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

Ana Carolina C. N. Mafra¹*, Régis Rodrigues Vieira¹, Camila N. Monteiro¹, Denise F. B. Cavalcante¹, João L. Miraglia¹, Daiana Bonfim¹, Danielle C. Palacio¹, Alessandra C. F. Martins¹, Letícia Yamawaka de Almeida¹, João Peres Neto¹

¹ Hospital Israelita Albert Einstein, São Paulo, São Paulo, Brazil.

* Corresponding author

Email: ana.mafra@einstein.br (ACCNM)
Abstract

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

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

Results: A total of 16,061 patients with confirmed COVID-19 were enrolled. Morbidities associated with a higher risk of SARS were obesity (OR=25.32) and immunodepression (OR=12.15). Morbidities associated with a higher risk of death were renal disease (OR=11.8) and obesity (OR=8.49), and clinical and demographic data were more important than the territory per se. Based on the data, a calculator was developed to predict the risk of death from COVID-19, with 92.2% accuracy in the development sample, 92.3% in the internal validation sample, and 80.0% in the temporal validation sample.
Conclusions: The risk factors for SARS and death in COVID-19 patients seeking health care, in order of relevance, were age, comorbidities, and socioeconomic factors, considering each discriminatory power.
Introduction

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

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

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

In the last five years, Brazil has undergone structural changes leading to more severe inequalities, which have strongly impacted the response to the pandemic. The following changes stand out: the spending ceiling that has frozen health care financing starting in 2016 through the next 20 years; changes in the National Primary Care Policy, in 2017, leading to uncertainty with flexibilization of the need for community health workers, and weakening of the Family Health Strategy, one of the pillars of PHC in the country. Additionally, in 2019 the program “Previne
Brasil” [Prevent Brazil] was launched and changed the PHC funding scenario, increasing doubts about the ability of SUS to keep its universal nature, at the risk of reducing PHC to a selective model, based on focal targets [4].

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

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

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

Despite the known social inequalities among countries and even regions within the same country, and the need for differentiated responses in highly vulnerable and resource-scarce settings, little is known about the specific realities of virus spread and the risk factors present in slums. The data extrapolated is from regions with vastly different conditions, which may not
correspond to the real world [11]. Hence, this study aimed to assess the space distribution and factors associated with the risk of SARS and death from COVID-19, based on routinely collected register data, as well as to develop and validate a predictive model of risk of death from COVID-19.

**Methods**

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

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

**Fig 1. Study territory.** A: Satellite image and photographs of the region. B: Distribution of the São Paulo Social Vulnerability Index. C: Distribution of the GeoSES Socioeconomic Index.
The distribution of the health care facilities, privat and public, accessed by the study population can be visualized in a map available in S1 Fig. Health care facilities were categorized as privat or public according to the National Registry of Health Care Facilities (CNES, in Portuguese), and was considered in the analyses as an indicator of health care access, since public health services are universally accessible, while only those with health insurance or high income have access to private services.

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

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

The independent variables were sociodemographic characteristics, comorbidities, and symptoms. The following inclusion criteria were used: to be a resident of the study region, aged 18 years and older, with laboratory confirmation of COVID-19 by PCR. When the same patient had multiple records, and ILI had been reported previously to SARS, only the SARS record was considered (Fig 2).
Results were presented as absolute and relative frequencies, overall and by outcome (death or SARS). For simple comparison between outcome groups, the chi-square test was used for categorical variables and Student’s t-test was used for numerical variables.

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

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

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

Since the multiple model presented a good discriminatory power, a predictive model for death was developed, with the main clinical and demographic characteristics of patients arriving at health care facilities for COVID-19 evaluation. In this analysis, the symptoms were also considered, and we followed the same method of variable selection presented above. The data was randomly split in two sets with a 2:1 ratio for training and testing, respectively. The predictive
performance of the model was assessed by accuracy measurements, discriminatory power and the Hosmer-Lemeshow test [16]. The temporal validation used data from patients over 18 years of age, with a positive COVID-19 test and who had symptoms onset between March and April of 2021.

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

## Results

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

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

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Missing rate (%)</th>
<th>SARS (yes)</th>
<th>C SARS</th>
<th>Death (yes)</th>
<th>C Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>16,061</td>
<td>1,925</td>
<td></td>
<td></td>
<td>375</td>
<td></td>
</tr>
<tr>
<td>Profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>Private</td>
<td>18 to 40 years</td>
<td>40 to 60 years</td>
<td>60 to 80 years</td>
<td>80 years and older</td>
<td>Sex</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Facility</td>
<td>Type of care</td>
<td>42.10 (14.94)</td>
<td>56.93 (16.77)**</td>
<td>68.08 (14.48)**</td>
<td>0.90</td>
<td>Age Range</td>
</tr>
<tr>
<td>Indig.</td>
<td>Native</td>
<td>7892 (49.1)</td>
<td>337 (17.5)**</td>
<td>18 (4.8)**</td>
<td>0.64</td>
<td>Race</td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
<td>6022 (37.5)</td>
<td>736 (38.2)</td>
<td>82 (21.9)</td>
<td>0.53</td>
<td>Age Range</td>
</tr>
<tr>
<td>Female</td>
<td>Male</td>
<td>1830 (11.4)</td>
<td>653 (33.9)</td>
<td>184 (49.1)</td>
<td>0.57</td>
<td>Race</td>
</tr>
<tr>
<td>Indigenous</td>
<td>Caucasian</td>
<td>317 (2.0)</td>
<td>199 (10.3)</td>
<td>91 (24.3)</td>
<td>0.55</td>
<td>Age Range</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>7230 (45.0)</td>
<td>1073 (55.7)</td>
<td>222 (59.2)</td>
<td>0.53</td>
<td>Race</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>6142 (50.9)</td>
<td>638 (48.6)</td>
<td>148 (50.7)</td>
<td>0.53</td>
<td>Age Range</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>5925 (49.1)</td>
<td>675 (51.4)</td>
<td>144 (49.3)</td>
<td>0.53</td>
<td>Race</td>
</tr>
<tr>
<td>Diabetes (y)</td>
<td>Yes</td>
<td>997 (6.5)</td>
<td>5.0</td>
<td>506 (44.9)**</td>
<td>0.71</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Obesity (y)</td>
<td>Yes</td>
<td>187 (1.2)</td>
<td>5.0</td>
<td>118 (10.5)**</td>
<td>0.55</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Immunodepression (y)</td>
<td>Yes</td>
<td>145 (0.9)</td>
<td>5.0</td>
<td>62 (5.5)**</td>
<td>0.52</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Heart disease (y)</td>
<td>Yes</td>
<td>1494 (9.8)</td>
<td>5.0</td>
<td>720 (63.8)**</td>
<td>0.79</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Renal disease (y)</td>
<td>Yes</td>
<td>122 (0.8)</td>
<td>5.0</td>
<td>85 (7.5)**</td>
<td>0.54</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Respiratory Disease (y)</td>
<td>Yes</td>
<td>418 (2.7)</td>
<td>5.0</td>
<td>125 (11.1)**</td>
<td>0.55</td>
<td>Symptoms</td>
</tr>
<tr>
<td>Loss of smell</td>
<td>Yes</td>
<td>1487 (10.2)</td>
<td>9.1</td>
<td>125 (27.3)**</td>
<td>0.59</td>
<td>Economic</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>Facility</td>
<td>1386 (9.5)</td>
<td>9.4</td>
<td>96 (22.8)**</td>
<td>0.57</td>
<td>Type of care</td>
</tr>
<tr>
<td>Private</td>
<td>Public</td>
<td>6999 (51.6)</td>
<td>15.5</td>
<td>1079 (58.3)**</td>
<td>0.54</td>
<td>Type of care</td>
</tr>
<tr>
<td>Public</td>
<td>Public</td>
<td>6570 (48.4)</td>
<td>772 (41.7)</td>
<td>199 (55.3)</td>
<td>0.53</td>
<td>Type of care</td>
</tr>
</tbody>
</table>
C: C-statistic, representing the area under the ROC curve. Measures reported as absolute frequency (%), when mean not indicated (standard deviation). * p value for hypothesis testing <0.05. ** p value for hypothesis testing <0.001. Tests performed: chi-square and Student’s t.

Geospatial distribution

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

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

In map “A”, Fig 3, there is clearly an area of higher density of positive COVID-19 cases, which overlaps with the area of the largest slums in the city of São Paulo, Paraisópolis, with a predominantly young and vulnerable population living in overcrowded households. However, the incidence of positive COVID-19 cases, SARS and deaths in this area are not high, when compared to other areas in the region (Figs “3B”, “3C” and “3D”).
The data on progression to SARS was reviewed for the 16,061 individuals included in the study. After excluding individuals with missing information on comorbidities, the multiple model was adjusted based on data from 15,235 patients.

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

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

<table>
<thead>
<tr>
<th>Factors</th>
<th>OR (95%CI)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>1.063 (1.057;1.069)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sex (Male)</td>
<td>1.61 (1.36;1.91)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Obesity (y)</td>
<td>25.32 (16.31;39.31)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Immunodepression (y)</td>
<td>12.15 (7.62;19.37)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Renal disease (y)</td>
<td>10.12 (5.81;17.62)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Heart disease (y)</td>
<td>9.79 (8.21;11.68)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Respiratory disease (y)</td>
<td>6.72 (4.81;9.40)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes (y)</td>
<td>6.70 (5.51;8.15)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>s (Longitude, Latitude)</td>
<td>Fig 4</td>
<td>0.721</td>
</tr>
</tbody>
</table>
Fig 4. Predicted probability of progression to SARS by the space-only model. N=16,601.

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

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

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

<table>
<thead>
<tr>
<th>Factors</th>
<th>OR (CI 95%)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>1.082 (1.071; 1.092)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sex (Male)</td>
<td>1.68 (1.27; 2.21)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Renal disease (y)</td>
<td>11.8 (7.1; 19.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Obesity (y)</td>
<td>8.49 (5.21; 13.85)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Immunodepression (y)</td>
<td>7.68 (4.22; 13.97)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Heart disease (y)</td>
<td>4.41 (3.26; 5.97)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Respiratory disease (y)</td>
<td>3.50 (2.23; 5.51)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes (y)</td>
<td>3.06 (2.29; 4.09)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
GeoSES (0.1 unit)        0.896 (0.844; 0.951)    <0.001
s (Longitude, Latitude)  Fig 5            0.657

Fig 5. Predicted probability of death (case fatality rate/100) by the space-only model.
N=15,795.

Predictive model for death from COVID-19

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

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

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

Predicted risk (or probability) = 1 / (1+ e^{-Risk score})

A risk calculator was developed, and it is accessible at: https://redcap.einstein.br/surveys/?s=PFLPC9JJ8F

To set a cutoff with a better balance between sensitivity, specificity and accuracy, the ROC curve was used to obtain a 0.0186 predicted probability with 92.2% specificity, 94.1% sensitivity and 92.2% accuracy. The C-statistic observed was 0.974 (95% CI: 0.967–0.981), demonstrating an excellent discrimination.
The training sample included 170 events, which is sufficient for the estimation of the coefficients of the 9 variables included in the final model. No collinearity between variables was evident, with the variance inflation factors ranging from 1.01 to 1.17. The maximum values for the estimated coefficient standard errors were 0.48 for the intercept, and 0.38 for immunodepression. The p value for the Hosmer-Lemeshow test was 0.485 indicating good model fit.

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

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

**Discussion**

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

The risk factors associated with severe acute respiratory syndrome (SARS) were age, obesity and immunodepression, a similar finding to those of systematic reviews that evaluated risk factors associated with severe COVID-19 [21-25]. However, none of these reviews evaluated the association of environmental factors, such as living in slums, overcrowded households or in
precarious sanitary conditions, with COVID-19 severity. Some observational studies have demonstrated a higher death prevalence among vulnerable and lower-income populations, but there is no solid evidence for this association [6,26].

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

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

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

In Paraisópolis, the largest slum of the study region, there was an evident higher density of cases, in agreement with other Brazilian studies that found greater virus spread in highly vulnerable and densely populated locations [28,29]. However, a high incidence of severe cases and deaths was found in two areas of low vulnerability also presenting low density and overall
incidence of cases, which is in accordance with the findings of the study by Bermudi et al. [29], showing high mortality in the richest areas of the city of São Paulo.

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

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

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

The proposed risk model for death had good performance, including a C-statistic of 97.4% (95% CI: 96.7%–98.1%), which is higher than the value found by another calculator for a population of elders in the USA, with a C-statistic of 85.3%, while needing more variables [38]. Booth et al. [39] proposed a different strategy to predict deaths associated with COVID-19 employing molecular biomarkers measured in laboratory samples used for PCR tests; however, this approach presents limitations related to cost and availability of results in a timely manner, in
addition a lower performance when compared to the model proposed by this study, with a C-
statistic of 93%, sensitivity of 91% and specificity of 91%. Additional studies have focused only
on the elderly or hospitalized individuals [40,41].

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

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

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

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

Acknowledgments

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

References


Supporting information

S1 Fig. Spatial distribution of the health care facilities.

S2 Text. STROBE Check-list.
Database

79,688 (up to Feb 9, 2021)

Official databases for reporting of suspect COVID-19 cases

SARS (SIVEP-Gripe): 4,053
ILL (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