

The spatial distribution and Calculator for Severity and Death for COVID-19

Camila Nascimento Monteiro (✉ c.nascimentomonteiro@gmail.com)

Hospital Israelita Albert Einstein

Ana Carolina Cintra Nunes Mafra

Hospital Israelita Albert Einstein

Regis Rodrigues Vieira

Hospital Israelita Albert Einstein

Denise de Fátima Barros Cavalcante

Hospital Israelita Albert Einstein

João Luiz Miraglia

Hospital Israelita Albert Einstein

Daiana Bonfim

Hospital Israelita Albert Einstein

Danielle da Costa Palacio

Hospital Israelita Albert Einstein

Alessandra Cristina Ferreira Martins

Hospital Israelita Albert Einstein

Ruchelli França de Lima

Hospital Israelita Albert Einstein

Etienne Larissa Duim

Hospital Israelita Albert Einstein

Leticia Yamawaka de Almeida

Hospital Israelita Albert Einstein

Lorrayne Belotti

Hospital Israelita Albert Einstein

Renata Soares Martins

Hospital Israelita Albert Einstein

João Peres Neto

Hospital Israelita Albert Einstein

Research Article

Keywords: COVID-19, mortality, SARS-CoV-2, prognostic factors

Posted Date: January 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1228607/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Latin America has one of the highest COVID-19 mortality rates in the world, driven by inequality in its population. We aimed to analyze the spatial distribution and factors associated with the risk of SARS and death in COVID-19 cases using routine record data and to develop and validate a prognostic tool for risk of death by Covid-19.

Methods: A cross-sectional study was conducted from March 2020 to April 2021 in the South Zone of the city of São Paulo, SP, Brazil, with 16,061 positive cases of COVID-19. The data were obtained from the records of the Brazilian Ministry of Health notification systems for flu-like syndrome (eSUS-VE) and hospitalized SARS (SIVEP-Gripe). The spatial distribution of the cases is represented in 2D kernel density. To assess the possible factors associated with the outcomes, generalized linear and generalized additive logistic models were adjusted. To assess the discriminatory power the C-statistic was used. A calculator was developed based on a prognostic model for the risk of death, validated with accuracy measures in the sample, internal validation and temporal validation.

Results: The average age of patients was 42.1 years. Evolved to SARS 925 (11.98%) and 375 (2.37%) died. The comorbidities associated with a higher risk of SARS were obesity (OR=25.32) and immunodepression (OR=12.15). The comorbidities associated with a higher risk of death were renal diseases (OR=11.8) and obesity (OR=8.49), with clinical and demographic information being more important than the spatial area. The COVID-19 risk of death calculator showed an accuracy of 92.2% in the sample build, 92.3% in the internal validation and 80% in the temporal validation.

Conclusions: Age and comorbidities were identified as the most associated factors to the severity of the disease. For the death analysis, the socioeconomic condition is included in addition to these factors. The calculator can be used and implemented in services of varying complexity because it contains easily accessible information to guide prevention and care.

Background

In December 2019 the first cases of SARS-CoV-2 infection were recorded in Wuhan, China. After 16 months, COVID-19 has affected over 220 million people and caused more than 4.6 million deaths globally¹. In Brazil, the first case occurred on February 26, 2020, and in merely six months it accounted for 100,000 deaths spread across every state in the nation. By September 14, 2021, the country had surpassed 21 million cases of the disease and over 500,000 deaths. In São Paulo, more than 1.4 million people were infected and made up about 6% of the registered deaths in the country^{1,2}.

Studies have been carried out to investigate the factors associated with COVID-19 infection, the severity of the disease and consequent death by COVID-19³. In a systematic review and meta-analysis, Du et al. show that the main factors associated with the severity of the illness and death globally are elderly and the presence of comorbidities such as chronic respiratory diseases, diabetes, hypertension, chronic kidney disease and cardiovascular diseases, but the focus of these studies were patients who had already been hospitalized.

Additionally, studies⁴ show that there is a large inequality in the distribution of the disease in different demographic and socioeconomic contexts, with poorer and younger populations⁵ being more affected. The populations with a higher risk of death are elderly, male⁶, black or brown race, lower income, and education levels, but assessed on an ecological, not individual⁷⁻⁹ basis.

In spite of the vast amount of literature on the factors associated with COVID-19, there are few studies that consider socioeconomic factors with individual data that can be controlled by age and the presence of comorbidities and that can aid decision-making in the context of population health. From this perspective, to assist in making clinical decisions, some predictive tools have been proposed to assess the risk of severity and death by COVID-19. However, the tools are restricted to elderly or hospitalized populations or they need laboratory support to assess the markers¹⁰⁻¹³.

We aimed to evaluate the risk factors for severity and death by COVID-19 in regions with social inequality and to develop a tool from individual data that helps in evaluating the risk of death by COVID-19.

Methods

A cross-sectional study was conducted from March 2020 to April 2021 in the South Zone of the city of São Paulo, SP, Brazil, including 16,061 men and women. The eligibility criteria included individuals aged 18 years or more with positive cases of COVID-19.

The area is composed of two administrative districts—Campo Limpo and Vila Andrade—that have an estimated population of nearly 400,000 inhabitants¹⁴ (Figure 1A). In this region, social inequality was measured by the Paulista Index of Social Vulnerability (Índice Paulista de Vulnerabilidade Social - IPVS)¹⁵, on a scale from 0, the lowest vulnerability (navy blue), to 6, high vulnerability (purple) (Figure 1B). The socioeconomic level was evaluated using the Geographic Socioeconomic Index for Health Studies (Índice Socioeconômico do Contexto Geográfico para Estudos em Saúde - GeoSES)¹⁶ in which lighter areas indicating lower values reflect conditions of lesser income, education levels and mobility (Figure 1C). This index ranges from -1 (lowest socioeconomic level) to 1 (highest socioeconomic level), and in the area of this study ranged from -0.56 and 0.3.

The data were obtained from the Brazilian Ministry of Health information systems: (1) mild cases of the disease was extracted from eSUS Epidemiological Monitoring System (eSUS-VE) for flu-like syndrome (FLS) notifications, considered to be and notified by any type of health care service; and (2) information about the Severe Acute Respiratory Syndrome (SARS), e.g., from patients who were hospitalized due to the disease, was extracted from Epidemiological Monitoring Information System for the Flu (SIVEP-Gripe), with, As the system was adapted during the period of the study, to include standardized fields for some symptoms and comorbidities, a search in the open fields for “other symptoms” and “other comorbidities” was performed to identify this information from instances prior to the standardization.

To analyze the use of different health care services sought by patients in the study, a map showing the services categorized as private or public was included in Supplementary Material 1. They were categorized according to registrations in the National Registry of Health Establishments (Cadastro Nacional de

Estabelecimentos de Saúde - CNES). This is an important indicator related to health care access because everyone has access to public services, but only individuals with medical insurance or good financial conditions can access private services.

The patients' residences were geocoded by the GISA/CEINFO - SMS, identified in the area, and made available by the Campo Limpo Health Monitoring Unit. The database was made anonymous therefore no information could be used to identify the patients in the study.

This project was approved by the Research Ethics Committee at the Hospital Israelita Albert Einstein (4.462.994) and the São Paulo Municipal Health Department (4.648.956).

The studied population were individuals with a confirmed COVID-19 diagnosis via the polymerase chain reaction (PCR) method. This criterion provided the scientific rigor necessary for this study, despite limited access to laboratory results during some periods of the pandemic. The criteria for inclusion in the study were: individuals residing in the area of the study, over age 18, and laboratory confirmation of COVID-19 via the PCR method. Duplicate notifications from the same patient were excluded when they had both a flu syndrome notification and a SARS notification; only the SARS notification was kept (Figure 2).

The outcome variables studied were: 1) evolution for SARS (categorized by FLS or SARS) and 2) death (discharged from hospital or death).

The independent variables considered were: socioeconomic characteristics, comorbidities and symptoms, obtained from the official notification systems; and the IPVS and GeoSES socioeconomic indicators, obtained from the patient's residential address, and the address geocoded in latitude and longitude.

The data was showed in absolute and relative frequency. For a simple comparison between outcome groups, the chi-squared hypothesis test and t-Student test were used. The spatial distribution of the studied cases is represented in 2D kernel density.

To assess the possible factors associated with the outcomes, generalized linear models and generalized logistic model were adjusted, the latter to include the spatial location of the patient's residence. First, the simple models and variables considered important were adjusted, with the p-value at a maximum of 0.2 with less than 10% missing, they were considered for the multiple models. Imputed data was not used; only variables with complete data were used in the analysis. With the aim of assessing factors prior to the disease to better direct public policies and preventative actions, e.g. vaccination, the symptoms were not considered in this model in order not to mix the results with other factors. The variables for the final model were selected using the backward stepwise selection method with a final p-value of <0.05. In the final model the quality of the adjustment was verified by the standard magnitude of errors, the influence of the inclusion or exclusion of variables in the estimated odds ratio, and the variance inflation factor, which did not surpass 1.2. To evaluate the discriminatory power of each variable and the final model, the C Statistic (the ROC curve area) was performed.

Since the multiple model reached good discrimination, a risk of death calculator was created using the logistic model for death with the principal clinical and demographic characteristics of the patients at the time that they accessed the health service. For this, the data was randomly separated into two datasets using a proportion of

2:1: a set for the build and a set for the internal validation of the model. In this analysis the symptoms were considered, for the purpose of guiding actions to be taken after the disease is identified. The final multiple model was obtained by following the same steps as the previous one, and the calculator was obtained by the probability predicted by the logistic model¹⁷.

The prognostic performance of the model was assessed by measures of accuracy, sensitivity, specificity, discriminatory power and the Hosmer and Lemeshow test¹⁷. For the temporal validation, data was used from COVID-19 positive patients over age 18 who began presenting symptoms between March and April 2021. This database contains information from 3,798 patients; 104 of whom died of COVID-19 during the period of the study.

R software, version 3.6.3 with the mgcv, ggplot2 and viridis packages, was used. A level of significance of 5% was adopted¹⁸⁻²¹. High-resolution Google Earth and Esri World Imagery (World Imagery (arcgis.com)) were also used.

Results

The study analyzed 16,061 patients 18-103 years old, 54.8% female with positive cases of COVID-19 between March 2020 and Evolved to SARS 925 (11.98%) and 375 (2.37%) died. February 2021. The average age of patients was 42.1 years (standard-deviation of 14.9) and a median age of 40.

Table 1 shows the demographic and socioeconomic characteristics of the patients, the missing, distribution of the outcomes, and C statistic.

Table 1
 Characteristics of the patients. N=16,061 positive COVID-19 patients.

	Overall	Missing (%)	SARS (yes)	C SARS	Death (yes)	C Death
N	16,061		1,925		375	
Profile						
Average age(sd)	42.10 (14.94)	0.0	56.93 (16.77)**	0.78	68.08 (14.48)**	0.90
Age Group (yrs)		0.0		0.75		0.86
18 to 40	7892 (49.1)		337 (17.5)**		18 (4.8)**	
40 to 60	6022 (37.5)		736 (38.2)		82 (21.9)	
60 to 80	1830 (11.4)		653 (33.9)		184 (49.1)	
80 or more	317 (2.0)		199 (10.3)		91 (24.3)	
Sex		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)	
Race		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)	
Asian	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)	
White		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)	

C: C statistic, which represents the area above the ROC curve. Measurements expressed in absolute frequency (%) when average is not indicated (standard-deviation). * P-value for hypotheses tests <0.05. ** P-value for hypotheses tests <0.001. Tests performed: Chi-squared and t-Student.

	Overall	Missing (%)	SARS (yes)	C SARS	Death (yes)	C Death
Morbidities						
Diabetes (Yes)	997 (6.5)	5.0	506 (44.9)**	0.71	154 (49.4)**	0.72
Obesity (Yes)	187 (1.2)	5.0	118 (10.5)**	0.55	44 (14.1)**	0.56
Immunodepression (Yes)	145 (0.9)	5.0	62 (5.5)**	0.52	26 (8.3)**	0.53
Heart disease (Yes)	1494 (9.8)	5.0	720 (63.8)**	0.79	215 (68.9)**	0.80
Renal Disease (Yes)	122 (0.8)	5.0	85 (7.5)**	0.54	53 (17.0)**	0.58
Respiratory disease (Yes)	418 (2.7)	5.0	125 (11.1)**	0.55	46 (14.7)**	0.56
Pregnant (Yes)	78 (0.5)	7.9	11 (1.7)**	0.51	0 (0.0)	0.50
Symptoms						
Fever (Yes)	6868 (43.8)	2.3	1211 (78.0)**	0.69	235 (73.2)**	0.65
Cough (Yes)	8354 (53.0)	1.9	1352 (83.3)**	0.67	271 (81.9)**	0.64
Sore throat (Yes)	3755 (24.6)	5.1	246 (22.3)	0.51	55 (20.3)	0.52
Dyspnea (Yes)	3567 (22.7)	2.0	1332 (82.7)**	0.83	293 (85.4)**	0.82
Diarrhea (Yes)	231 (1.5)	5.2	231 (21.1)**	0.61	65 (23.6)**	0.61
Vomit (Yes)	125 (0.8)	5.5	125 (12.0)**	0.56	32 (12.5)**	0.55
Abdominal pain (Yes)	43 (0.3)	9.5	43 (10.8)**	0.55	7 (7.7)**	0.53
Fatigue (Yes)	157 (1.1)	9.2	157 (35.4)**	0.68	18 (18.8)**	0.58
Loss of smell (Yes)	1487 (10.2)	9.1	125 (27.3)**	0.59	12 (12.5)	0.51
Loss of taste (Yes)	1386 (9.5)	9.4	96 (22.8)**	0.57	11 (12.1)	0.51
Socioeconomic						
Type of health service accessed						
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)	

C: C statistic, which represents the area above the ROC curve. Measurements expressed in absolute frequency (%) when average is not indicated (standard-deviation). * P-value for hypotheses tests <0.05. ** P-value for hypotheses tests <0.001. Tests performed: Chi-squared and t-Student.

	Overall	Missing (%)	SARS (yes)	C SARS	Death (yes)	C Death
IPVS avg(sd)	2.68 (1.60)	0.9	2.57 (1.54)*	0.52	2.78 (1.52)	0.53
0 to 3	12619 (79.3)	0.9	1555 (81.2)*	0.51	287 (76.7)	0.51
4 to 6	3302 (20.7)		360 (18.8)		87 (23.3)	
GeoSES avg (sd)	-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)	

C: C statistic, which represents the area above the ROC curve. Measurements expressed in absolute frequency (%) when average is not indicated (standard-deviation). * P-value for hypotheses tests <0.05. ** P-value for hypotheses tests <0.001. Tests performed: Chi-squared and t-Student.

The main variable that contributed with death and SARS was age, presence of symptoms of dyspnea, fever and cough, and the presence of morbidities heart disease and diabetes.

Figure 3 shows the locations with a higher density of cases identified in the area. In an area with one of the largest favelas in the city of São Paulo, Paraisópolis, a higher density of positive COVID-19 cases was seen. This area is characterized by a large, dense population that is vulnerable and predominantly young.

The data of evolution to SARS in 16,061 patients was studied. Patients without information on morbidities were excluded. The multiple model was adjusted based on information from 15,235 patients (94.9% completeness).

In the simple model, the information of the patient's address showed a p-value of <0.001, areas under curve (AUC) of 0.57 and an explained deviation of 0.74% (Figure 4).

In the final multiple model, the clinical and demographic information were more important than the area information, which ceased to be significant (Table 2). The morbidities associated with a higher risk of SARS were obesity and immunodepression. No socioeconomic information remained significant or had sufficient data for the multiple model. The model reached an AUC of 0.96 and an explained deviation of 50%.

Table 2

Multiple model of factors associated with SARS. N=15,235.

Factors	OR (CI 95%)	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)	Figure 4	0.721

The data of 15,795 patients were analyzed with the outcome of interest being the evolution to death. Excluding patients without information on morbidities, the multiple model was adjusted based on the information of 15,100 patients (completeness of 95.6%).

In the simple model, the information of the patient's residence showed a p-value of 0.001, AUC of 0.62 and explained deviance of 1.8% (Figure 5).

In the final multiple model, the clinical and demographic information were more important than the information about the area, which ceased to be significant (Table 3). The morbidities associated with a higher risk of death were renal diseases and obesity. In this model, the GeoSES index remained significant, indicating that the lower the socioeconomic level, the greater the risk of death. The complete model reached an AUC of 0.96 and explained deviance of 46.5%.

Tabel 3

Multiple model of factors associated with death. 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)	Figure 5	0.657

The COVID-19 risk of death calculator

To develop the 'COVID-19 risk of death calculator', a prognostic model using 9,931 observations was performed

The socioeconomic and geolocation data were not considered because of the difficulty of obtaining it at the time of the patient's clinical evaluation. The possibility of including symptoms was considered. The final model for the risk score is

Risk score = $-10.37 + (0.075 \times \text{Age, in years}) + (0.98 \times \text{Diabetes}) + (1.21 \times \text{Obesity}) + (1.89 \times \text{Immunodepression}) + (0.99 \times \text{Heart disease}) + (2.95 \times \text{Chronic renal disease}) + (1.18 \times \text{Respiratory disease}) + (0.78 \times \text{Fever}) + (1.81 \times \text{Dyspnea})$, in which the variables are input as 1 if there is the condition, and 0 if there is not.

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

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

The tool is available at: https://redcap.einstein.br/surveys/?s=DMDT7TCJ9R_

To establish a cutoff point with the best balance between sensitivity, specificity and accuracy, the ROC curve was used to obtain the point of 0.0186 of predicted probability with 92.2% specificity, 94.1% sensitivity and 92.2% accuracy. The observed AUC was 0.974 (95% CI: 0.967 - 0.981), indicating excellent discrimination.

Regarding the quality of the model, in the construction sample there were 170 events, sufficient for the inclusion of 17 variables, but only 9 were important, and thus, sufficient. The variance inflation factors were between 1.01 and 1.17, showing that there was no collinearity among the observed factors. The standard

errors of the estimated coefficients had a maximum value of 0.48 for the intercept and 0.38 for immunodepression. The Hosmer and Lemeshow test showed a p-value of 0.485, indicating that the predicted is not different from the observed in a significant way.

For the internal validation, a sample of 4,966 observations was used, in which 92 events occurred (10.2 events per variable). In this sample, the predicted model was applied, and the cut-off point was maintained, reaching a sensitivity of 92.2%, specificity of 98.9% and accuracy of 92.3%. The AUC was 0.985 (95% CI: 0.978 - 0.991), maintaining the discriminatory quality.

In the temporal validation, an accuracy of 80.04%, sensitivity of 98.08% and specificity of 79.53% were observed. The AUC was 0.936 (95% CI: 0.922 – 0.950).

Discussion

The study analyzed the spatial distribution and factors associated with the risk for SARS, risk of death due COVID-19 and a proposal for a risk of death calculator. The risk factors associated with SARS were age, obesity and immunodepression. Our results are similar to the systematic reviews that assessed the most prevalent risk factors for the aggravation of COVID-19²²⁻²⁶. However, none of these reviews included the environmental factor as a relevant aspect to the outcome of the disease, whether for the aggravation of or death by COVID-19. Some observational studies demonstrated a greater prevalence of mortality among vulnerable populations, but there is no solid evidence of this association^{27,28}.

The second result of this study was the increased risk of mortality, with the descending order of importance: age, comorbidities, sex, and situation of vulnerability. Among the comorbidities most associated with a greater risk of mortality, in order from highest to lowest, were: renal disease, obesity, immunodepression, heart disease, respiratory sickness and diabetes.

The distribution of the density of confirmed COVID-19 cases in the area suggests that the virus spreads independent of the status of vulnerability of the region, being greater in more populated and vulnerable areas. A reflection on the vulnerability of specific groups is necessary, since the impact of the pandemic on the area needs to be contextualized, due to the diversity of urban social vulnerability scores that affect the largest Brazilian city.

In Paraisópolis, one of the largest favela of São Paulo, a higher density of confirmed COVID-19 cases was observed, as is also reported by other Brazilian studies showing a greater spread of the virus in situations of high vulnerability and population density^{29,30}. However, we clearly see two areas with high socioeconomic levels that, in spite of low incidences, show a high chance of aggravation and death. This corroborates the study by Bermudi et al.³⁰ that shows a high mortality rate in areas with the worst social conditions in São Paulo.

The findings of the study corroborate with the importance of control of chronic conditions and care for the elderly in health care services for mortality reduction³¹. There is also evidence for the need to prioritize such actions in regions that present higher vulnerability and adverse regional conditions. Considering the fact that the difficulties inherent in the region of residence affect health conditions, causing unfavorable outcomes for

these populations, The Equity, one of the guidelines of the Unified Health System (SUS), should be implemented.

Despite of is plausible that poverty worsens health conditions³², this should not avoid controlling the risk factors that can undergo more immediate interventions, as well as strategies that provide isolation for the most vulnerable. SUS, free and universal, and the strong Primary HealthCare across the country are transformative for this reality^{33,34}. In countries who does not have universal healthcare coverage, such as the USA, there has been a high mortality rate for COVID-19³⁵.

In a prediction model for more serious forms of COVID-19 there is the recommendation for the inclusion of symptomology for medical support, resource planning and improved monitoring of COVID-19 patients¹³. The initial characteristics of the disease, including the number of symptoms, is predictive of its duration^{36,37}. Our study identified cough, fever, sore throat, dyspnea, anosmia, loss of taste, diarrhea and fatigue as the main symptoms of COVID-19, considering the area above the ROC curve (AUC) significant, for both aggravated SARS and death.

The third result is the risk of death calculator, based on the symptoms, comorbidities and age, with good performance parameters. The model presented a better performance (AUC of 97.4%) when compared to the parameters of the calculator proposed by the American study (AUC of 85.3%) that used a population restricted to elderly people and a larger number of variables¹⁰.

Booth et al.³⁹, also proposed a tool to predict death, considered molecular biomarkers in laboratory samples from PCR exams, but presented a limitation related to the cost and time to obtain the measurement of risk, including quality measures that are inferior to this study: AUC (93%), sensitivity (91%) and specificity (91%). Other studies focused only on the elderly or hospitalized cases^{13,39}. However, broadening the discussion to include adults over 18 years old makes it stronger for general population use.

Thus, the proposed tool is a powerful response to support managers and professionals in health care service planning and the prioritization of more directed and assertive strategies and actions, such as monitoring positive cases with a higher prediction of death and directing and distributing resources.

We highlight the importance of this calculator as a tool to aid any health care service, including Primary Health Care Service. Professionals at Primary Health Care Services often lack consolidated directives that can help decision making. Based on the indicators and results obtained by the calculator, professionals can estimate the risk of death and anticipate preventative measures to avoid this outcome. Since it was performed using data collected and available from the Brazilian Ministry of Health information systems, it can be easily validated externally and even recalibrated for different national contexts. In other countries that use different systems, it is also possible to validate and calibrate, as data on age, sex, comorbidities and symptoms are accessible.

In addition to the risk factors highlighted by the results of this study, many countries are in ethical, political and economic crises, such as Brazil has been experiencing regarding a failure to respond, management policies against social distancing, a lack of federal coordination, negationism and neoliberal policies that can impact the outcomes of the pandemic. Thus, the response to the pandemic in these contexts needs to be questioned

and larger studies are needed that incorporate these variables into the analyses, such as the need for interdisciplinary analyses^{40–42} considering clinical, demographic, socioeconomic and geospatial aspects, and incorporating the different aspects into a single model.

This study has several strengths, like the use of data from the Ministry of Health information systems on the records of patients with a confirmed COVID-19 diagnosis, including flu-like syndrome, GGeoSES, IPVS encompassing a large number of participants, and a proposal for a risk of death calculator, a tool to aid any health care service. Despite the richness of the data, this study has limitations. First, the study used data that had not been collected with a scientific purpose; healthcare professionals collected the data with a high percentage of missing data, however it is still robust enough to support our results. During the period of the study, the principal variant in circulation changed—in January the P1 variant was introduced and the vaccination campaign began. However, upon performing the temporal validation, the measures reduced, but still showed excellent discrimination. The risk calculator shows great potential to be evaluated regarding its implementation.

In conclusion, the study contributed to the global effort in the fight against COVID-19, with evidence that brings together various contexts in the spread of the virus, as well as risk factors present in areas of greater vulnerability, including those with a high population density and poverty, such as favelas.

Abbreviations

FHS – Family Health Strategy

IBGE – Brazilian Institute of Geography and Statistics

PCU – Primary Care Units

PHC – Primary Health Care

SBIBAE - *Sociedade Beneficente Israelita Brasileira Albert Einstein*

SMS/PMSP - Municipal Department of Health of São Paulo

GeoSES - *Índice Socioeconômico do Contexto Geográfico para Estudos em Saúde*

Declarations

Ethics approval and consent to participate

The study was performed in accordance with guidelines and regulations from Brazilian National Health Council. We confirm that we have been exempt applying the consent form because we used de-identifying and secondary data. All protocols and the research project were approved by the Research Ethics Committee of the Municipal Health Department of São Paulo, Brazil (Protocol CAAE 32368720.1.0000.0071) (project name: *COVID-19 e fatores de risco associados em uma população acompanhada pela Atenção Primária à Saúde no Sistema Único de Saúde*), according to Resolution 466/12 of the Brazilian National Health Council.

Consent for publication

Not applicable.

Availability of Data and Materials

The data that support the findings of this study are available from São Paulo Municipal Health Department, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Data are however available from the authors upon reasonable request and with permission of São Paulo Municipal Health Department and the Research Ethics Committee at the Hospital Israelita Albert Einstein and the Research Ethics Committee of the São Paulo Municipal Health Department.

Competing interests

The authors declare that they have no competing interests.

Funding

The author(s) received no specific funding for this work.

Authors' contributions

ACCNM, CNM, JLM, DB and JPN were responsible for conception and study design. ACCNM set up the database and did all statistical analyses, figures and maps. ACCNM and RRV took the photographs of the region. ACCNM, RRV, CNM, DFBC and JPN drafted the manuscript. JLM, DB, ACFM, RFL, ELD, LYA, RSM, LB and DCP commented on the data and its interpretation and critically revised the content. CNM verified the data and validated the analysis. All authors contributed to the final revision of the manuscript and approved the submission.

Acknowledgements

Not applicable

References

1. Map COVID-19 - Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html> (2021). Accessed 23 April 2021.
2. Boletim Diário COVID-19 no município de São Paulo de 22 de abril de 2021 (nº 392). https://www.prefeitura.sp.gov.br/cidade/secretarias/upload/saude/20210422_boletim_covid19_diario.pdf

- (2021). Accessed 23 April 2021.
3. Du P, Li D, Wang A, Shen S, Ma Z, Li X. A Systematic Review and Meta-Analysis of Risk Factors Associated with Severity and Death in COVID-19 Patients. *Can J Infect Dis Med Microbiol.* 2021;2021:6660930.
 4. Ahmed F, Ahmed N, Pissarides C, Stiglitz J. Why inequality could spread COVID-19. *The Lancet Public Health.* 2020;5(5):e240.
 5. Rocha R, Atun R, Massuda A, Rache B, Spinola P, Nunes L, et al. Effect of socioeconomic inequalities and vulnerabilities on health-system preparedness and response to COVID-19 in Brazil: a comprehensive analysis. *The Lancet Global Health.* 2021;9(6):782–792.
 6. Ribeiro KB, Ribeiro AF, Veras MASM, Castro MC. Social inequalities and COVID-19 mortality in the city of São Paulo, Brazil. *Int J Epidemiol.* 2021;50(3):732–742
 7. Li SL, Pereira RHM, Prete Jr CA, Zarebski AE, Emanuel L, Alves PJH, et al. Higher risk of death from COVID-19 in low-income and non-White populations of São Paulo, Brazil. *BMJ Global Health* 2021;6:e004959.
 8. Bernardo Y, do Rosario D, Conte-Junior C. COVID-19 Pandemic in Rio de Janeiro, Brazil: A Social Inequality Report. *Medicina (Kaunas).* 2021;57(6):596.
 9. Ferreira MC. Spatial association between the incidence rate of COVID-19 and poverty in the São Paulo municipality, Brazil. *Geospat Health.* 2020 Nov 26;15(2).
 10. Ioannou GN, Green P, Fan VS, et al. Development of COVIDVax Model to Estimate the Risk of SARS-CoV-2–Related Death Among 7.6 Million US Veterans for Use in Vaccination Prioritization. *JAMA New Open.* 2021;4(4):e214347.
 11. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Mod Pathol.* 2021;34(3):522–531.
 12. Haimovich AD, Ravindra NG, Stoytchev S, Young HP, Wilson FP, van Dijk D, Schulz WL, Taylor RA. Development and Validation of the Quick COVID-19 Severity Index: A Prognostic Tool for Early Clinical Decompensation. *Ann Emerg Med.* 2020;76(4):442–453.
 13. Sudre CH, Lee KA, Lochlain MN, Varsavsky T, Murray B, Graham MS et al. Symptom clusters in Covid-19: A potential clinical prediction tool from the COVID Symptom study app. *Sci Adv.* 2021;7(12):eabd4177.
 14. Fundação SEADE. São Paulo State Statistics Portal. <https://www.seade.gov.br/seade-disponibiliza-projecoes-de-populacao-e-domicilios-ate-2050> (2021). Accessed 23 April 2021.
 15. Pavarini SC, Barha EJ, Mendiondo MS, Filizola CL, Petrilli Filho JF, Santos AA. Family and social vulnerability: a study with octogenarians. *Rev Lat Am Enfermagem.* 2009 May-Jun;17(3):374–9.
 16. Barrozo LV, Fornacialli M, André ACS, Morais GAZ, Mansur G, et al. GeoSES: A socioeconomic index for health and social research in Brazil. *PLoS One.* 2020;15(4):e0232074.
 17. Royston P, Luas KGM, Altman D G, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009; 338:b604.
 18. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
 19. Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1):3–36.
 20. H. Wickham. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2016.

21. Simon Garnier (2018). viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1. <https://CRAN.R-project.org/package=viridis>.
22. Fadini GP, Morieri ML, Boscari F, Fioretto P, Maran A, Busetto L, et al. Newly-diagnosed diabetes and admission hyperglycemia predict COVID-19 severity by aggravating respiratory deterioration. *Diabetes Res Clin Pract.* 2020;168:108374.
23. Yang, J, Hu, J, Zhu, C. Obesity aggravates COVID-19: A systematic review and meta-analysis. *J Med Virol.* 2021; 93: 257–261.
24. Yang, Jing et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases*, Volume 94, 91–95.
25. Gao, Ya et al. Impacts of immunosuppression and immunodeficiency on COVID-19: A systematic review and meta-analysis. *Journal of Infection*, Volume 81, Issue 2, e93 - e95.
26. Geng, MJ., Wang, LP, Ren, X. *et al.* Risk factors for developing severe COVID-19 in China: an analysis of disease surveillance data. *Infect Dis Poverty* 10, 48 2021.
27. Khan M M A, Khan M N, Mustagir M G, Rana J, Islam M S, Kabir M I. Effects of underlying morbidities on the occurrence of deaths in COVID-19 patients: A systematic review and meta-analysis. *J Glob Health.* 2020 Dec;10(2):020503. doi: 10.7189/jogh.10.020503. PMID: 33110586; PMCID: PMC7567434.
28. Baqui, P et al. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *The Lancet Global Health*, Volume 8, Issue 8, e1018 - e1026.
29. Castro M C, Kim S, Barberia L, Ribeiro A F, Gunzerda S, Ribeiro K B et al. Spatiotemporal pattern of COVID-19 spread in Brasil. *Science.* 2021;372(6544):821–826.
30. Bermudi PMM, Lorenz C, Aguiar BS, Failla MA, Barrozo LV, Chiaravalloti-Neto F. Spatiotemporal ecological study of COVID-19 mortality in the city of São Paulo, Brazil: Shifting of the high mortality risk from areas with the best to those with the worst socio-economic conditions. *Travel Medicine and Infectious Disease.* 2021;39:101945.
31. Fatima M, O'Keefe KJ, Wei W, Arshad S, Gruebner O. Geospatial Analysis of COVID-19: A Scoping Review. *Int J Environ Res Public Health.* 2021;18(5):2336.
32. Tracking Universal Health Coverage: 2017 Global Monitoring Report – The World Bank. <https://www.worldbank.org/en/topic/universalhealthcoverage/publication/tracking-universal-health-coverage-2017-global-monitoring-report> (2017). Accessed 23 April 2021.
33. Mendonça MJ, Schilling C. Family Health Strategy, a strong model of Primary Health Care that brings results. *Debate Health.* 2018;42(8):18–37.
34. Ventura DFL, Ribeiro H, Giulio GM, Jaime PC, Nunes J, et al. Challenges of the COVID-19 pandemic: for a Brazilian research agenda on global health and sustainability. *Cad. Public Health.* 2020;36(4):e00040620.
35. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslei K, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London. doi: <https://doi.org/10.25561/77482>.
36. Surendra H, Elyazar IR, Djaafara BA, Ekawat LL, Saraswati K, et al. Clinical characteristics and mortality associated with COVID-19 in Jakarta, Indonesia: A hospital-based retrospective cohort study. *The Lancet Regional Health – Western Pacific.* 2021;9:100108.

37. Garcia-Vidal C, Cozar-Llistó A, Meira F, Dueñas G, Puerta-Alcalde P, et al. Trends in mortality of hospitalised COVID-19 patients: A single centre observational cohort study from Spain. *The Lancet Regional Health – Europe*. 2021;3:100041.
38. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Mod Pathol*. 2021;34(3):522–531.
39. Hallal, PC. SOS Brazil: science under attack. *The Lancet*. 2021;397(10272):373–374.
40. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. *Sci Total Environ*. 2020;739:140033.
41. Sisó-Almirall A, Kostov B, Mas-Heredia M, Vilanova-Rotllan S, Sequeira-Aymar E, Sans-Corrales M, et al. Prognostic factors in Spanish COVID-19 patients: A case series from Barcelona. *PLoS One*. 2020;15(8):e0237960.
42. Covino M, Matteis G, Santoro M, Sabia L, Simeoni B, Candelli M, Ojetti V, Franceschi F. Clinical characteristics and prognostic factors in COVID-19 patients aged ≥ 80 years. *Geriatr Gerontol Int*. 2020;20(7):704–708.

Figures

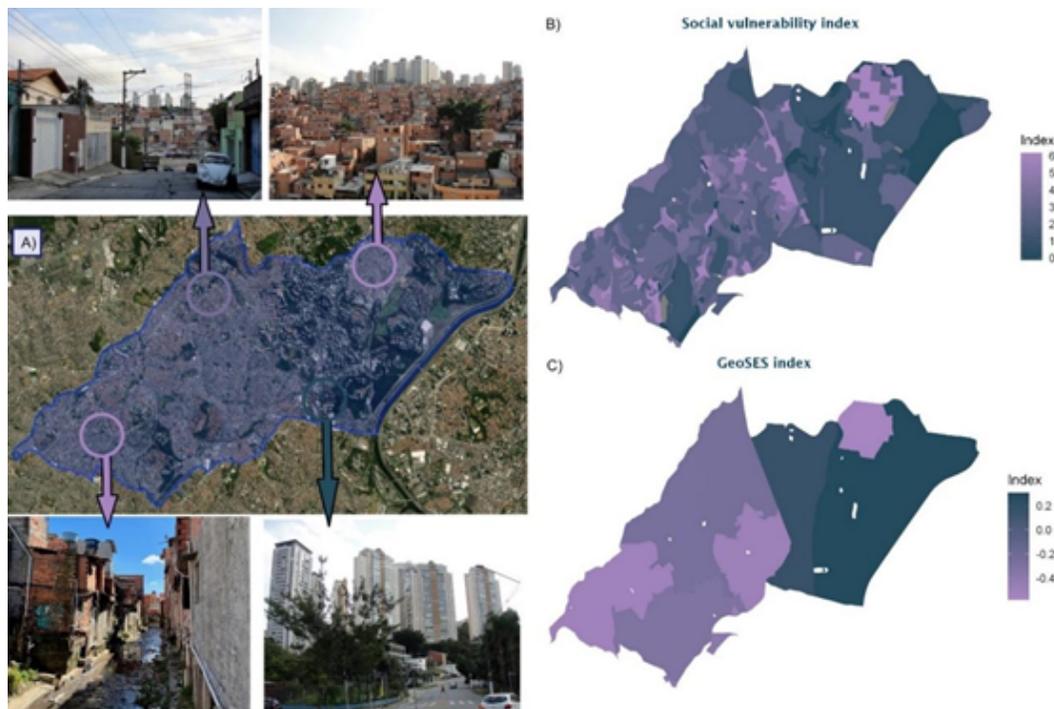


Figure 1

Area of the study. A: Satellite image and photos to portray the region. B: Distribution of the Paulista Index of Social Vulnerability. C: Distribution of the GeoSES socioeconomic index.

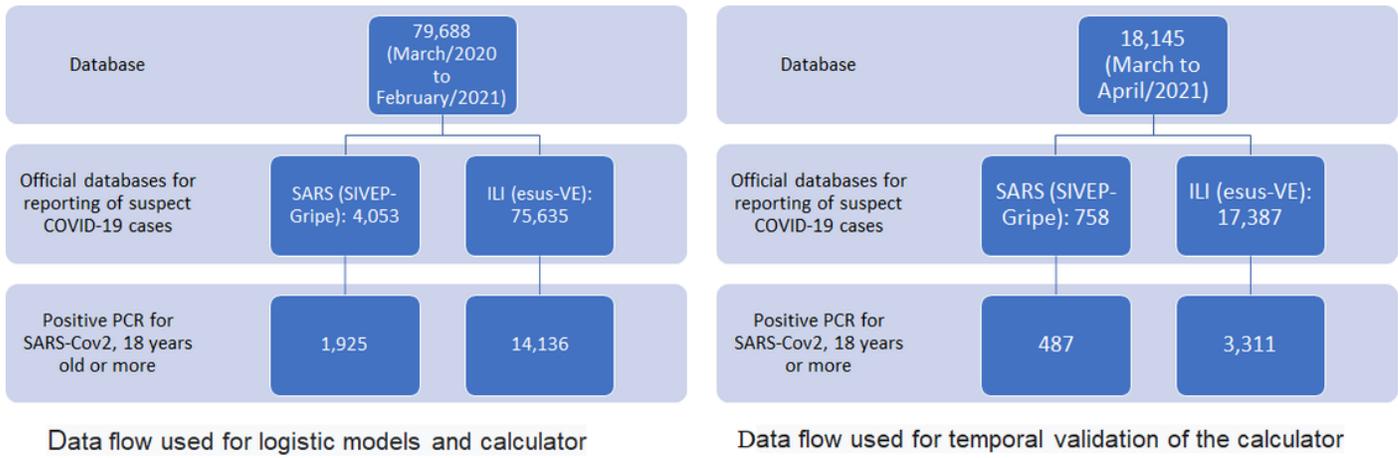


Figure 2

Flow of inclusion and exclusion of the patients.

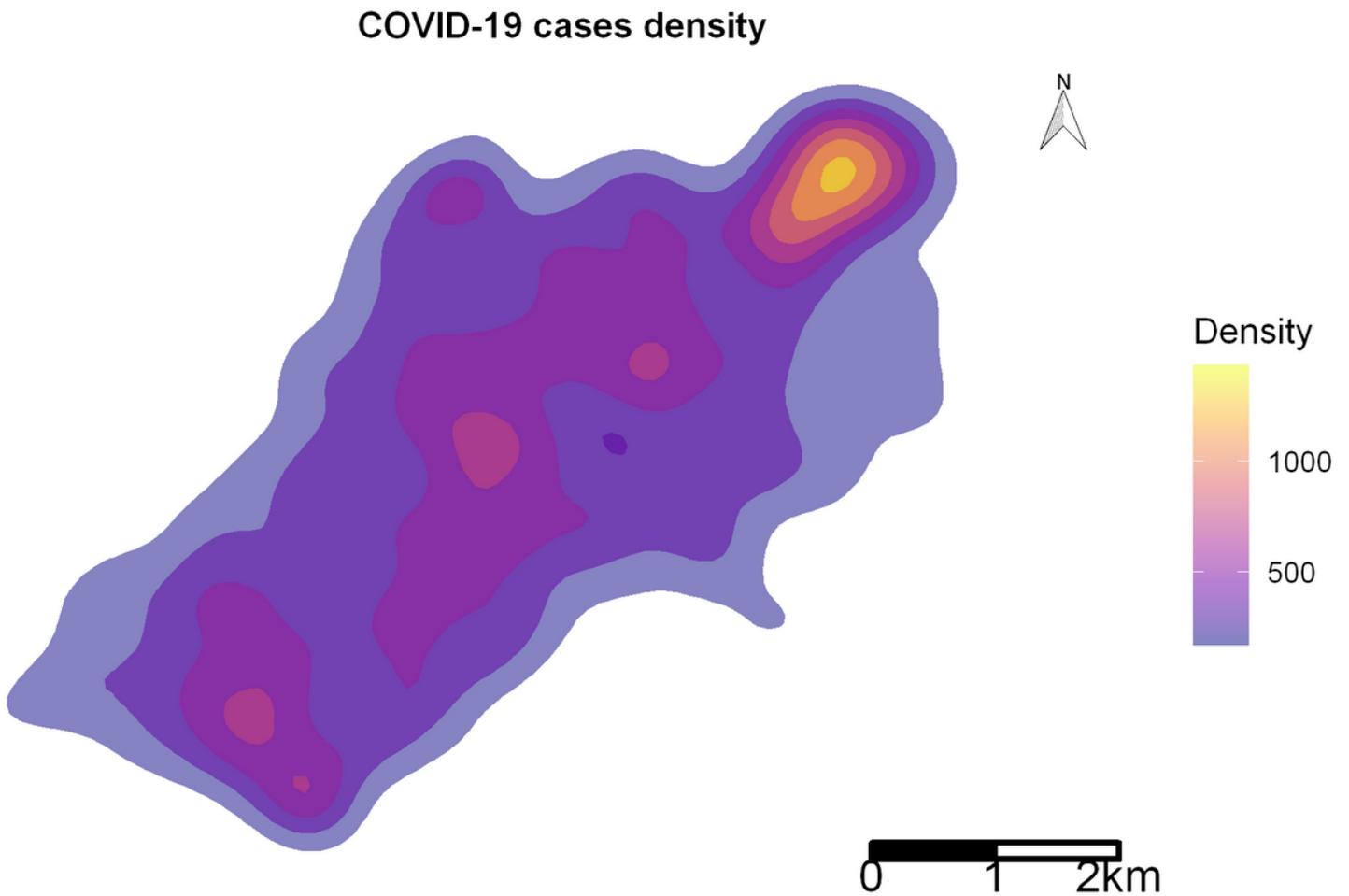


Figure 3

Density of positive COVID-19 cases in the area

Predicted probability of SARS by COVID-19

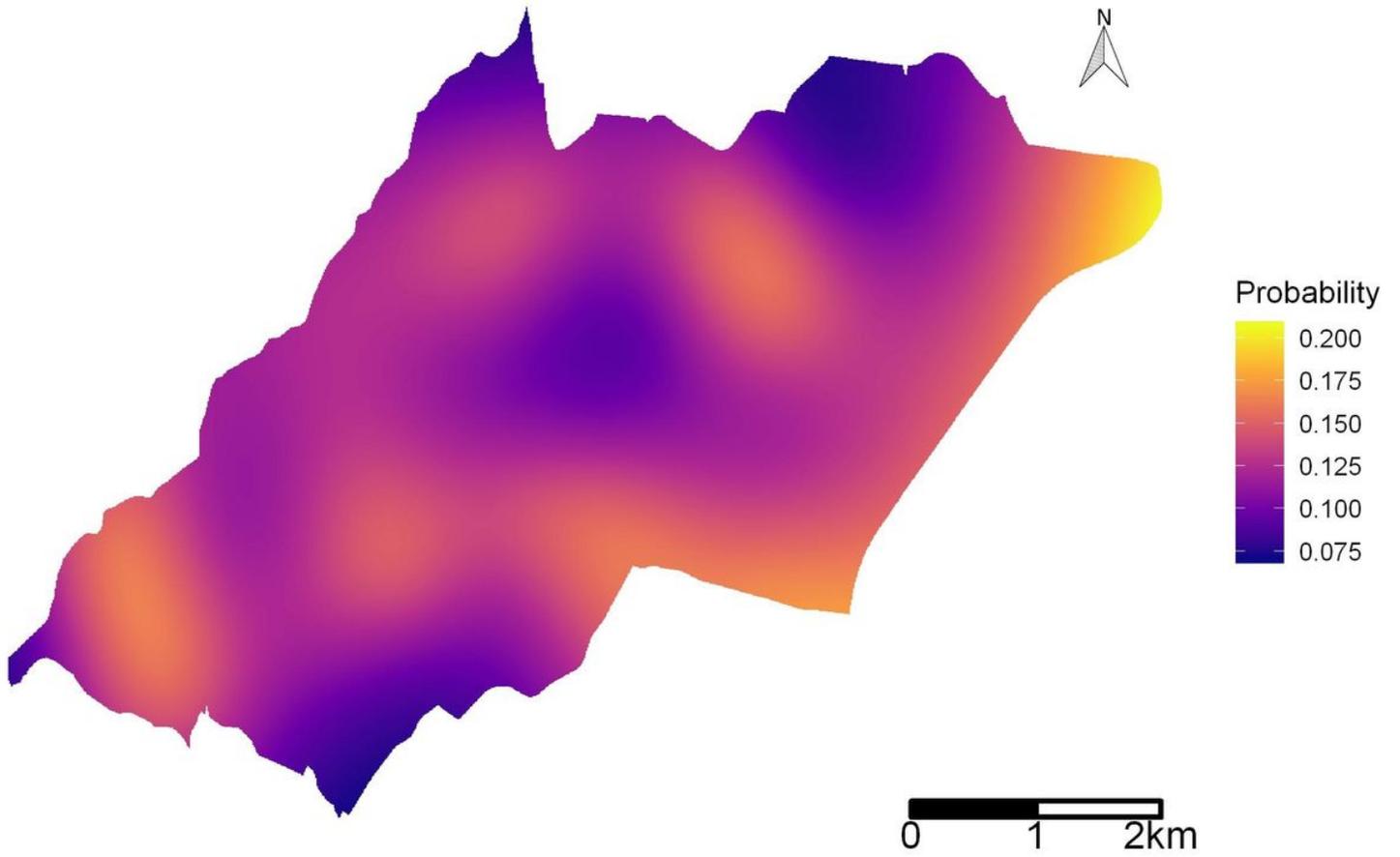


Figure 4

Predicted probability of evolution to SARS by the spatial-only model. N=16,601 .

Predicted probability of death by COVID-19

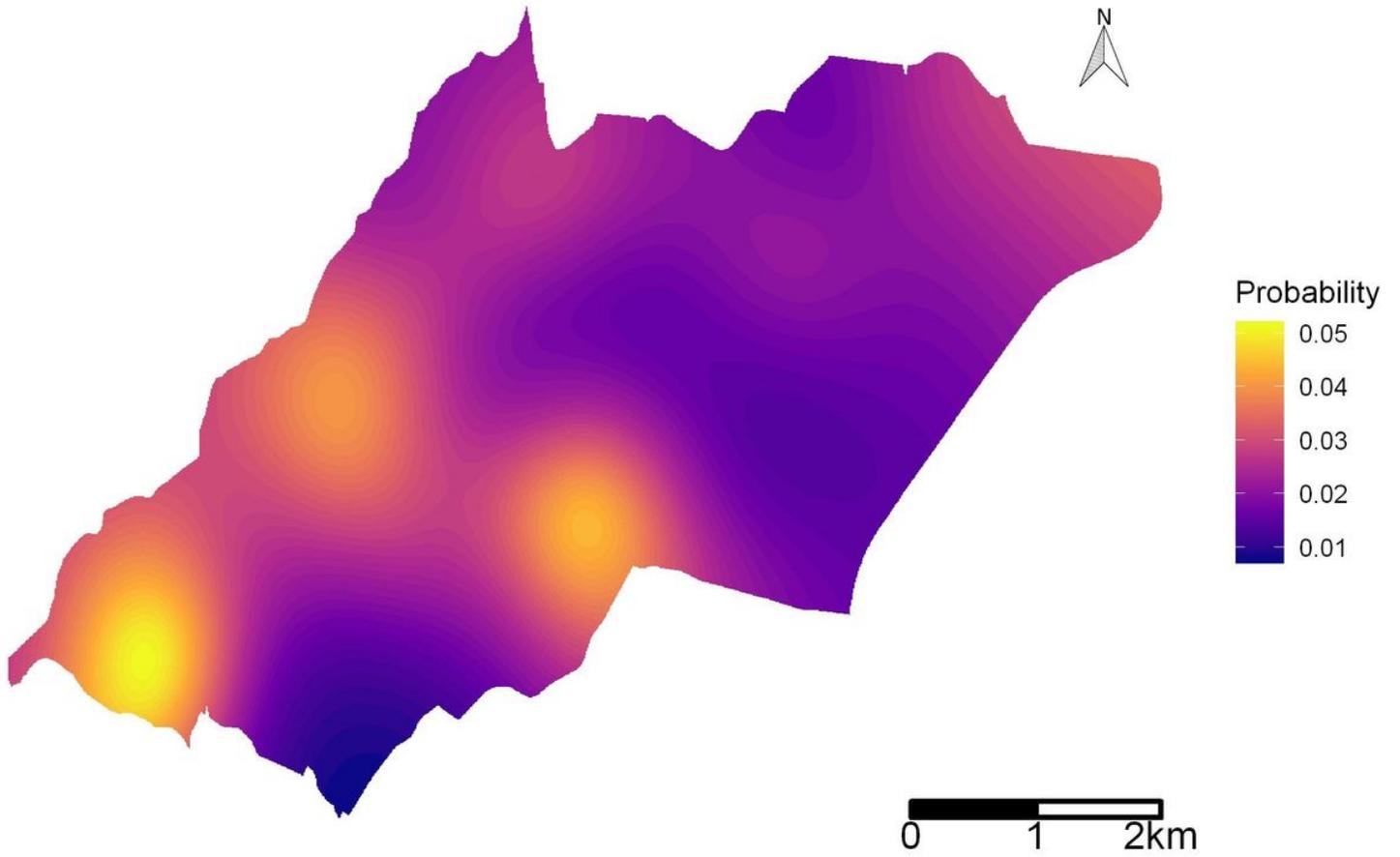


Figure 5

Predicted probability of death (fatality/100) by the spatial-only model. N = 15,795.