Spatial cluster analysis of COVID-19 in Malaysia (Mar-Sep, 2020)

Sami Ullah,[1] Nurul Hidayah Mohd Nor,[1] Hanita Daud,[1] Nooraini Zainuddin,[1] Muhammad Salim Javed Gandapur,[2] Iftikhar Ali,[3] Alamgir Khalil[4]

[1]Department of Fundamental & Applied Sciences, Universiti Teknologi Petronas, Seri Iskandar, Perak, Malaysia; [2]IIC University of Technologi, Kingdom of Cambodia; [3]Project Wing (PSDP), Planning Division & Special Initiatives, Govt of the Pakistan, Islamabad, Pakistan; [4]Department of Statistics, University of Peshawar, Peshawar, Pakistan

Abstract

Coronavirus disease 2019 (COVID-19) is the current worldwide pandemic as declared by the World Health Organization (WHO) in March 2020. Being part of the ongoing global pandemic, Malaysia has recorded a total of 8639 COVID-19 cases and 121 deaths as of 30th June 2020. This study aims to detect spatial clusters of COVID-19 in Malaysia using the Spatial Scan Statistic (SaTScan™) to guide control authorities on prioritizing locations for targeted interventions. The spatial analyses were conducted on a monthly basis at the state-level from March to September 2020. The results show that the most likely cluster of COVID-19 occurred in West Malaysia repeatedly from March to June, covering three counties (two federal territories and one neighbouring state) and moved to East Malaysia in July covering two other counties. The most likely cluster shows a tendency of having moved from the western part to the eastern part of the country. These results provide information that can be used for the evidence-based interventions to control the spread of COVID-19 in Malaysia.

Correspondence: Sami Ullah, Department of Fundamental & Applied Sciences, Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia.
Tel.: +60143074781.
E-mail: sami.khan3891@gmail.com

Key words: COVID-19; spatial clusters; Malaysia; SaTScan™.

Acknowledgements: the authors are grateful to Universiti Teknologi Petronas for providing facilities for conducting this study.

Funding: this article was funded by a Fundamental Research Grant Scheme (FRGS) from the Ministry of Education, Malaysia ((Ref: FRGS/1/2018/STG06/UTP/02/1)), and Murata Science Foundation, Japan (cost centre: 015ME0-107), received by Hanita Daud.

Conflict of interests: the authors declare no potential conflict of interests.

Received for publication: 20 December 2020.
Revision received: 20 February 2021.
Accepted for publication: 20 February 2021.

Geospatial Health 2021; 16:961
doi:10.4081/gh.2021.961

Introduction

A cluster of pneumonia cases infected with a coronavirus was reported in Wuhan, a city in Hubei province of China in December 2019 (Huang *et al*., 2020; Zhu *et al*., 2020). The Chinese authorities officially announced on 7 January 2020 that the aetiological agent was a novel coronavirus, now known as SARS-CoV-2, which is the cause of the acute respiratory coronavirus disease 2019 or COVID-19 (Zhou *et al*., 2020). The World Health Organization (WHO) declared COVID-19 outbreak, first as a Public Health Emergency of International Concern on 30 January 2020 and subsequently as a worldwide pandemic on 11 March 2020 (WHO, 2020, 2020a). Population density and intensity of social contacts are the main drivers for the propagation and amplification of this virus (Rocklöv & Sjödin, 2020).

COVID-19 was confirmed to have reached Malaysia in January 2020, when it was detected on travellers from China arriving *via* Singapore on 25 January 2020, following the first outbreak in Hubei Province, China (Reuters Staff, 2021). The number of reported cases remained first low (fewer than 30 cases) and were largely confined to imported cases until localized clusters began to emerge in March 2020; the largest cluster was linked to a religious gathering held in Sri Petaling, Kuala Lumpur in late February and early March 2020, leading to massive spikes of local cases and export of the infection to neighbouring countries by travel (Ng, 2021). Within a few weeks, Malaysia had recorded the largest cumulative number of confirmed COVID-19 infections in Southeast Asia (The Strait Times, 2021), breaching the number of 2000 active cases by the end of March 2020.

Few studies have been conducted focusing some of the characteristics of COVID-19 pandemic in Malaysia, such as impact of preventive measures on the infection dynamics (Khor *et al*., 2020; Salim *et al*., 2020), effect on mental health (Shanmugam *et al*., 2020) and behavioural changes in the general population (Hanafiah & Wan, 2020; Koh *et al*., 2020). However, spatial cluster analysis of COVID-19 cases in the country seems not to have been carried out so far. This would be useful by guiding control programs on prioritizing locations for targeted interventions, rapid testing and allocation of other limited resources.

One of the statistical methods for detecting spatial/geographical clusters of a disease is the Spatial Scan Statistic (Kulldorff, 1997), which is a freely available software called SaTScan™

(https://www.satscan.org/). It is widely used in public health and epidemiology (Amin *et al*., 2019; Omodior *et al*., 2019; Souza *et al*., 2019; Amin & Rivera, 2020; Desjardins *et al*., 2020; Ullah *et al*., 2020). There are several other cluster analysis algorithms in the literature, such as DBSCAN (Ester *et al*., 1996). It is understood that none of the cluster analysis methods is better than all the others. Each algorithm has some advantages for specific applications. We used SaTScan™ as it has the advantage of detecting cluster location, while also evaluating its statistical significance. This approach supplement and assist creation of basic maps of disease rates by relying on a variety of data models to determine whether the observed spatial or space-time patterns of a disease are due to chance or if they are randomly distributed. A useful function of SaTScathen™ that assist decision making is the possibility to integrate model outputs and application of geographic information systems (GIS) for a map-visualization.

This study aimed to identify statistically significant spatial clusters of COVID-19 cases at the state-level in Malaysia during the period from March to September 2020 using scan statistics under the discrete Poisson model. Spatial analysis was carried out for each month to identify the location of significant clusters in each month of the pandemic period.

Materials and methods

Study area and data

This study was conducted at the state-level in Malaysia. The total terrestrial area of Malaysia is divided into thirteen states and three federal territories, separated by the South China Sea into two regions, East Malaysia and West Malaysia. The geographical map of the study area is shown in Figure 1. West Malaysia shares a land and maritime border with Thailand and maritime borders with Singapore, Vietnam and Indonesia. East Malaysia shares land and maritime borders with Brunei and Indonesia and a maritime border with The Philippines and Vietnam. The monthly data on COVID-19 cases were collected for each of the sixteen counties (thirteen states and three federal territories) during the period from March to September 2020 from the Ministry of Health Malaysia (2020). These data are currently updated on a daily basis while the confirmed cases are spatially aggregated at the state-level on a monthly basis. The population counts for the respective states with respect to the second quarter 2019 were obtained from Demographics Statistics Malaysia (Statista, 2020).

Spatial scan statistic

Spatial scan statistics scans the region under study by a circular moving window defining a set of zones. The circular scanning window is centred on each county's centroids with a radius size continuously varying from zero up to a fixed value. The circle includes the counties whose centroids lie within the circle. Hence, a large number of overlapping circles of different sizes are obtained that mutually cover the total study area. Each circle denotes a possible candidate cluster.

When using this approach with the Poisson model, the case counts, x_i in each circular zone z are assumed to be distributed under the null hypothesis H_0 of spatial randomness (Eq. 1):

$$x_i \sim \mathrm{Pois}(\lambda_0) \qquad \text{Eq. 1}$$

The likelihood functions L_0 with the parameter restricted under the null hypothesis H_0 and L_1 with parameter unrestricted under the alternative hypothesis are then computed for each circular zone z. In each zone z, L_0 and L_1 have different parameters, from the heterogeneous population distribution. The method searches for zone z that maximizes the likelihood ratio (LR) as in Eq. 2:

$$\mathrm{LR}(z_i) = \left(\frac{L_1}{L_0}\right)_{z_i} \qquad \text{Eq. 2}$$

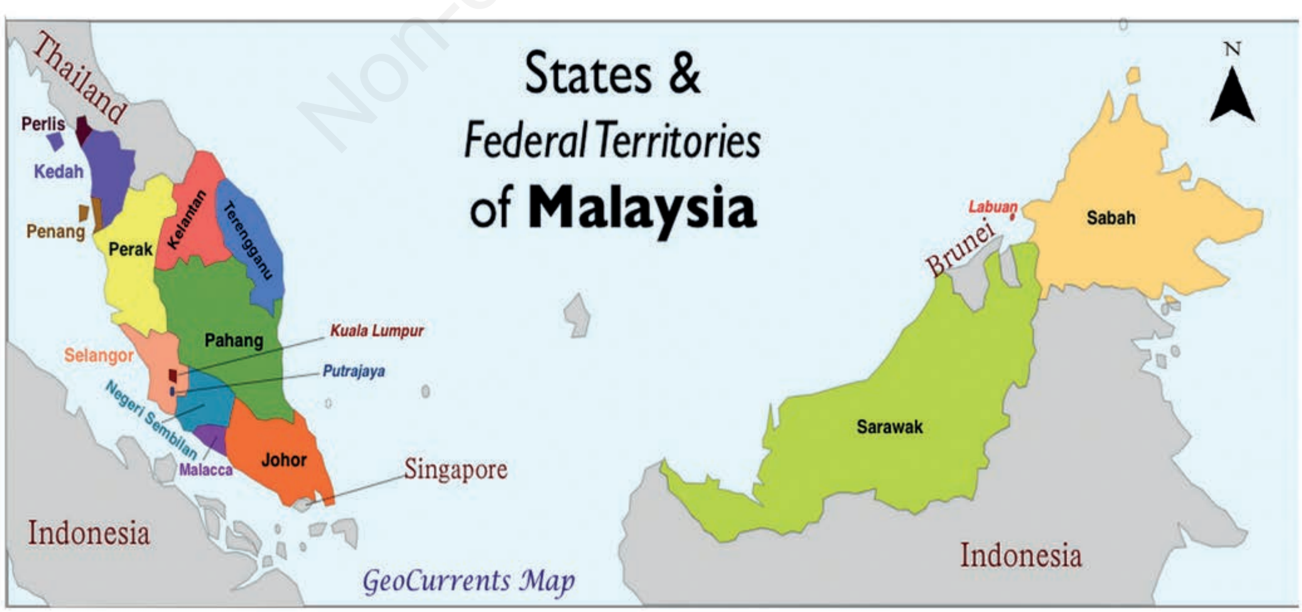

Figure 1. Geographical map of the study area (Wikipedia, 2020).

For each possible zone, the Log-likelihood ratio (LLR) is calculated and its statistical significance tested by Monte Carlo simulations (Kulldorff, 2018). For each potential cluster, a large number of replications of datasets (usually 999 or 9999) using the restricted parameters under H_0 is generated. The LLR is computed and a distribution is obtained. If the potential cluster has an LLR higher than, for instance, 95% of the datasets generated under the null hypothesis, the cluster is said to be statistically significant at 0.05 confidence level. The zone with the maximum LLR value and with the statistical significance is considered to be the most likely cluster, that is, the cluster least likely to have occurred by chance. The other zones with statistically significant LLR values are defined as the secondary clusters.

In this study, the spatial scan statistics was applied under the discrete Poisson Model using circular windows and with the default setting of maximum spatial window sizes *i.e.* <50% of the total population of the study area (Kulldorff, 2018). The clusters having a P-value of less than 0.05 were considered to be the significant clusters.

Results

Potential spatial clusters of COVID-19 cases in Malaysia were detected during each month from March to September 2020 (Table 1). In March, most likely clusters were seen in the three states (Negeri Sembilan, Putrajaya, Kuala Lumpur), and a secondary cluster in one state (Selangor). In April, the same states were detected as most likely clusters with Sarawak as the 1st secondary cluster, and Selangor as the 2nd secondary cluster. In May, the most likely clusters were again detected in same states, this time with a secondary cluster in only one state (Selangor). In June, the picture was again the same with respect to the most likely clusters, but this time the secondary cluster was in Melaka State. In July, the most likely cluster was detected in East Malaysia, covering two counties (Sarawak and Labuan), while two secondary clusters were detected in West Malaysia, each covering a single county (Kuala Lumpur and Johor). In August, the Most likely cluster was seen in Kuala Lumpur, 1st secondary cluster in two states (Perlis, Kedah), and the 2nd secondary cluster in a single county (Labuan). In September, the most likely cluster had moved back to the Eastern state Saba, while the secondary cluster was seen in the western state Kedah. All the clusters detected in March, April, and May were statistically significant at P<0.05. However, the secondary clusters in June (Melaka), in July (Johor), in August (Labuan) and in September (Kedah) did not reach statical significance at this level (Table 1).

The resulting geographical locations of COVID-19 clusters in the study area for each month from March to September are shown in MAPS IN Figures 2-8.

Discussion

The resuls show that the potential spatial clusters of COVID-19 tend to move from the Western part to the Eastern Part of the country. The most likely cluster has occurred repeatedly in each consecutive month from March until June, covering three states (Negeri Sembilan, Putrajaya, and Kuala Lumpur) in the western part of the country. The strong spread of COVID-19 in these states can be attributed to the cramped living conditions of foreign worker residences in factories and immigration detention centres (CodeBlue, 2020; Malaysiakini, 2020). The large number of COVID-19 can also be linked to a religious gathering at the Seri Petaling Mosque in Kuala Lumpur (BH Online, 2020; TheStar, 2020). However, in July, the most likely cluster moved from West Malaysia to East Malaysia. In August it moved to the West Malaysia for one month and then again back East Malaysia in September. The spike in COVID-19 cases in the eastern part of the country may be due to the expatriates from Indonesia crossing the border illegally. From the spatial point of view, the clusters were inconsistent during the whole study period *i.e.*, they moved from one place to another without producing a trend. The control measure of the government, such as movement control order (MCO) in the highly effected areas, and events of social gathering in other

Table 1. Spatial clusters of COVID-19 cases in Malaysia during each month from March to September.

Month	Cluster type	State	No. of states	Observed cases (no.)	Expected cases (no.)	LLR	P
March	Most likely	Negeri Sembilan, Putrajaya, Kuala Lumpur	3	637	256	230.282	<0.001
	1st secondary	Selangor	1	704	554	23.816	<0.001
April	Most likely	Negeri Sembilan, Putrajaya, Kuala Lumpur	3	1189	29	900.917	<0.001
	1st secondary	Sarawak	1	351	279	9.361	<0.001
	2nd secondary	Selangor	1	727	648	5.788	0.016
May	Most likely	Negeri Sembilan, Putrajaya, Kuala Lumpur	3	1157	168	1627.381	<0.001
	1st secondary	Selangor	1	478	364	20.799	<0.001
June	Most likely	Negeri Sembilan, Putrajaya, Kuala Lumpur	3	566	73	893.267	<0.001
	1st secondary	Melaka	1	28	23	0.598	0.983
July	Most likely	Sarawak, Labuan	2	109	31	70.402	<0.001
	1st secondary	Kuala Lumpur	1	40	19	9.706	<0.001
	2nd secondary	Johor	1	51	40	1.672	0.601
August	Most likely	Kuala Lumpur	1	111	20	113.008	<0.001
	1st secondary	Perlis, Kedah	2	120	27	99.452	<0.001
	2nd secondary	Labuan	1	4	1	2.251	0.369
September	Most likely	Sabah	1	1460	226	2148.832	<0.001
	1st secondary	Kedah	1	151	126	2.504	0.314

LLR, Log-likelihood ratio.

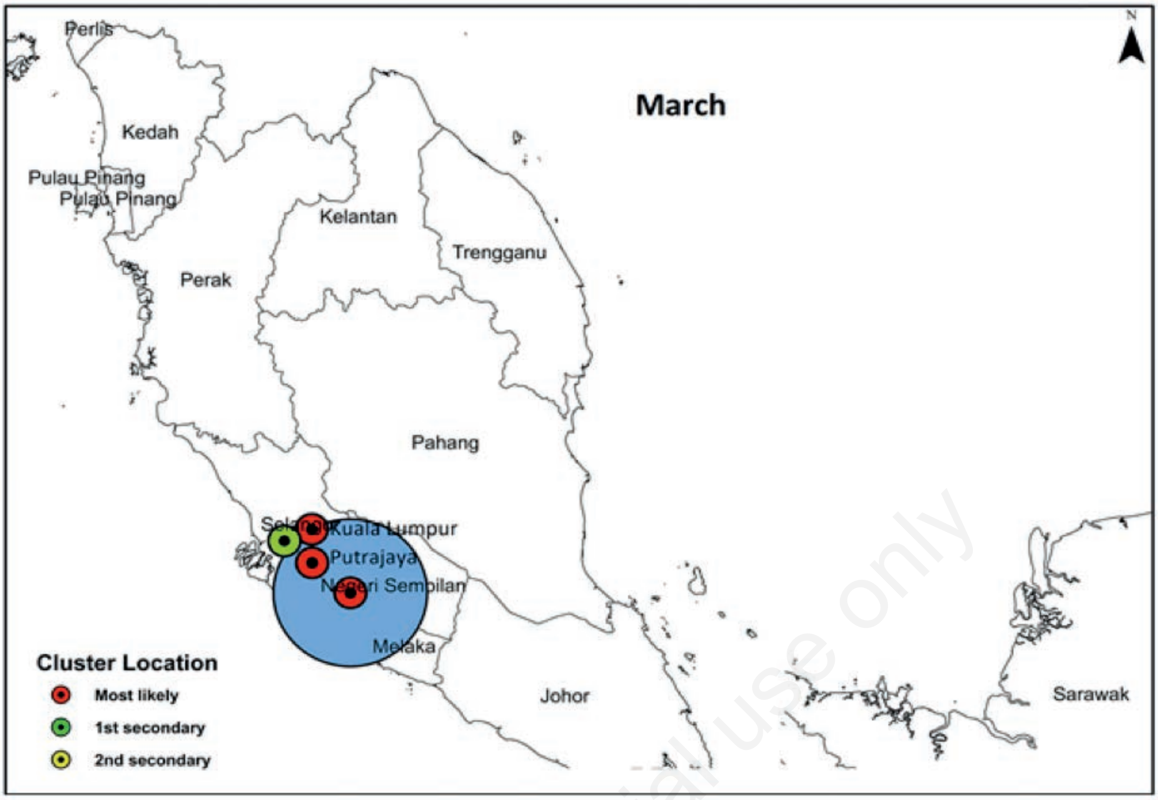

Figure 2. COVID-19 cluster map for March.

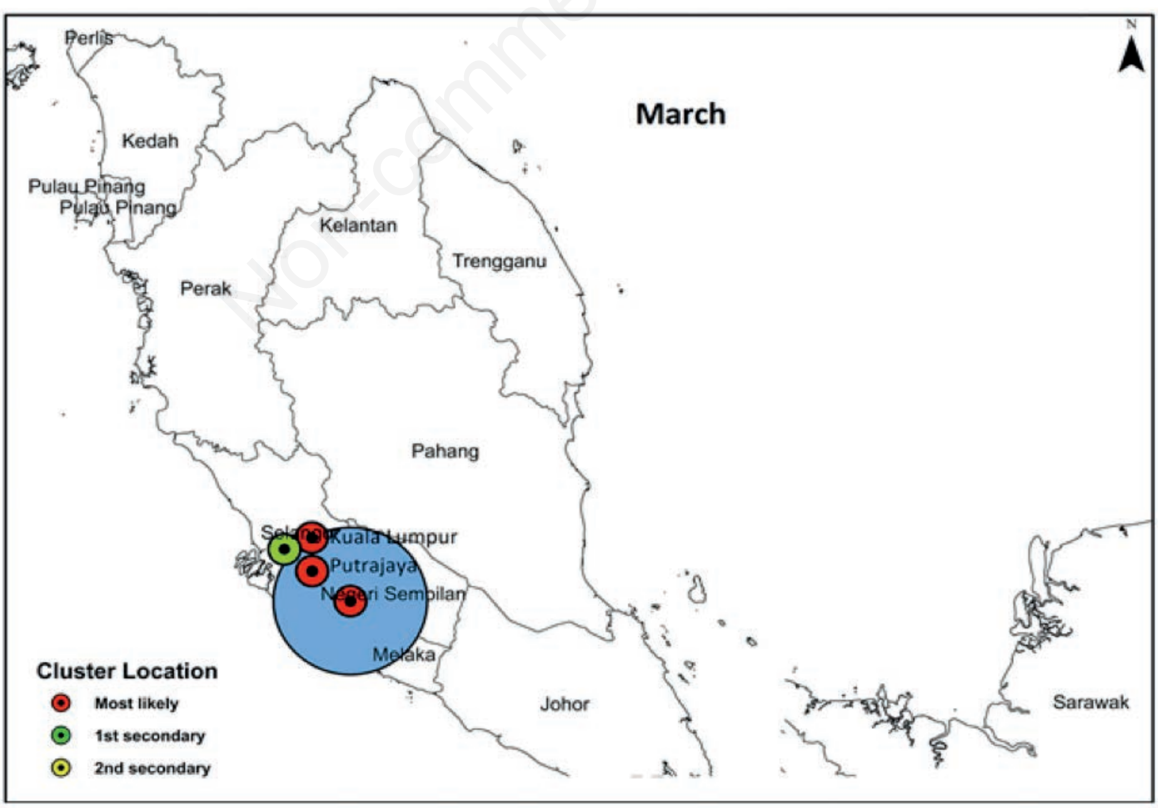

Figure 3. COVID-19 cluster map for April.

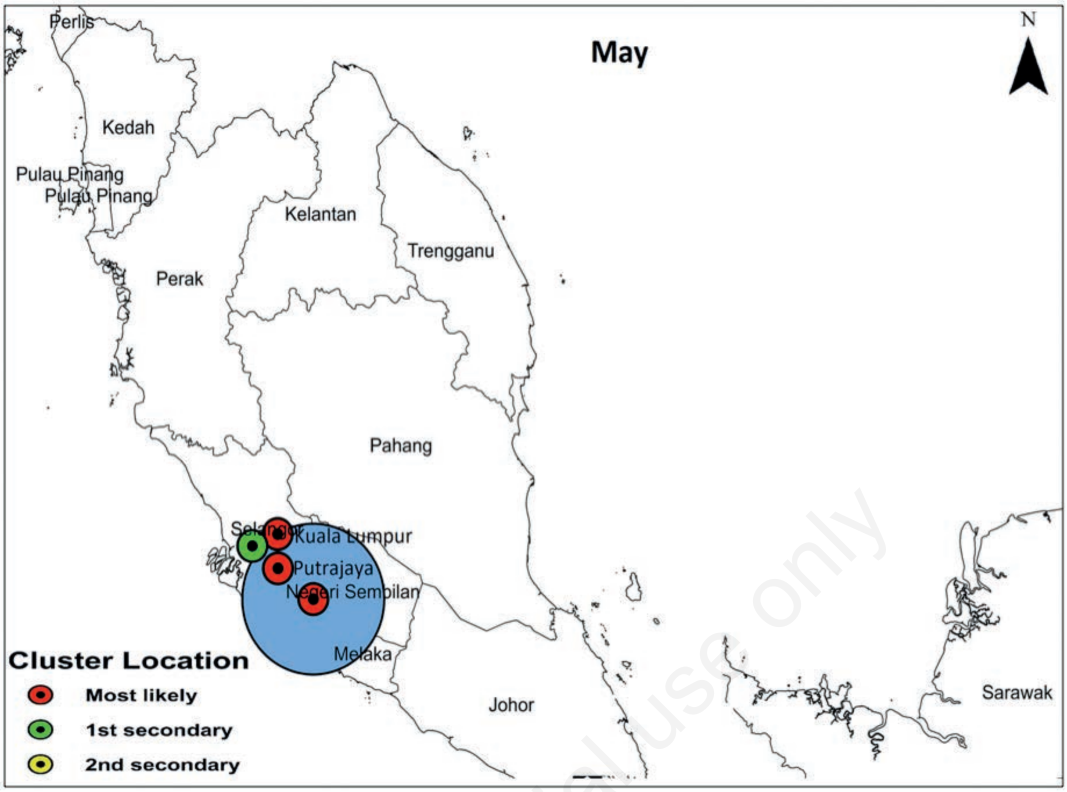

Figure 4. COVID-19 cluster map for May.

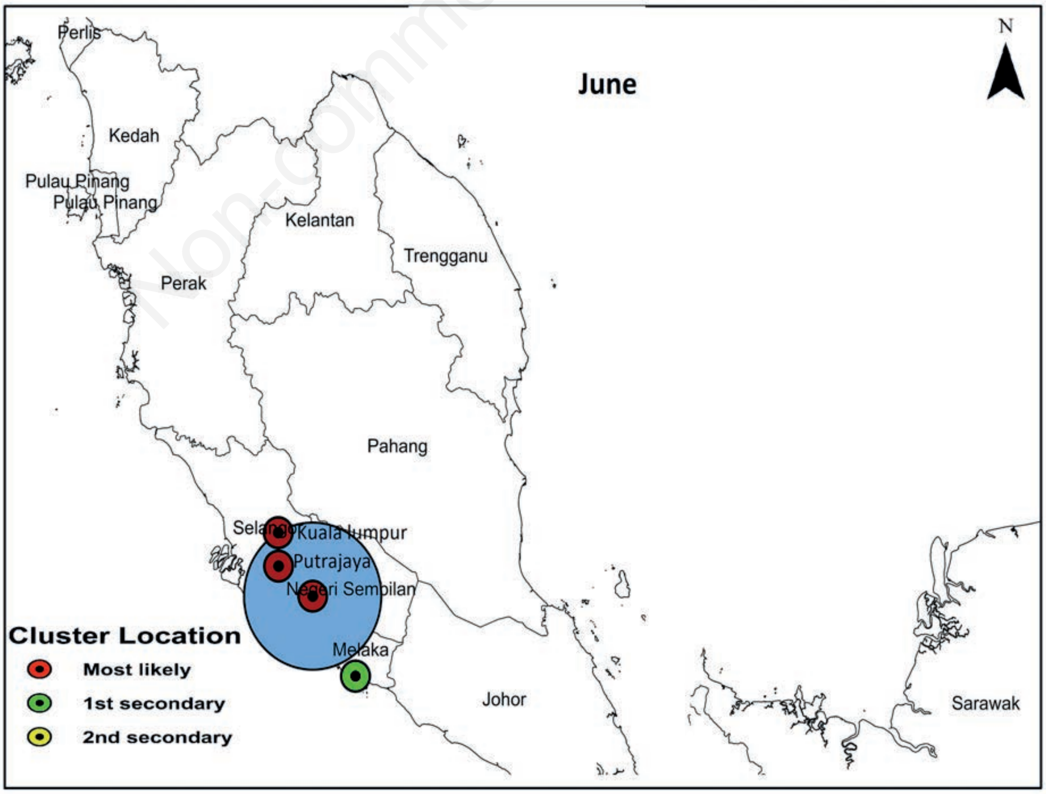

Figure 5. COVID-19 cluster map for June.

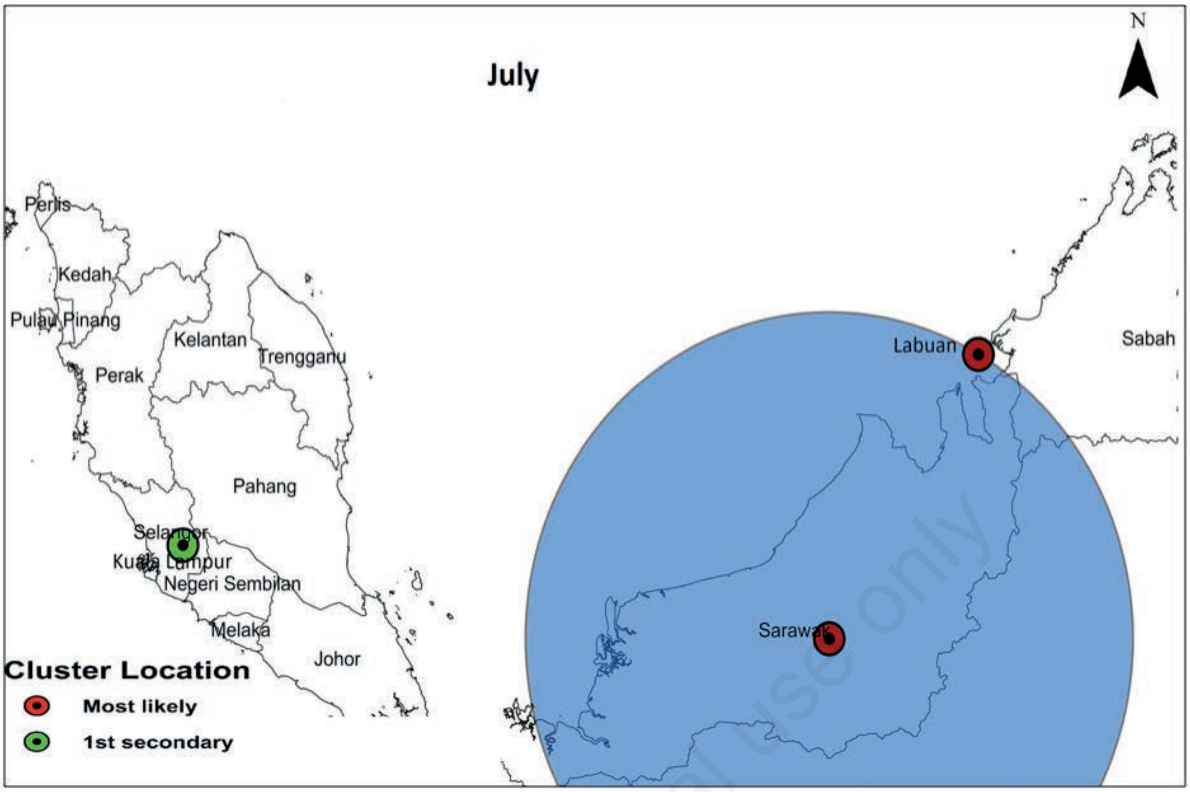

Figure 6. COVID-19 cluster map for July.

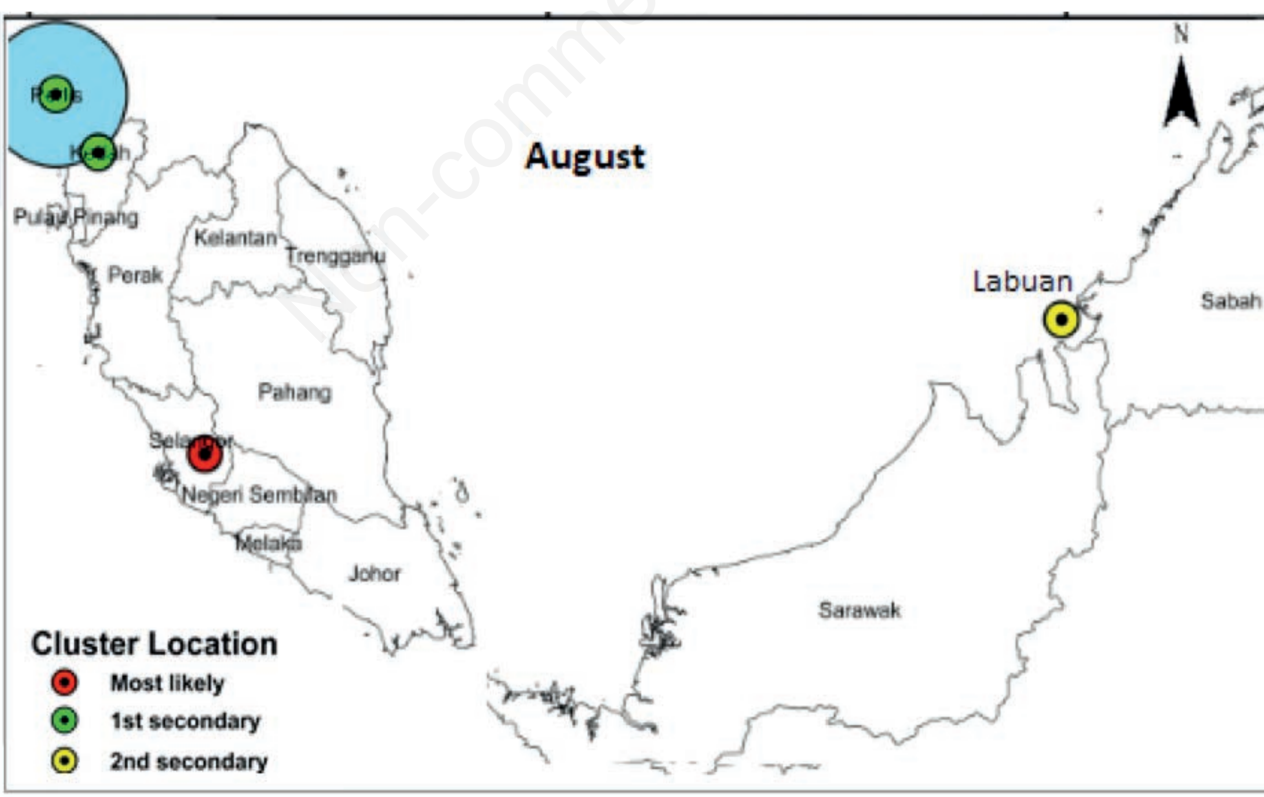

Figure 7. COVID-19 cluster map for August.

Figure 8. COVID-19 cluster map for September.

areas might have caused the spatial clusters of COVID-19 to change their locations in this way. However, the geographical cluster locations identified deserve increased public health and clinical attention as there is important to control the spread of COVID-19 in the country. Some important patterns in the spatial clusters of COVID-19 cases were identified. These patterns are of high significance for epidemiologists to find the environmental and social factors that possibly foster the spread of COVID-19 and other infectious diseases in an area.

Conclusions

This study contributes to the ongoing COVID-19 surveillance efforts by highlighting the potential spatial clusters for targeted interventions to mitigate further spread. The presence of spatial clusters of COVID-19 at the state-level in Malaysia during each month from March to September 2020 provides clues for policy-makers for better designing and implementing the control measures at the local level. With knowledge of the most likely cluster showing a tendency of moving from West Malaysia toward East Malaysia in September, opens the possibility to intervene pre-emptively. Our analysis also provides a good understanding of the temporal pattern of the spatial clusters of COVID-19 in the study area which should be useful for the surveillance of COVID-19 and other infectious diseases.

References

Amin RW, Fritsch BA, Retzloff JE, 2019. Spatial clusters of breast cancer mortality and incidence in the contiguous USA: 2000-2014. J. Gen Intern Med 34:412-9.

Amin RW, Rivera B, 2020. A spatial study of oral & pharynx cancer mortality and incidence in the USA: 2000-2015. Sci. Total Environ 713:136688.

BH Online, 2020. COVID-19: case in Masjid Seri Petaling new cluster; 22 Aug 2020. Available from: https://www.bharian.com.my/berita/nasional/2020/03/664769/covid-19-kes-di-masjid-seri-petaling-kluster-baharu Accesssed: 2 November 2020.

CodeBlue, 2020. Pedas cluster started by Malaysians who infected foreign workers: MOH. Available from: https://codeblue.galencentre.org/2020/06/22/pedas-cluster-started-by-malaysians-who-infected-foreign-workers-moh/ Accessed: 3 November 2020.

Desjardins MR, Hohl A, Delmelle EM, 2020. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr 118:102202.

Ester M, Kriegel HP, Sander J, Xu X, 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. InKdd 96:226-231.

Hanafiah KM, Wan CD, 2020. Public knowledge, perception and communication behavior surrounding COVID-19 in Malaysia. Preprint, Apr 2020. Available from: https://www.researchgate.net/profile/Khayriyyah_Mohd_Hanafiah/publication/340569667_Public_knowledge_perception_and_communication_behavior_surrounding_COVID-19_in_Malaysia/links/5eaf9c7ca6fdcc7050a86332/Public-knowledge-perception-and-communication-behavior-surrounding-COVID-19-in-Malaysia.pdf Accessed: 27 October 2020.

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395:497-506.

Khor V, Arunasalam A, Azli S, Khairul-Asri MG, Fahmy O, 2020. Experience from Malaysia during the COVID-19 movement control order. Urology 141:179-80.

Koh BYE, Pang TPN, Shoesmith WD, James S, Nor Hadi NM, Lin Loo JL, 2020. The behaviour changes in response to COVID-19 pandemic within Malaysia. Malaysian J Med Sci 27:45.

Kulldorff M, 1997. A spatial scan statistic. Commun Stat Theory Methods 26:1481-96.

Kulldorff M, 2018. SaTScan™ software for spatial, temporal and space-time scan statistics. User guide. Available from: https://www.satscan.org/. Accessed: 7 November 2020

Malaysiakini, 2020. Detention depot and Pedas clusters push new Covid-19 cases to 93. Available from: https://www.malaysiakini.com/news/528571 Accessed: 28 October 2020.

Ministry of Health Malaysia, 2020. Jumlah Kes COVID-19 Di Setiap Negeri Di Malaysia. Available from: http://covid-19.moh.gov.my/ Accessed: 2 November 2020.

Ng K, 2021. Coronavirus: Malaysia cases rise by 190 after mosque event as imams urge online services. The Independent; March 16 2021. Available from: https://www.independent.co.uk/news/world/asia/coronavirus-malaysia-cases-southeast-asia-mosque-islam-a9403816.html Accessed: 2 November 2020.

Omodior O, Kianersi S, Luetke M, 2019. Spatial clusters and non-spatial predictors of tick-borne disease diagnosis in Indiana. J. Community Health 44:1111-9.

Reuters Staff, 2021. Malaysia confirms first cases of coronavirus infection. Reuters; January 25. Available from: https://www.reuters.com/article/china-health-malaysia-idUSL4N29U03A Accessed: 2 November 2020.

Rocklöv J, Sjödin H, 2020. High population densities catalyse the spread of COVID-19. J Travel Med 7:taaa038.

Salim N, Chan WH, Mansor S, Bazin NEN, Amaran S, Faudzi AAM, Zainal A, Huspi SH, Khoo EJH, Shithil SM, 2020. COVID-19 epidemic in Malaysia: impact of lock-down on infection dynamics. Preprint, medRxiv. Available from: https://www.medrxiv.org/content/10.1101/2020.04.08.20057463v1 Accessed: 2 November 2020.

Shanmugam H, Juhari JA, Nair P, Ken CS, Guan NC, 2020. Impacts of COVID-19 pandemic on mental health in Malaysia: a single thread of hope. Malaysian J Psychiatry 29:78-84.

Souza RC, Assunção RM, Oliveira DM, Neill DB, Meira JRW, 2019. Where did I get dengue? Detecting spatial clusters of infection risk with social network data. Spat Spatiotemporal Epidemiol 29:163-75.

Statista, 2020. Demographics statistics Malaysia second quarter 2019. Available from: https://www.statista.com/statistics/1040670/malaysia-population-distribution-by-state/ Accessed: 3 November 2020.

The Strait Times, 2021. Coronavirus: Malaysia records eight deaths; 153 new cases bring total to 1,183. The Strait Times; March 21. Available from: https://www.straitstimes.com/asia/se-asia/malaysia-records-fourth-coronavirus-death Accessed: 4 November 2020.

TheStar, 2020. Covid-19: 77 patients linked to Sri Petaling mosque gathering as of Saturday; March 14. Available from: https://www.thestar.com.my/news/nation/2020/03/14/covid-19-77-patients-linked-to-sri-petaling-mosque-gathering-as-of-saturday-march-14 Accessed: 1 November 2020.

Ullah S, Daud H, Dass SC, Fanaee TH, Kausarian H, Khalil A, 2020. Space-time clustering characteristics of tuberculosis in Khyber Pakhtunkhwa province, Pakistan, 2015-2019. Int J Environ Res Public Health 17:1413.

Wikipedia, 2020. States and federal territories of Malaysia. Available from: https://en.wikipedia.org/wiki/States_and_federal_territories_of_Malaysia Accessed: 10 November 2020.

World Health Organization, 2020. Director-General's opening remarks at the media briefing on COVID-19. Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—-11-march-2020 Accessed: 12 November 2020.

Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270-3.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, 2020. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727-33.