		117 (40.1)	
Indigenous	15 (0.1)		1 (0.1)		0 (0.0)	
<b>Caucasian</b>		24.9		0.51		0.50
No	6142 (50.9)		638 (48.6)		148 (50.7)	
Yes	5925 (49.1)		675 (51.4)		144 (49.3)	
<b>Conditions</b>						
Diabetes (y)	997 (6.5)	5.0	506 (44.9)**	0.71	154 (49.4)**	0.72
Obesity (y)	187 (1.2)	5.0	118 (10.5)**	0.55	44 (14.1)**	0.56
Immunodepression (y)	145 (0.9)	5.0	62 (5.5)**	0.52	26 (8.3)**	0.53
Heart disease (y)	1494 (9.8)	5.0	720 (63.8)**	0.79	215 (68.9)**	0.80
Renal disease (y)	122 (0.8)	5.0	85 (7.5)**	0.54	53 (17.0)**	0.58
Respiratory Disease (y)	418 (2.7)	5.0	125 (11.1)**	0.55	46 (14.7)**	0.56
Pregnancy (y)	78 (0.5)	7.9	11 (1.7)**	0.51	0 (0.0)	0.50
<b>Symptoms</b>						
Fever	6868 (43.8)	2.3	1211 (78.0)**	0.69	235 (73.2)**	0.65
Cough	8354 (53.0)	1.9	1352 (83.3)**	0.67	271 (81.9)**	0.64
Sore throat	3755 (24.6)	5.1	246 (22.3)	0.51	55 (20.3)	0.52
Dyspnea	3567 (22.7)	2.0	1332 (82.7)**	0.83	293 (85.4)**	0.82
Diarrhea	231 (1.5)	5.2	231 (21.1)**	0.61	65 (23.6)**	0.61
Vomiting	125 (0.8)	5.5	125 (12.0)**	0.56	32 (12.5)**	0.55
Abdominal pain	43 (0.3)	9.5	43 (10.8)**	0.55	7 (7.7)**	0.53
Fatigue	157 (1.1)	9.2	157 (35.4)**	0.68	18 (18.8)**	0.58
Loss of smell	1487 (10.2)	9.1	125 (27.3)**	0.59	12 (12.5)	0.51
Loss of taste	1386 (9.5)	9.4	96 (22.8)**	0.57	11 (12.1)	0.51
<b>Socioeconomic</b>						
<b>Type of healthcare facility</b>						
Private	6999 (51.6)	15.5	1079 (58.3)**	0.54	161 (44.7)*	0.53
Public	6570 (48.4)		772 (41.7)		199 (55.3)	

<b>Mean IPVS (SD)</b>	2.68 (1.60)	0.9	2.57 (1.54)*	0.52	2.78 (1.52)	0.53
0-3	12619 (79.3)	0.9	1555 (81.2)*	0.51	287 (76.7)	0.51
4-6	3302 (20.7)		360 (18.8)		87 (23.3)	
<b>Mean GeoSES (SD)</b>	-0.20 (0.30)	0.0	-0.19 (0.30)*	0.52	-0.25 (0.28)*	0.54
<0	11438 (71.2)	0.0	1336 (69.4)	0.51	289 (77.1)*	0.52
>=0	4623 (28.8)		589 (30.6)		86 (22.9)	

191 C: C-statistic, representing the area under the ROC curve. Measures reported as absolute frequency  
 192 (%), when mean not indicated (standard deviation). \* p value for hypothesis testing <0.05. \*\* p  
 193 value for hypothesis testing <0.001. Tests performed: chi-square and Student's *t*.

194

### 195 **Geospatial distribution**

196 The density of COVID-19 cases was plotted to visualize the concentration of cases,  
 197 alongside the incidence, which considers the total population of each area (Fig 3). It is possible to  
 198 note on Figure 3 that disease severity was not correlated with the number of cases.

199

200 **Fig 3. Geospatial distribution of cases and outcomes.** Density of positive COVID-19 cases in  
 201 the region (A), incidence of cases in the population (B), SARS incidence among positive cases  
 202 (C), mortality among positive cases (D), by census sector, according to the place of residence of  
 203 cases.

204

205 In map "A", Fig 3, there is clearly an area of higher density of positive COVID-19 cases,  
 206 which overlaps with the area of the largest slums in the city of São Paulo, Paraisópolis, with a  
 207 predominantly young and vulnerable population living in overcrowded households. However, the  
 208 incidence of positive COVID-19 cases, SARS and deaths in this area are not high, when compared  
 209 to other areas in the region (Figs "3B", "3C" and "3D").

210 The data on progression to SARS was reviewed for the 16,061 individuals included in the  
211 study. After excluding individuals with missing information on comorbidities, the multiple model  
212 was adjusted based on data from 15,235 patients.

213 In the final multiple model (Table 2), clinical and demographic variables were more  
214 relevant than the region *per se*, which was not included in the final model, even though place of  
215 residence had a p value <0.001, C-statistic of 0.57, and 0.74% of deviance explained in the simple  
216 analysis (Fig 4). The comorbidities found to be associated with a higher risk of SARS were obesity  
217 and immunodepression. No socioeconomic variable remained significant to be included in the  
218 multiple model. The complete model presented a C-statistic of 0.96 and deviance explained of  
219 50%.

220

221 **Table 2. Results of the multiple model for SARS associated with COVID-19. N=15,235.**

Factors	OR (95%CI)	p value
Age (years)	1.063 (1.057;1.069)	<0.001
Sex (Male)	1.61 (1.36;1.91)	<0.001
Obesity (y)	25.32 (16.31;39.31)	<0.001
Immunodepression (y)	12.15 (7.62;19.37)	<0.001
Renal disease (y)	10.12 (5.81;17.62)	<0.001
Heart disease (y)	9.79 (8.21;11.68)	<0.001
Respiratory disease (y)	6.72 (4.81;9.40)	<0.001
Diabetes (y)	6.70 (5.51;8.15)	<0.001
s (Longitude, Latitude)	Fig 4	0.721

222

223

224 **Fig 4. Predicted probability of progression to SARS by the space-only model. N=16,601.**

225

226

227 The data of 15,795 individuals was reviewed for deaths. After excluding patients with no  
228 comorbidity data, the multiple model for death was adjusted based on information from 15,100  
229 patients.

230 As was the case for SARS, in the final multiple model (Table 3), clinical and demographic  
231 variables were more relevant than the region *per se*, which was not included in the final model,  
232 even though place of residence had a p value 0.001, C-statistic of 0.62, and 1.8% of deviance  
233 explained in the simple analysis (Fig 5). The comorbidities associated with a higher mortality risk  
234 were renal diseases and obesity. The complete model presented a C-statistic of 0.96 and deviance  
235 explained of 46.5%.

236

237 **Table 3. Results of the multiple model for deaths associated with COVID-19. N=15,100.**

Factors	OR (CI 95%)	p value
Age (years)	1.082 (1.071; 1.092)	<0.001
Sex (Male)	1.68 (1.27; 2.21)	<0.001
Renal disease (y)	11.8 (7.1; 19.7)	<0.001
Obesity (y)	8.49 (5.21; 13.85)	<0.001
Immunodepression (y)	7.68 (4.22; 13.97)	<0.001
Heart disease (y)	4.41 (3.26; 5.97)	<0.001
Respiratory disease (y)	3.50 (2.23; 5.51)	<0.001
Diabetes (y)	3.06 (2.29; 4.09)	<0.001

GeoSES (0.1 unit)	0.896 (0.844; 0.951)	<0.001
s (Longitude, Latitude)	Fig 5	0.657

---

238

239

240 **Fig 5. Predicted probability of death (case fatality rate/100) by the space-only model.**

241 **N=15,795.**

242

### 243 **Predictive model for death from COVID-19**

244 The development of the predictive model included 9,931 observations. Socioeconomic and  
245 geolocation data were not included, since this information is more difficult to obtain during the  
246 patient's clinical evaluation, but symptoms were included. The final predictive model included the  
247 following variables with their respective coefficients:

248 Risk score = -10.37 + (0.075 x age in years) + (0.98 x diabetes) + (1.21 x obesity) + (1.89 x  
249 immunodepression) + (0.99 x heart disease) + (2.95 x chronic renal disease) + (1.18 x respiratory  
250 disease) + (0.78 x fever) + (1.81 x dyspnea), where variables are coded as 1 when the condition is  
251 present, and 0 when absent.

252 Based on this formula, it is possible to calculate the predicted probability as follows:

253 Predicted risk (or probability) =  $1 / (1 + e^{-\text{Risk score}})$  A risk calculator was developed, and it  
254 is accessible at: <https://redcap.einstein.br/surveys/?s=PFLPC9JJ8F>

255 To set a cutoff with a better balance between sensitivity, specificity and accuracy, the ROC  
256 curve was used to obtain a 0.0186 predicted probability with 92.2% specificity, 94.1% sensitivity  
257 and 92.2% accuracy. The C-statistic observed was 0.974 (95% CI: 0.967–0.981), demonstrating  
258 an excellent discrimination.

259 The training sample included 170 events, which is sufficient for the estimation of the  
260 coefficients of the 9 variables included in the final model. No collinearity between variables was  
261 evident, with the variance inflation factors ranging from 1.01 to 1.17. The maximum values for  
262 the estimated coefficient standard errors were 0.48 for the intercept, and 0.38 for  
263 immunodepression. The p value for the Hosmer-Lemeshow test was 0.485 indicating good model  
264 fit.

265 For internal validation, a sample of 4,966 observations was used, in which 92 events  
266 occurred (10.2 events per variable). In this sample, by applying the predicted model and  
267 maintaining the same cutoff, 92.2% sensitivity, 98.9% specificity and 92.3% accuracy was  
268 achieved. The C-statistic observed was 0.985 (95%CI: 0.978 - 0.991), with the same  
269 discriminatory quality.

270 For temporal validation, the predicted model for death was applied to a sample of 3,798  
271 COVID-19 cases confirmed by PCR, notified between March and April 2021. 104 died from  
272 COVID-19 during the study period. The predictive analysis resulted in an accuracy of 80.04%,  
273 sensitivity of 98.08% and specificity of 79.53%. The C-statistic was 93.62 (95% CI: 92.24–95.00).  
274

## 275 **Discussion**

276 This study analyzed the space distribution and factors associated with the risk of SARS  
277 and death in positive COVID-19 cases and, derived a death risk calculator.

278 The risk factors associated with severe acute respiratory syndrome (SARS) were age,  
279 obesity and immunodepression, a similar finding to those of systematic reviews that evaluated risk  
280 factors associated with severe COVID-19 [21-25]. However, none of these reviews evaluated the  
281 association of environmental factors, such as living in slums, overcrowded households or in

282 precarious sanitary conditions, with COVID-19 severity. Some observational studies have  
283 demonstrated a higher death prevalence among vulnerable and lower-income populations, but  
284 there is no solid evidence for this association [6,26].

285           This study found a strong association between comorbidities and age with severe  
286 COVID-19, raising the question if the poor management of chronic conditions in a vulnerable  
287 population may have contributed to the severity of COVID-19. However, no significant  
288 association was found for living conditions, while regions with a lower prevalence of chronic  
289 diseases, even the highly vulnerable ones, had a lower risk of severe respiratory syndromes.

290           Another key finding of this study was the increased risk of death associated with (in  
291 order of importance) age, comorbidities, and vulnerability status. The comorbidities most often  
292 related to the risk of death, in decreasing order of association, were renal disease, obesity,  
293 immunodepression, heart disease, respiratory disease, diabetes, male sex and age. Better  
294 socioeconomic conditions had a protective effect on the risk of death.

295           The density, incidence, severity, and case fatality rate of COVID-19 in the study region  
296 suggested the virus could spread in both high- and low-vulnerable regions. It is important to  
297 consider the vulnerability of specific groups in order to understand the impact of the pandemic on  
298 the study region, part of the largest Brazilian city with 400 thousand inhabitants, and presenting a  
299 diverse range of urban social vulnerability scores [27].

300           In Paraisópolis, the largest slum of the study region, there was an evident higher density  
301 of cases, in agreement with other Brazilian studies that found greater virus spread in highly  
302 vulnerable and densely populated locations [28,29]. However, a high incidence of severe cases  
303 and deaths was found in two areas of low vulnerability also presenting low density and overall



304 incidence of cases, which is in accordance with the findings of the study by Bermudi *et al.* [29],  
305 showing high mortality in the richest areas of the city of São Paulo.

306           The finding of increased mortality indicate that efforts aimed at the management of  
307 chronic conditions and the care of the elderly should be the focus of public policies towards  
308 COVID-19 [30], for areas with high and low socioeconomic status.

309           Despite the extreme plausibility of the fact that poverty aggravates health conditions [31],  
310 this does not change the need for controlling risk factors that could be more easily modified, as  
311 well as seek strategies that would facilitate the implantation of social distancing for the most  
312 vulnerable. In Brazil, the SUS provides free and universal access to health services, with a large  
313 primary health care network present throughout the country which should be used as the  
314 transforming agents of this reality [32,33], given that countries without universal health care, such  
315 as the USA, suffered with high mortality rates from COVID-19 [34].

316           Predictive models for COVID-19 severe disease should include symptoms to help with  
317 medical decision making, resource planning and improvements in monitoring of COVID-19  
318 patients [35] (i.e., the number of symptoms is predictive of disease duration [36,37]). This study  
319 identified cough, fever, sore throat, dyspnea, anosmia, loss of taste, diarrhea and fatigue as  
320 symptoms significantly associated with COVID-19 SARS and death.

321           The proposed risk model for death had good performance, including a C-statistic of  
322 97.4% (95% CI: 96.7%–98.1%), which is higher than the value found by another calculator for a  
323 population of elders in the USA, with a C-statistic of 85.3%, while needing more variables [38].  
324 Booth *et al.* [39] proposed a different strategy to predict deaths associated with COVID-19  
325 employing molecular biomarkers measured in laboratory samples used for PCR tests; however,  
326 this approach presents limitations related to cost and availability of results in a timely manner, in

327 addition a lower performance when compared to the model proposed by this study, with a C-  
328 statistic of 93%, sensitivity of 91% and specificity of 91%. Additional studies have focused only  
329 on the elderly or hospitalized individuals [40,41].

330           The model proposed by this study is a powerful response to support managers and other  
331 professionals in planning patient care, prioritizing more targeted and assertive strategies and  
332 actions, such as monitoring of positive cases with higher prediction of death, stewardship, and  
333 distribution of resources at different care levels, regions and cities, including different vulnerability  
334 contexts.

335           Added to the risk factors highlighted in this study, many countries in the world are going  
336 through an economic, political and ethical crisis, including Brazil, with defective response, policies  
337 that go against social distancing, lack of country-level coordination, negationism, and neoliberal  
338 policies [42], which may affect the outcomes of this pandemic. Thus, the response to the pandemic  
339 in these settings must be challenged, and larger studies including these variables in the analyses  
340 are required, such as the need for interdisciplinary analyses [43-45], considering the clinical,  
341 demographic, socioeconomic and geospatial dimensions, and incorporating all of them into one  
342 single model.

343           This study had the strength, of using data derived from official databases of confirmed  
344 COVID-19 cases making it possible to include a large sample size, but at the same time is subject  
345 to the limitations of observational studies and of routinely-collected health data, with the  
346 possibility of the interference by multiple errors and biases (e.g., data linkage problems,  
347 misclassification bias and underreporting) During the study period the main SARS-CoV-2 variant  
348 in circulation changed, with the introduction of the P1 variant in January of 2021, in addition to  
349 the start of the vaccination campaign at the same month. Because of these factors, the temporal

350 validation was done finding some decrease in performance, but still finding and excellent  
351 discrimination.

352 In conclusion, this study contributed to the global effort to fight COVID-19, identifying  
353 risk factors associated with SARS and death among COVID-19 patients, and proposing a  
354 calculator for risk of death from COVID-19, that showed a satisfactory performance, both on a  
355 diverse setting including urban slums and extremely vulnerable populations.

356

## 357 **Acknowledgments**

358 We are grateful to all health professionals involved in combating the pandemic of COVID-  
359 19, especially those from Campo Limpo and Vila Andrade who, in addition to patient care,  
360 contributed with the information used in this study.

361

## 362 **References**

- 363 1. Map COVID-19. Johns Hopkins Coronavirus Resource Center. Available from  
364 <<https://coronavirus.jhu.edu/map.html>> Accessed on: 23 APR 2021.  
365
- 366 2. Boletim Diário COVID-19. Available from:  
367 <[https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/saude/20210422\\_boletim\\_covid19\\_diario.pdf](https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/saude/20210422_boletim_covid19_diario.pdf)> Accessed on: 23 APR 2021.  
368
- 369 3. Massuda A, Hone T, Leles FAG, de Castro MC, Atun R. The Brazilian health system at  
370 crossroads: progress, crisis and resilience. *BMJ Global Health*. 2018;3(4):e000829.  
371
- 372 4. Melo EA, Mendonça M H M, Oliveira J R, Andrade G C L. Changes in the National Primary  
373 Care Policy: between setbacks and challenges. *Health debate*. 2018;42(spe1):38-51.  
374
- 375 5. Saini, K S et al. Mortality in patients with cancer and coronavirus disease 2019: A systematic  
376 review and pooled analysis of 52 studies. *European Journal of Cancer*. 2020;139:43–50.  
377
- 378 6. Khan M M A, Khan M N, Mustagir M G, Rana J, Islam M S, Kabir M I. Effects of underlying  
379 morbidities on the occurrence of deaths in COVID-19 patients: A systematic review and meta-  
380 analysis. *J Glob Health*. 2020;10(2):020503.  
381

- 382  
383 7. Fahad A, et al. Early Report on Published Outcomes in Kidney Transplant Recipients  
384 Compared to Nontransplant Patients Infected with Coronavirus Disease 2019. *Transplantation*  
385 *Proceedings*. 2020. 52; 9:2659-2662.  
386  
387 8. Fadini, GP. Prevalência e impacto do diabetes entre pessoas infectadas com SARS-CoV-2. *J*  
388 *Endocrinol Invest*.2020; 43, 867–869.  
389  
390 9. Lau LL, Hung N, Wilson K. COVID-19 estratégias de resposta: considerando as  
391 desigualdades entre e dentro dos países. *Int J Equity Health*. 2020; 19, 137.  
392  
393 10. Horton R. Offline: COVID-19 is not a pandemic. *Lancet*. 2020; 26;396(10255):874.  
394  
395 11. Guimarães RB, Catão RC, Martinuci OS, Pugliesi EA, Matsumoto PSS. Geographical  
396 reasoning and the keys to reading Covid-19 in Brazilian territory. *Advanced Studies*, 34(99),  
397 119-140.  
398  
399 12. Alcântara E. Investigating spatiotemporal patterns of the COVID-19 in São Paulo State,  
400 Brazil. *Geospatial Health*.2020; 15(2).  
401  
402 13. Fundação SEADE. São Paulo State Statistics Portal. Available from <  
403 [https://www.seade.gov.br/seade-disponibiliza-projecoes-de-populacao-e-domicilios-ate-](https://www.seade.gov.br/seade-disponibiliza-projecoes-de-populacao-e-domicilios-ate-2050/)  
404 [2050/](https://www.seade.gov.br/seade-disponibiliza-projecoes-de-populacao-e-domicilios-ate-2050/)> Accessed on: 23 APR 2021.  
405  
406 14. Pavarini SC, Barha EJ, Mendiondo MS, Filizola CL, Petrilli Filho JF, Santos AA. Family  
407 and social vulnerability: a study with octogenarians. *Rev Lat Am Enfermagem*.  
408 2009;17(3):374-9.  
409  
410 15. Barrozo LV. GeoSES: A socioeconomic index for health and social research in Brazil. *PLoS*  
411 *One*. 2020;15(4):e0232074.  
412  
413 16. Royston P, Luas KGM, Altman D G, Vergouwe Y. Prognosis and prognostic research:  
414 Developing a prognostic model *BMJ* 2009; 338: b604  
415  
416 17. R Core Team (2019). R: A language and environment for statistical computing. R  
417 Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.  
418  
419 18. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of  
420 semiparametric generalized linear models. *Journal of the Royal Statistical Society*.  
421 2011;73(1):3-36.  
422  
423 19. H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York,  
424 2016.  
425  
426 20. Simon Garnier (2018). *viridis: Default Color Maps from 'matplotlib'*. R package version  
427 0.5.1. <https://CRAN.R-project.org/package=viridis>.

- 428  
429 21. Fadini GP, Morieri ML, Boscari F, Fioretto P, Maran A, Busetto L, et al. Newly-diagnosed  
430 diabetes and admission hyperglycemia predict COVID-19 severity by aggravating respiratory  
431 deterioration. *Diabetes Res Clin Pract.* 2020;168:108374.  
432
- 433 22. Yang, J, Hu, J, Zhu, C. Obesity aggravates COVID-19: A systematic review and meta-  
434 analysis. *J Med Virol.* 2021; 93: 257– 261.  
435
- 436 23. Yang, Jing et al. Prevalence of comorbidities and its effects in patients infected with SARS-  
437 CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases,*  
438 2020; 94, 91 – 95.  
439
- 440 24. Gao Ya. Impacts of immunosuppression and immunodeficiency on COVID-19: A  
441 systematic review and meta-analysis. *Journal of Infection.*2020;81, Issue 2, e93 - e95.  
442
- 443 25. Geng MJ, Wang LP, Ren X. Risk factors for developing severe COVID-19 in China: an  
444 analysis of disease surveillance data. *Infect Dis Poverty.*2021;48.  
445
- 446 26. Baqui P. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a  
447 cross-sectional observational study. *The Lancet Global Health.* 2020;8, e1018 - e1026.  
448
- 449 27. São Paulo, SEADE. Paulista Social Vulnerability Index version 2010.  
450 <[http://ipvs.seade.gov.br/view/pdf/ipvs/principais\\_resultados.pdf](http://ipvs.seade.gov.br/view/pdf/ipvs/principais_resultados.pdf)> Accessed on: 23 APR 2021.  
451
- 452 28. Castro MC, Kim S, Barberia L, Ribeiro AF, Gunzerda S, Ribeiro KB. Spatiotemporal pattern  
453 of COVID-19 spread in Brazil. *Science.* 2021;eabh1558.  
454
- 455 29. Bermudi PMM, Lorenz C, Aguiar BS, Failla MA, Barrozo LV, Chiaravalloti-Neto F.  
456 Spatiotemporal ecological study of COVID-19 mortality in the city of São Paulo, Brazil:  
457 Shifting of the high mortality risk from areas with the best to those with the worst socio-  
458 economic conditions. *Travel Medicine and Infectious Disease.*2021;39: 101945.  
459
- 460 30. Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial Analysis of COVID-19:  
461 A Scoping Review. *Int J Environ Res Public Health.* 2021 Feb 27;18(5):2336.  
462
- 463 31. Tracking Universal Health Coverage: 2017 Global Monitoring Report – The World Bank.  
464 Available from:  
465 <[https://www.worldbank.org/en/topic/universalhealthcoverage/publication/tracking-universal-](https://www.worldbank.org/en/topic/universalhealthcoverage/publication/tracking-universal-health-coverage-2017-global-monitoring-report)  
466 [health-coverage-2017-global-monitoring-report](https://www.worldbank.org/en/topic/universalhealthcoverage/publication/tracking-universal-health-coverage-2017-global-monitoring-report) > Accessed on: 23 APR 2021.  
467
- 468 32. Mendonça MJ, Schilling C. Family Health Strategy, a strong model of Primary Health Care  
469 that brings results. *Debate Health.* 2018;42: 18-37.  
470
- 471 33. Ventura, D F L, et al. Challenges of the COVID-19 pandemic: for a Brazilian research  
472 agenda on global health and sustainability. *Cad. Public Health.*2020;36,4.  
473

- 474 34. Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani. Impact of non-pharmaceutical  
475 interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College  
476 London.2020.  
477
- 478 35. Sudre CH. Symptom clusters in Covid19: A potential clinical prediction tool from the  
479 COVID Symptom study app. MedRxiv 2020.06.12.20129056.  
480
- 481 36. Surendra, H.Clinical characteristics and mortality associated with COVID-19 in Jakarta,  
482 Indonesia: A hospital-based retrospective cohort study Surendra, Henry et al. The Lancet  
483 Regional Health. 2020;9: 100108.  
484
- 485 37. Garcia-Vidal, C. Trends in mortality of hospitalised COVID-19 patients: A single centre  
486 observational cohort study from Spain. The Lancet Regional Health.2020;3, 100041.  
487
- 488 38. Ioannou GN, Green P, Fan VS. Development of COVIDVax Model to Estimate the Risk of  
489 SARS-CoV-2–Related Death Among 7.6 Million US Veterans for Use in Vaccination  
490 Prioritization. JAMA Netw Open. 2021;4(4):e214347.  
491
- 492 39. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in  
493 COVID-19 infection using machine learning. Mod Pathol. 2021 Mar;34(3):522-531.  
494
- 495 40. Haimovich AD, Ravindra NG, Stoytchev S, Young HP, Wilson FP, van Dijk D, Schulz WL,  
496 Taylor RA. Development and Validation of the Quick COVID-19 Severity Index: A Prognostic  
497 Tool for Early Clinical Decompensation. Ann Emerg Med. 2020;76(4):442-453.  
498
- 499 41.Sudre CH, Lee KA, Lochlain MN, Varsavsky T, Murray B, Graham MS et al. Symptom  
500 clusters in Covid-19: A potential clinical prediction tool from the COVID Symptom study app.  
501 MedRxiv 2020; 129056.  
502
- 503 42. Hallal, PC. SOS Brazil: science under attack. The Lancet. 2020;397:373–374.  
504
- 505 43. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the  
506 study of COVID-19. A review. Sci Total Environ. 2020. 15;739:140033.  
507
- 508 44. Sisó-Almirall A, Kostov B, Mas-Heredia M, Vilanova-Rotllan S, Sequeira-Aymar E, Sans-  
509 Corrales M, Sant-Arderiu E, Cayuelas-Redondo L, Martínez-Pérez A, García-Plana N,  
510 Anguita-Guimet A, Benavent-Àreu J. Prognostic factors in Spanish COVID-19 patients: A case  
511 series from Barcelona. PLoS One. 2020;15(8):e0237960.  
512
- 513 45. Covino M, De Matteis G, Santoro M, Sabia L, Simeoni B, Candelli M, Ojetti V, Franceschi  
514 F. Clinical characteristics and prognostic factors in COVID-19 patients aged  $\geq 80$  years. Geriatr  
515 Gerontol Int. 2020;20(7):704-708.  
516  
517  
518  
519

520 **Supporting information**

521

522 **S1 Fig. Spatial distribution of the health care facilities.**

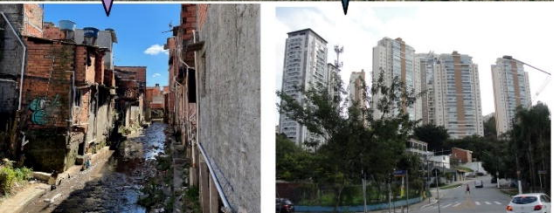
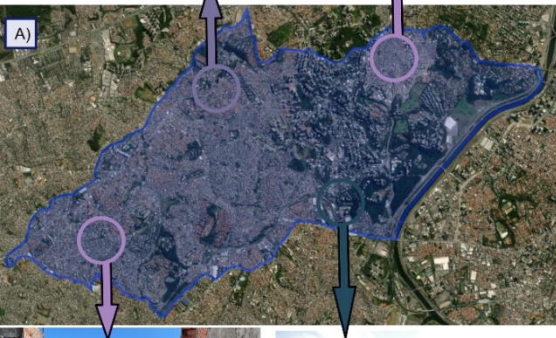
523

524 **S2 Text. STROBE Check-list.**

525

526





B) Social vulnerability index



C) GeoSES index





Database

79,688 (up to  
Feb 9, 2021)

Official databases for  
reporting of suspect  
COVID-19 cases

SARS (SIVEP-  
Gripe): 4,053

ILI (esus-VE):  
75,635

Positive PCR for SARS-  
CoV-2, single, 18 years  
and older

1,925

14,136

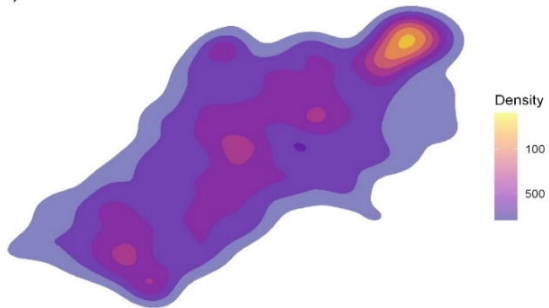
With outcome of  
interest available

discharge/death:  
1,659

discharge/death:  
14,136

A)

COVID-19 cases density



B)

COVID-19 incidence (%)



C)

SARS cases (%)



D)

Lethality (%)



# Predicted probability of SRAG by COVID-19



# Predicted probability of death by COVID-19

