Volume 41 **Number** 9, 2025, **Pages** e22594
DOI: 10.22146/bkm.v41i09.22594

Utilizing "Google Trends" data to support early detection of epidemic outbreaks: a preliminary study

Frisca Rahmadina[1], Bagas Suryo Bintoro[2], Aditya Lia Ramadona[2]*

Submitted:
June 26th, 2025
Accepted:
September 25th, 2025
Published:
September 29th, 2025

[1]Department of Environmental Health, Faculty of Public Health, Universitas Sriwijaya, South Sumatra, Indonesia

[2]Department of Health Behavior, Environment, and Social Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

***Correspondence:**
alramadona@ugm.ac.id

Abstract

Purpose: This study examined the potential application of Google Trends in supporting early epidemic detection and health campaigns, using the COVID-19 pandemic in Indonesia as a case study. **Method:** COVID-19 case data from 2020 to 2022 were collected. Search patterns were analyzed using Indonesian keywords for symptoms: "demam", "sakit kepala", "pilek", "bersin", "sakit tenggorokan", "perut", "batuk", "nafsu makan", "muntah", "lesu", "mual", and "diare." The search patterns were then compared to the COVID-19 case data. **Results:** We observed a pattern alignment between Google Trends and COVID-19 case peaks. Additionally, differences in lag time were identified between search trends and case peaks across SARS-CoV-2 variants. For instance, the peaks of "sakit tenggorokan" and "batuk" searches lagged about one week for Omicron, around two weeks for Delta, and more than two weeks for Alpha. **Conclusion:** Internet search activity can support early detection of epidemics and inform timely health campaigns. Moreover, search trends might offer a novel approach to estimate disease incubation periods.

Keywords: COVID-19; early detection; Google Trends; health campaigns; internet search activity

INTRODUCTION

The rapid advancement of digital technologies has transformed the way health information is accessed, shared, and utilized globally. Among these technologies, internet search activity represents a dynamic and immediate reflection of public interest and behavior toward health issues [1,2]. The field of digital epidemiology has emerged from this phenomenon, utilizing non-traditional data sources such as search engines, social media, and mobile apps to complement traditional disease surveillance systems [3,4]. One of the most critical applications of internet search data is its potential for early detection of epidemic outbreaks. Previous studies have shown that search query analysis can detect disease trends days to weeks earlier than conventional reporting systems [3,5]. For instance, Google Trends data successfully predicted influenza outbreaks and dengue fever trends across various regions [6,7].

During the COVID-19 pandemic, these digital tools became particularly relevant. Search engines, notably Google, with a 97.3% market share in Indonesia [8], serves as primary tools for seeking health-related information, offering insights into variant-specific symptomatology and opportunities for early detection. Beyond surveillance, Google Trends data also holds promise for informing health campaigns. By understanding the timing, geographic distribution, and intensity of search behavior, health authorities can design targeted communication strategies to address public concerns, correct misinformation, and promote preventive behaviors [9]. Strategic health campaigns that align with search behavior patterns may enhance the effectiveness of health promotion efforts, especially in contexts with rapidly evolving epidemics [10].

Despite these potentials, several gaps persist in current research. Most existing studies have focused on developed countries, which limits the generalizability of their findings to low- and middle-income countries, such as Indonesia, where digital literacy, internet access, and health information-seeking behavior differ substantially [11]. Furthermore, there is limited integration between real-time search analytics and actionable public health interventions. Given these gaps, this study aims to explore the utilization of Google Trends data to support the early detection of epidemic waves and inform timely public health communication during the COVID-19 pandemic in Indonesia. By contextualizing internet search behavior within a developing country framework, this research aims to advance the digital epidemiology literature and provide practical insights for enhancing epidemic preparedness and health communication strategies.

METHODS

This study employed a descriptive observational research design, utilizing secondary data. The research was conducted in Indonesia, covering the period from January 2020 to December 2022. The study used publicly accessible data sources, specifically Google Trends search indices and official COVID-19 case reports from the Indonesian government's COVID-19 information portal.

The selection criteria for data involved keywords representing common symptoms associated with COVID-19. The keywords were chosen in the Indonesian language to reflect local search behaviors. They included: "demam" (fever), "sakit kepala" (headache), "pilek" (cold), "bersin" (sneeze), "sakit tenggorokan" (sore throat), "perut" (stomach), "batuk" (cough), "nafsu makan" (appetite), "muntah" (vomit), "lesu" (fatigue), "mual" (nausea), and "diare" (diarrhea). These keywords were selected as they represent the most common symptoms of COVID-19 reported by official health authorities, such as the WHO, during the early phases of the pandemic [12]. These keywords were searched individually in Google Trends (https://trends.google.com/trends), with search settings restricted to Indonesia and a timeframe of January 2020 to December 2022.

Google Trends presents the popularity of a search term on a scale from 0 to 100, where 100 represents the highest search interest and 0 represents minimal or no search interest during the specified period. COVID-19 case data were collected based on daily reports published by the Indonesian government, providing the number of confirmed positive cases. Data collection involved downloading and compiling COVID-19 daily case numbers, then aligning them with the daily Google Trends search indices for each symptom keyword. No human respondents were directly involved in this study; therefore, no ethical approval or informed consent was required. The data collected is publicly available and anonymized.

Operational definitions were established for each keyword to ensure consistent interpretation during analysis. "Demam" referred to elevated body temperature; "sakit kepala" referred to pain localized in the head region; "pilek" referred to nasal congestion or runny nose; "bersin" referred to sudden expulsion of air from the nose; "sakit tenggorokan" referred to pain or irritation in the throat; "sakit perut" referred to symptoms associated with abdominal discomfort; "batuk" referred to forced coughing episodes; "nafsu makan" referred to decreased appetite; "muntah" referred to vomiting incidents; "lesu" referred to extreme fatigue; "mual" referred to feelings of nausea; and "diare" referred to increased frequency and liquidity of bowel movements.

Data analysis involved a visual examination and descriptive comparison of the trends in symptom-related search indices and COVID-19 case incidence. The primary focus was to identify patterns of search volume peaks and their temporal relationship to peaks in COVID-19 case counts. In particular, a lag time was observed between search interest peaks and case peaks, suggesting the potential use of Google Trends in early epidemic detection. Statistical software was not used for complex modeling in this phase; the study remained exploratory, focusing on pattern identification through graphical analysis.

RESULTS

The trends of COVID-19 cases in Indonesia from 2021 to 2022 showed three major peaks. The first significant peak occurred in early 2021, coinciding with the initial widespread transmission and the emergence of the Alpha variant. The second, and most severe, peak was driven by the Delta variant, which surged dramatically between June and August 2021. Finally, the third major peak corresponded to the Omicron variant, which began in late 2021 and peaked in early 2022.

Public interest in COVID-19 symptoms, as reflected by search activity on Google Trends, began to emerge in January 2020. A significant relationship was observed between public search behavior and the trend of confirmed COVID-19 cases.

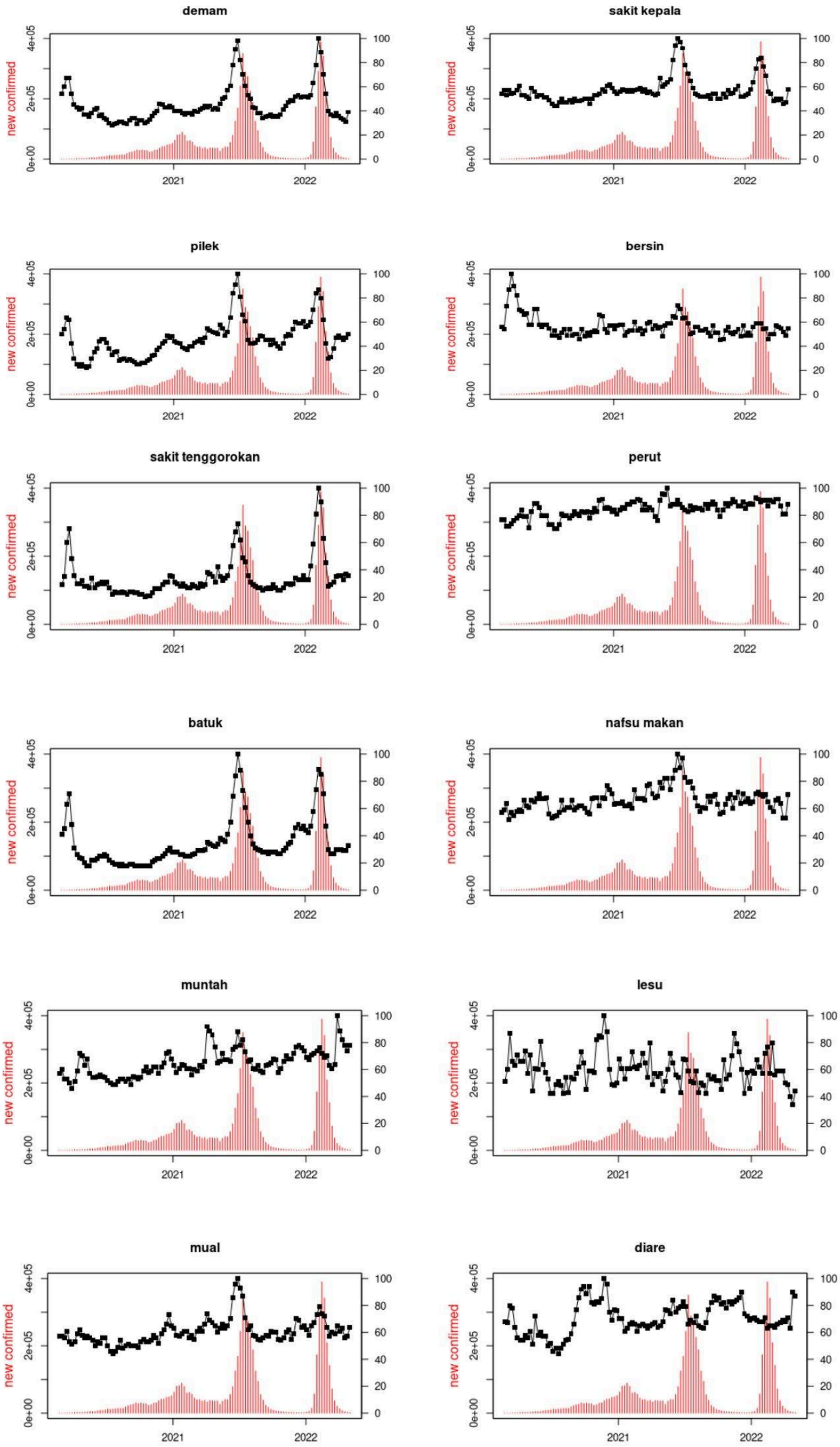

Figure 1. Trends of COVID-19 cases and symptom-related searches in Indonesia (2020–2022)

Figure 1 shows that keywords such as "demam" (fever), "pilek" (cold), "bersin" (sneeze), "sakit tenggorokan" (sore throat), "batuk" (cough), "muntah" (vomiting), "lesu" (fatigue), and "diare" (diarrhea) reached their first peak of search activity during the early stages of the pandemic in Indonesia.

A closer examination of Figure 1 reveals distinct search patterns for different symptoms, particularly in relation to the major outbreak waves. During the Delta variant wave in mid-2021, which was the most severe peak in terms of confirmed cases, search interest for several specific symptoms also reached its highest point. As noted, keywords such as "sakit kepala" (headache), "nafsu makan" (appetite loss), and "mual" (nausea) showed a pronounced spike that closely mirrored the sharp increase in COVID-19 cases.

This suggests that these particular systemic symptoms may have been especially prominent or concerning to the public during the period when the Delta variant was dominant. In contrast, core respiratory symptoms, such as "batuk" (cough) and "sakit tenggorokan" (sore throat), exhibited significant search peaks during all three major waves, highlighting their consistent role as key public health indicators throughout the pandemic. Meanwhile, searches for "bersin" (sneeze) saw their highest relative interest in the very early stages of the pandemic, likely driven by initial public awareness campaigns, before peaking again, albeit less dramatically, during the Omicron wave.

Another key observation from the data was the presence of lag time between search trends and the peaks of COVID-19 cases. Keywords related to respiratory symptoms, particularly "sakit tenggorokan" and "batuk", showed different lag patterns depending on the COVID-19 variant. For the Omicron variant, search interest for these symptoms peaked approximately one week before the peak in confirmed cases.

For the Delta variant, the lag time was approximately two weeks, while for the Alpha variant, the lag exceeded two weeks. This variation in lag time suggests that changes in virus characteristics, public awareness, and the severity of symptoms could influence how early individuals seek information online before official case numbers rise. These findings suggest the potential utility of Google Trends as an early indicator of COVID-19 epidemic surges in Indonesia.

DISCUSSION

This study demonstrated the significant potential of digital epidemiology as a disease surveillance tool in lower-middle-income countries, such as Indonesia. By leveraging Google Trends data, we identified patterns of health information-seeking behavior that can serve as an early warning signal for epidemic waves, aligning with previous research suggesting internet search behaviors are early indicators of disease transmission [13].

The most important finding is the observation of a consistent, variant-dependent lag time between the peak of symptom searches and the peak of confirmed cases. Searches for key symptoms such as “sakit tenggorokan” and “batuk” peaked approximately one week before the Omicron case peak, two weeks before the Delta peak, and more than two weeks before the Alpha peak. This result is consistent with the shorter incubation period of Omicron [14] and supports findings that associate it with milder, upper respiratory tract symptoms such as rhinorrhea, sneezing, and sore throat [15,16]. This suggests that digital search data can not only reflect the temporal trends of disease spread but also provide real-time clues about changing clinical characteristics resulting from viral mutations, highlighting its potential to contribute to variant differentiation based on symptomology [17]. The use of local Indonesian language terms, such as “sakit kepala” and “sakit tenggorokan”, which have been incorporated in only a few prior analyses [13,18], enhances the contextual relevance of these findings.

These findings have substantial practical implications for health policy. The ability to predict case surges up to two weeks in advance offers a crucial window of opportunity for health authorities to conduct proactive interventions [19]. Real-time monitoring of symptom-specific queries allows for the design of targeted and timely risk communication campaigns [20,21]. For example, a sudden surge in searches for “sakit tenggorokan” or “demam” in a specific region could prompt tailored campaigns emphasizing symptom recognition, testing, and preventive measures in those high-search areas. Integrating search trend analysis into health promotion strategies allows for adaptive interventions that are responsive to dynamic community needs, ultimately enhancing the effectiveness of epidemic control efforts [22,23].

Socially, the application of this tool has two sides. On one hand, as a low-cost surveillance tool, Google Trends has the potential to improve health equity by providing data from regions where traditional surveillance may be weak or slow [24]. However, an over-reliance on this data risks exacerbating the digital divide [25]. Policies designed based on the behavior of internet users could inadvertently neglect vulnerable populations without digital access, such as older people, rural communities, and low-income groups [26]. Therefore, the integration of this data must be done carefully, as a complement, not a substitute for conventional epidemiological data, to ensure a fair and equitable health response.

Although promising, the validity and reliability of digital epidemiology require continuous improvement. Future research should focus on integrating multiple digital data sources, such as social media platforms and electronic health records, to enhance the capabilities for detecting outbreaks. Furthermore, the development of methodologies, such as advanced keyword mapping and adjustment for media effects, is necessary to refine digital epidemiology models. This study has several strengths. First, it utilizes publicly available, real-time, and low-cost data, a particularly valuable approach for surveillance in resource-limited settings. Its primary strength lies in analyzing three distinct epidemic waves within a developing country context, utilizing local language keywords to enhance contextual relevance. To the best of our knowledge, this study systematically compares the lag time of symptom searches across different COVID-19 variants in Indonesia, offering clear and actionable insights for public health authorities.

Nevertheless, several limitations must be acknowledged. First, Google Trends data reflect only the search behaviors of internet users, potentially excluding rural or low-resource populations. Second, search motivations are multifactorial; not all searches are symptom-driven, with media coverage and public announcements likely influencing trends. Third, variations in spelling, synonyms, and keyword selection may have affected the comprehensiveness of search data collected. These limitations reaffirm that digital epidemiology should be viewed as a powerful complementary tool, rather than an absolute replacement for traditional public health surveillance systems.

CONCLUSION

This study highlights the potential of Google Trends as a complementary tool to traditional surveillance systems for monitoring COVID-19 symptom patterns in Indonesia. Real-time search data related to symptoms such as fever, sore throat, cough, and diarrhea proved to be early indicators preceding increases in confirmed cases. The findings suggest that digital epidemiology can strengthen early warning systems. However, it cannot fully replace conventional surveillance due to issues such as internet accessibility, information-seeking behaviors, and the influence of misinformation on search trends. To optimize the use of Google Trends, future public health strategies should consider regional disparities in internet use and infrastructure across Indonesia. Incorporating digital surveillance alongside conventional methods could improve outbreak prediction and response, particularly in regions with higher COVID-19 prevalence. Further research is recommended to analyze regional differences and to establish stronger correlations between real-time symptom search patterns and actual case data, which may enhance the precision and reliability of digital epidemiology in pandemic preparedness.

REFERENCES

1. Rushendran R, Chitra V. Exploring infodemiology: unraveling the intricate relationships among stress, headaches, migraines, and suicide through Google Trends analysis. Frontiers in Big Data. 2025;7: 1365417.
2. Gholamzadeh M, Asadi Gharabaghi M, Abtahi H. Public interest in online searching of asthma information: insights from a Google Trends analysis. BMC Pulmonary Medicine. 2025;25(1).
3. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveillance. 2020;25(10):2000199.
4. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity - a comparison with awareness campaigns for lung and prostate cancer. BMC Cancer. 2011;11:442.
5. Smith KC, Rimal RN, Sandberg H, Storey JD, Lagasse L, Maulsby C, et al. Understanding newsworthiness of an emerging pandemic: International newspaper coverage of the H1N1 outbreak. Influenza and other Respiratory Viruses. 2013;7(5):847–853.
6. Foroughi F, Lam AKY, Lim MSC, Saremi N, Ahmadvand A. “Googling” for cancer: An infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. JMIR Cancer. 2016;2(1):e5.
7. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012–1014.
8. Statcounter Global Stats. Search engine market share in Indonesia - June 2022 [Internet]. 2022. Available from: [Website]
9. Kisa S, Kisa A. A Comprehensive analysis of covid-19 misinformation, public health impacts, and communication strategies: scoping review. Journal of Medical Internet Research. 2024;26: e56931.

10. Ahmed MM, Okesanya OJ, Olakeke NO, Adigun OA, Adebayo UO, Oso TA, Eshun G, Lucero-Prisno III DE. Integrating digital health innovations to achieve universal health coverage: promoting health outcomes and quality through global public health equity. Healthcare. 2025;13(1060):22.
11. Poulsen A, Song YJC, Fosch-Villaronga E, LaMonica HM, Iannelli O, Alam M, et al. Digital rights and mobile health in Southeast Asia: a scoping review. Digital Health. 2024;10:20552076241257058.
12. World Health Organization. COVID-19 symptoms [Internet]. 2020. Available from: [Website]
13. Szilagyi IS, Ullrich T, Lang-Illievich K, Klivinyi C, Schittek GA, Simonis H, et al. Google Trends for pain search terms in the world's most populated regions before and after the first recorded COVID-19 case: Infodemiological study. Journal of Medical Internet Research. 2021;23(4):e27214.
14. Jansen L, Tegomoh B, Lange K, Showalter K, Figliomeni J, Abdalhamid B, et al. Investigation of a SARS-CoV-2 B.1.1.529 (Omicron) Variant Cluster — Nebraska, november–december 2021. MMWR Morbidity and Mortality Weekly Report. 2021;70:1782–1784.
15. Mahase E. COVID-19: hospital admission 50-70% less likely with omicron than delta, but transmission a major concern. BMJ. 2021;375:n3151.
16. Iacobucci G. COVID-19: runny nose, headache, and fatigue are commonest symptoms of omicron, early data show. BMJ. 2021;375:n3103.
17. Islam R, Hossain J. Detection of SARS‐CoV‐2 Omicron (B.1.1.529) variant has created panic among the people across the world: What should we do right now?. Journal of Medical Virology. 2022;94: 1768–1769.
18. Lu T, Reis BY. Internet search patterns reveal the clinical course of COVID-19 disease progression and pandemic spread across 32 countries. NPJ Digital Medicine. 2021;4(1):22.
19. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika virus disease outbreak. JMIR Public Health Surveillance. 2016;2(1):e30.
20. Kostkova P, Saigí-Rubió F, Eguia H, Borbolla D, Verschuuren M, Hamilton C, et al. Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. Frontiers in Digital Health. 2021;3:707902.
21. Chen J, Wang Y. Social media use for health purposes: Systematic review. Journal of Medical Internet Research. 2021;23(5):e17917.
22. Zhang L, Guo W, Lv C. Modern technologies and solutions to enhance surveillance and response systems for emerging zoonotic diseases. Science in One Health. 2024;3:100061.
23. Corbin JH, Oyene UE, Manoncourt E, Onya H, Kwamboka M, Amuyunzu-Nyamongo M, et al. A health promotion approach to emergency management: effective community engagement strategies from five cases. Health Promotion International. 2021;36(Supplement 1):i24–i38.
24. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Global Health Action. 2018;12:1552652.
25. Sanders AK, Scanlon E. The digital divide is a human rights issue: advancing social inclusion through social work advocacy. Journal of Human Rights and Social Work. 2021;6:130-143.
26. Djatmiko GH, Sinaga O, Pawirosumarto S. Digital transformation and social inclusion in public services: a qualitative analysis of e-government adoption for marginalized communities in sustainable governance. Sustainability. 2025;17; 2098.