http://jhmi.sums.ac.ir
Health Man & Info Sci 2022, 9(1), 16–21

HMIS
Health Management and Information Science

Original Article

Google Trend as an Early Warning System for Corona Outbreak Investigation in Iran

Sajjad Nozari[1], Seyedeh Leila Dehghani[2], Razieh Chabok[3], Behrooz Moloudpour[4], Zahra Moradi Vastegani[5], Somayeh Moalemi[6], Masoumeh Sadat Mousavi[1,7]*

[1]Assistant professor, Department of Epidemiology and Biostatistics, School of Public Health, Shahrekord University of Medical Sciences, Shahrekord, Iran
[2]Department of Public Health Behbahan Faculty of Medical Sciences, Behbahan, Iran
[3]Student Research Committee, School of Nursing and Midwifery, Sabzevar University of Medical Sciences, Sabzevar, Iran
[4]Student Research Committee, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran
[5]Alimentary Tract Research Center, Imam Khomeini Hospital Clinical Research Development Unit, Jundishapur University of Medical Sciences, Ahvaz, Iran
[6]School of Management and Medical Informatics, Shiraz University of Medical Sciences, Shiraz, Iran
[7]Assistant professor, Modeling in Health Research Center, Shahrekord University of Medical Sciences, Shahrekord, Iran

Abstract

Introduction: Digital epidemiology is introduced as a major aspect of epidemiology; its sources are digital data and it uses spaces such as Google, YouTube and Twitter as databases. In the recent Covid-19 pandemic, the use of digital epidemiology, as an early warning system, has been considered. This study aimed to investigate the context of Google Trend as an early warning system in the study of coronavirus outbreaks in Iran.

Methods: The coronavirus epidemic in Iran started on February 24, 2020, and with some differences to consider the rumors in the community, we consider the date before the announcement of all by February 16, 2020 until November 16, 2021. We searched using keywords related to symptoms such as "fever", "cough" and "sore throat" and the keyword "corona symptoms"; information was extracted and entered in Microsoft Excel and the keyword chart was drawn according to the date of each wave. Spearman correlation test was performed to find the correlation between keywords in SPSS version 18.

Results: The trend chart of the keywords "fever", "cough" and "sore throat" and the keyword "corona symptoms" in different waves of coronavirus in Iran showed an increase in keyword searches before the onset of the corona epidemic wave. Spearman correlation coefficient between sore throat and fever was 0.645, sore throat and cough 0.775, sore throat and corona symptoms 0.684, between fever and cough keywords 0.435, fever and corona symptoms 0.779 and between keyword cough and corona symptoms 0.503. In all these coefficients, the level of error of the first type was 0.05 significant (P<0.001)

Conclusion: Google Trend, a digital epidemiology tool, can be used as an effective early warning system to control the corona pandemic, and this field of epidemiological knowledge with all its limitations needs further research.

Keywords: Google trend, Early warning system, outbreak investigation, Digital epidemiology, Iran, Covid 19

Article History:
Received: 12 December 2021
Accepted: 01 February 2022

Please cite this paper as:
Nozari S, Dehghani SL, Chabok R, Moloudpour B, Moradi Vastegani Z, Moalemi S, Mousavi MS. Google Trend as an Early Warning System for Corona Outbreak Investigation in Iran. Health Man & Info Sci. 2022; 9(1): 16-21. doi: 10.30476/JHMI.2022.94812.1122.

***Correspondence to:**
Masoumeh Sadat Mousavi, Modeling in Health Research Center, Shahrekord University of Medical Sciences, Shahrekord, Iran
Email: mousavi68.ma@yahoo.com
ma@skuns.ac.ir

Introduction

With the advancement of science in various dimensions and levels, the gates of technology and communication have been opened to human beings and the human environment more than in previous decades; the global village and, consequently, the health of society and human health depend on adaptation to the stage of development of human societies (1). Today, people go to experts and scientific books or magazines for their questions less than the seventies and try to answer their questions with a very limited keywords on the Internet (2, 3).

The prevalence of digital media-related behaviors in the world is growing rapidly, so that 3.8 billion people in the world are actively and continuously using the Internet (4). For example, people use the Internet to diagnose the diseases based on the symptoms

and problems and to know the treatment process or which doctor has a better performance in the field of their disease and after that time reservation for a visit. These activities depend on the country or region and on the Internet and the common cyberspace in the country and the applications under the famous network in the country. Information from the needs of the community is very valuable and expensive and needs to be considered as a database(5) .

The health of the community and its members is highly dependent on the management of the health department and the discussion of disease control. For proper health policy, we need appropriate information in the field of health. Health sector information in Iran is collected in a traditional way from the community with traditional epidemiological methods such as interviews, questionnaires, etc. and periodic disease reporting or care training in health centers, or hospital data and scientific documents. (6) this data does not include all members of society and a large part of society without a share in this information is not considered. Failure to consider this part of the community causes an error in estimating and wasting resources and facilities(7).

The Internet, is an information revolution that today, with tools such as personal computers and smartphones, etc., plays a role in people's daily lives. Therefore, the Internet, as a database, needs more research and data mining in epidemiology(8).

Digital epidemiology, according to Marcel Salathe, is »the science of epidemiology, which is based on digital data and is formed outside the structure of public health« and uses epidemiological methods to collect and analyze the data (9). Various sources of information such as social networks (Telegram, Instagram, Facebook, etc.), websites, search engines, online service centers, and online stores are very valuable data platforms in digital epidemiology(10).

Digital epidemiology, as a new aspect of epidemiology in recent years, has been based on the principle that the data formed on the Internet can be used in various areas of health, such as:

1) Predicting the prevalence of disease and health consequences in complex technical and social systems with mathematical calculations

2) Helping health management by predicting epidemics and outbreaks with high-realistic data in large-scale modeling

3) Digital evidence-based hypothesis for health sector policy(11, 12)

The Internet and its information have become very important and widely used in the discussion of Corona health education. People then search the Internet for " the symptoms of coronavirus" after feeling the symptoms and deviations in their health status to determine if they are infected or not (13).

It can be hypothesized that the presence of digital epidemiology in the design of a care system in the corona pandemic can reduce the burden of disease through an early warning system (14). The purpose of this study was to evaluate the feasibility of using Google Trend as an early warning system for rapid detection of pandemics and epidemics in Iran.

Materials and Methods

This study was conducted in the Google Trends Database to show a correlation and relationship between daily incidence and search volume for keywords related to coronavirus disease. The findings of this study follow the introduction of Google Trend as an Early Warning system for coronavirus disease. The date of the announcement of the first cases of Corona in Iran was announced on February 19, 2020, and Iran came out of the white situation on March 24. About a week later, on March 29, Mike Ryan, executive director of the World Health Organization's emergency medical program, said that Iran had only two cases last week, and today it has confirmed up to 245 cases of coronavirus infection.

Frequently used keywords about corona symptoms in Iran used in this study were corona symptoms, fever, sore throat, cough. The increase in the incidence of coronavirus in Iran is referred to as "wave" and unfortunately in Iran, a reliable and specific source has not published a diagram of the incidence of the disease since the beginning of the pandemic and the start date of each wave has been extracted from news sources affiliated with the Ministry of Health.

The data for each studied keyword was extracted from Google Trends and classified in Microsoft Excel. The Google Trend index data for each coronavirus epidemic wave was extracted in a time-dependent graph. Spearman correlation test was used to find the correlation between keywords due to abnormal distribution, and correlation coefficients with P. Value are presented in the Table. The time of detection of the first case of the disease in Iran was February 19, 2020, and the time of the announcement of the epidemic by the Ministry of Health was February 24, 2020. Considering the rumors and news that existed before the announcement of the epidemic by the Ministry of Health, the start of the study was scheduled for February 16, 2020.

Results

The first wave of the corona epidemic occurred after the uptrend of keyword examination, corona

symptoms. About a week before the start of the epidemic wave, the upward trend in Google search had begun. In the second, third, and fifth waves, there is a correlation between the wave of coronavirus cases and the increase of GTI index (Figure 1).

Waves of coronavirus infection in Iran:

The first wave: 26/2/2020 to 29/4/2020

The second wave: 30/5/2020 to 31/8/2020

Third wave: 26/9/2020 to 28/12/2020

Fourth wave: 26/3/2012 to 25/5/2012

Fifth wave 27/6/2021 to 6/10/2012. According to Figure 2, the keywords "fever" in the first, second, and fifth waves of the corona have an apparent correlation with the keyword "corona symptoms".

According to Figures 3 and 4, which are the key words for "sore throat" and "cough" as symptoms of coronavirus. Adaptation in the upward trend of the search for the word sore throat and cough are almost similar to the keywords of Corona Symptoms and fever. Spearman correlation coefficient was calculated for GTI index data for each of the keywords (sore throat, fever, cough, corona symptoms), and these coefficients show a high correlation in keyword search since the corona epidemic in Iran (Table 1).

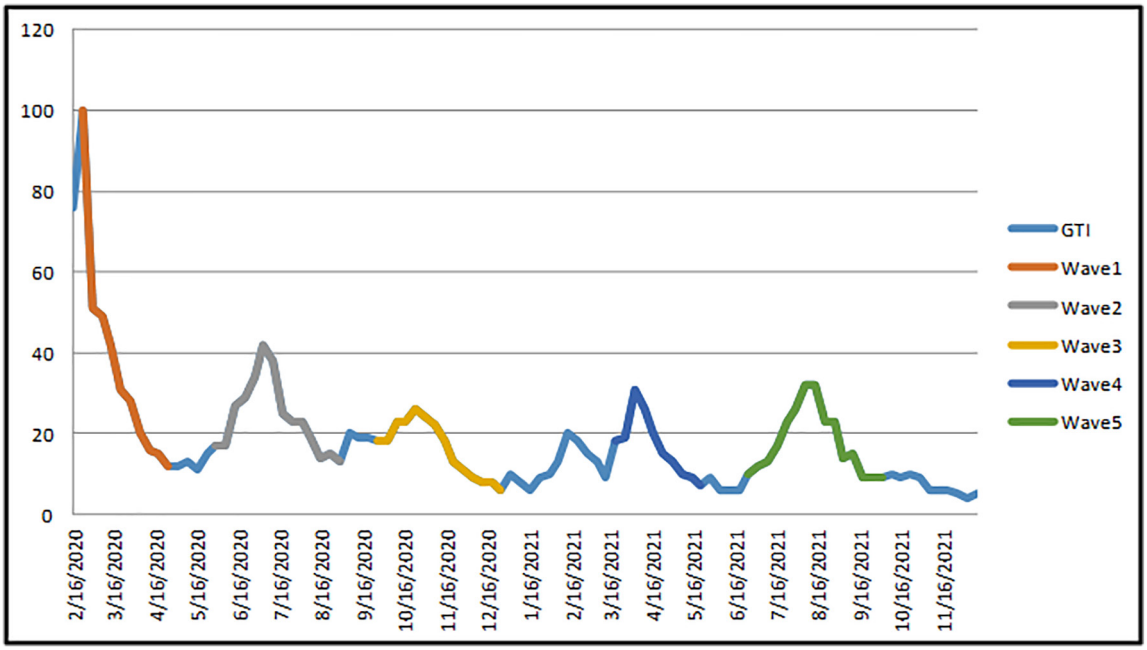

Figure 1: GTI Chart with Keyword "Corona Symptoms" in Iran - February 16, 2020 - WeeklyChart

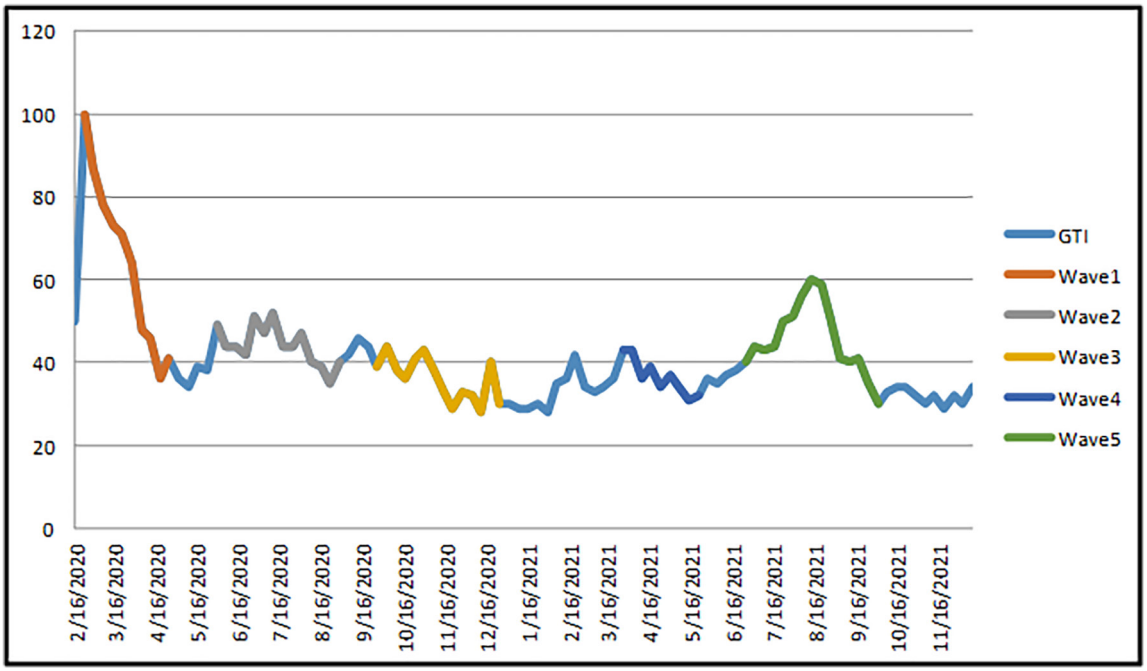

Figure 2: GTI chart for the keyword "fever" in Iran - February 16, 2020_ Weekly chart

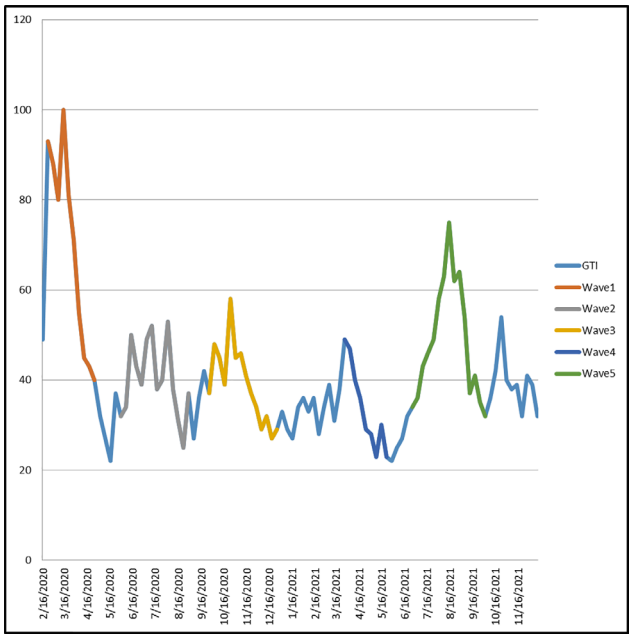

Figure 3: GTI chart for the keyword "Sore throat" in Iran - February 16, 2020_ Weekly chart

Discussion

By integrating digital epidemiology, as a tool in disease control and care based on the data obtained from this study, it can be hoped that Google Trend can be an Early Warning system for identifying outbreaks related to corona disease. Our study showed that the search volume of keywords such as "corona symptoms", "fever", "cough" and "sore throat" in Google increased before or with the upward trend of cases in Iran. According to the results of the correlation between keywords, a strong and moderate significant correlation was observed between the volume of keyword searches at a given time.

Fever is the most severe increase in search among other coronavirus symptoms, but it should be noted that fever is associated with other infectious diseases, and this can reduce the share of fever search volume in the search for coronavirus symptoms. Cough, as one of the main clinical symptoms in corona, has a similar trend to fever and keyword corona symptoms and a high correlation with the other keywords like fever and corona symptoms; this can indicate the relationship between increased search volume and disease outbreak. Our findings are consistent with those of other studies. For example, a study by Giuseppe Lippi in Italy found that the number of weekly searches for the keywords fever, cough, and dyspnea peaked one week before the corona pandemic, and there was a significant correlation between the keywords (15). The study carried out by Walker et al. found that there was a correlation between corona mortality and the use of Google search terms throughout Italy, Spain, the United Kingdom, the United States, Germany,

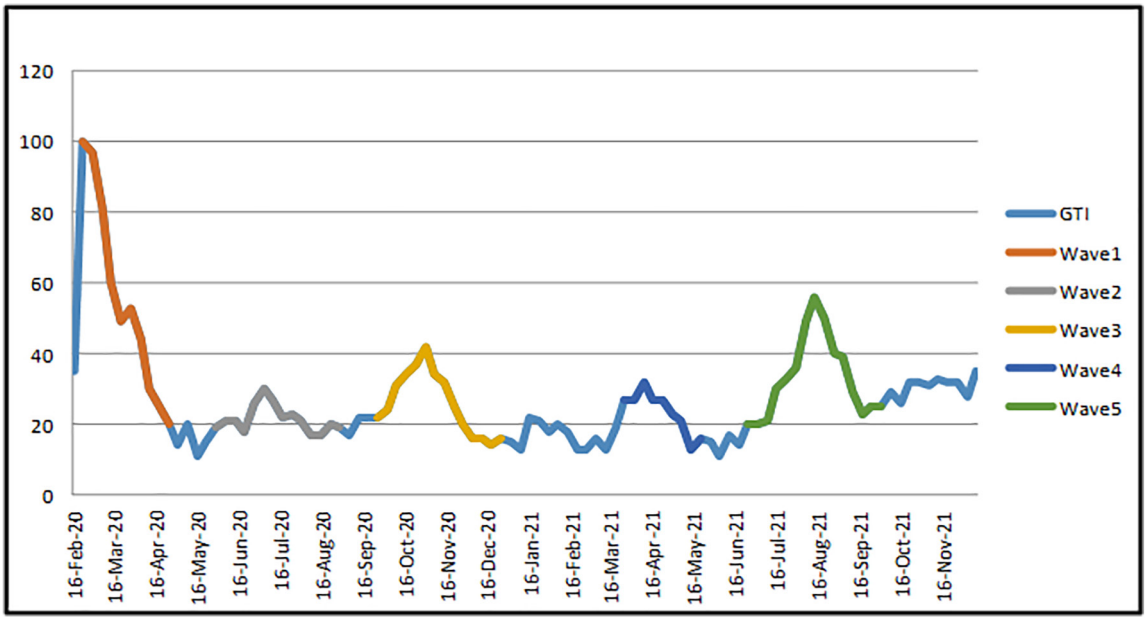

Figure 4: GTI chart for the keyword "Cough" in Iran - February 16, 2020_ Weekly chart

Table 1: Spearman correlation coefficient for keywords fever, sore throat, cough, and corona symptoms

	Sore throat	Fever	Cough	Corona symptoms
Sore throat		0.645	0.775	0.684
Fever			0.435	0.779
Cough				0.503

France, Iran, and the Netherlands (16). The study of Tamás Ferenci et al. also showed a correlation between the incidence of coronavirus and Google Trend data in European countries, which is consistent with the results of this study.

The use of these data in Iran is not without its drawbacks. For example, compared to the study mentioned in Italy, the Iranian health care system does not have an online database for daily cases of coronavirus for research (17). The Ministry of Health of Iran does not specify the exact time of the onset of different waves of coronavirus, and the start date of different waves of the pandemic in Iran is not exactly announced.

To explain the diagram of different keywords and the large volume of keyword searches in the first wave and the beginning of the corona pandemic in Iran compared to subsequent waves with even more deaths and higher incidence rates, we can say that people have a low level of information about this disease at the beginning of the pandemics, and increasing knowledge of people about the symptoms of coronavirus through various sources reduced the volume of search over time.

The reason for not choosing "Corona" as the keyword of the early warning system is that in different months of the year during the epidemic, a part of people's social life was combined with the keywords such as "Corona vaccine", "Corona loan", "Corona quarantine" which takes up part of the search volume but cannot help us predict the course of the disease, and it is the cause of the wrong diagnosis of the next wave.

It should also be noted that part of the volume of Google searches for research purposes and knowledge acquisition does not help to investigate an outbreak. According to studies, a research unit in the context of digital epidemiology in the health research department in disease care control can act as an early warning system and identify subsequent waves of disease before the daily number of patients increases.

Conclusion

Based on the findings of this study, it can be concluded that there is valuable data in the context of the Internet and Google Trend that can be used in epidemiological studies and used to make decisions at the level of health management. Of course, this data requires validation and accuracy as to how to choose the search the keywords to avoid misjudgment.

Limitations of the Study

The use of network data and pre-indexed data of Google Trends is one of the serious limitations of this study; the other problem of this study includes the inaccurate date of corona outbreak waves in Iran.

Authors' Contribution

S N and M M were the main investigators of the study. SN , MM , SB , R CH, BM and ZM were involved in preparing the concept and design. S N and MM revisited the manuscript and critically evaluated the intellectual contents. All authors participated in preparing the final draft of the manuscript, revised the manuscript, and critically evaluated the intellectual contents. All authors have read and approved the final version of the manuscript.

Ethical Considerations

Considering the use of Internet data from the context of Google Trends and not facing the human phase, this study does not require ethics code. However, this study has been registered in the research management system of Shahrekord University of Medical Sciences with the code5778.

Funding/Support

This study did not receive support from any financial or organizational sources

Conflict of Interest: None declared.

References

1. Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. *Nat Rev Gastroenterol Hepatol*. 2019;16(5):312-21. doi: 10.1038/s41575-019-0102-5.
2. Lu SQ, Xie G, Chen Z, Han X, editors. The management of application of big data in internet of thing in environmental protection in China. 13 August 2015. Redwood: 2015 IEEE First International Conference on Big Data Computing Service and Applications; 2015. doi: 10.1109/BigDataService.2015.68.
3. Sarowar MG, Kamal MS, Dey N. Internet of Things and its impacts in computing intelligence: a comprehensive review–IoT application for big data. *Big Data Analytics for Smart and Connected Cities*. 2019:103-36. doi: 10.4018/978-1-5225-6207-8.ch005.
4. Eckmanns T, Fuller H, Roberts SL. Digital epidemiology and global health security; an interdisciplinary conversation. *Life Sci Soc Policy*. 2019;15(1):2. doi: 10.1186/s40504-019-0091-8.
5. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*.

2013;309(13):1351-2. doi: 10.1001/jama.2013.393.
6. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health*. 2017;7(3):185-9. doi: 10.1016/j.jegh.2017.06.001.
7. Raub M. Bots, bias and big data: artificial intelligence, algorithmic bias and disparate impact liability in hiring practices. *Ark L Rev*. 2018;71:529.
8. Griffin GP, Mulhall M, Simek C, Riggs WW. Mitigating bias in big data for transportation. *Journal of Big Data Analytics in Transportation*. 2020;2(1):49-59. doi: 10.1007/s42421-020-00013-0.
9. Lippi G, Mattiuzzi C, Cervellin G. Is Digital Epidemiology the Future of Clinical Epidemiology? *J Epidemiol Glob Health*. 2019;9(2):146. doi: 10.2991/jegh.k.190314.003.
10. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. *Nat Med*. 2020;26(8):1183-92. doi: 10.1038/s41591-020-1011-4.
11. Tarkoma S, Alghnam S, Howell MD. Fighting pandemics with digital epidemiology. *EClinicalMedicine*. 2020;26:100512. doi: 10.1016/j.eclinm.2020.100512.
12. Khan N, Faisal S. Epidemiology of Corona virus in the world and its effects on the China economy. *Available at SSRN 3548292*. 2020. doi: 10.2139/ssrn.3548292.
13. Park YE. Research evidence for reshaping global energy strategy based on trend-based approach of big data analytics in the corona era. *Energy Strategy Reviews*. 2022;41:100835. doi: 10.1016/j.esr.2022.100835.
14. Salathe M, Freifeld CC, Mekaru SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. *N Engl J Med*. 2013;369(5):401-4. doi: 10.1056/NEJMp1307752.
15. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo G, Huiart L. Digital health strategies to fight COVID-19 around the globe: challenges and recommendations. *J Med Internet Res*. 2020;46(3):1-8. doi: 10.2196/preprints.19284.
16. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. *Int Forum Allergy Rhinol*. 2020;10(7):839-47. doi: 10.1002/alr.22580.
17. Lippi G, Mattiuzzi C, Cervellin G. Google search volume predicts the emergence of COVID-19 outbreaks. *Acta Biomed*. 2020;91(3):e2020006. doi: 10.23750/abm.v91i3.10030.