Article

Not peer-reviewed version

Space-time analysis of the COVID-19 pandemic and its relationship with socioeconomic and demographic variables in the Metropolitan Region of São Paulo, Brazil

Keila Valente de Souza de Santana * , Aluízio Marino , Gabriela Rosa Martins , Pedro Henrique Barbosa Muniz Lima , Pedro Henrique Rezende Mendonça , Raquel Rolnik

Posted Date: 16 August 2024

doi: 10.20944/preprints202408.1082.v1

Keywords: COVID-19, space-time cluster, the Metropolitan Region of São Paulo, population density, socioeconomic variables

Article

Space-Time Analysis of the COVID-19 Pandemic and Its Relationship with Socioeconomic and Demographic Variables in the Metropolitan Region of São Paulo, Brazil

Keila Valente de Souza de Santana [1,*], Aluízio Marino [1], Gabriela Rosa Martins [2], Pedro Henrique Barbosa Muniz Lima [1], Pedro Henrique Rezende Mendonça [3] and Raquel Rolnik [1]

[1] Faculty of Architecture and Urban Planning, University of São Paulo, São Paulo 05508-080, Brazil; aluizio.marino@usp.br (A.M.); pedro.barbosa.lima@usp.br (P.H.B.M.L.); raquelrolnik@usp.br (R.R.)
[2] Faculty of Philosophy, Literature and Human Sciences of the University of São Paulo, University of São Paulo, São Paulo 05508-080, Brazil; gmartins@usp.br
[3] Institute of Mathematics and Statistics, University of São Paulo, São Paulo 05508-090, Brazil; pedro.rezende.mendonca@alumni.usp.br
* Correspondence: keilla@usp.br

Abstract: Geographic Information Systems (GIS) played an important role in understanding the dynamics of the territorial spread of COVID-19 during the pandemic. Even today, it is important to understand such dynamics in view of the imminent risk of new epidemics and pandemics. This study aimed to identify space-time clusters of incidence and mortality due to COVID-19 during the pandemic, analyzing socioeconomic and demographic characteristics in order to highlight the priority areas for control and surveillance actions. This study was conducted in the Metropolitan Region of São Paulo (MRSP) from March 2020 to February 2022. To detect the clusters, a multidimensional punctual process was created to perform multiple tests for every geographic point using the SaTScan™ software. Socioeconomic and demographic differences were analyzed using nonparametric Mann-Whitney and Kruskal-Wallis tests. High-risk clusters of incidences were observed in May 2020 (RR=2; p-value <0.001) and March to June 2021 (RR=2.6; p-value <0.001) in the capital of São Paulo and neighboring municipalities, with statistically significant differences between the socioeconomic variables analyzed. No low-risk cluster of COVID-19 incidence was found, but low-risk clusters of mortality from COVID-19 were found in the period from July to December 2020 in the central region of the capital (RR=0.33; p-value <0.001), which concentrates the highest incomes and the lowest percentages of Black, Brown (Mixed-race), and Indigenous people in the MRSP.

Keywords: COVID-19; space-time cluster; the Metropolitan Region of São Paulo; population density; socioeconomic variables

1. Introduction

The pandemic of COVID-19 (coronavirus disease 2019), an infectious disease caused by the highly transmissible novel coronavirus SARS-CoV-2 (Severe Acute Respiratory Syndrome - Coronavirus - 2), was first detected in December 2019 in the People's Republic of China [1]. In Brazil, the first case was identified in the Metropolitan Region of São Paulo (MRSP) in February 2020, with a rapid progression of cases throughout the state [2].

During and after the COVID-19 pandemic, as in other epidemics, Geographic Information Systems (GIS) can play an important role in understanding spatial clusters and trends of virus transmission in various parts of the world [3]. Also, GIS can contribute to the visualization and correlation of information on socioeconomic, demographic, and health factors.

This study sought to identify space-time clusters of incidence and mortality from COVID-19 in the MRSP, identifying high- and low-risk areas over time. The detection of space-time clusters has been used in health surveillance to support the planning and evaluation of health services [4]. After identifying space-time clusters of incidence and mortality, the socioeconomic and demographic characteristics of these clusters were also analyzed in order to highlight priority areas for COVID-19 control and surveillance actions. Then, this study seeks to contribute to the development of territorial scenarios that can guide and enable strategic actions. If they had been created opportunely, these scenarios could have reduced the problems resulting from the spread of the disease during the pandemic.

2. Study Site

The MRSP is the largest urban cluster in South America and the sixth in the world, according to a 2014 United Nations (UN) report [5]. The MRSP is made up of 39 municipalities (Figure 1), including São Paulo, the capital of the state of São Paulo, which are distributed in the five sub-regions below:

i. North: Caieiras, Cajamar, Francisco Morato, Franco da Rocha, and Mairiporã.
ii. East: Arujá, Biritiba-Mirim, Ferraz de Vasconcelos, Guararema, Guarulhos, Itaquaquecetuba, Mogi das Cruzes, Poá, Salesópolis, Santa Isabel, and Suzano.
iii. Southeast: Diadema, Mauá, Ribeirão Pires, Rio Grande da Serra, Santo André, São Bernardo do Campo, and São Caetano do Sul.
iv. Southwest: Cotia, Embu, Embu-Guaçu, Itapecerica da Serra, Juquitiba, São Lourenço da Serra, Taboão da Serra, and Vargem Grande Paulista.
v. West: Barueri, Carapicuíba, Itapevi, Jandira, Osasco, Pirapora do Bom Jesus, and Santana de Parnaíba.

The capital is home to 12,330,000 inhabitants; it is considered a global city and is the largest Brazilian metropolis. In addition to São Paulo, the municipalities of Guarulhos, São Bernardo, Santo André, and Osasco have significant populations exceeding 700,000 inhabitants. Two cities in the region have the highest population densities in Brazil: Taboão da Serra and Diadema, with 13,400 and 12,800 inhabitants per km², respectively [6].

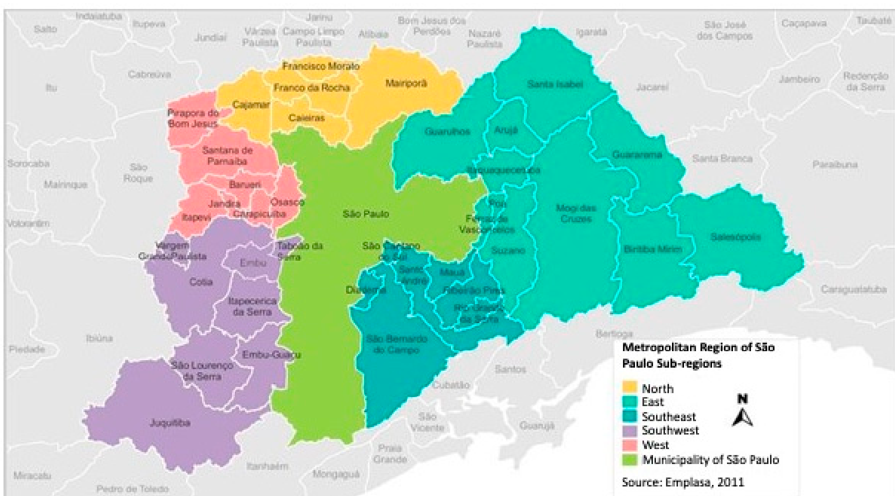


Figure 1. Map showing the municipalities of the MRSP.

3. Materials and Methods

This is an ecological and descriptive study assessing secondary data about the incidence and mortality of COVID-19 in the 633 weighting areas of the 39 municipalities that comprise the MRSP. Weighting areas (WA) are territorial units identified by sets of contiguous census sectors belonging to the same district, for the purpose of weighting the results of the population census sample questionnaire. A census sector is a territorial unit established for survey control purposes, comprising a continuous area located in a single urban or rural block, with a size and number of households that allow the survey by a census agent [7].

Information from the period February 2020 to March 2022 about the date of notification, sex, age, evolution, and postal code of each patient with COVID-19 who recovered or died, was accessed through a partnership with the Data Center of the State of São Paulo (CDESP), which provided data from the Epidemiological Surveillance System (SIVEP-Gripe) of the State Epidemiological Surveillance Center. These data were grouped by postal code (first five digits of the postal code) and georeferenced using the postal code database of the *Centro de Estudos da Metrópole* (Center for Metropolitan Studies) [8]. Linear geometries of the postal code grouping system were intersected with the weighting areas from the 2010 IBGE Census, with cases assigned proportionally to the length of the intersected lines. This study was approved by the Research Ethics Committee of the School of Psychology, Universidade de São Paulo, report number CAAE: 71605223.2.0000.5561, August 14, 2023.

The socioeconomic variables – per capita income, persons per household, and percentage of Black, Brown (Mixed-race), and Indigenous people (BBIP) – by PA were built based on data from the Brazilian Institute of Geography and Statistics (IBGE), according to the 2010 census, the most recent census available. The variables were selected according to bibliographic references that analyzed the relationship between COVID-19 spread and socioeconomic factors [9,10].

Dasymetric mapping techniques were used to analyze the population density, which subdivide areas of origin into smaller spatial units so that there is greater internal consistency of the variable being mapped [11]. In this study, the variable of population density was calculated by dividing the number of inhabitants in WA by the total area built for residential purposes in that area. This analysis used Google Open Buildings, a large-scale open dataset that contains the vectorization of building roof contours generated from a deep learning model that was trained to determine building areas from high-resolution satellite images. Data is available under the Creative Commons Attribution license (CC BY-4.0) and the Open Data Commons Open Database License (ODbL) v1.0 [12].

COVID-19 incidence and mortality rates were obtained for the 39 municipalities of the MRSP from March 2020 to February 2021 and from March 2021 to February 2022. These rates were standardized by age and sex using the 2010 population of the MRSP. Space-time and time variations in the clusters were determined using scan statistics, a multidimensional point process that performs multiple tests for each geographical point analyzed, in SaTScan v10.0 [13]. To analyze space-time variations, COVID-19 cases and deaths were grouped by month. The method uses a discrete Poisson distribution, counting cases and deaths in space and time [14]. The number of expected cases is obtained by multiplying the population of the area by the overall incidence rate. However, in order to calculate the COVID-19 incidence and mortality rates in this study, the adjustment covariates of sex and age were considered in order to eliminate the confounding factor of these variables. Once the covariates were selected, the expected value was calculated considering the age and sex structure of the population.

The proportion of the population considered for cluster detection was optimized using the Gini index option in SatScan for purely spatial analysis, with a maximum population size of 10% for the spatial scan window. This option encourages the search for smaller true clusters and can be characterized as a population inequality coefficient [4]. The scan window calculated by the Gini index is optimized according to the characteristics of the population. The p-value of the clusters was obtained using the Monte Carlo hypothesis test with 999 replications [13]. The significance level was set at 5%. Statistical tests were calculated using the likelihood ratio. Windows are elliptical for space-time analyses.

When the population size significantly differs between the areas studied, adjustments must be made before modeling, either by excluding areas with large populations or by subdividing these areas into smaller populations. In this study, the WAs of the municipality of São Paulo and nearby municipalities are smaller due to the population size, but it does not invalidate the quality of data presented in this study.

After identifying the space-time clusters, we statistically compared the values of the demographic and socioeconomic variables of the group of WAs belonging to high-risk clusters to low-risk clusters of mortality and between high-risk clusters of incidences.

 doi:10.20944/preprints202408.1082.v1

Then, we compared the values of the groups for each demographic and socioeconomic variable using the Mann-Whitney and Kruskal-Wallis non-parametric tests for non-normal distribution of data. The null hypothesis was that the medians and interquartile ranges for the same variable were equal, with a significance level of 5%.

SaTScan™ version 10.0.1 (Kulldorff, Harvard Medical School, Boston, MA, USA), which uses geographical coordinates [14], was used to identify cases grouped in space-time and time. Maps with significant clusters and their relative risks from the space-time analyses were generated in QGIS 3.28. Temporal trends were obtained in SatScan. The significance level was set at p=0.05. R 4.3.2 for Mac was used for database manipulation and statistical analysis.

4. Results

A total of 191,083 cases of COVID-19 were reported in the MRSP between March 2020 and February 2022. According to the evolution of the patient's condition, the notifications were either recovery or death. This study was conducted in two periods, from March 2020 to February 2021 and from March 2021 to February 2022. In the first period of the pandemic, 100,587 cases were reported, with an annual rate of 512.9 cases per 100,000 inhabitants. In the second period of the pandemic, 89,783 cases were reported, with an annual rate of 457.8 cases per 100,000 inhabitants.

4.1. Incidence Clusters

In temporal terms (Figure 2A), a high-risk cluster was identified in the first period of the pandemic in May (relative risk [RR]=3.07), while a low-risk cluster was present from August to October 2020 (RR=0.47). The high-risk interval identified in the second period of the pandemic was from March to June, with RR ranging from 1.78 to 3.58. A low-risk period was also identified beginning in July (RR=0.87) (Figure 2B).

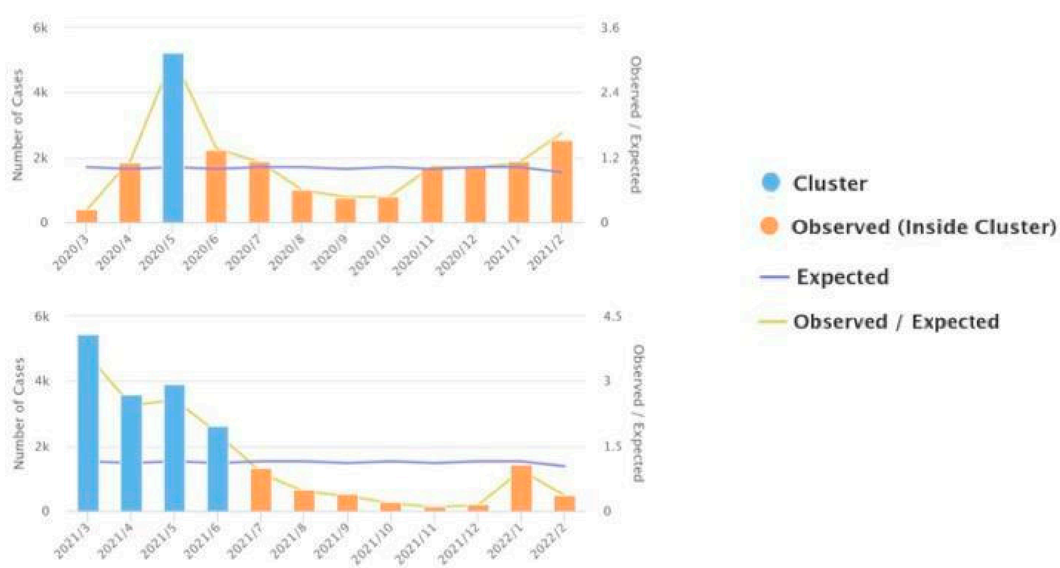


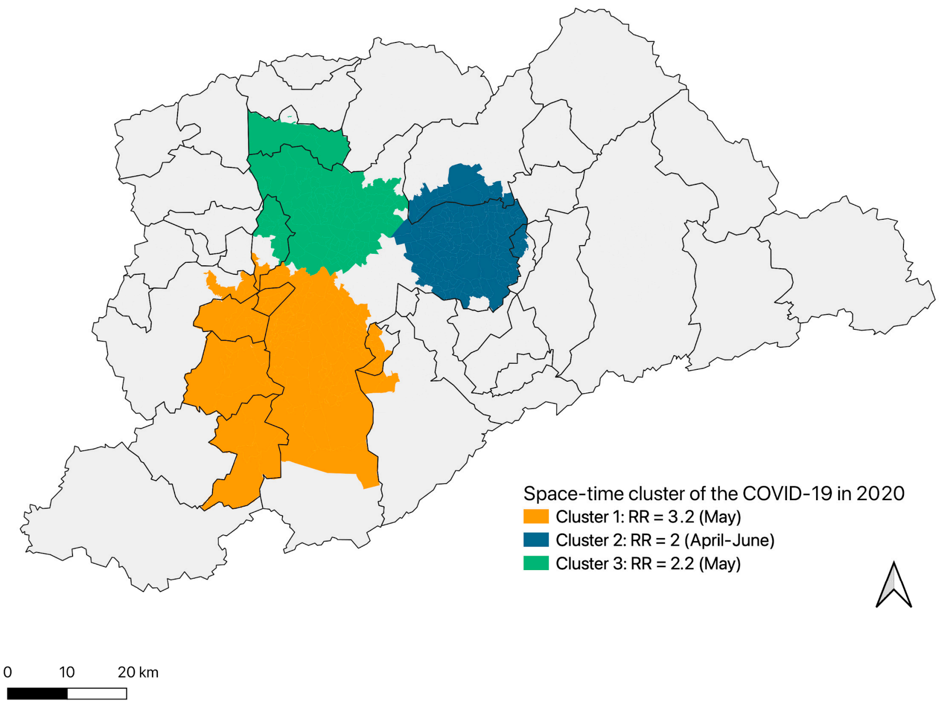
Figure 2. Distribution of COVID-19 cases in the MRSP. (A) Seasonal analysis showing the months of low and high risk in 2020-2021. (B) Seasonal analysis showing the months of low and high risk in 2021-2022.

In the first period of the pandemic, three significant high-risk clusters were identified using the space-time scan statistics of the total number of cases (Figure 3A). All were located in the capital, the municipality of São Paulo, and neighboring cities, particularly in May 2020. Cluster 1, located to the south region of the city of São Paulo and including the cities of Taboão, Embu, Itapecerica da Serra, Embu-Guaçu, Diadema, and São Bernardo do Campo, showed the highest relative risk (RR=3.2; p-

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2024 doi:10.20944/preprints202408.1082.v1

value <0.001); followed by cluster 3 (RR=2.2; p-value <0.001), located in the northwest region of the capital and including part of the cities of Cajamar and Osasco; and cluster 2 (RR=2; p-value <0.001), located in the northeast region of the capital and including part of the city of Guarulhos and a smaller part of the city of Ferraz de Vasconcelos. No low-risk cluster was identified in this period.

In the second period of the pandemic, three space-time clusters located in the capital São Paulo and neighboring municipalities were also observed, from March to June 2021 (Figure 3B). Cluster 1, located in the south region of São Paulo and including the entire municipality of Diadema and part of Embu, Itapecerica da Serra, Embu-Guaçu and São Bernardo do Campo, with a high relative risk (RR=2.9; p-value <0.001); followed by cluster 2 (RR=2.9; p-value <0.001), in the east region of the capital and including part of Guarulhos and a small portion of Ferraz de Vasconcelos; and cluster 3 (RR=2.6; p-value <0.001) located in the northwest region of the capital, but in this period including all municipalities of Barueri, Carapicuíba, and Osasco, as well as a large part of Cajamar, Santana de Parnaíba, and Taboão da Serra, and to a smaller extent, the cities of Cotia and Itapevi.



Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 August 2024 doi:10.20944/preprints202408.1082.v1

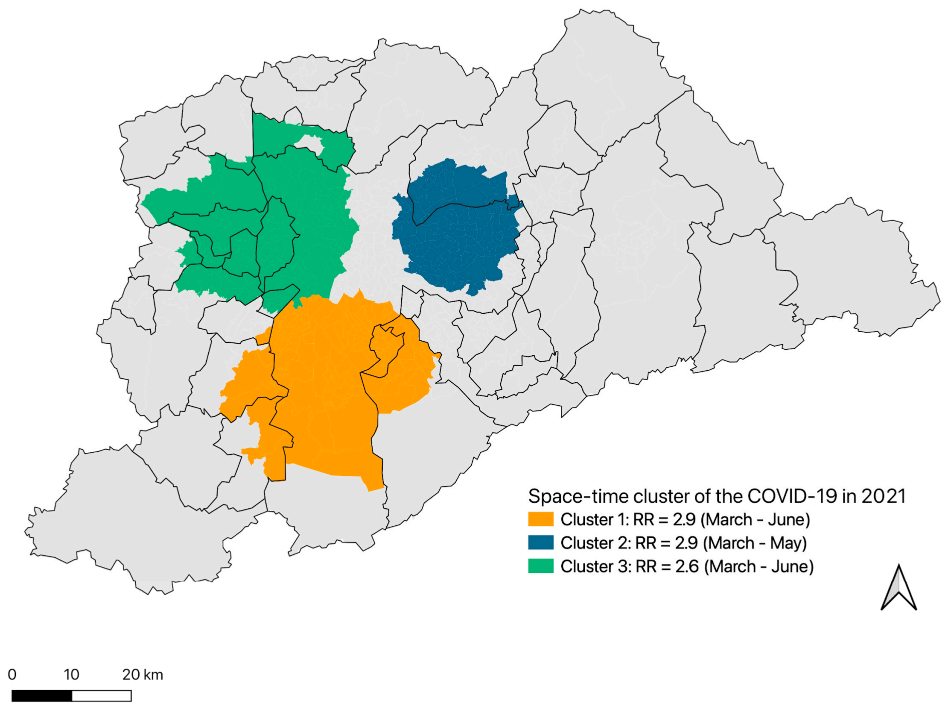


Figure 3. (A) Space-time clusters of COVID-19 incidence in the first year of the pandemic, from March 2020 to April 2021. (B): Space-time clusters of COVID-19 incidence in the second year of the pandemic, from March 2021 to April 2022.

The statistical comparison of the demographic and socioeconomic variables from the weighting areas presented in the space-time high-risk clusters showed important differences between the three groups (Table 1). In the first period of the pandemic, all analyses resulted in significant p-values, except for population density. The areas in clusters 1 and 2 had a lower per capita income, a higher percentage of BBIP, and a higher number of people per household. The median per capita income of cluster 3 was more than 50% higher than those of clusters 1 and 2. The interquartile range of cluster 3 was also much higher than the first and especially the third quartile of clusters 1 and 2.

In the second period of the pandemic, all analyses resulted in significant p-values, except for the number of people per household. The areas in clusters 1 and 2 had lower per capita income when compared to cluster 3. The median per capita income in cluster 3 remained higher than those of clusters 1 and 2, just over 10%. The interquartile range of the clusters was wide, demonstrating the heterogeneity of this group in relation to the variables of percentage of BBIP and population density.

Table 1. Median and interquartile range of socioeconomic indicators of the weighting areas in the MRSP from March 2020 to February 2021. Method: Kruskal-Wallis.

Variable	High-risk cluster COVID-19 in 2020			High-risk cluster COVID-19 in 2021		
	1	2	3	1	2	3
Per capita income	551 (461 - 881)	559 (442 - 719)	989 (611 - 1918)*	567 (469 - 892)	661 (513 - 867)	739 (521 - 1179)*
BBIP (%)	52 (40 -56)	45 (35 - 54)	29 (18 - 42)*	50 (37 - 56)	40 (30 - 49,3)	43 (28,3 - 50)*
People per household	3.4 (3.2 - 3.5)	3.4 (3.3 - 3.5)	3.1 (2.8 - 3.4)*	3.3 (3.2 - 3.4)	3.3 (3.2 - 3.3)	3.3 (3.1 - 3.4)

 doi:10.20944/preprints202408.1082.v1

Population density	328 (253 - 405)	297 (260 - 346)	324 (238 - 411)	334 (255 - 407)	283 (251 - 329)	318 (220 - 379)*

4.2. Mortality Clusters

In temporal terms (Figure 4A), a high-risk cluster was identified in the first year of the pandemic in May (relative risk [RR]=3.7), while a low-risk cluster was present from August 2020 to January 2021 (RR=0.53). The high-risk interval identified in the second period of the pandemic was in March and April, with RR ranging from 5.47 to 2.77. A low-risk period was also identified beginning in July (RR=0.67) (Figure 3B).

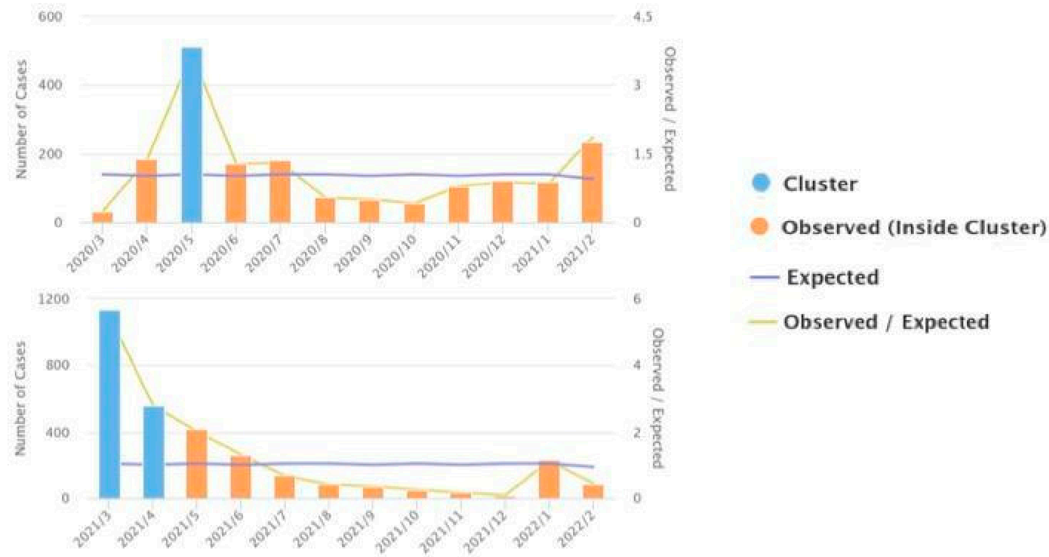
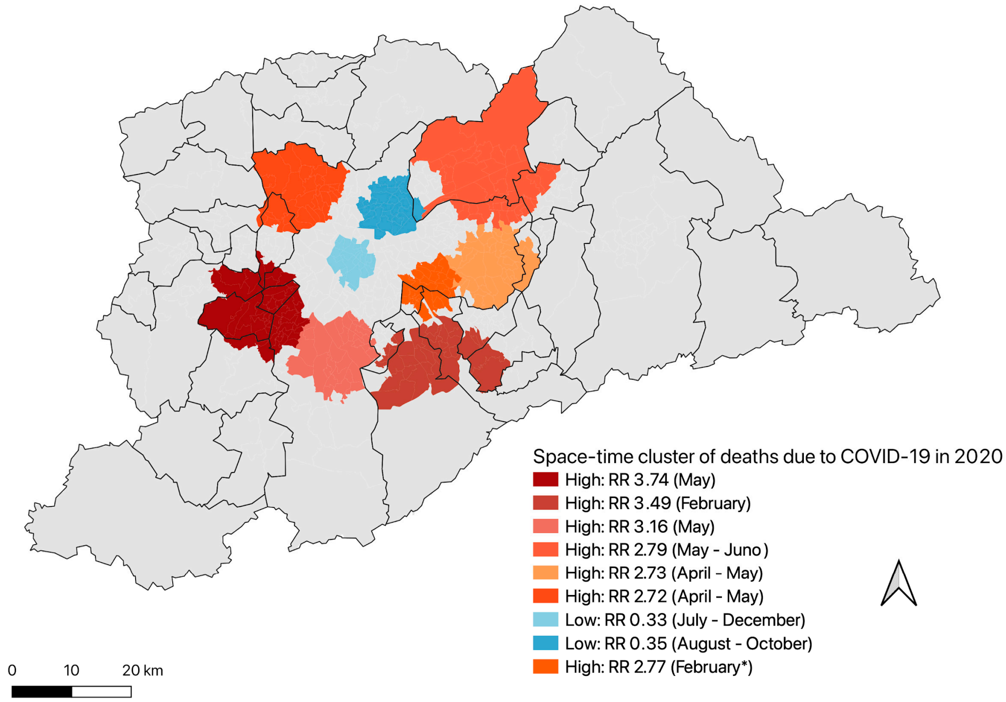


Figure 4. Distribution of deaths due to COVID-19 in the MRSP. (A) Seasonal analysis showing the months of low- and high-risk in 2020-2021. (B) Seasonal analysis showing the months of low- and high-risk in 2021-2022.

In the first year of the pandemic, seven significant high-risk clusters were identified using the space-time scan statistics of total deaths (Figure 5A). All clusters were located in the capital or neighboring cities, in particular in May 2020. In all high-risk clusters, the relative risk was higher than 2.7. Two low-risk clusters were identified in the capital, one in the central area (RR=0.33; p-value <0.001) and one in the north region of the capital, near Guarulhos, from July to December 2020 (RR=0.35; p-value <0.001).

In the second year of the pandemic, four space-time high-risk clusters were observed in the capital and neighboring cities, from March to May 2021 (Figure 5B). In all high-risk clusters, the relative risk was higher than 3.4. A low-risk cluster was identified in the central-south region of the capital from July to December 2021.



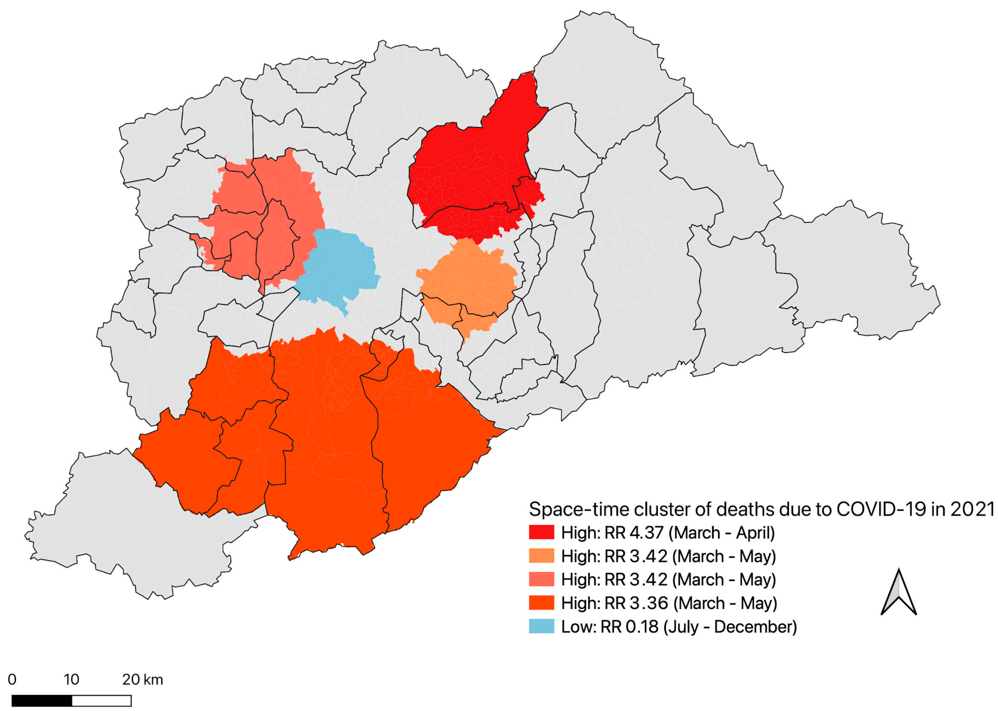


Figure 5. Space-time clusters of COVID-19 deaths in the first year of the pandemic. (4A) First year of the pandemic, from March 2020 to April 2021. *Year: 2021. (4B) Second year of the pandemic, from March 2021 to April 2022.

The statistical comparison of the demographic and socioeconomic variables of the weighting areas included in the space-time high-risk clusters for COVID-19 mortality showed important differences between the high- and low-risk groups (Table 2). In the first year of the pandemic, all

 doi:10.20944/preprints202408.1082.v1

analyses resulted in significant p-values, except for the variable of population density. The areas in the high-risk clusters had lower per capita income, a higher percentage of BBIP, and a higher number of people per household. The median per capita income of low-risk clusters was three times higher than those of the high-risk clusters.

Table 2. Median and interquartile range of socioeconomic indicators of the weighting areas in the MRSP, from March 2021 to February 2022. Method: Kruskal-Wallis.

Variable	COVID-19 death clusters in 2020		COVID-19 death clusters in 2021	
	High-risk	Low-risk cluster	High-risk	Low-risk cluster
Per capita income	548 (452 - 745)	1943 (989 - 3490)*	558 (454 - 718)	3362 (2223 - 3956)*
BBIP (%)	49 (38 - 54)	18 (9 - 20)*	47 (38 - 45)	12 (9 - 20)*
People per household	3.4 (3.3 - 3.5)	2.6 (2.4 - 3.1)*	3.4 (3.3 - 3.5)	2.5 (2.3 - 2.7)*
Population density	314 (243 - 396)	349 (268 - 424)	300 (238 - 361)	355 (256 - 440)*

In the second year of the pandemic, all analyses resulted in significant p-values. The areas in the high-risk clusters had lower per capita income than the low-risk clusters. High-risk clusters also had a higher percentage of BBIP and a higher number of people per household. Population density in high-risk clusters had a lower median than in low-risk clusters; however, this significant value may have been influenced by data amplitude.

5. Discussion

With the scan method to identify space-time clusters, the study showed that the notification of cases and deaths was not random or homogeneous, and that high-risk and low-risk areas were concentrated in specific regions along the months of the pandemic in the MRSP. In Brazil, states and, in some cases, municipalities adopted measures to restrict human movement during the pandemic, which may have led to different epidemiological scenarios in the area studied. However, in a metropolitan region such as São Paulo, which has a high level of internal movement, exclusively municipal restrictions may have had a lower impact.

As reported in other studies, the onset of COVID-19 cases occurred in the capital, São Paulo, spreading by spatial contiguity shortly after the start of the pandemic in March 2020 [3,15]. In our study, conducted on a more detailed level, the clusters were concentrated in the WAs of São Paulo and neighboring municipalities, indicating that the capital was a zone of influence and convergence at all times during the pandemic. The main clusters of cases and deaths, in the first period, were observed in May 2020 in the capital and neighboring municipalities, while in the second period they occurred between March and June 2021, remaining in the capital, but with increased coverage of neighboring municipalities. Therefore, we see the need for a regional analysis, since the dynamics of the virus spread may be associated with regional axes of circulation involving the capital and neighboring municipalities.

As in other places, the initial strategies to control the virus defined restrictions of human movement, but maintained the provision of essential services. In the MRSP, the implementation of control measures and social distancing through quarantine in March 2020 allowed only the operation of essential services, such as food, supplies, health, banking, cleaning, and security [16]. However, there were no policies or actions to protect the population that worked in these essential services and had to use public transport and who mostly lived in the outskirts of the city. The public transport system in the city of São Paulo has a highly radial model, structured to take passengers from the outskirts to the central area, or from the neighborhoods to the radial transport lines [17]. In the study on urban mobility in the MRSP conducted by Pilotto and Novasky (2022), the areas with the highest mobility rates are located in the central region of the capital, while the areas with the highest immobility rates are located in the suburbs and neighboring municipalities [18].

 doi:10.20944/preprints202408.1082.v1

Our hypothesis is that the poor conditions of public transportation in peripheral regions may have influenced the dynamics of contamination of the local population through community transmission of the coronavirus. Among the measures adopted for public transportation during the pandemic was the reduction of the vehicle fleet to avoid economic losses for the companies providing the service [17,19]. This measure was adopted due to the drop in the number of passengers, who started working remotely. However, for those people whose work did not allow them to stay at home, the reduction of the fleet increased the waiting time for travel and, at times, the resulting crowds may have favored the virus spread [17,19].

Studies indicate that the spatial distribution of COVID-19 cases was well explained by the variable of human mobility at the beginning of the pandemic [20,21]. The identification of a space-time high-risk cluster (cluster 2), from April to June 2020, in the area where Guarulhos International Airport is located, in a city that is adjacent to São Paulo, is consistent with studies that demonstrate the influence of mobility on the spread of the SARS-CoV-2 virus [22]. Guarulhos International Airport is the largest airport in Brazil and the second busiest in Latin America in terms of passenger and cargo transportation [23].

Incidence and mortality rates were higher in the first months of the second period of the pandemic. It may have been the result of the greater flexibility of control measures, the crowds in the pre-election and election periods in November, travel and celebrations during the holiday season, and the start of the school term in January 2021 [3]. Also, new variants of SARS-CoV-2 were circulating in the second study period—the Alpha and Gamma variants in November 2020 and the Delta variant in May 2021. It may have influenced the increase in the number of cases and the high transmission rate in the population.

Social behaviors, often driven by economic survival needs, were critical for the pattern of virus transmission [10]. In both the first and second periods of the pandemic, the statistically significant differences in the socioeconomic variables of per capita income and percentage of BBIP showed that cluster 3—located in a central region of the city of São Paulo—although apparently less socioeconomically vulnerable, had a high risk of COVID-19 incidence. However, in the second period, the cluster also began to include adjacent and non-adjacent municipalities, which also led to a smaller difference between the socioeconomic variables analyzed. There were no significant differences between the population densities of the clusters in relation to the values found, with strong internal heterogeneity identified within the groups.

Regarding mortality from COVID-19 in the first period, high-risk clusters were observed in the most peripheral region of the capital São Paulo and in neighboring municipalities, particularly in May 2020. In the second year of the pandemic, clusters persisted in the same locations, with a large high-risk cluster located in the south region of the capital and municipalities in the southwest and southeast of the MRSP. We observed that although no low-risk clusters of disease incidence were detected, low-risk clusters of mortality from COVID-19 were found in both periods analyzed, from July to December 2020, in the central area and especially in the southwest-central region of the city of São Paulo, which has the highest incomes in the MRSP. These low-risk clusters had a higher average per capita income, a lower percentage of BBIP, and a lower number of people per household. According to a study conducted by Pasternak (2015), the social structure in the MRSP has a model in which the upper classes are preferentially located in the central region and the popular classes in the periphery [24].

A general decrease was observed in both cases and deaths across the region beginning in the second half of 2021, probably because of the mass vaccination campaign, which prioritized the elderly and health professionals, but did not consider the population that is socioeconomically and territorially more vulnerable and exposed to the virus due to the greater use of public transportation. Despite this fact, other factors may be associated with the space-time inequality of cases and deaths, such as the prevalence of comorbidities in the population, risk behaviors like smoking and alcohol consumption, as well as conditions of vulnerability.

The method used in this study allowed the detection of clusters which provided relative risks, calculated by the incidence within the cluster in relation to the incidence outside the cluster. The tests

 doi:10.20944/preprints202408.1082.v1

applied to the regions that comprised the cluster detected areas with low and high rates within the cluster, reducing the possibility of detecting a cluster with only one unit of analysis. The method has been widely used to detect statistically significant space-time clusters of diseases and calculate relative risks, contributing to real-time geographic surveillance of diseases and early detection of epidemics and retrospective analysis [3,4,25,26].

Considering the demographic and socioeconomic dynamics, our results identified the locations where the first clusters of local transmission of COVID-19 occurred in the MRSP. It also provided an initial insight, which should be explored in terms of how urban mobility contributed as a factor of exposure to the coronavirus and vulnerability to the disease so that policies and actions can be formulated to mitigate the problem. The techniques used in this study also helped detect locations with high and low risk of mortality due to the disease, which may also contribute better management of future outbreaks.

Author Contributions: Conceptualization, Keila Santana, Pedro Mendonça, Pedro Lima, Aluízio Marino, Raquel Rolnik; methodology, Keila Santana, Pedro Mendonça, Gabriela Martins and Pedro Lima; software, Keila Santana, Pedro Mendonça, Gabriela Martins and Pedro Lima; formal analysis, Keila Santana; resources, Raquel Rolnik; writing—original draft preparation, Keila Santana; writing—review and editing, Pedro Mendonça and Raquel Rolnik; visualization, Pedro Mendonça, Pedro Lima, Gabriela Martins, Aluízio Marino, Raquel Rolnik; supervision, Raquel Rolnik; project administration, Aluízio Marino and Raquel Rolnik; funding acquisition, Raquel Rolnik. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by São Paulo State Research Support Foundation (Fapesp) nº 2023/03355-6, nº 2021/08276-1.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 73. 2020. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed on 08/01/2024).
2. UNASUS. Coronavírus: Brasil confirma primeiro caso da doença. Available online: https://www.unasus.gov.br/noticia/coronavirus-brasil-confirma-primeiro-caso-da-doenca#:~:text=O%20Minist%C3%A9rio%20da%20Sa%C3%BAde%20confirmou,para%20It%C3%A1lia%2C%20regi%C3%A3o%20da%20Lombardia.2020 (accessed on 08/01/2024).
3. Palasio, R.G.S.; Lorenz, C.; Lucas, P.C.; Nielsen, L.; Masuda, E.T.; Trevisan, C.M. Spatial, spatio-temporal, and origin-destination flow analyses of patients with severe acute respiratory syndrome hospitalized for COVID-19 in Southeastern Brazil, 2020-2021. Revista do Instituto de Medicina Tropical de São Paulo **2023**, 65.
4. Han, J.; Zhu, L.; Kulldorff, M.; Hostovich, S.; Stinchcomb, D.G.; Tatalovich, Z. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. International journal of health geographics **2016**, 15, 1-11.
5. São Paulo. Plano de Desenvolvimento Urbano Integrado da Região Metropolitana de São Paulo. Available online: https://rmsp.pdui.sp.gov.br/?page_id=127#:~:text=A%20Regi%C3%A3o%20Metropolitana%20de%20S%C3%A3o,Unidas%20(ONU)%20de%202014.2023 (accessed on 08/05/2024).
6. IBGE. Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2022. Available online:https://www.ibge.gov.br/estatisticas/sociais/trabalho/22827-censo-demografico-2022.html2022 (accessed on 08/05/2024).
7. IBGE. Censo Populacional 2010. Available online:https://cidades.ibge.gov.br/brasil/sp/araraquara/panorama (accessed on 08/02/2024).
8. METRÓPOLE CDED. Base Cartográfica Digital Georreferenciada de Logradouros da Região Metropolitana de São Paulo - Edição 2020. Available online: https://centrodametropole.fflch.usp.br/pt-br/node/9838 (accessed on 08/03/2024).
9. Aguilar-Palacio, I.; Maldonado, L.; Malo, S.; Sánchez-Recio, R.; Marcos-Campos, I.; Magallón-Botaya, R. COVID-19 inequalities: individual and area socioeconomic factors (Aragón, Spain). International journal of environmental research and public health **2021**, 18, 6607.
10. Chiaravalloti Neto, F.; Bermudi, P.M.M.; Aguiar, B.S.; Failla, M.A.; Barrozo, L.V.; Toporcov, T.N. Covid-19 hospital mortality using spatial hierarchical models: cohort design with 74,994 registers. Revista de Saúde Pública **2023**, 57, 2s.

doi:10.20944/preprints202408.1082.v1

11. Petrov, A. One hundred years of dasymetric mapping: back to the origin. The Cartographic Journal **2012**, 49, 256-64.
12. Research Google. Open Buildings: a dataset of buildings footprint to support social good applications. Available online: https://sites.research.google/open-buildings/2023 (accessed on 01/02/2024).
13. Kulldorff, M. SaTScanTM user guide. March; 2018. Available online: https://www.satscan.org/SaTScan_TM_Manual_do_Usu%C3%A1rio_Portugues.pdf (accessed on 01/03/2024).
14. Kulldorff, M. A spatial scan statistic. Communications in Statistics-Theory and methods **1997**, 26,1481-96.
15. Fortaleza, C.M.C.B.; Guimarães, R.B.; Catão, R.C.; Ferreira, C.P.; Almeida, G.B.; Vilches, T.N. The use of health geography modeling to understand early dispersion of COVID-19 in São Paulo, Brazil. PloS one. **2021**, 16, e0245051.
16. São Paulo. Governo de SP determina quarentena em todo o Estado. Available online: https://www.saopaulo.sp.gov.br/spnoticias/ultimas-noticias/ao-vivo-governo-de-sp-anuncia-novas-medidas-para-combate-ao-coronavirus-no-estado/#:~:text=resumo%20em%203%20t%C3%B3picos&text=O%20Governo%20de%20S%C3%A3o%20Paulo,%2C%20bancos%2C%20limpeza%20e%20seguran%C3%A7a.2020 (accessed on 10/06/2024).
17. Mendonça, P.H.R.; Rolnik, R.; Yeuw, T.T.; Marino, A. Mobilidade na cidade de São Paulo: lições das transformações durante a pandemia de COVID-19, ENANPUR. Belém, Brazil, 2023.
18. Pilotto, A.S.; Novaski, M.A.M. Indicadores de mobilidade urbana na RMSP a partir da pesquisa OD-Metrô. Cadernos Metrópole **2022**, 25, 229-54.
19. Silva, R.B. Vidas no sufoco nos transportes na pandemia: um App de mapeamento colaborativo para alerta de lotação na Região Metropolitana de São Paulo (RMSP). Confins Revue franco-brésilienne de géographie/Revista franco-brasileira de geografia **2023**, 58.
20. Celum, C.; Barnabas, R.; Cohen, M.S.; Collier, A.; El-Sadr, W.; Holmes, K.K. COVID-19, Ebola, and HIV Leveraging lessons to maximize impact. New England Journal of Medicine **2020**; 383, e106.
21. Leveau, C.M. Spatiotemporal variations in mortality from COVID-19 in neighborhoods of the Autonomous City of Buenos Aires, Argentina. Spatiotemporal variations in mortality from COVID-19 in neighborhoods of the Autonomous City of Buenos Aires, Argentina **2020**.
22. Rex, F.E.; Borges, C.A.S.; Käfer PS. Spatial analysis of the COVID-19 distribution pattern in São Paulo State, Brazil. Ciencia & saude coletiva. 2020;25:3377-84.
23. OGLOBO. Entre os maiores do mundo, aeroporto de Guarulhos é o segundo mais pontual 2017. Available online: https://oglobo.globo.com/boa-viagem/entre-os-maiores-do-mundo-aeroporto-de-guarulhos-o-segundo-mais-pontual-20728827 (accessed on 06/06/2024).
24. Pasternak, S. São Paulo: transformações na ordem urbana, Letra Capital Editora LTDA; Brazil, 2015.
25. Lacerda, A.B.; Lorenz, C.; Azevedo, T.S.; Cândido, D.M.; Wen, F.H.; Eloy, L.J. Detection of areas vulnerable to scorpionism and its association with environmental factors in São Paulo, Brazil. Acta tropica **2022**, 230, 106390.
26. Barrozo, L.V.; Serafim, M.B.; Moraes, S.L.; Mansur, G. Monitoramento espaço-temporal das áreas de alto risco de COVID-19 nos municípios do Brasil. Hygeia: Revista Brasileira de Geografia Médica e da Saúde **2020**, 16.