CORRELATION ANALYSIS BETWEEN COVID-19 CASES AND GOOGLE TRENDS DATA: A SITUATIONAL ANALYSIS BASED ON STATES IN MALAYSIA

ABSTRACT

By Siti Nur Nafisah Binti Hafidzan (2018425456)

Coronavirus disease (COVID-19) in Malaysia evolved from January 2020 until recently. The purpose of this research is to study state-by-state association between COVID-19 incidence and Google Trends (GT) in Malaysia. This correlational research is conducted by measuring relationship between daily COVID-19 cases and 10 Google Keywords' relative search volumes (RSV) in Malaysia; 13 states and 3 federal territories. In this study, COVID-19 cases development, keywords popularity over time and keyword association towards COVID-19 event based on sub regions were discovered through descriptive analysis and statistical analysis method. Finding of this study shows there is 3 disparate keyword popularity trends upon COVID-19 case increment. Also, study result discovered that COVID-19 cases are weakly to moderately associated (r_s = -0.481 to r_s = 0.679) with keyword search during study period. As for state-by-state analysis finding, dual-association; negative to positive association (r_s = -0.371 to r_s = 0.407) is observed between COVID-19 cases and Google Trends data. In addition, this study provides an overview of Malaysians' awareness level towards COVID-19 issue based on states. To conclude, this study proposes there is an insufficient approach of risk communication that utilize online platform to combat COVID-19 in Malaysia. Data presented in this study is valuable to design relevant strategy using digital medium as a risk communication tool. Further research can be performed on time-series analysis for prior prediction of COVID-19 outbreak in states of Malaysia.

Keywords: COVID-19, Google Trends, correlation, state by state, Malaysia

1.0 INTRODUCTION

A previously unidentified coronavirus emerged from Wuhan, China at the end of December 2019. The infection previously known as the 2019 novel coronavirus. Subsequently, World Health Organization (WHO) officially announced the disease as COVID-19. Fever, dry cough, dyspnea, loose stool, nausea and vomiting are clinical symptoms of COVID-19 whereases, infectious agent responsible for COVID-19 is severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2 (Guan et al., 2020; Shereen et al., 2020). Person-to-person transmission via inhalation and direct contact are recognized as the causal of infection outbreak across the globe including Thailand, Republic of Korea, Japan, United States, Philippines and Vietnam (Lotfi et al., 2020; Shereen et al., 2020).

In Malaysia, first coronavirus infection was reported on 25 January 2020, coming from three infected Chinese nationals due to travelling history with closed contact. Since then, COVID-19 cases developed and distributed rapidly throughout Malaysia creating numerous clusters, morbidity and mortality (Shah et al., 2020). Meanwhile, important factors affecting spatio-temporal spread of COVID-19 outbreak within Malaysia are travelling history to high-risk area, social gathering, close contact, linkages between clusters and human mobility (Mat et al., 2020; Shah et al., 2020). Generally, COVID-19 incidences from 25 January 2020 to 30 November 2020 is divided into three waves; January-February, February-June and September until present (Shah et al., 2020). In order to control this infection, Government of Malaysia enforced series of strict movement control order based on area affected with COVID-19 cases. Movement control approaches that has been implemented are

Movement Control Order (MCO), Conditional Movement Control Order (CMCO) and Recovery Movement Control Order (RMCO) (Aziz et al., 2020).

Along with MCO approach, studies on immunology, serology, vaccine and test kit development has been conducted to identify effective treatment for COVID-19 (Anand et al., 2020; Grenache et al., 2020; Haveri et al., 2020; Xu et al., 2020; Zhou et al., 2020). To enhance study scopes, researchers explore on important role of digital search network and social media to assess public awareness, risk communication efficiency and information sharing behaviour (Chen et al., 2020; Hu et al., 2020; Husain et al., 2020; Husnayain et al., 2020; Magade, 2020). Findings of these studies provide impactful insights on cyber-based capabilities during pandemic. However, correlational study on state-by-state analysis of COVID-19 cases and internet search trend in Malaysia is limited and scarce.

In this regard, present study is addressed to determine the correlation between COVID-19 cases and Google Trends (GT) data in Malaysia. State-by-state correlation analysis is performed based on 10 different keywords related to coronavirus disease. It is necessary to establish correlation between COVID-19 cases and GT data to monitor public search interest on COVID-19 in Malaysia. Also, situational analysis is conducted to reflect risk communication effectiveness in Malaysia. Apart from that, study findings provide brief overview on awareness among Malaysian population on COVID-19 during pandemic phase. As for research opportunity, state-by-state time-series analysis may be conducted referring present study to predict prior COVID-19 contagion in different states of Malaysia based on searched keywords on Google platform.

2.0 MATERIAL AND METHODS

2.1 Study population

Present study purposed to assess popularity of queries searched among Malaysian during first case infection reported by Ministry of Health (MOH) until recent cases; 65,697 counts. Study population emphasized on sub-regional area of East Malaysia and West Malaysia; known as Peninsular Malaysia. East Malaysia is made up of the two largest states and one federal territory; Sabah, Sarawak and Federal Territory of Labuan. The East sub-regions are located on the popular island of Borneo while Peninsular of Malaysia consist of two federal territories namely Federal Territories of Kuala Lumpur as capital city and Federal Territories of Putrajaya as administrative centre. Majority states are located in West Malaysia with four regions; Northern Region (Perlis, Kedah, Penang and Perak), Eastern Region (Kelantan, Terengganu and Pahang), Central Region (Selangor, Federal Territory of Kuala Lumpur, Federal Territory of Putrajaya), Southern Region (Negeri Sembilan, Malacca and Johor). In short, this study observed total of 13 states and 3 federal territories for COVID-19 cases and GT data in Malaysia.

2.2 Study design

Correlational study between COVID-19 cases and GT data in Malaysia is performed through secondary data analysis. In general, this study is adapted from Kurian et al. (2020). Secondary data utilized for this study are COVID-19 cases in Malaysia and GT data. These data were readily accessible and digitally collected at reliable sources. Both secondary data are verified in advance as quantitative data implying data value is measured in the form of numbers or counts, with a unique

numerical value associated with each data set. Duration of study period is fixed from 25 January 2020 until 30 November 2020. Table I is tabulated to summarized the study design on correlation analysis between COVID-19 cases and GT data in Malaysia. Further, COVID-19 cases data and GT data of 10 distinct keywords represent dependent variable of study while Malaysia sub regions represent independent variable of the study. Each variable is interpreted in state-by-state basis throughout Malaysia which comprises of 13 states and 3 federal territories.

Table I
Summary on study variables of secondary data analysis including objectives, inclusion criteria and exclusion criteria.

Objectives	Dependent Variables	Independent Variables	Inclusion Criteria	Exclusion Criteria
▪ Determine COVID-19 cases in Malaysia	▪ COVID-19 cases data in Malaysia	▪ 13 states and 3 federal territories; Johor, Kedah, Kelantan, Pahang, Perak, Selangor, Terengganu, Perlis, Negeri Sembilan, Malacca, Penang, Sarawak, Sabah, Kuala Lumpur, Putrajaya, Labuan	▪ Daily data of COVID-19 cases and	▪ COVID-19 case reported before and after study period
▪ Determine GT data trend in Malaysia	▪ 10 Google keywords; *coronavirus symptom, cluster, coronavirus, COVID+death, face mask, hand sanitizer, quarantine, sore throat+shortness of breath+fatigue, swab test, vaccine*		▪ Weekly GT data based on 10 keywords	▪ Keywords that are unrelated with coronavirus disease ▪ GT data of other countries
▪ Determine correlation of COVID-19 cases and GT data in Malaysia	▪ COVID-19 cases versus GT data		▪ Data collection from 25 Jan 2020 to 30 Nov 2020	▪ Study data over Malaysia population

*COVID-19 = coronavirus disease, GT = Google Trends

2.3 Data collection and management

2.3.1 Data sources

Quantitative data gathered for this study consists of COVID-19 cases in Malaysia and GT data. Both data are electronically extracted from authentic web-based medium. First study data is COVID-19 cases in Malaysia. Information on daily positive cases and cumulative cases of COVID-19 is generated by MOH on official COVID-19 website (http:/covid-19.moh.gov.my/) and MOH online newspaper articles (https://kpkesihatan.com/category/communicable-disease/). Unrefined COVID-19 data retrieved from the websites consist of confirmed positive case; daily case and cumulative cases. The data is collected based on 13 states and 3 federal territories in Malaysia. Total COVID-19 case data recorded from 25 January 2020 to 30 November 2020 will be sorted into 311 days.

Next, Google Trends (GT) data. In practical terms, GT data best describes as relative searched volume (RSV) while GT is the analytical tool producing RSV. RSV can be extracted exclusive on GT site (https://trends.google.com). GT processes the magnitude of web searches performed for a specified keyword among other searches, providing the RSV for each keyword. This standardized value is calculated by dividing the total number of searches for a keyword by the total searches of the geography and time range. Data are collected in a time series presented on a normalized scale of 0 to 100, where 0 represents no search and 100 represents the peak search activity for a particular keyword or string. Trend shifts can be observed online for time series of interest. Keywords can be filtered by location; worldwide, country, state, city and by time span (Kurian et al., 2020).

In addition, Google Trends data can be downloaded as a ".csv" (comma-separated values) file and data extracted in this study started from 25 January 2020 until 30 November 2020. Weekly data based on Malaysia calendar rules is generated on GT therefore, data from week 4 until week 48 is retrieved. In total, 10 keywords related to COVID-19 were chosen on the basis of popularity and distribution pattern. The following keywords were selected in this study are *coronavirus symptom, cluster, coronavirus, COVID+death, face mask, hand sanitizer, quarantine, sore throat+shortness of breath+fatigue, swab test* and *vaccine*. Categories of keywords used in this study focused on COVID-19 related terms which consists of signs and symptom, control and prevention and potential remedies.

GT data content is validated by matching these keywords with terms located in Google's search engine's "related queries". If these keywords shows similar scope as the terms in the "related queries" category, the result data remains valid (Limilia & Pratamawaty, 2020). The output of the content validity test showed that the terms "related queries" were equivalent and related to the keywords used for coronavirus disease, therefore keywords chosen are reliable and valid. GT data is retrieved by individual states in Malaysia. In total, data for 13 states and 3 federal territories were obtained for each preferred keyword. Datasets information for this study are compacted in Table II including sources, specifications, attributes, value and pre-processed procedure.

Table II
Datasets of study on state-by-state correlation analysis between COVID-19 cases and Google Trends data in Malaysia.

Datasets	Sources	Specifications	Attributes	Value	Being Pre-processed
COVID-19 cases data in Malaysia	Official website for updated information on COVID-19 at http:/covid-19.moh.gov.my/	Daily reported positive COVID-19 cases	Data time interval	January 25, 2020 to November 30, 2020	Data tabulation into daily and cumulative cases for each state in Malaysia
			Day	1-311 days	
	Digital newspaper report on COVID-19 at https://kpkesihatan.com/category/communicable-disease/	Cumulative reported positive COVID-19 cases	States	13 states, 3 fed. ter.	
			Coronavirus disease	Positive cases	
Google Trends data	Primary site for collecting Google Trends data at https://trends.google.Com	Weekly RSV of keywords popularity related to coronavirus infections	Data time interval	January 25, 2020 – November 30, 2020	Data tabulation into weekly series for each state in Malaysia
			Week	4-48 weeks	
			States	13 states, 3 fed. ter.	
			10 keywords	RSV	

*COVID-19 = coronavirus disease, fed. ter. = federal territories, RSV = relative search volume

2.3.2 Data processing

Data processing is a method of gathering and converting raw data into usable information. It is performed in a step-by-step process. Data processing cycle in this study compromised of data collection from reliable sources, data preparation or data cleaning by removing unnecessary or inaccurate data, and data arrangement.

I) COVID-19 cases data in Malaysia

To process COVID-19 case data, daily and cumulative case infection are retrieved from targeted sources. Data such as newspaper article, infographic of COVID-19 case and COVID-19 posters are downloaded. Then, data is arranged in chronological order from January 2020 to November 2020. COVID-19 case data that has been arranged are transferred into a spreadsheet; Microsoft Excel. Unrelated data such as epidemiological information and sociodemographic data are removed. Daily and cumulative reported case data are arranged following each states and federal territories in Malaysia. A tabulated dataset of 311 days x 16 subregions (13 states and 3 federal territories) in Malaysia is produced.

II) Google Trends data

Google Trends data management initiated with keywords selection and validity confirmation. Data collection procedure is conducted by entering keywords into 'search term' section on GT website. Filter features is applied in GT based on country; Malaysia while time range filter has to be modified by computing daily range date for each week. GT automatically filter redundant and non-essential data. Although, GT did not provide the absolute number of searches. For that reason, decent

filtering technique is vital to achieve generation of fine-grained data and data accuracy. There are two methods to filter GT data; real time and non-real time. Real time is a random sample of queries from the last seven days, while non-real time is another arbitrary sample of the entire Google data collection that can be retrieved anywhere from 2004 to 36 hours of current time. GT data in this study is filtered using non-real time procedure. Afterwards, data is exported in ".csv" format and readily interpreted by spreadsheet. Further, GT data for each states and federal territory of Malaysia are organized in weekly manner.

III) Pairing of COVID-19 cases data and Google Trends data

COVID-19 cases data and GT data that has been managed is adjoined in Statistical Package for the Social Sciences (SPSS) Software Version 25. Both data is transferred into ".sav" file. SAV file extension usually perceived as a generic extension used to store files and data. Total of 10 ".sav" files are created based on 10 chosen keywords. Each ".sav" consist of daily COVID-19 cases and RSV of keyword based on 13 states and 3 federal territories.

2.3.3 Data analysis

Data analysis is the process of collecting and organizing data in order to draw scientific conclusions of the study. Data analysis uses analytical and logical reasoning to gain information from the processed. Present study on assessing state-by-state correlation analysis between COVID-19 case and GT data in Malaysia require R package for statistical computing and graphics generation. Daily COVID-19 cases and RSV of keyword for each sub regions are analysed using Spearman's correlation coefficient as data variables are widely spread. Correlation value and 95% confidence

interval (95% CI) of state-by-state output data is presented in table form and radar graph. To observe keywords' trending pattern during study period, RSV for each keyword is illustrated in line graph together with overall cumulative COVID-19 case in Malaysia. Additional scatterplot chart is plotted between COVID-19 cases and relative search volume of 10 keywords to discover keywords popularity against number of COVID-19 cases.

3.0 RESULT

3.1 COVID-19 cases in Malaysia and Google keywords' relatives search volumes (RSVs) trend

This study was undertaken in all 13 states and 3 federal territories of Malaysia highlighting COVID-19 cases in Malaysia and relative search volumes of Google keywords. Line graph is generated to visualize COVID-19 cases over time as well as keyword search popularity. The line chart consists of a horizontal x-axis and a vertical x-axis. X-axis indicates time frame; from January 2020 to November 2020 and y-axis indicate a quantity in numbers; COVID-19 cases and relative search volumes (RSV) of Google keyword. Also, line graphs use data point "markers" that are connected by straight lines to aid in visualization.

First graph produced in this study is COVID-19 incidents in Malaysia. Overall, 10 months-time-frame data on coronavirus infection is transformed into line graph. Based on Figure I, three distinct COVID-19 waves from January 2020 until November 2020 are observed. Early phase of COVID-19 outbreak shows stagnant progress of coronavirus case within January to February. This phase is formerly known as first wave of COVID-19 infection; starting from 25 January 2020 until

16 February 2020. Further, reported daily case gradually evolved from mid-March and case steadily present below 20,000 counts until late of September. This period represents second wave of COVID-19 in Malaysia. Ultimately, sudden case increment from 20,000 cases to approximately 70,000 cases indicate third wave outbreak in Malaysia; between mid-September to the end of November 2020.

Next, in order to compare popularity searches pattern of 10 keywords related to COVID-19, line chart is produced from keyword's RSV value within study period. RSV numbers represent the search interest relative to the highest point on the chart. RSV values distributed from 0 RSV to 100 RSV. Graph is compacted in Figure I together with COVID-19 cases graph; started on January 2020 until November 2020. In general, three types of spiked trends are observed from keyword's linear diagram which are; presence of spike on early stage of COVID-19 pandemic, peaked pattern near the end of study period and fluctuated pattern within 10 months duration. Graph from 3 keywords; *coronavirus symptom*, *coronavirus*, *hand sanitizer* and *quarantine* show that search popularity concentrated on the beginning of COVID-19 outbreak followed by sudden inclination trend. Meanwhile keywords with delayed popularity are *cluster* and *swab test*. Similar peaked pattern of term search on Mac, August and October is observed in *COVID+death, sore throat+shortness of breath+ fatigue+cough* and *face mask* keywords. Lastly, keyword with significant search interest in each pandemic wave is *vaccine* keyword. Peaked range between 80 RSV to 100 RSV is observed in *vaccine* RSV diagram.

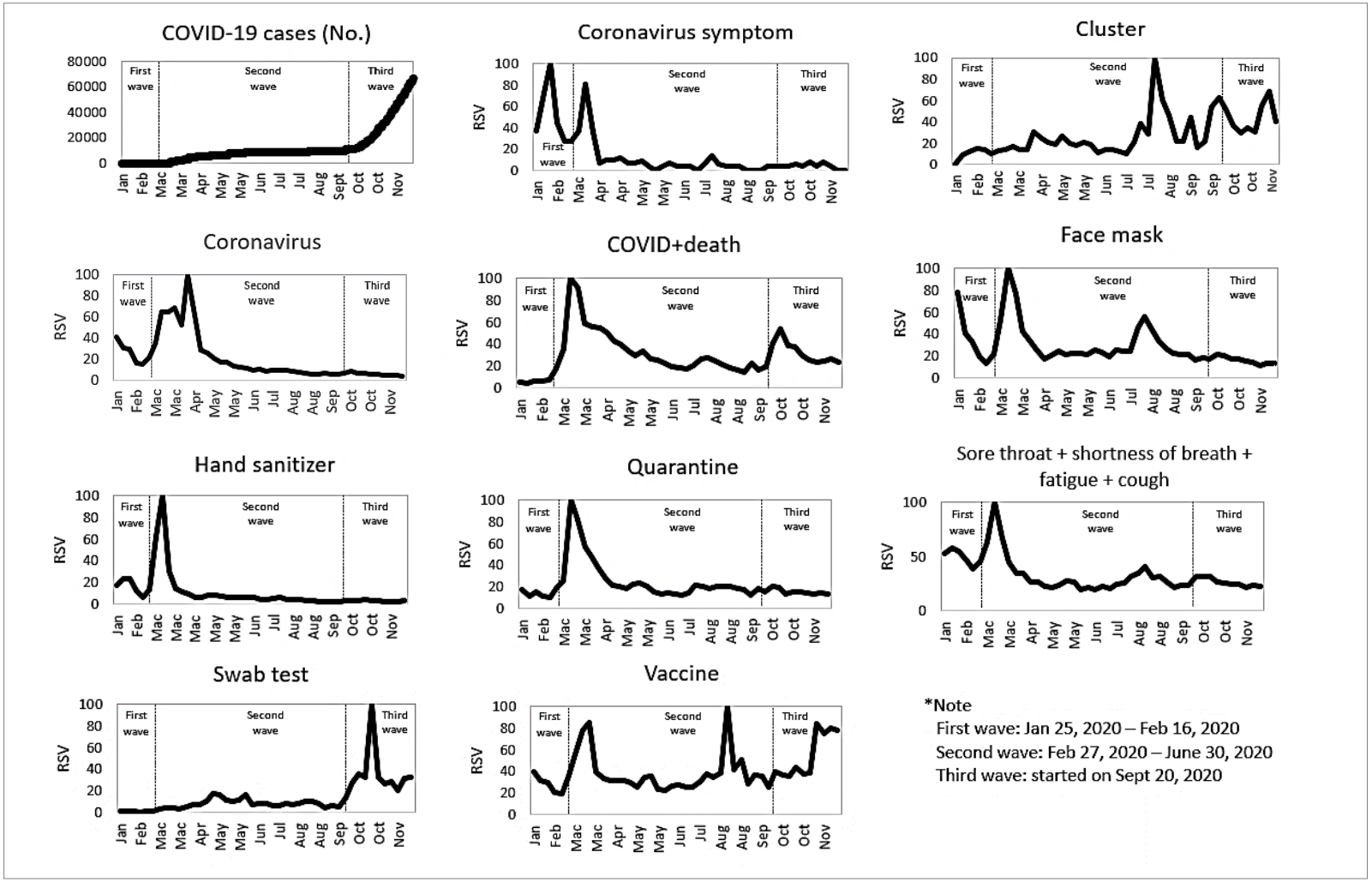

Figure I COVID-19 cases and trending of relative search volume (RSV) for 10 keywords since January 2020 through November 2020. *COVID/COVID-19 = coronavirus disease 2019

3.2 COVID-19 cases in Malaysia and Google keywords' relatives search volumes (RSVs) distribution

Present study combined data between COVID-19 cases progression and keyword searched in Malaysia to provides plot pattern. Combination of these variables are visualized in scatterplot diagram as in Figure II. Scatterplot is a type of plot or mathematical diagram that uses Cartesian coordinates to demonstrate values for two variables of a data set. Also, scatter graph suggests correlations between variables. Correlation is interpreted either positive correlation (rising pattern), negative correlation (falling pattern) or null correlation (no distinctive pattern). Scatter graph is observed based on direction of dots slopes, from lower left to upper right. Nevertheless, correlation does not imply causation between variables.

Figure II represents overall distribution pattern of keyword search in Malaysia from zero COVID-19 case until estimated 70,000 cases. Based on clustered dots formation in the diagram, almost all keywords massively browsed before COVID-19 case reach 30,000 counts. Relative search volume for 10 keywords against COVID-19 cases range varies from 0-20 RSV (low popularity), 20-60 RSV (medium to half popularity) and 20-100 RSV (low to highest popularity). Briefly, a rising pattern of correlation between COVID-19 cases and keyword search is discovered in 4 Google keywords; *cluster, COVID+death, swab test* and *vaccine*. This pattern stipulates a positive correlation between COVID-19 cases in Malaysia and browsed term on Google online network.

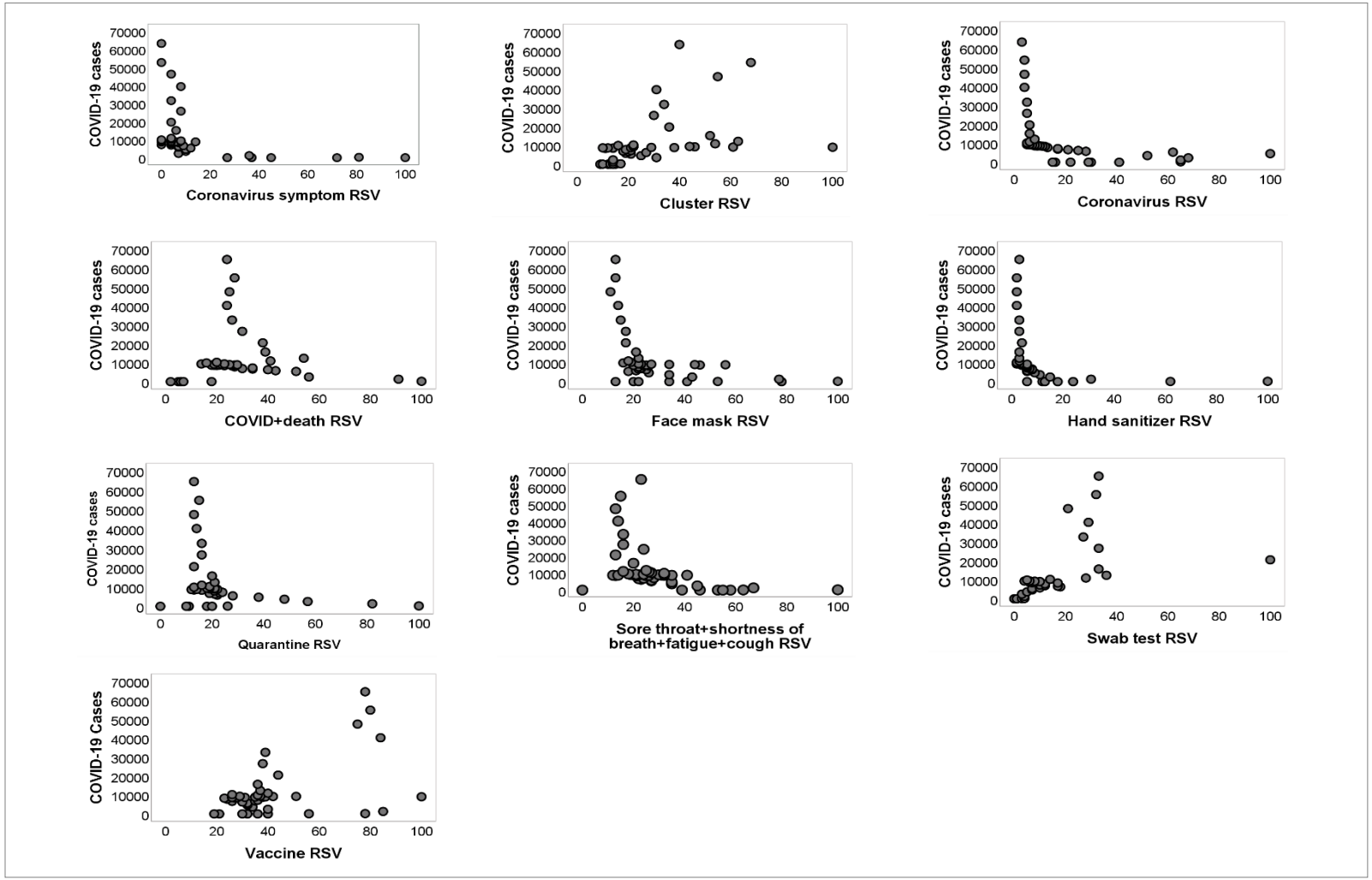

Figure II Scatterplot of coronavirus disease 2019 against relative search volume (RSV) of 10 Google keywords from January 2020 through November 2020. *COVID/COVID-19 = coronavirus disease 2019

3.3 Correlation analysis of COVID-19 cases and Google Trends data in states and overall Malaysia

Current study implies general correlation between COVID-19 cases and Google Trends data in Malaysia as a whole from 25 January 2020 to 30 November 2020. Bivariate analysis known as Spearman's correlation produced r_s value portraying the strength of association between variables. Keyword correlation in overall Malaysia ranged from r_s = -0.481 to r_s = 0.679. Value of r_s in Table III shows association diversion of selected keywords. Subsequently, moderate to strong positive correlation is detected in 4 keywords marked with (*) symbol; r_s = 0.391 (*vaccine*), r_s = 0.574 (*COVID+death*), r_s = 0.585 (*cluster*) and r_s = 0.679 (*swab test*). Also, a keyword with strong negative correlation significantly observed at r_s = -0.481 (*face mask*).

Table III
Correlation analysis between 10 Google keywords and daily COVID-19 case in Malaysia from January 25, 2020 until November 30, 2020.

	Google keyword	r_s	95% Confidence Interval	
			Upper	Lower
R1	Coronavirus symptom	-0.396	-0.486	-0.298
R2	Cluster*	0.585	0.507	0.654
R3	Coronavirus	-0.347	-0.441	-0.245
R4	COVID+death*	0.574	0.494	0.644
R5	Face mask	-0.481	-0.562	-0.391
R6	Hand sanitizer	-0.425	-0.512	-0.329
R7	Quarantine	0.168	0.058	0.274
R8	Sore throat+shortness of breath+fatigue	-0.264	-0.365	-0.157
R9	Swab test*	0.679	0.614	0.735
R10	Vaccine*	0.391	0.293	0.481

*COVID = coronavirus disease 2019

This study further explores on state level correlation analysis to enhance understanding on COVID-19 daily cases and Malaysian interest search by states. State-by-state analysis result is tabulated as supplementary data. Actual r_s value for each states and federal territories may be referred in Appendix. Correlation data is visualized into radar chart as in Figure III. A radar diagram is a two-dimension diagram and it is a visual guidance in which different axes emerge from the same central point. Based on Figure III, distorted black line in each diagram indicates correlation value distribution in 13 states and 3 federal territories.

Generally, weak to moderate positive correlation (r_s = 0.024 to r_s = 0.525) is obtained through states analysis while remarkable state with five keywords of positive association is Kelantan state. Keywords detected are *cluster, COVID+death, hand sanitizer, quarantine and swab test*. In contrast, Federal Territory of Kuala Lumpur implies insignificant correlation to moderate negative association (r_s = -0.057 to r_s = -0.371) for all Google keywords. Keyword with overall positive correlation from majority states and federal territories is *COVID+death* keyword. The positive correlation portrays that Malaysian continuously search on COVID-19 and the disease death toll during COVID-19 pandemic from 25 January 2020 until 30 November 2020.

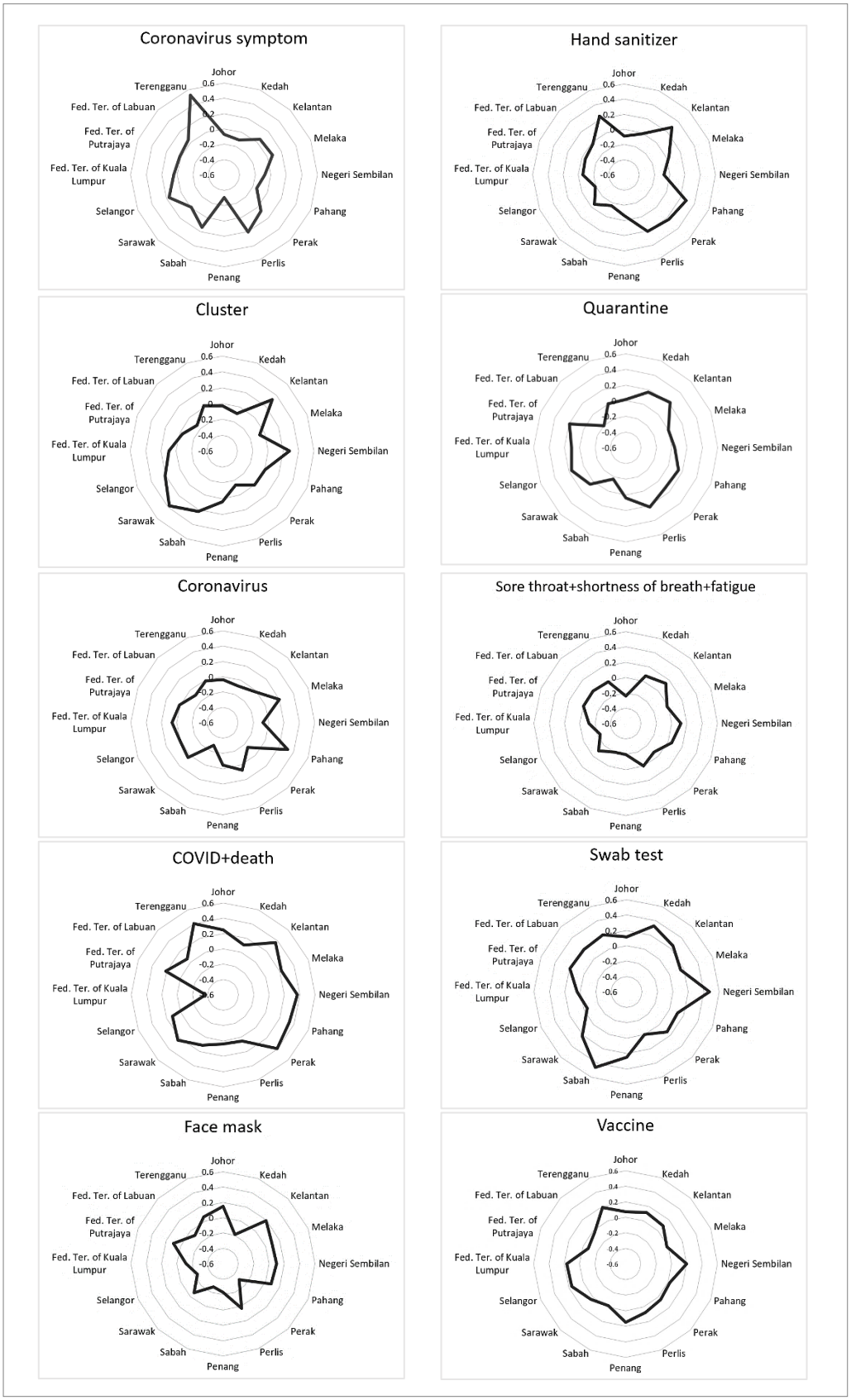

Figure III Data distribution of correlation analysis between daily COVID-19 cases and 10 selected Google keywords from 25 January 2020 until 30 November 2020.
*COVID = coronavirus disease 2019, Fed. Ter. = Federal Territory

4.0 DISCUSSION

Coronavirus outbreak in Malaysia initially penetrated from Johor Bahru, Johor before infection distributed throughout regions of Malaysia. Declining and resurgence of COVID-19 transmission from 25 January 2020 until 30 November 2020 caused outbreak division into three waves. First COVID-19 wave consists of 22 cumulative cases which is considered as small case size and roughly lasted for three weeks. Consequently, after 11 days of zero reported case, second wave infection is introduced, coming from large religious gathering in Kuala Lumpur (Sheng et al., 2020). Since then, COVID-19 daily case is constantly present until a sudden rise of cases occurs between mid-September 2020 to end of November 2020. This phase is recognized as the third wave outbreak concerning transmission of COVID-19 linkage from positive cases and new active clusters (Kaos, 2020; Yildiz, 2020).

Present study aimed to scrutinize Google keywords search trend in Malaysia for 10 months. Three different patterns are observed during study period which are early-peaked pattern, fluctuated pattern and end-peaked pattern. Study result on *coronavirus* keyword shows it is highly browsed during early COVID-19 infection in Malaysia before steep deceleration on April 2020 followed by plateau state until November 2020. Trend with early peaked pattern presumably related to WHO declaration of disease nomenclature changes from coronavirus disease to COVID-19 since 11 February 2020. Previous study found that search queries for coronavirus had initially draw public interest followed by a gradual shift of interest to new topics or term (Filho et al., 2020; Yau et al., 2020).

Further, declining trend of certain keywords is probably associated with public interest over time. This similar pattern is observed in *hand sanitizer* keyword implying significant reduction of term searched starting from mid-Mac towards end of November. At first, peaked pattern is observed probably implying public has massive interest on searching for ways to avoid COVID-19 transmission through usage of hand sanitizer. Over time, the trend went down might be due to Malaysian has been purchasing the item and carry hand sanitizer with them whenever they go as recommended by MOH. As time goes by, Malaysian gradually shift their attention towards more interesting issue. To strengthen this assumption, a previous study found a correlation between public interest over time and popularity pattern where the more attentive public are towards COVID-19, the more searches volume recorded in GT and vice versa. (Husain et al. 2020).

In addition, media manipulation indirectly influences keywords popularity as Malaysians have become more reliant on the internet for everyday activities. Irregular trend for *cluster*, *COVID+death* and *vaccine* keywords presumably occur regarding trigger by media advocacy especially during enforcement of MCO series, emergence of new cluster and current issue that contribute to rising of COVID-19 cases. For example, cluster-related keywords consistently search based on breaking news of Tabligh cluster due to religious gathering, Sivagangga cluster due to incompliance of self-quarantine and Teratai cluster due to close contact transmission of workers living in employees' dormitory area. (Muhamad, 2020; MOH, 2020; Salim, 2020). These trending topics are delivered by media via online platform such as electronic daily COVID-19 report provided by MOH and online newspaper article such as Astro Awani, Malaysia Kini and The Star. Theory related with this statement is the agenda-

setting theory which indicates the media agenda primarily from news outlet cause alteration of keyword search trend (Ho et al., 2018; Manaf et al., 2017; Mccombs & Shaw, 1972).

To confirm association between COVID-19 incident and Google Trend data, a correlation analysis had been conducted in overall Malaysia, with additional state-by-state analysis value. Weak to moderate positive association strength is observed between keyword search by states. Five subregions in Malaysia portray persistence interest in COVID-19 related topic; Kelantan, Terengganu, Pahang, Perlis and Negeri Sembilan. Finding of this study indirectly portray regional health-seeking behaviour as well as awareness of state population towards COVID-19 issue. In comparison from previous studies, strong to high positive correlation is determined between COVID-19 case and Google Trend data. These studies were conducted both globally and in United States (Effenberger et al., 2020; Kurian et al., 2020). The positive finding between current and previous studies verify presumption on alertness and behavioral pattern of Malaysian on gathering COVID-19 information.

Next, this study is conducted as a situational analysis on public searched terms in Malaysia during COVID-19 pandemic. Issue found in current study shows presence of negative association in certain localities. For instance, result for Federal Territory of Labuan correlation shows that COVID-19 cases did not correspond to search interest which means search on coronavirus topic decreasing upon rising of COVID-19 cases. Speculation on density population arises as it might be the main reason for the negative association result. In parallel, Department of Statistics Malaysia categorized Federal Territory of Labuan under least populated states in Malaysia (DOSM, 2011). To support this statement, past study discovered population density

indirectly impacts internet user and internet searching activity in certain location (Ho et al., 2018).

In addition, moderate to strong negative correlation of keywords is observed in few states of Malaysia such as Federal Territory of Kuala Lumpur, Selangor and Sabah. It is debatable whether lack of awareness is main cause of declining trend against cases increment as these localities were detected with high infection rate over time (MOH, 2020). The possible explanation for this situation is public response toward COVID-19 prevention practice and public abidance on following Government order. Deputy Prime Minister, Dato' Sri Ismail Sabri Yaakob stated that public compliance on wearing face mask is encouraging but physical distancing obedience is yet unmet with Government expectation (Bernama, 2021). Also, MOH consistently urge public to follow social distancing SOP provided by Government. MOH also stated that continuous distribution and dispersion of COVID-19 infection in Malaysia mainly caused by failure to comply with physical distancing.

Last but not least, efficacy of risk communication approach carried out by Government can be reflected based on study findings of COVID-19 cases and search trend in Malaysia. Centers of Disease Control and Prevention (CDC) define risk communication as the distribution of health threats and incidents information to the public, such as disease outbreaks and recommendations for improving the conduct of risk mitigation approaches. Situational analysis based on Google keywords search in Malaysia shows weak correlation between term browsed and COVID-19 cases where some of the Malaysia states show significant negative associations. It can be assumed that strategy on information dissemination based on digital platform did not reach to a level where public understand clearly about COVID-19 criticalness thus, they have

negligible concern to explore more regarding COVID-19 and alternating keywords popularity pattern. The effectiveness of risk communication will not only flatten infection curves but it will ensure public abidance on following SOP of the new norm. Based on MOH digital report on *Tinjauan Amalan Norma Baharu KKM*, public with sufficient understanding and knowledge of COVID-19 is complying with MOH new norm SOP and encourage others to abide with the SOP such are wearing face mask, frequent use of hand sanitizer and most importantly, practicing social distancing.

To summarize, COVID-19 cases progression and keyword trending pattern from in Malaysia are discovered in this study. Correlational analysis for overall Malaysia and discrete state-level correlation are observe in diverse range value; weak to moderate and positive to negative association value. Important key that influence keywords popularity during COVID-19 phase are public interest over time, media manipulation, population density, health-seeking information behaviour and response, lastly and most importantly, efficiency of risk communication. Based on result and discussion on this study, a new cyber-related risk communication strategy shall be constructed in order to ensure success of COVID-19 information dissemination and clarification toward public. Constraint issue of the present study are Google Trends data is exclusiveness of Google platform where it does not contain information on other internet search networks site as a whole. Other than that, the parameters used in this study might not cover all public search keywords because Malaysia has multiple accessible online platforms and communications networks either than Google. Further research time-series analysis between states in Malaysia can be conducted to forecasts COVID-19 incidence within subregions. The finding will be remarkable to improve disease surveillance system in Malaysia.

5.0 CONCLUSION

Conclusively, COVID-19 case and keywords popularity pattern has been described based on peak point of RSV for 10 keywords over 10 months in Malaysia. General correlation of COVID-19 cases and Google Trends data in overall Malaysia ranged from weak to moderate positive correlation in strength while state-by-state correlation analysis finding is divided into two categories; positive association and negative association. Study result provide situational analysis on behavioral response, awareness and performance of messages circulation in Malaysian population towards COVID-19. This suggests that public awareness of COVID-19 was not strong enough and inefficient risk communication strategy is discovered. Thus, more effective measures should be taken by Malaysia Government to strengthen the propaganda of combating COVID-19. Effective utilization of digital platform in risk communication will enhance public awareness, improve public response and behaviour and finally control the spread of COVID-19 in Malaysia. Also, Google Trends capabilities in monitoring and prevention of COVID-19 should not be overlooked. Direction of research opportunity suggested is the exploration on cross-correlation analysis in states scale to predict COVID-19 outbreak in Malaysia.

7.0 REFERENCES

Anand, S., Montez, R., Han, J., Bozeman, J., Kerschmann, R., Beyer, P., Parsonnet, J., & Chertow, G. (2020). Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. *The Lancet*, *396*(10259), 1335–1344.

Aziz, N. A., Othman, J., Lugova, H., & Suleiman, A. (2020). Malaysia's approach in handling COVID-19 onslaught: Report on the Movement Control Order (MCO) and targeted screening to reduce community infection rate and impact on public health and economy. *Journal of Infection and Public Health, 13*(12), 1823-1829.

Bernama. (2021, January 24). Ismail Sabri: Failure to observe physical distancing contributes to rise in Covid-19 cases. *The Star.* https://www.thestar.com.my/news/nation/2021/01/24/ismail-sabri-failure-to-observe-physical-distancing-contributes-to-rise-in-covid-19-cases

Centers of Disease Control and Prevention, CDC. (2020). *Risk communication.* https://www.cdc.gov/healthcommunication/risks/index.html

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance*, *6*(2), e19273.

Department of Statistics Malaysia, DOSM. (2011). *Population distribution and basic demographic characteristic report 2010*. https://www.dosm.gov.my/

Effenberger, M., Kronbichler, A., Shin, J. I., Mayer, G., Tilg, H., & Perco, P. (2020). Association of the COVID-19 pandemic with internet search volumes: A Google TrendsTM analysis. *International Journal of Infectious Diseases*, *95*, 192–197.

Filho, P. R. M., Araújo, A. A. de S., Júnior, L. J. Q., & Santos, V. S. (2020). Digital public interest on coronavirus information and social distancing: a Brazilian nationwide analysis using an infodemiology approach. *Journal of Travel Medicine*, *Sept*(2020), taaa154.

Google. (2020). *Google Trends*. https://trends.google.com

Grenache, D. G., Sever, C., Mathur, G., & Mathur, S. (2020). Antibody testing for COVID-19. *American Journal of Clinical Pathology*, *154*(3), 425–426.

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D. S. C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, J., Liang, Z., Peng, Y., Wei, L., Liu, Y., Hu, Y., Peng, P., Wang, J., Liu, J., Chen, Z., Li, G., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S. Zhong, N. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine, 382*(18), 1708–1720.

Haveri, A., Smura, T., Kuivanen, S., Österlund, P., Hepojoki, J., Ikonen, N., Pitkäpaasi, M., Blomqvist, S., Rönkkö, E., Kantele, A., Strandin, T., Kallio, K. H., Mannonen, L., Lappalainen, M., Broas, M., Jiang, M., Siira, L., Salminen, M., Puumalainen, T., Sane, J., Melin, M., Vapalahti, O., & Savolainen, K. C.

(2020). Serological and molecular findings during SARS-CoV-2 infection: The first case study in Finland, January to February 2020. *Eurosurveillance, 25*(11), 1560-7917.

Ho, H. T., Carvajal, T. M., Bautista, R. J., Capistrano, J. D. R., Viacrusis, K. M., Hernandez, L. F. T., & Watanabe, K. (2018). Using Google Trends to examine the spatio-temporal incidence and behavioral patterns of dengue disease: A case study in metropolitan Manila, Philippines. *Tropical Medicine and Infectious Disease, 3*(4), 118-124.

Hu, D., Lou, X., Xu, Z., Meng, N., Xie, Q., Zhang, M., Zou, Y., Liu, J., Sun, G. P., & Wang, F. (2020). More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. *Journal of Global Health, 10*(1), 011003.

Husain, I., Briggs, B., Lefebvre, C., Cline, D. M., Stopyra, J. P., O'Brien, M. C., Vaithi, R., Gilmore, S., & Countryman, C. (2020). Fluctuation of public interest in COVID-19 in the United States: Retrospective Analysis of Google Trends Search Data. *JMIR Public Health and Surveillance, 6*(3), e19969.

Husnayain, A., Fuad, A., & Su, E. C. Y. (2020). Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. *International Journal of Infectious Diseases, 95*, 221–223.

Kaos, J. J. (2020, September 19). COVID-19: 101 new cases, four new clusters detected. The Star. https://www.thestar.com.my news/nation/2020/09/29/covid-19-101-new-cases-four-new-clusters-detected

Kurian, S. J., Bhatti, A. R., Alvi, M. A., Ting, H. H., Storlie, C., Wilson, P. M., Shah, N. D., Liu, H., & Bydon, M. (2020). Correlations between COVID-19 cases and Google Trends data in the United States: A state-by-state analysis. *Mayo Clinic Proceedings, 95*(11), 2370-2381.

Limilia, P., & Pratamawaty, B. B. (2020). Google Trends and information seeking trend of COVID-19 in Indonesia. *Jurnal ASPIKOM, 5*(2), 188-205.

Lotfi, M., Hamblin, M. R., & Rezaei, N. (2020). COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clinica Chimica Acta. 508*(2020), 254–266.

Magade, M. E. (2020). Hashtags, comment threads and re-tweets: The emergence of young voices and alternative communicative practices for socio-political change at the time of COVID-19 in South Africa. *Global Media Journal, 18*(35), 211-221.

Manaf, A. M. A., Taibi, M., & Manan, K. A. (2017). Media agenda and public agenda: A study of issues during the 13th general election. *Malaysian Journal of Communication, 33*(2), 13–26.

Mat, N. F. C., Edi nur, H. A., Razab, M. K. A. A., & Safuan, S. (2020). A single mass gathering resulted in massive transmission of COVID-19 infections in Malaysia with further international spread. *Journal of Travel Medicine, 27*(3).

Mccombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly, 36*(2), 176–187.

Ministry of Health, MOH. (2020). *COVID-19 – Kementerian Kesihatan Malaysia.* http://covid-19.moh.gov.my/

Ministry of Health, MOH. (2020). *Communicable disease: from the desk of the Director-General of Health Malaysia.* https://kpkesihatan.com/category/communicable-disease/

Ministry of Health, MOH. (2020, August 5). *Kes positif baharu dikesan dalam kluster PUI Sivagangga.* http://covid-19.moh.gov.my/sorotan/kes-positif-baharu-dikesan-dalam-kluster-pui-sivagangga

Ministry of Health, MOH. (2020, December 25). *Tinjauan Amalan Norma Baharu KKM.* http://covid-19.moh.gov.my/semasa-kkm/122020/tinjauan-amalan-norma-baharu-kkm

Ministry of Health, MOH. (2020, November 28). *Patuhi Prosedur Amalan Standard (SOP) ketika menghadiri sebarang jenis perhimpunan dan ketika berada di tempat awam.* http://covid-19.moh.gov.my/sorotan/112020/patuhi-prosedur-amalan-standard-sop-ketika-menghadiri-sebarang-jenis-perhimpunan-dan-ketika-berada-di-tempat-awam

Muhamad, H. (2020, May 19). Sri Petaling tabligh rally cases contribute almost half of COVID-19 cases in Malaysia - Dr Noor Hisham. *Astro Awani.* https://www.astroawani.com/berita-malaysia/sri-petaling-tabligh-rally-cases-contribute-almost-half-of-covid19-cases-in-malaysia-dr-noor-hisham-243611

Salim, S. (2020, November 24). Covid-19: 1,511 new cases at Teratai cluster originate from Top Glove's dorms, total swells to 4,036. *The Edge Markets.* https://www.theedgemarkets.com/article/covid19-teratai-cluster-linked-top-glove-dorms-records-1511-new-cases-says-moh

Shah, A. U. M., Safri, S. N. A., Thevadas, R., Noordin, N. K., Rahman, A. A., Sekawi, Z., Ideris, A., & Sultan, M. T. H. (2020). COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *International Journal of Infectious Diseases, 97*(2020), 108–116.

Sheng, N. C. F., Seposo, X. T., Moi, M. L., Tajudin, M. A. B. A., Madaniyazi, L., & Sahani, M. (2020). Characteristics of the COVID-19 epidemic and control measures to curb transmission in Malaysia. *International Journal of Infectious Diseases, 101*(2020), 409–411.

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research, 24*(2020), 91–98.

World Health Organization, WHO. (2019). *Naming the coronavirus disease (COVID-19) and the virus that causes it.* https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

Xu, R., Zhou, Y., Hu, D., Yue, Y., Li, Q., & Xia, J. (2020). Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: a modelling study using mobile phone data. *The Lancet Digital Health, 2*(8), 417–424.

Yau, E. K. B., Ping, N. P. T., Shoesmith, W. D., James, S., Hadi, N. M. N., & Lin, L. J. (2020). The behaviour changes in response to COVID-19 pandemic within Malaysia. *Malaysian Journal of Medical Sciences, 27*(2), 45–50.

Yildiz, O. F. (2020, December 19). Malaysia: 1,359 health workers catch virus in 3rd wave. *Anadolu Agency*. https://www.aa.com.tr/en/asia-pacific/malaysia-1-359-health-workers-catch-virus-in-3rd-wave/2082068

Zhou, P., Li, Z., Xie, L., An, D., Fan, Y., Wang, X., Li, Y., Liu, X., Wu, J., Li, G., & Li, Q. (2020). Research progress and challenges to coronavirus vaccine development. *Journal of Medical Virology, 93*(2), 741-754.

APPENDIX

State-by-state correlation analysis value between daily COVID-19 cases in Malaysia and relative search volume of 10 selected Google keywords.

State	(R1)	(R2)	(R3)	(R4)	(R5)	(R6)	(R7)	(R8)	(R9)	(R10)
Johor	-0.071	-0.026	-0.036	0.248	0.149	-0.090	0.016	-0.241	0.113	0.070
Kedah	-0.107	-0.088	-0.079	0.101	-0.190	-0.017	0.168	0.075	0.327	0.116
Kelantan	0.053	0.322	-0.022	0.363	0.193	0.287	0.217	0.142	0.247	0.091
Fed. Ter. of Kuala Lumpur	0.044	0.097	0.063	-0.371	-0.121	-0.057	0.104	-0.117	0.043	0.167
Fed. Ter. of Labuan	0.049	-0.136	-0.095	0.066	-0.078	-0.019	-0.198	-	0.179	-0.031
Malacca	0.076	-0.070	0.199	0.219	0.092	0.039	0.000	-0.017	0.154	-0.017
Negeri Sembilan	-0.077	0.277	-0.079	0.369	0.101	-0.075	0.038	0.121	0.474	0.200
Pahang	-0.148	0.009	0.315	0.334	0.077	0.281	0.143	0.050	0.112	0.027
Perak	0.068	-0.001	-0.143	0.394	-0.304	0.230	0.125	-0.078	0.136	0.048
Perlis	0.207	-0.137	0.069	0.052	0.028	0.204	0.217	-	-	0.077
Penang	-0.307	0.039	-0.048	0.041	-0.229	-0.062	0.038	-0.195	0.248	0.148
Fed. Ter. of Putrajaya	0.016	-0.038	0.011	0.214	0.098	-0.051	0.196	-0.002	0.192	-0.073
Sabah	0.142	0.221	-0.284	0.112	-0.277	-0.163	-0.167	-0.199	0.460	-0.018
Sarawak	-0.002	0.374	0.041	0.235	-0.066	-0.052	0.059	-0.100	0.213	0.043
Selangor	0.166	0.207	0.007	0.120	-0.240	-0.190	0.162	-0.239	-0.047	0.158
Terengganu	0.525	0.022	-0.011	0.407	0.063	0.242	0.006	-0.008	0.205	0.186

*R1 = *coronavirus symptom*, R2 = *cluster*, R3 = *coronavirus*, R4 = *COVID+death*, R5 = *face mask*, R6 = *hand sanitizer*, R7 = *quarantine*, R8 = *sore throat+shortness of breath+fatigue*, R9 = *swab test*, R10 = *vaccine*, Fed.Ter. = Federal Territory