

Original Paper

Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study

Muhammad Syamsuddin, PhD; Muhammad Fakhruddin, MSc; Jane Theresa Marlen Sahetapy-Engel, BSc; Edy Soewono, Prof Dr

Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia

Corresponding Author:

Edy Soewono, Prof Dr
Faculty of Mathematics and Natural Sciences
Institut Teknologi Bandung
Jalan Ganesha 10
Bandung, 40132
Indonesia
Phone: 62 8122000534
Email: esoewono@math.itb.ac.id

Abstract

Background: The popularity of dengue can be inferred from Google Trends that summarizes Google searches of related topics. Both the disease and its Google Trends have a similar source of causation in the dengue virus, leading us to hypothesize that dengue incidence and Google Trends results have a long-run equilibrium.

Objective: This research aimed to investigate the properties of this long-run equilibrium in the hope of using the information derived from Google Trends for the early detection of upcoming dengue outbreaks.

Methods: This research used the cointegration method to assess a long-run equilibrium between dengue incidence and Google Trends results. The long-run equilibrium was characterized by their linear combination that generated a stationary process. The Dickey-Fuller test was adopted to check the stationarity of the processes. An error correction model (ECM) was then adopted to measure deviations from the long-run equilibrium to examine the short-term and long-term effects. The resulting models were used to determine the Granger causality between the two processes. Additional information about the two processes was obtained by examining the impulse response function and variance decomposition.

Results: The Dickey-Fuller test supported an implicit null hypothesis that the dengue incidence and Google Trends results are nonstationary processes ($P=.01$). A further test showed that the processes were cointegrated ($P=.01$), indicating that their particular linear combination is a stationary process. These results permitted us to construct ECMs. The model showed the direction of causality of the two processes, indicating that Google Trends results will Granger-cause dengue incidence (not in the reverse order).

Conclusions: Various hypothesis testing results in this research concluded that Google Trends results can be used as an initial indicator of upcoming dengue outbreaks.

(*J Med Internet Res* 2020;22(7):e17633) doi: [10.2196/17633](https://doi.org/10.2196/17633)

KEYWORDS

dengue; Google Trends; infodemiology; infoveillance; vector error correction model; Granger causality

Introduction

Dengue is known as an infectious disease, which is caused by the dengue virus from *Flaviviridae* and genus *Flavivirus* families. This virus has four serotypes, namely DEN-1, DEN-2, DEN-3, and DEN-4 [1-3]. Infection by one of these four

serotypes does not give cross-protective immunity. Hence, people who live in endemic areas can be reinfected by the other three serotypes throughout their lifetime [4]. *Aedes aegypti* and *Aedes albopictus* mosquitoes are vector transmitters of dengue. The disease is transmitted by mosquitoes through arthropod vectors in tropical and subtropical areas around the world [5]. As the most rapidly spreading mosquito-borne disease in the

world, dengue fever has affected the lives of approximately 1.8 billion people in Southeast Asia alone. In the dengue-endemic region, Indonesia is one of the largest countries, with a population of 267 million [6]. Since the first dengue incidents reported in 1968, the number and range of dengue incidents in Indonesia have increased nationwide [7]. All 34 provinces in Indonesia have been reported to have dengue cases, showing the extensive range of the disease [8]. Early detection of disease activity can reduce the impact of the disease [9].

Bandung is one of the crowded cities in Indonesia. It has the highest dengue incidence, especially in West Java. Daily habits, landscape structures, weather, and the ecosystem in the city play roles in dengue vector breeding as primary factors for dengue transmission. The climate in Bandung is a mountainous climate (humid and cold), with an average temperature of 23.5°C. The average rainfall is 200.4 mm, and there are on average 21.3 rainy days per month. It is an ideal environment for *Aedes aegypti*. According to the Extraordinary Early Childhood Awareness System (SKDKLB-DBD) report, Bandung had the highest dengue incidence from 2002 to 2006, with a total of 22,335 infected people. In January 2019, the West Java Provincial Health Office recorded 236 dengue cases in Bandung. For these reasons, we selected Bandung as our study area to investigate and analyze the association between dengue data from Google Trends and dengue incidence data from a reputable hospital in Bandung.

In this modern world, it is impossible to say that technology, especially the internet, does not influence human lives. Over the years, research has been performed to investigate the accuracy of using internet search engine data to predict real-life phenomena, such as influenza epidemics and flu trends [9], stock markets [10-12], house prices [13,14], and tourism demand [15-17]. Google Trends is a public website belonging to Google Inc that offers data based on Google Search, which shows how frequently a particular search term is entered.

According to StatCounter, in 2016, Google was the most used text search engine in Indonesia. About 97% of people who use the internet in Indonesia use Google. It is assumed that Indonesian people show the trends to find information about dengue on the internet [18]. Therefore, we hypothesized that the popularity of dengue on Google has a correlation with the dengue incidence in Indonesia. Several studies about Google Trends results and the relationship with various diseases have been carried out, such as a study on dengue fever in Indonesia by using moving average analysis [19]. Other Google Trends-related work is presented in the following sentences. Dengue in several countries (Bolivia, Brazil, India, Indonesia, and Singapore) has been studied to maximize a fitting model by using a univariate linear model [20]. A spatiotemporal analysis of dengue incidence has been performed by using an exponential generalized autoregressive condition heteroscedastic model [21]. In other studies, we can see a web-based search for the early identification of the disease prevalence of coronary heart disease [22], forecasting of influenza cases using internet

data [23], the use of Google Trends in health care research by using correlation analysis [24], infodemiology and infoveillance [25-27], a framework of social media data and quality assessment for a reporting standard [28], the spread of pertussis in Europe [29], and the spread of AIDS in the United States [30]. This led us to hypothesize that dengue incidence and Google Trends results have a long-run equilibrium.

We started our research with an initial hypothesis that the popularity of dengue on Google correlates with dengue cases in Bandung. We then investigated the relationship between these two data by using the Dicky-Fuller test, error correction model (ECM), impulse response function, and variance decomposition. We hoped that information from Google Trends can be used for the early detection of upcoming dengue outbreaks so that policymakers can prepare for the early prevention or control of the epidemic.

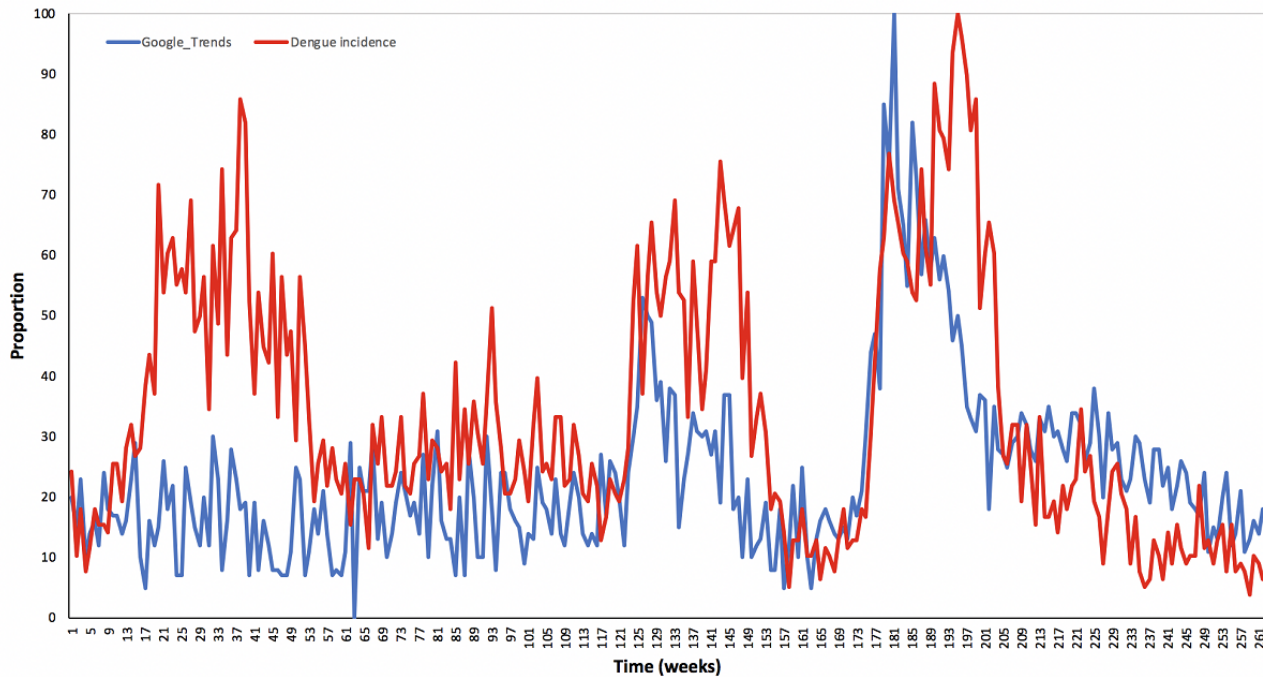
Methods

Collected Data

Google Trends is a website that analyzes the popularity of a topic in various countries and various languages based on search requests. The data source is over the internet and open source and can be easily accessed by everyone. In Google Trends, a user can enter a keyword in the form of words or phrases related to the selected topic or cases. Google Trends is not case sensitive but takes into account spelling errors that might occur. Users can specify the duration of time they want to review by selecting a time range or specifying a date. In addition, users can specify the area to be reviewed by selecting the appropriate country, city, or province or state. They can also see the popularity of these keywords globally by selecting the option *worldwide*.

Data used in this study are time-series data of dengue incidence from Santo Borromeus Hospital in Bandung, as well as popularity data taken from Google Trends via the website (Figure 1). The time range used in this paper is from September 9, 2012 until September 7, 2017. We consider weekly data over the time interval, and there were 261 data points. We assumed that the incidence data represent all dengue cases in Bandung owing to the location of the hospital in the city center. Google Trends data were obtained from the Google Trends website on google [31] by entering the keyword “demam berdarah dengue” plus “dbd.” With regard to the technique of taking data, we followed previous guidelines [32] when using Google Trends for valid results in our study by selecting the appropriate keyword(s), region(s), period, and category. Quotation marks were used so that the search results only showed the popularity of keywords in that exact order. The data represented the popularity of the keyword “demam berdarah dengue” plus “dbd” found on Google. Google Trends normalizes its popularity data by dividing each data point by the total amount of searches at a given time and location. This results in a proportion scaled in a range from 0 to 100. This scale shows the popularity of specific keywords relative to the time and location of the query.

Figure 1. Dengue data plot from Google Trends and reported cases in Bandung.



Stationary Test

We performed a stationary test for the time series data of Google Trends (X_t) and the dengue incidence (Y_t). A time series $\{y_t\}$ is said to be stationary if it satisfies the following conditions: (1) $E[y_t] = \mu < \infty$; (2) $Var[y_t] = \sigma^2 < \infty$; and (3) $Cov[y_t, y_{t+s}] = \gamma_s < \infty$, for $s > 0$.

Differencing a series produces another set of observations, such as the first differenced values, where $\Delta y_t = y_t - y_{t-1}$. Generalizing this operation and performing the difference operation as much as n times can be written as $\Delta^n y_t = y_t - y_{t-n}$. If a series is stationary without any differencing, it is said to be integrated of the order 0 or $I(0)$. However, if it is stationary only after differencing once, it is said to be integrated of the order 1 or $I(1)$. The Dickey-Fuller test was used to detect the presence of a unit root and determine the stationarity of Google Trends and dengue fever incidence series.

Cointegration Test

For cointegration, Engle and Granger [33] used the stationarity test of the residual series obtained from the long-run equilibrium equation. If the residual series, denoted by $\{e_t\}$, is stationary, given that y_t and z_t are first-order stationary, they are cointegrated.

After finding the Google Trends and dengue incidence series to be first-order difference stationary, the long-run equilibrium relationship can be stated in the following form:

$$Y_t = \beta_0 + \beta_1 X_t + e_t \quad (1)$$

where e_t denotes the residual.

Let $\{\hat{e}_t\}$ be the residual sequence. The series $\{\hat{e}_t\}$ contains the estimated values of deviations from the long-run relationship.

By using the Dickey-Fuller test to check its stationarity, it was found that the level values of $\{\hat{e}_t\}$ were stationary.

Estimation and Analysis of a Vector ECM

After a cointegrating relationship has been established, an ECM can be built to establish the short-run relationship between two variables. A likelihood ratio test can be used to determine the time lag of the vector ECM or the value of p . The regression equation for an ECM is as follows:

$$\Delta Y_t = \alpha_1 + \alpha_Y \hat{e}_{t-1} + \sum_{i=1}^p \alpha_{11}(i) \Delta Y_{t-i} + \sum_{i=1}^p \alpha_{12}(i) \Delta X_{t-i} + \epsilon_{Yt} \quad (2)$$

$$\Delta X_t = \alpha_2 + \alpha_X \hat{e}_{t-1} + \sum_{i=1}^p \alpha_{21}(i) \Delta Y_{t-i} + \sum_{i=1}^p \alpha_{22}(i) \Delta X_{t-i} + \epsilon_{Xt} \quad (3)$$

Analysis of cointegration shows that Google Trends and dengue incidence have a long-run equilibrium relationship. However, they are in disequilibrium in the short term. View equations 2 and 3 as a vector autoregression (VAR) model as follows:

$$Z_t = \begin{pmatrix} \Delta Y_t \\ \Delta X_t \end{pmatrix} \quad (4)$$

Hence, the vector ECM at hand can be written as a VAR model as follows:

$$Z_t = \alpha + \hat{\alpha} \hat{e}_{t-1} + A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \epsilon_t \quad (5)$$

Before estimating the vector ECM, the optimal lag order is first determined.

Causal Relationship Between Google Trends and Dengue Incidence

One way to test causality is to see whether the time lag of one variable is relevant for another variable. In a two-equation

system with stationary variables y_t and z_t with p lags, it is said that $\{y_t\}$ does not Granger-cause $\{z_t\}$ if and only if the coefficients of y_t in the equation for z_t are equal to zero. In other words, if $\{y_t\}$ does not provide improvement for the forecasting performance of $\{z_t\}$, $\{y_t\}$ does not Granger-cause $\{z_t\}$. Granger causality only shows the effects of past values of $\{y_t\}$ toward the current values of $\{z_t\}$.

In order to test Granger causality, a standard F test of the restriction $a_{21}(1) = a_{21}(2) = \dots = a_{21}(p) = 0$ is performed.

In a cointegrated system, X_t does not Granger-cause Y_t if the values of ΔX_{t-i} do not enter in the equation of ΔY_t and if Y_t does not respond to deviation from the long-run equilibrium.

Impulse Response Function and Variance Decomposition

To analyze the dynamic effects of the model in response to shocks and the effects on the two variables, the impulse response function and variance decomposition were examined.

Results

Stationary Test

The stationary test results can be seen in Table 1, Figure 2A, and Figure 2B. The table shows that the level values of the Google Trends and dengue incidence series were nonstationary. However, Google Trends and dengue incidence data were found to be stationary after being differentiated once. This was done to reduce the fluctuations in the data.

Table 1. Dickey-Fuller test for Google Trends data, dengue incidence data, first differenced Google Trends data, and first differenced dengue incidence data.

Variable	Dickey-Fuller test statistic (value)	Dickey-Fuller critical value (N=250)
X_t ^a	-2.42 (.02)	-2.58
Y_t ^b	-2.24 (.03)	-2.58
ΔX_t ^c	-21.76 (.01)	-2.58
ΔY_t ^d	-27.85 (.01)	-2.58

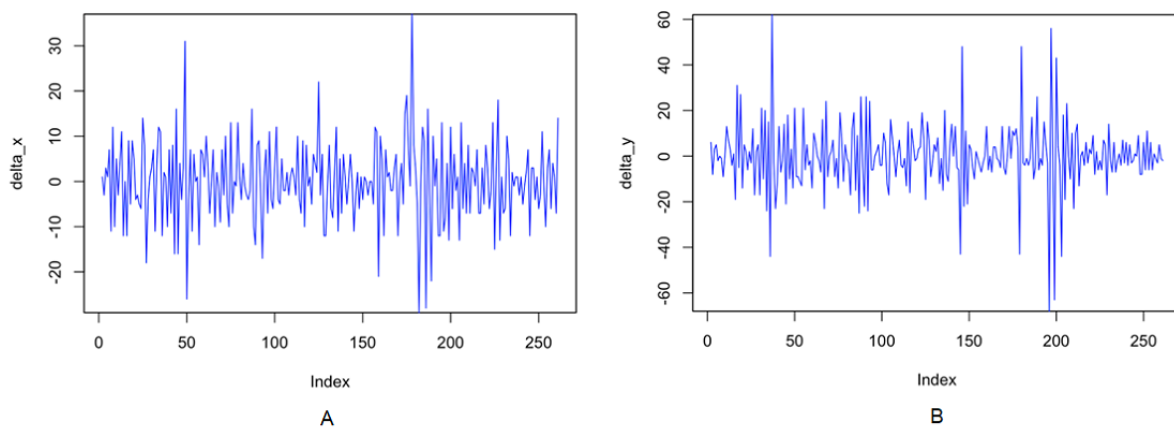
^aGoogle Trends data.

^bDengue incidence data.

^cFirst differenced Google Trends data.

^dFirst differenced dengue incidence data.

Figure 2. (A) $\{X_t\}$ and (B) $\{Y_t\}$ plots with one-time difference. X_t : Google Trends data; Y_t : dengue incidence data.



Cointegration Test

The cointegration test results of ordinary least squares regression yielded that the long-run equilibrium relationship can be shown as follows:

$$Y_t = \frac{12.609}{(1.84)} + \frac{0.455}{(0.06)} X_t + e_t \tag{6}$$

with e_t denoting the residual.

Let $\{\hat{e}_t\}$ be the residual sequence. The series $\{\hat{e}_t\}$ contains the estimated values of deviations from the long-run relationship. By using the Dickey-Fuller test to check its stationarity, it was found that the level values of $\{\hat{e}_t\}$ were stationary. The results are shown in Table 2.

From previous results, it was seen that $\{X_t\}$ and $\{Y_t\}$ are $I(1)$, and because the $\{\hat{e}_t\}$ series is stationary, $\{X_t\}$ and $\{Y_t\}$ are cointegrated. Hence, a vector ECM can be constructed.

Table 2. Dickey-Fuller test for the residual sequence.

Variable	Dickey-Fuller test statistic	Dickey-Fuller critical value (N=250)
e_t^a	-8.77	-2.58

^aresidual estimated as follows: $e_t = Y_t - 12.609 - 0.455X_t$

Likelihood Ratio Test to Find the Time Lag

The longest feasible lag length was set as 8 weeks. Thereafter, the value of the determinant of the variance-covariance matrix of a model with lag length eight was examined (denoted as Σ_8) and compared with that of a model with lag length seven (denoted as Σ_7). The likelihood ratio is $(T - c)(\ln \Sigma_7 - \ln \Sigma_8)$, where T is the number of observations and c is the number of parameters that are estimated in each equation of the unrestricted

system. In the case of comparing the eight-lag model to the seven-lag model, the value of c is $1 + 8n$, with n being the number of variables, which is two in this case. If the likelihood ratio is smaller than the critical value (χ^2_4 at a significance of $\alpha=1\%$), the null hypothesis of the restriction $A_8=0$ is rejected. This is done until lag 1.

The results of this test are shown in [Table 3](#). The likelihood ratio test showed that the optimal number of lags needed for this vector ECM is three.

Table 3. Likelihood ratio test for lag length.

Number	H_0	H_1	Likelihood ratio	χ^2_4	Verdict
1	$A_8=0$	$A_8 \neq 0$	7.655	13.277	H_0 rejected
2	$A_7=0$	$A_7 \neq 0 A_8=0$	3.291	13.277	H_0 rejected
3	$A_6=0$	$A_6 \neq 0 A_8=A_7=0$	0.221	13.277	H_0 rejected
4	$A_5=0$	$A_5 \neq 0 A_8=\dots A_6=0$	2.543	13.277	H_0 rejected
5	$A_4=0$	$A_4 \neq 0 A_8=\dots A_5=0$	6.191	13.277	H_0 rejected
6	$A_3=0$	$A_3 \neq 0 A_8=\dots A_4=0$	19.666	13.277	H_0 rejected
7	$A_2=0$	$A_2 \neq 0 A_8=\dots A_3=0$	27.887	13.277	H_0 rejected
8	$A_1=0$	$A_1 \neq 0 A_8=\dots A_2=0$	60.361	13.277	H_0 rejected

Estimation of the ECM

After finding the optimal number of lags, an ECM model was built. The estimated vector ECM is as follows:

$$\begin{pmatrix} \Delta Y_t \\ \Delta X_t \end{pmatrix} = \begin{pmatrix} -0.1816 \\ -0.0267 \end{pmatrix} e_{t-1} + \begin{pmatrix} -0.6054 & -0.0654 \\ 0.0523 & -0.3726 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-1} \\ \Delta X_{t-1} \end{pmatrix} + \begin{pmatrix} -0.4001 & 0.1859 \\ 0.0672 & -0.2283 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-2} \\ \Delta X_{t-2} \end{pmatrix} + \begin{pmatrix} -0.2283 & 0.0524 \\ 0.1047 & -0.0095 \end{pmatrix} \begin{pmatrix} \Delta Y_{t-3} \\ \Delta X_{t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{Y_t} \\ \varepsilon_{X_t} \end{pmatrix} \quad (7)$$

From the equation, it is seen that the speed of the adjustment parameter is -0.1816 for $\{Y_t\}$ and -0.0267 for $\{X_t\}$. This means that when there is a deviation of 1 from the long-run equilibrium in the period $t - 1$, the number of dengue incidences will decrease by 0.1816 and dengue popularity in Google will decrease by 0.0267.

The speed of adjustment parameter for dengue incidence was nine times larger than the value for Google Trends, meaning that dengue incidence is more responsive to deviations from

the long-run equilibrium. On the other hand, Google Trends only responds slightly to the aforementioned deviation.

It was found that this model has an R-squared value of 0.4128 for the ΔX_t equation and 0.1511 for the ΔY_t equation, as well as an Akaike Information Criterion (AIC) value of 2370.2. Since the R-squared value is quite low, it can be said that the model cannot explain the data at hand accurately.

Causal Relationship Between Google Trends and Dengue Incidence

Based on the vector ECM in equation 7, Granger causality was tested between Google Trends and dengue incidence. It was noted that at lag 2 and 3, X_t Granger-causes Y_t at a significance of $\alpha=5\%$, since the P obtained was similar (.04). However, Y_t does not Granger-cause X_t . This means that information from past values of Google Trends regarding dengue at a lag of 2 and 3 weeks is useful for explaining the present value of dengue incidence. The results are shown in [Textbox 1](#).

Textbox 1. Granger causality test for Google Trends data and dengue incidence data.

Part 1: Does Google Trends Granger-cause incidence?

$$\Delta Y_t = \sum_{i=1} \alpha_i \Delta Y_{t-i} + \sum_{j=1} \beta_j \Delta X_{t-j} + c_1 + e_{1t}$$

$H_0: \beta_1 = \dots = \beta_j = 0$ (Google Trends does not Granger-cause incidence)

Lag	F statistic	P value
1	0.752 (1.259)	.39
2	3.175 (2.259)	.04
3	2.769 (3.259)	.04

Part 2: Does incidence Granger-cause Google Trends?

$$\Delta X_t = \sum_{i=1} \alpha_i \Delta X_{t-i} + \sum_{j=1} \beta_j \Delta Y_{t-j} + c_2 + e_{2t}$$

$H_0: \beta_1 = \dots = \beta_j = 0$ (incidence does not Granger-cause Google Trends)

Lag	F statistic	P value
1	0.141 (1.259)	.71
2	0.049 (2.259)	.95
3	1.877 (3.259)	.13

Impulse Response Function and Variance Decomposition

The results for 12 periods (3 months) are obtained as presented below.

Impulse Response Function

As shown in Figure 3B, a positive shock in dengue popularity on Google Trends has a relevant impact on dengue incidence. Dengue incidence shows a large increase after two periods. Thereafter, it shows a slight decrease, but then, it increases again slowly. Its effects seem to be long term, since the incidence keeps increasing until the end of the 12 periods. This phenomenon suggests that shocks in dengue fever popularity on Google have a relevant impact on dengue fever incidence.

On the other hand, through analysis of the response of dengue incidence to a positive shock, it was found that dengue

popularity increases slightly and then remains constant. This behavior is presented in Figure 3A. After a positive shock in dengue incidence, there is little fluctuation in its popularity on Google Trends. This suggests that shocks in dengue incidence do not have a relevant effect on its popularity on Google Trends. Its effects are only short term and do not remain in the long run.

Generally, the impulse response function shows that Google Trends has a relevant impact on dengue fever incidence and has a long-term effect. On the contrary, dengue incidence has only a short-term and small effect on the popularity of dengue on Google.

Variance Decomposition

Variance decomposition estimates the contribution of shocks in a variable toward the response of another variable. As shown in Figure 3D, the contribution of dengue incidence to its variance gradually declines in the first two periods. Thereafter, it declines

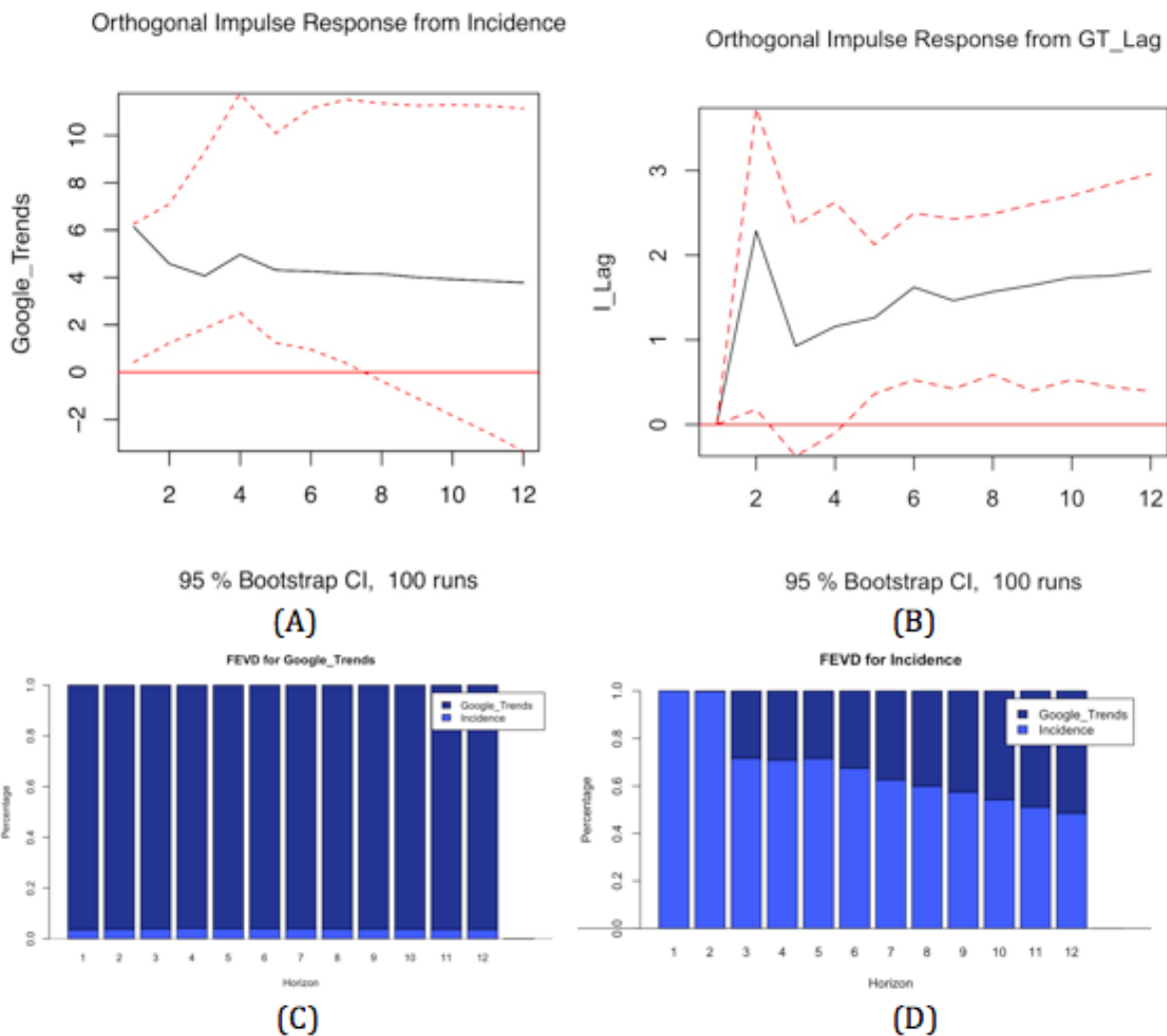
further until the contribution of dengue incidence is finally only around 40%. In the first period, Google Trends only has a small contribution to dengue incidence variance with only 0%. Thereafter, during the second period, it increases to around 28% and then continues to climb gradually. After the 12 periods, Google Trends has up to 60% contribution to dengue incidence variance.

On the other hand, Figure 3C shows that Google Trends variance mainly comes from itself, where dengue incidence only contributes at a rate of approximately 4%. This rate increases in the first four periods (from 3.6% in the first period to 4.2%

in the fourth period). Thereafter, the contribution rate from dengue incidence remains around 3.7%. This means that dengue incidence very slightly influences Google Trends in the short term, but does not influence the popularity of Google Trends in the long term.

In summary, it can be seen that Google Trends influences dengue incidence in the long term, but dengue incidence only influences Google Trends in the short term and not in the long term. As presented in the model, dengue incidence is related to not only the popularity of dengue in Google but also its lagged value of up to 1 week.

Figure 3. Impulse response function of (A) dengue data from Google Trends with respect to reported cases and (B) reported cases with respect to dengue data from Google Trends. Forecast error variance decomposition of (C) dengue data from Google Trends and (D) reported cases.



Discussion

Principal Findings

Our results show that there is indeed a causal relationship between dengue popularity in Google Trends and dengue incidence in Bandung. A Granger cointegrated relationship between dengue popularity in Google Trends and dengue incidence in Bandung was noted. This is justified because both

data sets were found to be $I(1)$, and the residual from the ordinary least squares regression was also found to be stationary.

Based on the ECM, it can be seen that there is a relationship between Google Trends results and dengue incidence. Through Granger analysis, it was seen that Google Trends Granger-causes dengue incidence in Bandung at a lag of 2 and 3 weeks. This was further supported by the impulse response function, where shocks in dengue popularity in Google cause dengue incidence to increase. It was also supported by the variance decomposition, where after 1 week, the contribution from Google Trends to

dengue incidence variance increases. Granger analysis also showed that dengue incidence does not Granger-cause its popularity in Google.

The vector ECM also showed that dengue incidence is more responsive to deviations from the long-run equilibrium, since it has a larger value of the speed of adjustment, which is nine times the value for Google Trends.

Limitations

The results showed a causal relationship between dengue popularity in Google Trends and dengue incidence in Bandung. However, this exact ECM cannot be used for forecasting or early detection owing to the low R-squared values of 0.4128 for the Google Trends equation and 0.1511 for the dengue incidence equation. A further improved model will need to be built for future forecasting.

The results of this study can help provide a more real-time indication of dengue outbreaks in Bandung. Owing to Indonesia's standard and traditional approach to dengue surveillance, the data of dengue cases have several weaknesses, such as low accuracy and timeliness [20]. In addition, data

available from Santo Borromeus Hospital were only from 2008 until 2017. Owing to the limitations of Google Trends, it was preferable to use weekly data with a 5-year period (giving 260 data points) rather than data from 2008 to 2014 on a monthly basis with only 84 data points. Another limitation is that people searching for dengue-related information may not necessarily have the disease, as they could be searching because a relative or friend is ill. Besides, a search in Bandung does not necessarily mean that the intended sick person is in Bandung. The individual could be searching for someone else who is ill in another city.

Our proposed model used strong assumptions, such as the behavior of the use of gadgets and social media in the community, which is quite high, and a good internet signal in the observation area (Bandung in this case). Therefore, it is risky to implement the findings in areas with low internet access.

Conclusions

Google Trends data may be used as an initial indicator of a dengue outbreak in Bandung. However, further improvements to the ECM need to be made by using more data points to gain more extensive insights.

Acknowledgments

Part of this work was supported by the Indonesian RistekDikti Grant 2020. The second author gratefully acknowledges the financial support provided by the Indonesia Ministry of Research and Technology through the Pendidikan Magister menuju Doktor untuk Sarjana Unggul (PMDSU) Program.

Authors' Contributions

MS performed the statistical analysis, review, and interpretation. MF and JTMSE performed numerical simulation and constructed this paper. ES interpreted and reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Halstead SB, Heinz FX, Barrett A, Roehrig JT. Dengue virus: molecular basis of cell entry and pathogenesis, 25-27 June 2003, Vienna, Austria. *Vaccine* 2005 Jan 04;23(7):849-856. [doi: [10.1016/j.vaccine.2004.03.069](https://doi.org/10.1016/j.vaccine.2004.03.069)] [Medline: [15603884](https://pubmed.ncbi.nlm.nih.gov/15603884/)]
2. Malavige GN, Fernando S, Fernando DJ, Seneviratne SL. Dengue viral infections. *Postgrad Med J* 2004 Oct;80(948):588-601 [FREE Full text] [doi: [10.1136/pgmj.2004.019638](https://doi.org/10.1136/pgmj.2004.019638)] [Medline: [15466994](https://pubmed.ncbi.nlm.nih.gov/15466994/)]
3. Fakhruddin M, Putra PS, Wijaya KP, Sopaheluwakan A, Satyaningsih R, Komalasari KE, Mamenun, Sumiati, et al. Assessing the interplay between dengue incidence and weather in Jakarta via a clustering integrated multiple regression model. *Ecological Complexity* 2019 Aug;39:100768. [doi: [10.1016/j.ecocom.2019.100768](https://doi.org/10.1016/j.ecocom.2019.100768)]
4. Westaway E, Brinton M, Gaidamovich SYa, Horzinek M, Igarashi A, Kääriäinen L, et al. Flaviviridae. *Intervirology* 1985;24(4):183-192. [doi: [10.1159/000149642](https://doi.org/10.1159/000149642)] [Medline: [3000978](https://pubmed.ncbi.nlm.nih.gov/3000978/)]
5. Lu L, Lin H, Tian L, Yang W, Sun J, Liu Q. Time series analysis of dengue fever and weather in Guangzhou, China. *BMC Public Health* 2009 Oct 27;9:395 [FREE Full text] [doi: [10.1186/1471-2458-9-395](https://doi.org/10.1186/1471-2458-9-395)] [Medline: [19860867](https://pubmed.ncbi.nlm.nih.gov/19860867/)]
6. Karyanti MR, Uiterwaal CS, Kusriastuti R, Hadinegoro SR, Rovers MM, Heesterbeek H, et al. The changing incidence of dengue haemorrhagic fever in Indonesia: a 45-year registry-based analysis. *BMC Infect Dis* 2014 Jul 26;14:412 [FREE Full text] [doi: [10.1186/1471-2334-14-412](https://doi.org/10.1186/1471-2334-14-412)] [Medline: [25064368](https://pubmed.ncbi.nlm.nih.gov/25064368/)]
7. Wahyono TY, Nealon J, Beucher S, Prayitno A, Moureau A, Nawawi S, et al. Indonesian dengue burden estimates: review of evidence by an expert panel. *Epidemiol Infect* 2017 Aug;145(11):2324-2329 [FREE Full text] [doi: [10.1017/S0950268817001030](https://doi.org/10.1017/S0950268817001030)] [Medline: [28545598](https://pubmed.ncbi.nlm.nih.gov/28545598/)]
8. Wardhani P, Aryati A, Yohan B, Trimarsanto H, Setianingsih TY, Puspitasari D, et al. Clinical and virological characteristics of dengue in Surabaya, Indonesia. *PLoS One* 2017;12(6):e0178443 [FREE Full text] [doi: [10.1371/journal.pone.0178443](https://doi.org/10.1371/journal.pone.0178443)] [Medline: [28575000](https://pubmed.ncbi.nlm.nih.gov/28575000/)]
9. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009 Feb 19;457(7232):1012-1014. [doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634)] [Medline: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)]

10. Da Z, Engelberg J, Gao P. In search of attention. *The Journal of Finance* 2011 Oct;66(5):1461-1499. [doi: [10.1111/j.1540-6261.2011.01679.x](https://doi.org/10.1111/j.1540-6261.2011.01679.x)]
11. Preis T, Moat HS, Stanley HE. Quantifying trading behavior in financial markets using Google Trends. *Sci Rep* 2013;3:1684 [FREE Full text] [doi: [10.1038/srep01684](https://doi.org/10.1038/srep01684)] [Medline: [23619126](https://pubmed.ncbi.nlm.nih.gov/23619126/)]
12. Hamid A, Heiden M. Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization* 2015 Sep;117:62-81. [doi: [10.1016/j.jebo.2015.06.005](https://doi.org/10.1016/j.jebo.2015.06.005)]
13. Wu L, Brynjolfsson E. *The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales*. United States: University of Chicago Press; Apr 2015:89-118.
14. Askitas N, Zimmermann KF. Detecting mortgage delinquencies. *Social Science Research Network* 2011 Jul:1-16 [FREE Full text]
15. Song H, Liu H. Predicting Tourist Demand Using Big Data. In: *Analytics in Smart Tourism Design*. Switzerland: Springer; Oct 2016:13-29.
16. Pan B, Chenguang Wu D, Song H. Forecasting hotel room demand using search engine data. *JHTT* 2012 Sep 28;3(3):196-210. [doi: [10.1108/17579881211264486](https://doi.org/10.1108/17579881211264486)]
17. Yang X, Pan B, Evans JA, Lv B. Forecasting Chinese tourist volume with search engine data. *Tourism Management* 2015 Feb;46:386-397. [doi: [10.1016/j.tourman.2014.07.019](https://doi.org/10.1016/j.tourman.2014.07.019)]
18. Prabowo H, Hamdani R, Sanusi Z. The New Face of People Power: An Exploratory Study on the Potential of Social Media for Combating Corruption in Indonesia. *AABFJ* 2018;12(3):19-40. [doi: [10.14453/aabfj.v12i3.3](https://doi.org/10.14453/aabfj.v12i3.3)]
19. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. *Glob Health Action* 2019;12(1):1552652 [FREE Full text] [doi: [10.1080/16549716.2018.1552652](https://doi.org/10.1080/16549716.2018.1552652)] [Medline: [31154985](https://pubmed.ncbi.nlm.nih.gov/31154985/)]
20. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* 2011 May;5(5):e1206 [FREE Full text] [doi: [10.1371/journal.pntd.0001206](https://doi.org/10.1371/journal.pntd.0001206)] [Medline: [21647308](https://pubmed.ncbi.nlm.nih.gov/21647308/)]
21. Risteski D, Davcev D. Can we use daily Internet search query data to improve predicting power of EGARCH models for financial time series volatility. In: *Proceedings of the International Conference on Computer Science and Information Systems (ICSIS)*. 2014 Oct Presented at: International Conference on Computer Science and Information Systems (ICSIS); October 17-18, 2014; Dubai, United Arab Emirates. [doi: [10.15242/icie.e1014066](https://doi.org/10.15242/icie.e1014066)]
22. Senecal C, Widmer RJ, Lerman LO, Lerman A. Association of Search Engine Queries for Chest Pain With Coronary Heart Disease Epidemiology. *JAMA Cardiol* 2018 Dec 01;3(12):1218-1221 [FREE Full text] [doi: [10.1001/jamacardio.2018.3459](https://doi.org/10.1001/jamacardio.2018.3459)] [Medline: [30422176](https://pubmed.ncbi.nlm.nih.gov/30422176/)]
23. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. *JMIR Public Health Surveill* 2018 Jan 09;4(1):e4 [FREE Full text] [doi: [10.2196/publichealth.8950](https://doi.org/10.2196/publichealth.8950)] [Medline: [29317382](https://pubmed.ncbi.nlm.nih.gov/29317382/)]
24. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. *PLoS One* 2014;9(10):e109583 [FREE Full text] [doi: [10.1371/journal.pone.0109583](https://doi.org/10.1371/journal.pone.0109583)] [Medline: [25337815](https://pubmed.ncbi.nlm.nih.gov/25337815/)]
25. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
26. Eysenbach G. Infodemiology: the epidemiology of (mis)information. *The American Journal of Medicine* 2002 Dec;113(9):763-765. [doi: [10.1016/s0002-9343\(02\)01473-0](https://doi.org/10.1016/s0002-9343(02)01473-0)]
27. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med* 2011 May;40(5 Suppl 2):S154-S158. [doi: [10.1016/j.amepre.2011.02.006](https://doi.org/10.1016/j.amepre.2011.02.006)] [Medline: [21521589](https://pubmed.ncbi.nlm.nih.gov/21521589/)]
28. Kim Y, Huang J, Emery S. Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. *J Med Internet Res* 2016 Feb 26;18(2):e41 [FREE Full text] [doi: [10.2196/jmir.4738](https://doi.org/10.2196/jmir.4738)] [Medline: [26920122](https://pubmed.ncbi.nlm.nih.gov/26920122/)]
29. Gianfredi V, Bragazzi N, Mahamid M, Bisharat B, Mahroum N, Amital H, et al. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. *Public Health* 2018 Dec;165:9-15. [doi: [10.1016/j.puhe.2018.09.001](https://doi.org/10.1016/j.puhe.2018.09.001)] [Medline: [30342281](https://pubmed.ncbi.nlm.nih.gov/30342281/)]
30. Mavragani A, Ochoa G. Forecasting AIDS prevalence in the United States using online search traffic data. *J Big Data* 2018 May 19;5(1). [doi: [10.1186/s40537-018-0126-7](https://doi.org/10.1186/s40537-018-0126-7)]
31. Google Trends. URL: <https://trends.google.co.id/trends/?geo=ID> [accessed 2018-05-10]
32. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health Surveill* 2019 May 29;5(2):e13439 [FREE Full text] [doi: [10.2196/13439](https://doi.org/10.2196/13439)] [Medline: [31144671](https://pubmed.ncbi.nlm.nih.gov/31144671/)]
33. Engle RF, Granger CW. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 1987 Mar;55(2):251. [doi: [10.2307/1913236](https://doi.org/10.2307/1913236)]

Abbreviations**ECM:** error correction model**VAR:** vector autoregression

Edited by G Eysenbach; submitted 30.12.19; peer-reviewed by H Susanto, A Mavragani, S Kardes; comments to author 10.03.20; revised version received 03.05.20; accepted 20.05.20; published 24.07.20

*Please cite as:**Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E**Causality Analysis of Google Trends and Dengue Incidence in Bandung, Indonesia With Linkage of Digital Data Modeling: Longitudinal Observational Study**J Med Internet Res 2020;22(7):e17633*URL: <http://www.jmir.org/2020/7/e17633/>doi: [10.2196/17633](https://doi.org/10.2196/17633)PMID: [32706682](https://pubmed.ncbi.nlm.nih.gov/32706682/)

©Muhammad Syamsuddin, Muhammad Fakhruddin, Jane Theresa Marlen Sahetapy-Engel, Edy Soewono. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 24.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.