

INCIDENCE AND LETHALITY OF COVID-19 CLUSTERS IN BRAZIL VIA CIRCULAR SCAN METHOD

Henrique José de Paula ALVES¹
Felipe Augusto FERNANDES²
Kelly Pereira de LIMA²
Ben Dêivide de Oliveira BATISTA³
Tales Jesus FERNANDES²

■ **ABSTRACT:** The COVID-19 pandemic has spread rapidly around the world in a frightening way. In Brazil, the third country with the highest number of infected and deaths from the disease, it is important for government health authorities to identify the federation units that stand out in cases and deaths from this disease to target resources. The circular scan statistic proposed by Martin Kulldorff allows to identify with some statistical significance the units of the federation that stand out in relation to the number of cases and deaths of COVID-19 in Brazil. Such units of federation are known as clusters. Once these clusters were identified, we used the coefficients of incidence and lethality to better describe the behavior of these clusters during three phases of the pandemic: the initial phase, the peak phase, and also the stability and fall phase. We observed changes in the location of the clusters identified in these three phases and used the R software and also the SaTScan software to obtain the maps and results, which were consistent with what was reported by the Brazilian media.

■ **KEYWORDS:** COVID-19; Pandemic; Brazil; Kulldorff; incidence; lethality.

¹Instituto de Pesquisa Econômica Aplicada - IPEA, CEP: 20071-900, Rio de Janeiro, RJ, Brazil. E-mail: jpahenrique@gmail.com

²Universidade Federal de Lavras - UFLA, Departamento de Estatística, CEP: 37200-900, Lavras, MG, Brasil. E-mail: fernandesfelipe@gmail.com, kelly.lima.88@gmail.com, tales.jfernandes@ufla.br

³Universidade Federal de São João del-Rei - UFSJ, Departamento de Física e Matemática, CEP: 36307-352, Ouro Branco, MG, Brasil. E-mail: ben.deivide@ufs.edu.br

1 Introduction

At the end of 2019, a local outbreak of SARS-CoV-2 pneumonia was identified in Wuhan (Hubei, China), a new coronavirus. Since then, contagion has spread to all provinces in mainland China and 27 other countries and regions, with more than 70,000 cases confirmed as the 17 February date of 2020 (DONG *et al.*, 2020).

Coronaviruses (CoV) are a sizeable viral family known around the mid-1960s, causing respiratory infections in humans and animals. These viruses are large, enveloped, positive-stranded RNA that can be divided into four genera: alpha, beta, delta, and gamma, of which CoVs alpha and beta infect humans (WILD *et al.*, 2017; FERNANDES *et al.*, 2020). This infection causes significant damage to public health, which is aggravated by the lack of specific vaccines and medications to treat COVID-19 (KHAILANY *et al.*, 2020; NASSIRI, 2020; VELAVAN AND MEYER, 2020; SALATHÉ *et al.*, 2020). Besides, many critically ill patients need hospital assistance and mechanical respirators, so there is a need for social distance, and these characteristics made COVID-19 quickly considered a pandemic by the WHO (SALATHÉ *et al.*, 2020). Cordes *et al.* (2020) used the scan statistic to identify clusters in which COVID-19 stood out in New York City, aiming to identify clusters with high and low positive rates and also a high positive proportion of COVID-19 infection. Hohl *et al.* (2020) used the circular scan statistic proposed by Martin Kulldorff in the U.S. using the proposed Poisson temporal space model.

In Brazil, the first documented case was in the city of São Paulo at 26 in February of 2020 and spread to reach 27 federal units (F.U.'S). Since then, until the 16 October 2020, 5,169,386 cases were confirmed and 152,460 deaths in national territory (SAÚDE, 2020). Given the above, it is evident that Brazil's situation is worrying. Action to combat rapid contagion and a detailed survey of how the pandemic took place must be carried out in the country. Isolated measures to combat COVID-19 can and should be implemented by organs responsible bodies in each Brazilian F.U.. However, the population heterogeneity of each Brazilian F.U., in addition to the number of beds in intensive care units (ICU) available for emergency care and also the socio-economic vulnerability conditions, make it difficult to indicate which of these F.U.'S needs more resources to combat COVID-19.

One strategy is to identify F.U.'S, whose occurrence of any factor related to this pandemic is inconsistent with the others, which can be just one F.U. or a set of F.U.'S connected, as was done in the work of (ALVES *et al.*, 2020). In this sense, spatial methods of cluster detection become useful tools as they aim to identify different patterns of spatial association (PINTO *et al.*, 2014). Some of these methods can be seen in Choynowsky (1959); Naus (1965a,b); Clayton and Kaldor (1987); Whittemore *et al.* (1987); Openshaw *et al.* (1988); Turnbull *et al.* (1989); Birch *et al.* (2009). Other more recent proposals are found in Duarte *et al.* (2017); Kulldorff (2001); Duczmal and Assuncao (2004); Duczmal *et al.* (2006); Li *et al.* (2020); Novee *et al.* (2020); Abolhassani *et al.* (2020).

However, we highlight the proposed circular scan method Kulldorff (1997),

and, as an objective described here, we use this method to find the most plausible cluster concerning COVID-19 among the Brazilian F.U.'S, in three phases: the initial stage, the peak period, and the degree of stability and fall of the disease in Brazil. We also calculated the coefficients of incidence and lethality for each of the clusters detected in each phase considered.

2 Kulldorff's circular scan method

To assign statistical significance to a cluster detection method, Kulldorff (1997) proposed a statistical significance test that has been widely used. The idea is to build circles and gradually increase its radius. For each new circle created, the number of cases and the population at risk are updated, calculating a likelihood ratio involving the relative risk inside and outside the zone, according to some pre-established probability model. The circle radius variation is "pseudo-arbitrary", being limited to up to 50% of the total map population. The most likely cluster will be the one formed by the areas whose likelihood ratio is maximum. There are several probability distribution models associated with this test statistic: Poisson, Bernoulli, Binomial, Hypergeometric, Normal, and space-time permutation. Here, as we are dealing with new case counts and further deaths from COVID-19, a natural choice would be the Poisson model.

2.1 The Poisson model

Consider the problem of testing the following hypotheses associated with the dimension of the space:

$$H_0 : p = q \quad \text{against} \quad H_1 : p > q, \quad (1)$$

where p is the probability that an individual will be a case of some specific characteristic within a zone and q is and the probability that an individual will be a case of that same characteristic outside that zone. Thus, the null hypothesis is equivalent to the statement that the number of cases is randomly distributed in the region of interest. The alternative hypothesis, on the other hand, confronts the existence of this randomness in at least one area from the same region. Consider N to be the total number of individuals in the region of interest and have the total number of cases in that region be denoted by C . Also, let Z be the total number of all possible zones that can be formed in that region of interest and make z be a specific zone. Let n_z and c_z be the number of individuals and the number of cases of some characteristic within the zone z , respectively, and make μ_z be the number of cases expected within the zone z . Under H_0 , we have $\mu_z = pn_z$ as the number of cases expected within the zone z and, consequently, $\mu_{\bar{z}} = p(N - n_z)$ as the number of cases expected outside the z zone. Thus, according to Kulldorff (1997), where we adapt some notations, the support function under the restricted parametric space

is denoted by:

$$L_0(Z, p) = \frac{\mu_z^{c_z} e^{-\mu_z}}{c_z!} \times \frac{\mu_{\bar{z}}^{c_{\bar{z}}} e^{-\mu_{\bar{z}}}}{c_{\bar{z}}!} \\ = \frac{(pn_z)^{(c_z)} e^{-pn_z}}{c_z!} \times \frac{[p(N - n_{\bar{z}})]^{(C - c_z)} e^{p(N - n_{\bar{z}})}}{C - c_{\bar{z}}!}. \quad (2)$$

In this case, the log-likelihood is given by

$$l_0(Z, p) = c_z[\log(p) + \log(n_z)] - pn_z - \log(c_z!) + \\ (C - c_z)[\log(p) + \log(N - n_z)] - p(N - n_z) - \log[(C - c_z)!], \quad (3)$$

whose estimator is $\hat{p} = C/N$. Substituting in the expression (2), we get

$$L_0(Z) = \frac{\mu_z^{c_z} (C - \mu_z)^{(C - c_z)} e^{-C}}{c_z!(C - c_z)!}. \quad (4)$$

Under H_1 , where a case is more likely to occur in any area of the region of interest, the expected number of cases within the z zone will be $\mu_z = pn_z$ and the expected number of cases outside that zone same zone will be $\mu_{\bar{z}} = q(N - n_z)$. Thus, the support function under the unrestricted parametric space is given by the expression

$$L(Z, p, q) = \frac{(pn_z)^{(c_z)} e^{-pn_z}}{c_z!} \times \frac{[q(N - n_z)]^{(C - c_z)} e^{q(N - n_z)}}{(C - c_z)!} \quad (5)$$

and the log-likelihood is given by the following expression

$$l(Z, p, q) = c_z[\log(p) + \log(n_z)] - pn_z - \log(c_z!) + \\ (C - c_z)[\log(q) + \log(N - n_z)] - q(N - n_z) - \log[(C - c_z)!], \quad (6)$$

whose estimators are $\hat{p} = c_z/n_z$ and $\hat{q} = (C - c_z)/(N - n_z)$. Substituting in the expression (5), we get the expression

$$L(Z) = \frac{c_z^{c_z} (C - c_z)^{(C - c_z)} e^{-C}}{c_z!(C - c_z)!}. \quad (7)$$

Make $RR = c_z/\mu_z$ and $rr = (C - c_z)/(C - \mu_z)$. Thus, the likelihood ratio, according to Kulldorff (1997), where we adapted some notations, is given by the following expression

$$\lambda = \frac{L(Z)}{L_0(Z)} = \begin{cases} (RR)^{c_z} (rr)^{(C - c_z)} & , RR > rr \\ 1 & , \text{otherwise} \end{cases} \quad (8)$$

We are only interested in high-rate clusters and, therefore, this is the justification for the case where $\lambda = 1$. This method helps us to find the most likely cluster, which does not always correspond to the actual cluster, making this test statistic provide an estimate for the position and radius of the actual cluster.

2.1.1 Algorithm for cluster detection using the Kulldorff method

To detect the most likely cluster, choose the zone \hat{z} that maximizes λ , as in the expression (8). However, its distribution is unknown and is then obtained via Monte Carlo simulation. The algorithm for detecting clusters follows the following steps:

1. a centroid is chosen in each region of the map;
2. the distance between the chosen centroid and the other centroids is calculated, sorting them in ascending order and storing them in a vector;
3. for each centroid of the study region, steps 1 and 2 are repeated;
4. From that chosen centroid a circular window is then created, and its radius is continuously increased according to the distances obtained in step 2;
5. each time a new centroid is reached, the number of cases and the population at risk within the circle (zone) z is updated;
6. then the value of the test statistic λ is calculated, as in the expression (8), for each pair (c_z, n_z) ;
7. the zone with the highest value of λ is stored and this process is repeated until the threshold for the radius of that circle is reached;
8. repeat steps 3 – 6 for each of the centroids on the map;
9. the significance of the test is obtained via Monte Carlo simulations.

The spatial scan statistic applied has some limitations regarding its use, such as cluster overlap and even the choice of the optimal size for the scanning radius (the Gini coefficient is a possible tool to correct this problem, see (HAN, JUNHEE et al., 2016)). However, as stated by Assunção, there is no better technique for detecting spatial clusters than the technique described here (COSTA *et al.*, 2012).

Although an important temporal division is considered, the study carried out here is very simple due to the various application alternatives, such as, for example, the use of the Bernoulli model considering the space-time interaction and also the inclusion of qualitative covariates in the study, we believe that other applications specifically for this data would not be justified, accorded Kulldorff (1997). First, because the use of Bernoulli's spatio-temporal model requires the condition that the number of cases from the previous (prospective) or later (retrospective) period be attributed to the population at risk. In the first case, for the nature of the data in question, it would be inconsistent to use such a model, especially in the peak period of the disease, where the number of cases and deaths tends to increase a lot every day, thus allowing the future number of cases to be higher. than the population at risk, which does not make sense. In the second case, it is not in our interest to evaluate in this way, because what we want is to evaluate the evolution

of the disease. Second, the inclusion of qualitative covariates, such as length of stay, severe respiratory disease, smoking, comorbidities, among others, in this study is not possible due to lack of information. The competent bodies are not making this data available.

3 Diagnostic epidemiological measures

It is of great importance for the bodies responsible for public health in some locality to know the rate of manifestation of a specific disease (incidence) to assess situations of cause and effect and the number of people who were affected by such disease (lethality) by estimating the severity of the disease. We calculated these coefficients for the weighted detected clusters for each 100,000 inhabitants.

3.1 The incidence coefficient

The incidence coefficient (IC) expresses the risk of new cases of disease occurring in a population over a period of time. Its expression is given by

$$IC = \frac{|NC|}{|POP|} * 100,000, \quad (9)$$

where $|NC|$ expresses the new cases and $|POP|$ the population at risk.

3.2 The lethality coefficient

The lethality coefficient (LC) provides the risk of death from a disease in a population over a period of time. Its expression is given by

$$LC = \frac{|CD|}{|CAS|} * 100,000, \quad (10)$$

where $|CD|$ expresses accumulated deaths and $|CAS|$ the number of accumulated cases in this population.

4 Results

In this section, we present our results and a discussion of our perceptions through these results.

4.1 The dataset

The data under study are available from the Ministry of Health of Brazil and can be found at https://susanalitico.saude.gov.br/extensions/covid-19_html/covid-19_html.html. We chose to analyze the evolution of COVID-19 in Brazil, by F.U., in three different periods: 2020-02-04 to 2020-04-01, 2020-04-02 to

2020-07-01, and 2020-07-02 to 2020-09-15. We justify our choice for three reasons: onset of contagion of the disease, peak of contagion of the illness, and phase of stability and fall of the disease, in Brazil (see Figure 1).

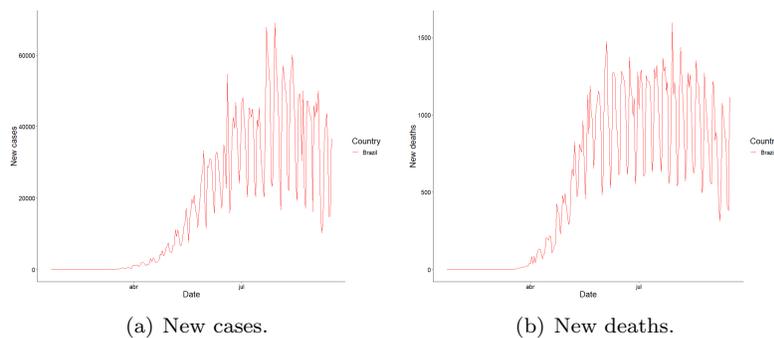


Figure 1 - Trend line of new COVID-19 cases and deaths in Brazil: 2020-02-04 to 2020-09-15.

We only considered the data counted on the last day of each period: 2020-04-01, 2020-07-01, and 2020-09-15, respectively, for the variables number of new cases and new deaths. We justify the choice of these dates because we want to identify, at the end of each period, via the circular scan method, the most plausible cluster. So we chose to analyze the new cases and deaths.

The Brazilian F.U.'S are represented on the maps obtained in the section 4.2 by the classification code according to the Brazilian Institute of Geography and Statistics (IBGE): 11 – Rondônia (RO), 12 – Acre (AC), 13 – Amazonas (AM), 14 – Roraima (RR), 15 – Pará (PA), 16 – Amapá (AP), 17 – Tocantins (TO), 21 – Maranhão (MA), 22 – Piauí (PI), 23 – Ceará (CE), 24 – Rio Grande do Norte (RN), 25 – Paraíba (PB), 26 – Pernambuco (PE), 27 – Alagoas (AL), 28 – Sergipe (SE), 29 – Bahia (BA), 31 – Minas Gerais (MG), 32 – Espírito Santo (ES), 33 – Rio de Janeiro (RJ), 35 – São Paulo (SP), 41 – Paraná (PR), 42 – Santa Catarina (SC), 43 – Rio Grande do Sul (RS), 50 – Mato Grosso do Sul (MS), 51 – Mato Grosso (MT), 52 – Goiás (GO), 53 – Distrito Federal (DF).

We then use the circular scan method proposed by Kulldorff (1997) to find the most plausible cluster for the occurrence of new cases and new deaths from COVID-19, where we choose the maximum radius size of this circle up to 50% of the size of the total population of the map.

4.2 The identified clusters

Unlike what was presented by Cordes *et al.* (2020) and Hohl *et al.* (2020), we seek to identify clusters of new infection cases and new deaths by COVID-19 in Brazilian F.U.'S. Besides, we calculate the incidence and lethality coefficients in the detected clusters. The objective is to describe how worrying the situation

is in these clusters. We believe that these coefficients are an indicator of Organ's competent bodies' intervention with public health policies.

4.2.1 Initial period of the disease: 2020-02-04 to 2020-04-01

In the period of initial disease contagion, Figure 2(a) and 2(b) shows the location of the primary cluster for new cases and also further deaths, where this cluster is formed only by the state of São Paulo (35) ($p\text{-value} = 1 \times 10^{-8}$). Besides, Figures 2(c) and 2(d) present the description of this cluster concerning new cases and new deaths from COVID-19 in the initial stage of the disease.

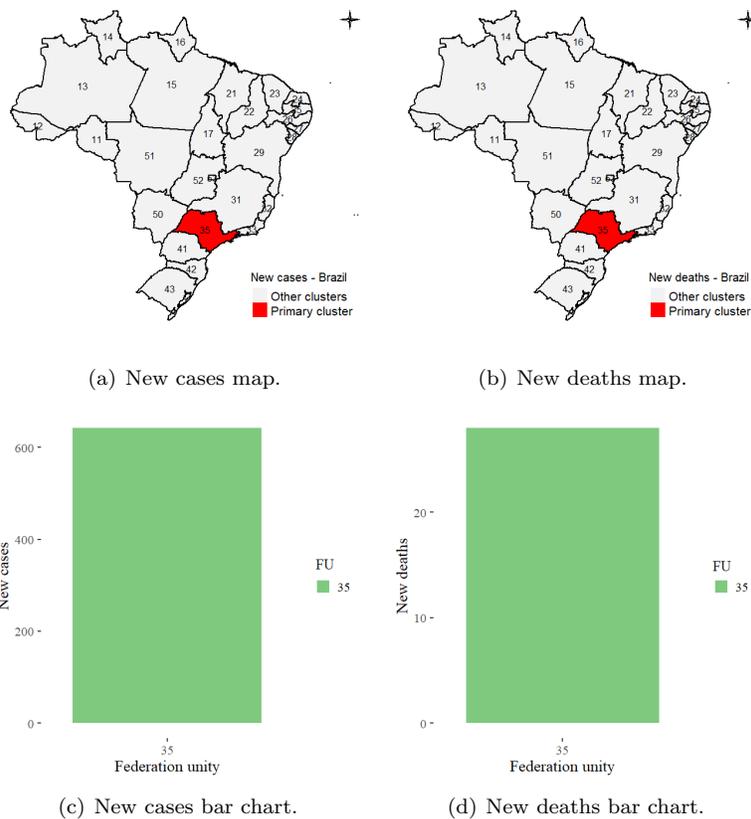


Figure 2 - The cluster of new cases, new deaths, and description measured in a bar chart of COVID-19 in the Brazilian FU's: 2020-02-04 to 2020-04-01.

We mentioned earlier that we are considering three distinct periods of study: the initial period of contagion, the picon period of contagion of the disease, and the period of stability and fall. Then, we used Kuldorff's circular scan method

considering the disease's peak period to verify if the cluster detected in that period remains.

4.2.2 Peak illness period: 2020-04-02 to 2020-07-01

We noted in Figures 3(a) and 3(b) that the most plausible cluster by the circular scan method is now the state of Roraima (14) ($p\text{-value} = 1 \times 10^{-8}$) considering the new cases and new deaths recorded on 01-07-2020. Figures 3(c) and 3(d) illustrate the absolute values of new cases and new deaths for this cluster.

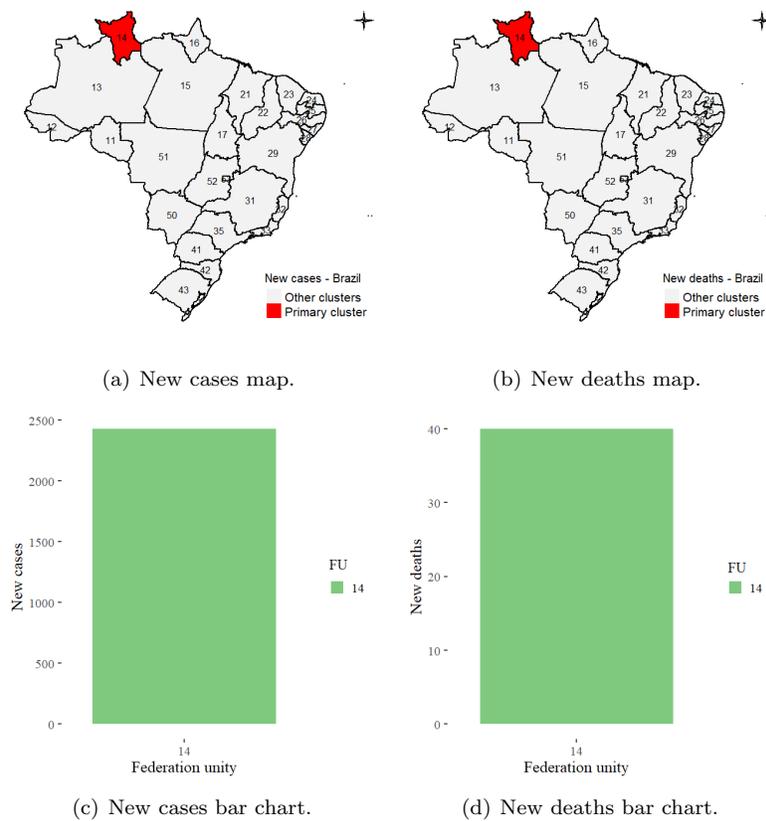
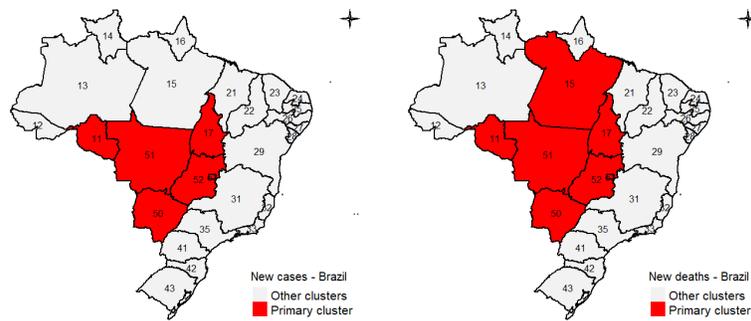


Figure 3 - The cluster of new cases, new deaths, and description measured in a bar chart of COVID-19 in the Brazilian FU's: 2020-04-02 to 2020-07-01.

We then applied the same method in the third period of infection of the disease: stability and fall.

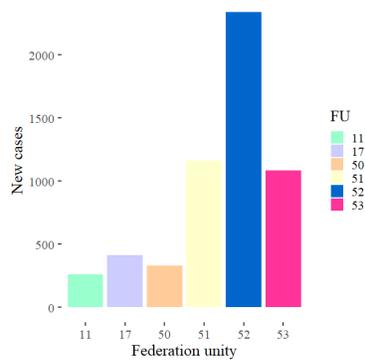
4.2.3 Period of stability and fall of the disease: 2020-07-02 to 2020-09-15

In the period of stability and fall of the disease, Figures 4(a) and 4(b) shows the spatial distribution of the primary cluster found for new cases and also for further deaths, and these clusters are formed by F.U.'S Rondônia (11), Mato Grosso (51), Tocantins (17), Goiás (52), Distrito Federal (53), and Mato Grosso do Sul (50), in addition to Pará (15) which it belongs only to the cluster related to new deaths ($p\text{-value} = 1 \times 10^{-8}$). Besides, Figures 4(c) and 4(d) show heterogeneity in the absolute values assumed by such variables in the F.U.'S that form each of these clusters.

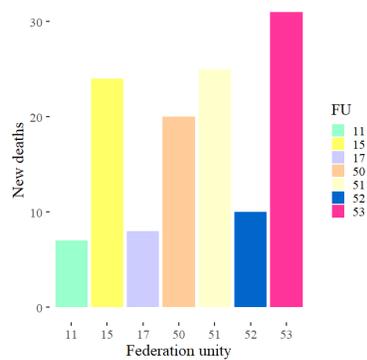


(a) New cases map.

(b) New deaths map.



(c) New cases bar chart.



(d) New deaths bar chart.

Figure 4 - The cluster of new cases, new deaths, and description measured in a bar chart of COVID-19 in the Brazilian FU's: 2020-07-02 to 2020-09-15.

We then calculated the coefficients of incidence and mortality for each of the clusters detected in each evaluated period. We emphasize that, although clusters are detected, this does not mean that the manifestation (incidence) and severity (lethality) of COVID-19 are the most serious in the Brazilian FU's. We only

evaluated these measures in the detected clusters.

4.3 The coefficients: incidence and lethality.

Recently, (GOUVEIA *et al.*, 2020) analyzed the incidence and lethality of COVID-19 in Brazil through a historical series. The results showed an increasing trend of incidence in all states, except Ceará, and the highest lethality found was in the state of Piauí. (MACIEL *et al.*, 2020) analyzed the spatial distribution of the incidence of COVID-19 in Brazil and correlated it with the municipal human development index (MHDI) of the municipalities in Ceará.

Unlike, we, once we identified distinct clusters in the three periods mentioned above, to verify these clusters' behavior in each of these periods, we calculated the coefficients of incidence and lethality (see section 3) each contagion period of COVID-19 for the respective detected clusters. The justification for choosing these coefficients is that they measure the rate of manifestation of the disease (incidence) and also the cause and effect situations by estimating the severity of the illness (lethality) (see section 3).

Barbosa *et al.* (2020) evaluated Brazilian FU's that had up to 50 deaths from COVID-19 regarding the accumulated incidence, mortality, and lethality among the elderly Brazilian population, in the period between the beginning of the pandemic in Brazil until 2020-05-25. The variables considered were services and health professionals' offer, demographic, income, and development indicators. As a result, the authors concluded that Pará (15) had the highest accumulated incidence and mortality rates (763.37, 219.06) among the elderly, for every 100,000 inhabitants. The highest accumulated mortality rates were observed in Bahia (29), Rio de Janeiro (33), and Pernambuco (26).

We, in the same period, according to Table 1, considered all Brazilian federative units and not only the elderly population, but all individuals at risk, including children, and observed that São Paulo (35), which was the cluster detected in this period via circular scan statistics, showed an incidence of 1.39 and lethality of 0.06 for every 100,000 inhabitants.

Table 1 - Each cluster's incidence and lethality coefficients were found in each corresponding period (Date) for every 100,000 inhabitants. The term FU means the federation unit, where each code is by the classification of the Brazilian Institute of Geography and Statistics (IBGE)

Date	Cluster (F.U.)	Incidence	Cluster (F.U.)	Lethality
2020-04-01	35	1.3869	35	0.0605
2020-07-01	14	384.9926	14	6.3373
2020-09-15	11, 17, 50 51, 52, 53	27.8015	11, 15, 17, 50 51, 52, 53	0.4373

Unlike Barbosa *et al.* (2020), we went ahead and evaluated the incidence and

lethality in the clusters detected in the other periods: peak and stability and fall. Table 1 shows that Roraima (14), which was the cluster detected during the peak phase of contagion of the disease (2020-07-01), had the highest coefficients for each 100,000 inhabitants. The incidence coefficient for that cluster in that period was 384,9926, indicating that for every 100,000 individuals at risk, 385 became new cases of COVID-19.

This cluster's lethality coefficient was 6.3373 considering only new deaths, indicating that for every 100,000 individuals at risk, approximately 7 individuals died daily. Both coefficients were obtained at the end of the peak contagion period for this disease.

Conclusions

The circular scan method proposed by Martin Kulldorff corresponded to the detection of the most plausible clusters in each period considered. We note that these clusters were consistent with what was reported by Brazilian media and also government health agencies throughout each period considered. We recommend using this method to detect clusters emerging from georeferenced phenomena of any nature.

Acknowledgments

We thank CNPq for the financial support, and reviewers and editors for their comments and suggestions.

ALVES, H.J.P; FERNANDES, F.A.; LIMA, K.P. de; BATISTA, B. D. de O.; FERNANDES, T. J.. Incidência e letalidade dos clusters de COVID-19 no Brasil via método scan circular. *Rev. Bras. Biom.*, Lavras, v.39, n.4, p.556-570, 2021.

- *RESUMO: A pandemia da COVID-19 se alastrou rapidamente no mundo de forma assustadora. No Brasil, o terceiro país do mundo com maior número de infectados e mortes pela doença, é importante para as autoridades governamentais de saúde identificar as unidades da federação que se destacam em casos e mortes dessa doença para o direcionamento de recursos. O método scan circular de Kulldorff permite identificar com alguma significância estatística as unidades da federação que mais se destacam, conhecidas como clusters. Uma vez identificado esses clusters, nós utilizamos os coeficientes de incidência e letalidade com o intuito de descrever melhor o comportamento desses clusters durante três fases da pandemia: a fase inicial, a fase de pico e também a fase de estabilidade e queda. Nós observamos mudanças de localidade dos clusters identificados nessas três fases e utilizamos o software R e também o software SaTScan na obtenção dos mapas e resultados.*
- *PALAVRAS-CHAVE: COVID-19; pandemia; Brasil; Kulldorff; incidência; letalidade.*

References

- ALBOLHASSANI, A.; PRATES, M. O.; CASTELLARES, F.; MAHMOODI, S. Zero-inflated bell scan: A more flexible spatial scan statistic. *Spatial Statistics*, p.100433, 2020.
- ALVES, H. J. de P.; FERNANDES, F. A.; LIMA, K. P. de; BATISTA, B. D. de O.; FERNANDES, T. J. The COVID-19 pandemic in Brazil: an application of the k-means clustering method. *Research, Society and Development*, v.9, n.10, e5829109059, 2020.
- BARBOSA, I. R.; GALVÃO, M. H. R.; SOUZA, T. A. de; GOMES, S. M.; MEDEIROS, A. A.; LIMA, K. C. de. Incidência e mortalidade por COVID-19 na população idosa brasileira e sua relação com indicadores contextuais: um estudo ecológico. *Revista Brasileira de Geriatria e Gerontologia*, v.23, n.1, 2020.
- BIRCH, C.; CHIKUKWA, A.; HYDER, K.; VILAS, V. Spatial distribution of the active surveillance of sheep scrapie in Great Britain: An exploratory analysis. *BMC veterinary research*, v.5, n.23, p.1-14, 2009.
- CHOYNOWSKI, M. Maps based on probabilities. *Journal of the American Statistical Association*, v.54, n.286, p.385-388, 1959.
- CLAYTON, D.; KALDOR, J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, p.671-681, 1987.
- CORDES, J.; CASTRO, M. C. Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Elsevier: Spatial and Spatio-temporal Epidemiology*, n.34, p.100355, 2020.
- COSTA, M. A.; ASSUNÇÃO, R. M.; KULLDORFF, M. Constrained spanning tree algorithms for irregularly-shaped spatial clustering. *Computational Statistics & Data Analysis*, p.1771-1783, 2012.
- DONG, E.; D.U., H.; GARDNER, L. Um painel interativo baseado na web para rastrear COVID-19 em tempo real. *Doenças infecciosas da Lancet*, v.20, n.5, p.533-534, 2020.
- DUARTE, A. R.; SILVA, S. B. D.; OLIVEIRA, F. L. P. D.; RIBEIRO, M. C.; CANÇADO, A. L. F.; MOURA, F. D. R. A weighted non-connectivity penalty for detection and inference of irregularly shaped clusters. *Revista Brasileira de Biometria*, v.35, n.1, p.160-173, 2017.
- DUCZMAL, L.; ASSUNÇÃO, R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis*, v.45, n.2, p.269-286, 2004.

- DUCZMAL, L.; KULLDORFF, M.; HUANG, L. Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics*, v.15, n.2, p.428-442, 2006.
- FERNANDES, F. A.; ALVES, H. J. P. de; FERNANDES T. J.; MUNIZ, J. A. Panorama da fase inicial do crescimento dos números de casos e óbitos causados pela COVID-19 no Brasil. *Research, Society and Development*, v.9, n. 10, p.1-19, 2020.
- GOUVEIA, G. P.; GOUVEIA, S. S. V.; VASCONCELOS, S. S.; SILVA, E. S. M.; MEDIEROS, R. G.; DA SILVA, R. L. Historical series of cases of COVID-19 in Brazilian regions in the year of 2020. *Journal of Health & Biological Sciences*, v.8, n.1, p.1-7, 2020.
- HAN, J.; ZHU, L.; KULLDORFF, M.; HOSTOVICH, S.; STINCHCOMB, D. G.; TATALOVICH, Z.; LEWIS, D. R.; FEUER, E. J. Using Gini coefficient to determine optimal cluster reporting sizes for spatial scan statistics. *International journal of health Geographics*, v.15, n.1, p.1-11, 2016.
- HOHL, A.; DELMELLE, E. M.; DESJARDINS, M. R.; LAN, Yu. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Elsevier: Spatial and Spatio-temporal Epidemiology*, n.34, p.100354, 2020.
- KHAILANY, R. A.; SAFDAR, M.; OZASLAN, M. Genomic characterization of a novel SARS-CoV-2. *Gene reports*, p.100682, 2020.
- KULLDORFF, M. A spatial scan statistic. *Communications in Statistics-Theory and Methods*, v.26, n.6, p.1481-1496, 1997.
- KULLDORFF, M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, v.164, n.1, p.61-72, 2001.
- KULLDORFF, M.; HEFFERNAN, R.; HARTMAN, J.; ASSUNÇÃO, R.; MOSTASHARI, F. A space-time permutation scan statistic for disease outbreak detection. *Plos med*, v.2, n.3, p.e59, 2005.
- LETKO, M.; MARZI, A.; MUNSTER, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B beta coronaviruses. *Nat Microbiol: PUBMED*, n.5, p.562-569, 2020.
- LI, T.; GAO, Y. WANG, S. Escip: an expansion-based spatial clustering method for inhomogeneous point processes. *Annals of the American Association of Geographers*, v.110, n.1, p.259-276, 2020.
- MACIEL, J. A. C.; CASTRO-SILVA, I. I.; FARIAS, M. R. de. Análise inicial da correlação espacial entre a incidência de COVID-19 e o desenvolvimento humano nos municípios do estado do Ceará no Brasil. *SciELO Public Health: Revista Brasileira de Epidemiologia*, n.23, p. e200057, 2020.

- NASSIRI, R. Perspective on Wuhan viral pneumonia. *Advances in Public Health, Community and Tropical Medicine*, v.2, 2020.
- NAUS, J. L. Clustering of random points in two dimensions. *Biometrika*, v.52, n.1-2, p.263-266, 1965a.
- NAUS, J. I. The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*, v.60, n.310, p.532-538, 1965b.
- NOVEE, L. C. L.; PALATINO, M. C. Detecting local clusters of under-5 malnutrition in the province of Marinduque, Philippines using spatial scan statistics. *Nutrition and Metabolic Insights*, v.13, e1178638820940670, 2020.
- OPENSHAW, S.; CHARLTON, M.; CRAFT, A. W.; BIRCH, J. M. Investigation of leukemia clusters by use of a geographical analysis machine. *The Lancet*, v.331 n.8580, p.272-273, 1988.
- PINTO, E.; SANTOS, G.; OLIVEIRA, F. L. P. de. Análise espaço-temporal aplicada às ocorrências de hipertensão e diabetes nos municípios do estado de Minas Gerais. *Revista Brasileira de Biometria*, v32, n.01, p.238-266, 2014.
- SALATHÉ, M.; ALTHAUS, C. L.; NEHER, R.; STRINGHINI, S.; HODCROFT, E.; FELLAY, J.; ZWAHLEN, M.; SENTI, G.; BATTEGAY, M.; WILDER-SMITH, A. Covid-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss medical weekly*, v.150, n.w20225, 2020.
- SAÚDE, Ministério da. Painel Coronavírus. Recuperado de <https://covid.saude.gov.br/>. 2020.
- TURNBULL, B. W.; IWANO, E. J.; BURNETT, W. S.; HOWE, H. L.; CLARK, L. C. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology*, v.32, n. suppl, p.136-143, 1989.
- VELAVAN, T.; MEYER, C. La epidemia de covid-19. *Trop Med Int Health*, 2020.
- WHITTEMORE, A. S.; FRIEND, N.; BROWN Jr, B. W.; HOLLY, E. A. A test to detect clusters of disease. *Biometrika*, v.74, n.3, p.631-635, 1987.
- WILDE, A. H.; SNIJDER, E. J.; KIKKERT, M.; van HEMERT, M. J. Host factors in coronavirus replication. In: *Roles of Host Gene and Non-coding RNA Expression in Virus Infection*. Springer, p.1-42, 2020.

Received on 23.10.2020.

Approved after revised on 05.04.2021.