

Major Article

Utilizing prospective space-time scan statistics to discover the dynamics of coronavirus disease 2019 clusters in the State of São Paulo, Brazil

Ricardo Vicente Ferreira^[1] , Marcos Roberto Martines^[2] , Rogério Hartung Toppa^[3] ,
Luiza Maria de Assunção^[4] , Michael Richard Desjardins^[5]  and Eric Delmelle^[6] 

[1]. Universidade Federal do Triângulo Mineiro, Programa de Pós-graduação *Stricto Sensu* em Ciência e Tecnologia Ambiental, Uberaba, MG, Brasil.

[2]. Universidade Federal de São Carlos, Centro de Ciências Humanas e Biológicas, Sorocaba, SP, Brasil.

[3]. Universidade Federal de São Carlos, Departamento de Ciências Ambientais, Sorocaba, SP, Brasil.

[4]. Universidade do Estado de Minas Gerais, Faculdade de Ciências Jurídicas, Ituiutaba, MG, Brasil.

[5]. Department of Epidemiology, Spatial Science for Public Health Center, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

[6]. University of North Carolina-Charlotte, Center for Applied Geographic Information Science, Department of Geography and Earth Sciences, Charlotte, NC, USA.

ABSTRACT

Background: The number of deaths and people infected with coronavirus disease 2019 (COVID-19) in Brazil has steadily increased in the first few months of the pandemic. Despite the underreporting of coronavirus cases by government agencies across the country, São Paulo has the highest rate among all Brazilian states.

Methods: To identify the highest-risk municipalities during the initial outbreak, we utilized daily confirmed case data from official reports between February 25 and May 5, 2020, which were aggregated to the municipality level. A prospective space-time scan statistic was conducted to detect active clusters in three different time periods.

Results: Our findings suggest that approximately 4.6 times more municipalities belong to a significant space-time cluster with a relative risk (RR) > 1 on May 5, 2020.

Conclusions: Our study demonstrated the applicability of the space-time scan statistic for the detection of emerging clusters of COVID-19. In particular, we identified the clusters and RR of municipalities in the initial months of the pandemic, explaining the spatiotemporal patterns of COVID-19 transmission in the state of São Paulo. These results can be used to improve disease monitoring and facilitate targeted interventions.

Keywords: Relative risk. Space-time statistics. SARS-CoV-2. Geographic information systems. Disease surveillance.

INTRODUCTION

Space-time surveillance of coronavirus disease 2019 (COVID-19) can be a powerful tool for mitigating the risk of transmission¹ and can be useful for state and local health departments to monitor outbreaks in a timely manner². In Brazil, the virus began to spread at the beginning of February 2020³; however, the Ministry of Health of Brazil officially reported that the first case occurred on February 25, 2020, in the state of São Paulo⁴. Large cities, such as São Paulo

and Rio de Janeiro, were the main epicenters of the epidemic⁵. There were indications of rapid transmission within the state of São Paulo⁶ in municipalities with small populations and limited access to health care resources.

Monitoring the spread of COVID-19 is essential for the development of health strategies such as the allocation of tests and hospital beds. Prospective space-time scan statistics⁷ are powerful exploratory approaches for detecting active and emerging

Corresponding author: PhD Ricardo Vicente Ferreira. **e-mail:** ricardo.ferreira@uftm.edu.br

Authors' contribution: RVF: Conceptualization, Methodology, Formal analysis, Data curation, Writing - original draft; RHT, MRM: Conceptualization, Methodology, Data curation, Formal analysis, Writing - review & editing; LMA: Formal analysis, Writing - review & editing; MRD, ED: Metodology, Writing - review & editing.

Conflict of Interest: The authors have declared no conflict of interest.

Financial Support: This study was financed in part by the Pro Reitoria de Pesquisa e Pós Graduação of Universidade Federal do Triângulo Mineiro.

Received 8 December 2021 | **Accepted** 6 June 2022

disease clusters. This approach determines whether the observed spatiotemporal patterns of the disease are randomly distributed or show a statistically significant cluster during the most recent period of analysis⁸.

In this study, we analyzed the initial outbreak of COVID-19 in the state of São Paulo, which exhibited the highest number of cases during the first months of the pandemic⁹. Using a space-time scan statistic, our study aimed to detect emerging COVID-19 clusters in municipalities in the state of São Paulo. We demonstrate the evolution of the relative risk (RR) and statistically significant clusters, considering three time periods across São Paulo municipalities.

METHODS

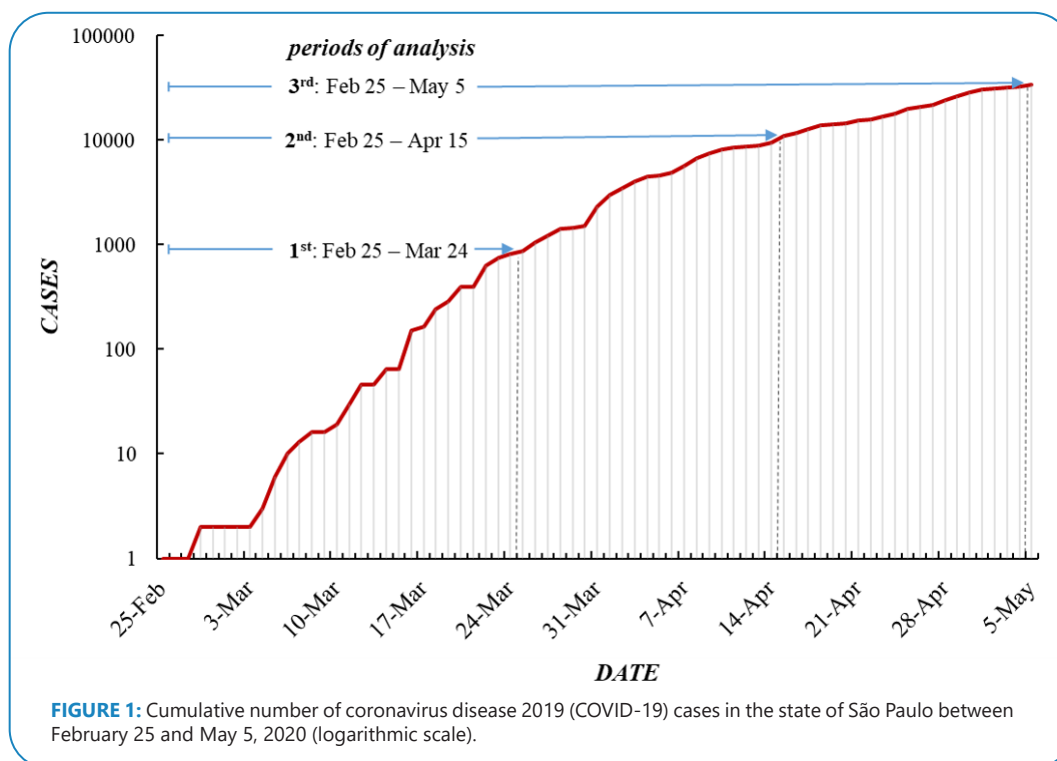
On March 1, 2020, the Health Surveillance Secretary of the Ministry of Health of Brazil identified the first confirmed COVID-19 case by real-time reverse-transcriptase polymerase chain reaction (RT-PCR), following the protocol of the University Charité (Berlin, Germany). In Brazil, the notification system is decentralized, and municipalities can conduct tests in private or public laboratories. The results are then notified by the state health system, which then notifies the national government.

The state of São Paulo has approximately 46.6 million inhabitants in 2021 and an area of 248,219.5 square kilometers. On March 24, 2020, the governor of São Paulo mandated social distancing to prevent the spread of the novel coronavirus. As the incubation period for the novel COVID-19 is well supported by evidence to be approximately 14 days¹⁰, we selected three separate space-time analysis periods, from February 25 to March 24, February 25 to April 15, and February 25 to May 5, 2020, which included five incubation periods of onset of the most current COVID-19 case in the dataset (**Figure 1**).

The COVID-19 cases were retrieved from the Brazil.io project website¹¹. This project compiles data from daily COVID-19 case reports by the municipality in the 27 Brazilian Federation Units and is available in a raw format, which has been organized in a format that SaTScan supports¹². In the state of São Paulo, daily reports are prepared using the state data analysis system (<https://www.seade.gov.br/coronavirus/>) and contain the addresses of confirmed cases. The data include information on 645 municipalities in the state of São Paulo. In addition, we used records between February 25 and May 5, 2020, and a geographic information system (GIS) to geocode the locations¹³ of COVID-19 cases and aggregated them to the municipality level. Municipality-level GIS files were retrieved from the Instituto Brasileiro de Geografia e Estatística (IBGE).

Active clusters were detected using the Poisson space-time scan statistic method⁸, which was integrated into the SaTScan software. The prospective space-time statistic detects active clusters of disease, which is the excess incidence of cases during the last period of analysis¹⁴. The statistic systematically implements moving cylinders to scan the study area. The cylinders were centered on the centroid of the municipalities, the base of the cylinder was the spatial scanning window, and the height represented the temporal scanning window. The cylinders were expanded until the maximum spatial and temporal upper bounds were reached. We defined the upper bounds as having a maximum spatial and temporal scanning window size of 10% of the at-risk population to avoid extremely large clusters and 50% of the study period, respectively. In addition, each cluster's duration was set to a minimum of 2 days, and each cluster contained a minimum of five confirmed cases of COVID-19¹⁵.

The Poisson model detects space-time clusters that were still occurring or active on the last day of the analysis^{7,8,15}. We assumed that the data on COVID-19 cases in our study area followed the



Poisson distribution. The null hypothesis states that the model reflects a constant risk, such as the absence of anomalous clustering of COVID-19. The alternative hypothesis states that the number of observed cases exceeds expected cases derived from the null model. The expected cases were calculated by multiplying the population in the cylinder by the total COVID-19 rate in the cylinder ($p_i * C/P$), with p_i the population in i ; C the total COVID-19 cases in the state of São Paulo, and P the total estimated population in the state of São Paulo. We assumed that the study population was static during the study period.

In SatSan, a maximum likelihood ratio test was used to identify cylinders with an elevated risk of contracting COVID-19. The cylinder has an elevated risk when the likelihood ratio is > 1 , where the rate of cases inside the cylinder is greater than that outside the cylinder. To derive statistical significance, 999 Monte Carlo simulations were computed, and clusters were reported at the 95% confidence level ($P < 0.05$). Thus, 999 likelihood ratios were computed for each cylinder, where each cylinder was a potential

cluster. To circumvent the assumption that the municipalities belonging to a cluster were homogenous, we also reported and mapped the RR for each municipality during each study period. RR was defined as the estimated risk of COVID-19 within a municipality divided by the risk outside the municipality (i.e., everywhere else).

RESULTS

First period of analysis: February 25 to March 24, 2020

We did not detect any statistically significant emerging space-time clusters of COVID-19 between the first period of analysis (28 days) and two incubation periods (**Figure 2A**). However, we observed five municipalities with an $RR > 1$ outside the cluster, with more observed than that of the expected cases, for the first period. The highest RR was observed in São Paulo municipality ($RR = 28.03$), followed by São Caetano do Sul ($RR = 3.42$), Jaguariúna ($RR = 2.38$), Santana de Parnaíba ($RR = 1.97$), and Cotia ($RR = 1.1$) municipalities.

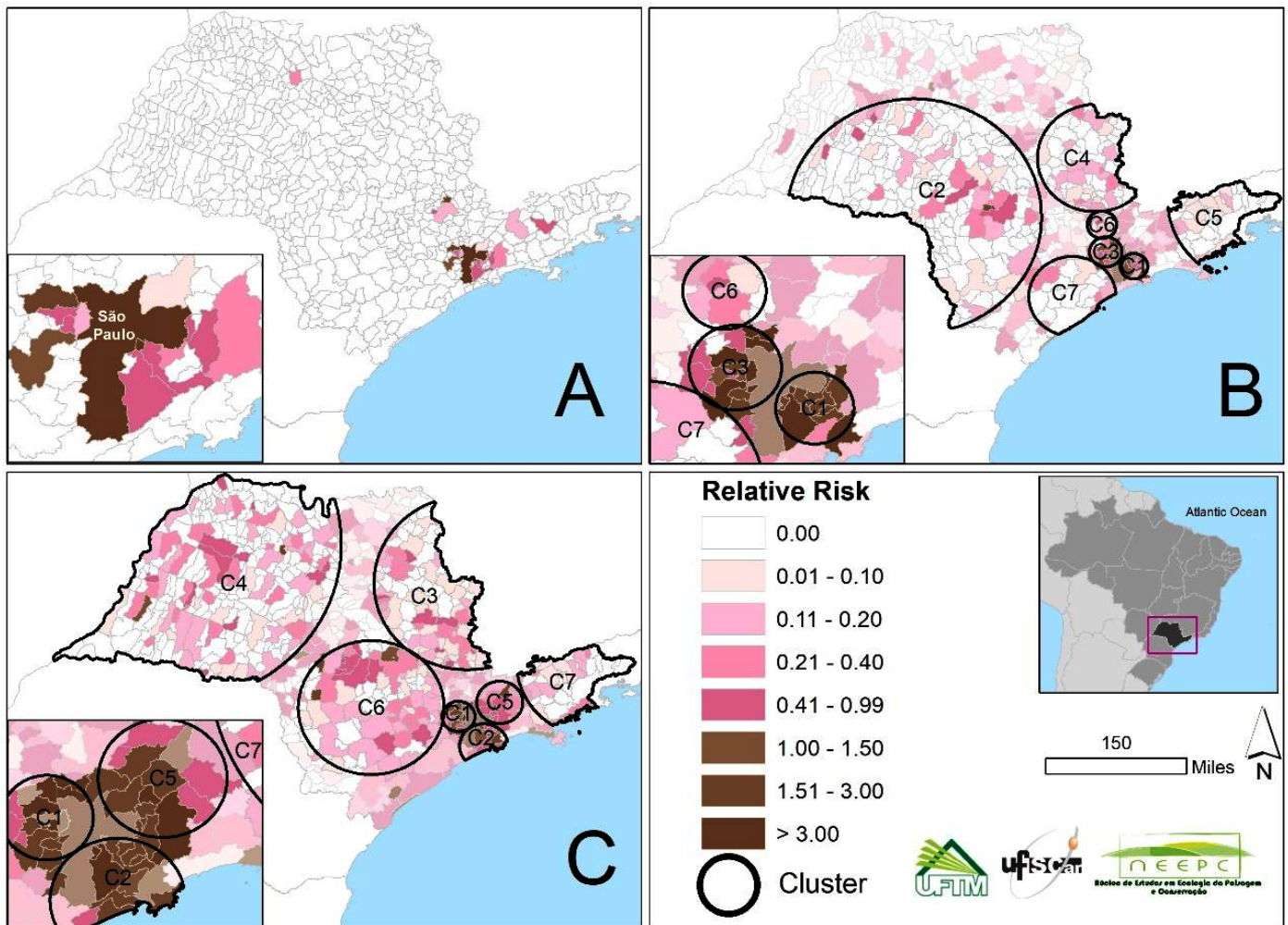


FIGURE 2: Spatial distribution of emerging space-time clusters of COVID-19 for the municipalities in the state of São Paulo, Brazil. **A:** First period of analysis: February 25 to March 24, 2020 (28 days). **B:** Second period of analysis: February 25 to April 15, 2020 (50 days; C1, $RR = 3.29$; C3, $RR = 2.64$; the other clusters did not show $RR > 1$). **C:** Third period of analysis: February 25 to May 5, 2020 (70 days; C1, $RR = 3.71$; C2, $RR = 3.12$; C5, $RR = 2.68$; the other clusters did not show $RR > 1$). **C1:** Cluster 1; **C2:** Cluster 2; **C3:** Cluster 3; **RR:** relative risk.

Second period of analysis: February 25 to April 15, 2020

We detected seven statistically significant emerging space-time clusters of COVID-19 in the state of São Paulo for the second period of analysis (50 days, approximately three incubation periods), which included 353 municipalities. In the second period, there were two observed clusters with an RR > 1 (cluster 1: RR = 3.29; cluster 3: RR = 2.64), including 11 and 17 municipalities, respectively (**Table 1**). Both were located in the eastern region in the state of São Paulo, near São Paulo municipality (**Figure 2B**).

Cluster 1 had 943 observed cases, with the highest RR in São Caetano do Sul municipality (RR = 2.23, with 85 observed cases for 38.34 expected cases), followed by Santos municipality (RR = 2.12, with 216 observed cases for 103.11 expected cases). Cluster 1 also included São Bernardo do Campo (RR = 1.16) and Santo André municipalities (RR = 1.09), with an RR > 1. Finally, cluster 3 contained 17 municipalities, including 6 with an RR > 1 and 641 observed cases (**Table 1**). The highest RR was in the Caieiras municipality (RR = 1.74), followed by Taboão da Serra (RR = 1.34), Santana de Parnaíba (RR = 1.24), Cotia (RR = 1.18), Franco da Rocha (RR = 1.12), and Osasco (RR = 1.04) municipalities.

Third period of analysis: February 25 to May 5, 2020

We detected seven statistically significant emerging space-time clusters of COVID-19 in our third period of analysis (70 days, five

incubation periods), covering 513 municipalities in the state of São Paulo. Clusters 1, 2, and 5 showed the highest RR (**Table 2**) with RR = 3.71, RR = 3.12, and RR = 2.68, respectively. All of these were located near São Paulo municipality in the eastern region of the state (**Figure 2C**).

Cluster 1 included 11 municipalities with RR > 1, among which the highest RR was observed in the Barueri municipality (RR = 2.03). Cluster 2 had five municipalities with an RR > 1, among which the highest RR was in São Caetano do Sul municipality (RR = 1.84). Finally, cluster 5 had two municipalities with an RR > 1, in which the highest RR was in Igaratá municipality (RR = 1.84). All the previously mentioned municipalities border São Paulo municipality, including Barueri and São Caetano do Sul, except for Igaratá municipality, which is located 96 km away from São Paulo by road.

DISCUSSION

In this study, we adapted the methodology used by Hohl et al.¹⁶, using three periods to detect emerging clusters of COVID-19 in the state of São Paulo, Brazil, utilizing confirmed case data from official reports. It is important to highlight that the RR throughout São Paulo increased during the three periods analyzed. In the first period (February 25 to March 24, 2020), within 28 days from the first recorded case, we detected only 5 municipalities with an

TABLE 1: Emerging space-time clusters of coronavirus disease 2019 (COVID-19) from February 25 to April 15, 2020, including relative risk (RR) estimates for the clusters identified in São Paulo, Brazil.

Cluster	Duration	Observed cases	Expected cases	RR	n of municipalities	Population	n of municipalities with RR > 1
1	3/30–4/15	943	305.05	3.3	11	3,845,800	4
2	3/22–4/15	118	522.04	0.2	203	4,475,331	2
3	3/31–4/15	641	252.45	2.6	17	3,381,536	6
4	3/22–4/15	86	421.67	0.2	61	3,614,906	0
5	3/22–4/15	22	168.99	0.1	33	1,448,732	0
6	3/22–4/15	36	127.66	0.3	9	1,094,410	0
7	3/22–4/15	21	96.37	0.2	19	826,174	0

P-value < .001; n: number; RR: relative risk.

TABLE 2: Emerging space-time clusters of COVID-19 from February 25 to May 5, 2020, including relative risk (RR) estimates for the clusters identified in São Paulo, Brazil.

Cluster	Duration	Observed cases	Expected cases	RR	n of municipalities	Population	n of municipalities with RR > 1
1	4/15–5/5	2,598	741.98	3.7	17	3,381,536	11
2	5/15–5/5	2,899	985.48	3.1	14	4,491,304	5
3	5/1–5/5	534	1,641.67	0.3	87	4,489,123	0
4	4/1–5/5	583	1,660.57	0.3	263	4,540,784	2
5	4/24–5/5	1,118	425.91	2.7	15	3,396,868	2
6	4/1–5/5	758	1,635.26	0.4	85	4,471,575	3
7	4/1–5/5	108	485.36	0.2	32	1,327,200	0

P-value < .001; n: number; RR: relative risk.

RR > 1, whereas in the third period (February 25 to May 5, 2020), within 70 days from the first recorded case, we detected 23 municipalities with an RR > 1. These results show that within our study period, there were 4.6 times more municipalities with more observed than that of the expected cases in the state of São Paulo. Our results show that São Paulo municipality was a hotspot of COVID-19, which is highlighted by the location of the clusters with an RR > 1 in the second and third periods. However, 70 days after the spread of the virus in the state of São Paulo, we observed municipalities with an RR > 1 in almost all emerging clusters.

In the first period analyzed (February 25 to March 24, 2020), 20 municipalities had an RR for COVID-19. However, no clusters were detected, while all municipalities with reported COVID-19 incidence had a population greater than 50,000 inhabitants. In the second (February 25 to April 15, 2020) and third periods (February 25 to May 5, 2020), clusters with an RR > 1 were located around the municipality of São Paulo. However, in the second period, 25 municipalities with a population of less than 50,000 inhabitants began to exhibit COVID-19 incidence. In the third period, the number of municipalities with reported COVID-19 incidence increased to 117, demonstrating the spread to the countryside in the state¹².

Regarding the limitations of this study, we acknowledge that there are inherent biases in the dataset we used, similar to other countries since tests were not always readily accessible when the pandemic arrived in Brazil, which led to the underreporting of cases and deaths. Another limitation is the determination of the SaTScan parameters¹⁷. The default configuration rarely produces useful and informative results because some clusters occupy a large proportion of the study area¹⁸. The task of determining the most suitable parameters is context-dependent and influenced by the geographic scale, processes that generate the clusters, and objectives of the study.

This study presented an analysis of the space-time transmission dynamics of daily COVID-19 cases with the intent of identifying emerging space-time clusters active in the municipalities of the state of São Paulo in three different time periods. Three significant active clusters were detected in the state of São Paulo on May 5, 2020. Thus, this space-time approach is useful for infectious disease surveillance and identifying statistically significant clusters of cases and the RR of locations that belong to a cluster to facilitate further targeted interventions at finer spatial scales. We encourage researchers to utilize this approach for continued surveillance of COVID-19 worldwide, which can be strengthened by adjusting testing efforts and relevant covariates, such as poverty levels, access to healthcare, and vaccination rates.

REFERENCES

1. Yang C, Sha D, Liu Q, Li Y, Lan H, Guan WW, et al. Taking the pulse of COVID-19: A spatiotemporal perspective. *Int J Digit Earth*. 2020;13(10):1186-211.
2. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Appl Geogr*. 2020;118:102202.
3. Delatorre E, Mir D, Gräf T, Bello G. Tracking the onset date of the community spread of SARS-CoV-2 in western countries. *Mem Inst Oswaldo Cruz*. 2020;115:e200183.
4. Ministério da Saúde (MS). Brasil. Available from: <https://covid.saude.gov.br/>. 2020
5. Lancet, T. COVID-19 in Brazil: "So what?". *Lancet* (London, England). 2020;395(10235):1461.
6. Codeço CT, Villela D, Coelho FC, Bastos LS, Carvalho LM, Gomes MF, et al. Estimativa de risco de espalhamento da COVID-19 no Brasil e avaliação da vulnerabilidade socioeconômica nas microrregiões brasileiras. 2020;2.
7. Kulldorff, M. A spatial scan statistic. *Communications in Statistics-Theory and methods*. 1997;26(6):1481-96.
8. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *J R Statist Soc: Series A (Statistics in Society)*. 2001;164(1):61-72.
9. Ministério da Saude. Coronavírus Brasil. Available: <https://covid.saude.gov.br/>. 2021.
10. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med*. 2020;172(9):577-82.
11. Brasil.IO: COVID-19 epidemiological bulletins by municipality per day. 2020. Available from: <https://brasil.io/dataset/covid19/> (last actualization: MaDEy, 05 2020).
12. Martines MR, Ferreira RV, Toppa RH, Assunção LM, Desjardins MR, Delmelle EM. Detecting space-time clusters of COVID-19 in Brazil: mortality, inequality, socioeconomic vulnerability, and the relative risk of the disease in Brazilian municipalities. *J Geogr Syst*. 2021;23(1):7-36.
13. Delmelle EM, Desjardins MR, Jung P, Owusu C, Lan Y, Hohl A, et al. Uncertainty in geospatial health: challenges and opportunities ahead. *Ann Epidemiol*. 2022;65:15-30.
14. Jones RC, Liberatore M, Fernandez JR, Gerber SI. Use of a prospective space-time scan statistic to prioritize shigellosis case investigations in an urban jurisdiction. *Public Health Rep*. 2006;121(2):133-9.
15. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*. 2020;118:102202.
16. Hohl A, Delmelle EM, Desjardins MR, Lan Y. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spat Spatiotemporal Epidemiol*. 2020;34:100354.
17. Chen J, Roth RE, Naito AT, Lengerich EJ, MacEachren AM. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int J Health Geogr*. 2008;7:57.
18. Haining RRP: *Spatial Data Analysis: Theory and Practice*. 1st ed. Cambridge: Cambridge University Press; 2003; 454 p.