The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine

eJIFCC

Ups and downs of COVID-19: can we predict the future? Local analysis with Google Trends for forecasting the burden of COVID-19 in Pakistan

Sibtain Ahmed[1], Muhammad Abbas Abid[2], Maria Helena Santos de Oliveira[3], Zeeshan Ansar Ahmed[1], Ayra Siddiqui[4], Imran Siddiqui[1], Lena Jafri[1], Giuseppe Lippi[5]

[1] *Department of Pathology and Laboratory Medicine, Aga Khan University, Karachi, Pakistan*
[2] *Section of Clinical Chemistry, Department of Pathology and Laboratory Medicine, Aga Khan University, Karachi, Pakistan*
[3] *Biostatistics Master's Program, Maringá State University, Paraná, Brazil*
[4] *Medical College, Aga Khan University. Stadium Road, Karachi, Pakistan*
[5] *Section of Clinical Biochemistry, Department of Neuroscience, Biomedicine and Movement, University of Verona, Verona, Italy*

ARTICLE INFO

Corresponding author:
Dr. Lena Jafri
Associate Professor
& Section Head Chemical Pathology
Department of Pathology
and Laboratory Medicine
The Aga Khan University
Pakistan
Phone: 92-213-4861927
E-mail: lena.jafri@aku.edu

Key words:
Google Trends, COVID-19, Pakistan, pandemic, prediction

ABSTRACT

Background

We aim to study the utility of Google Trends search history data for demonstrating if a correlation may exist between web-based information and actual coronavirus disease 2019 (COVID-19) cases, as well as if such data can be used to forecast patterns of disease spikes.

Patients & methods

Weekly data of COVID-19 cases in Pakistan was retrieved from online COVID-19 data banks for a period of 60 weeks. Search history related to COVID-19, coronavirus and the most common symptoms of disease was retrieved from Google Trends during the

same period. Statistical analysis was performed to analyze the correlation between the two data sets. Search terms were adjusted for time-lag over weeks, to find the highest cross-correlation for each of the search terms.

Results

Search terms of 'fever' and 'cough' were the most commonly searched online, followed by coronavirus and COVID. The highest peak correlations with the weekly case series, with a 1-week backlog, was noted for loss of smell and loss of taste. The combined model yielded a modest performance for forecasting positive cases. The linear regression model revealed loss of smell (adjusted R^2 of 0.7) with significant 1-week, 2-week and 3-week lagged time series, as the best predictor of weekly positive case counts.

Conclusions

Our local analysis of Pakistan-based data seemingly confirms that Google trends can be used as an important tool for anticipating and predicting pandemic patterns and pre-hand preparedness in such unprecedented pandemic crisis.

❖❖❖❖❖

INTRODUCTION

Pakistan is a low-income country in the subcontinent and also one of the most overpopulated countries in the world, with a high prevalence of communicable diseases. Pakistan ranks 154th out of 189 countries, with Human Development Index value of 0.557 (1). Due to high population densities, lack of skilled medical personnel and resources, low literacy rate and budgetary constraints, among other reasons, Pakistan's healthcare system is especially vulnerable to epidemic infectious disease, and coronavirus disease 2019 (COVID-19) is unfortunately not an exception to this rule.

Pakistan has earlier struggled with controlling other life-threatening infectious diseases such as dengue, hepatitis, acquired immunodeficiency syndrome (AIDS) and so forth, and has lost thousands of lives because of this (2,3,4). This has led researchers to believe that the risk for COVID-19 morbidity and mortality may be higher in Pakistan compared to other worldwide countries (5). The lack of resources, such as test assays, has made it imperative for low-income countries (like Pakistan) to identify reliable alternatives to mass COVID-19 testing, so that the spread of disease could be curbed before becoming unmanageable for society, government and healthcare system.

Since COVID-19 has put more pressure on an already overburdened and underfunded healthcare system, diagnostic testing for severe acute respiratory disease coronavirus 2 (SARS COV-2) infections has been vastly limited. This has potentially led to an underestimation of COVID-19 prevalence in the country (6). However, this problem is not only limited to Pakistan and other developing countries. Most clinical laboratories worldwide, up to 80%, have reported facing difficulties in SARS-COV-2 testing, while more than half reported shortage of supplies needed for routine molecular testing (52%) (7). Hence, a substitute for diagnostics is needed, that could accurately mirror COVID-19 epidemiology in specific geographical areas. In a study conducted by Ginsberg et al, internet search-engine query data was used to predict the course of influenza in the United States (8). This surveillance method provided positive results in non-English speaking countries as well. Studies have correlated the data from Europe, China, Korea and Taiwan with Google Trends for COVID-19 and other epidemic diseases, such as influenza (9, 10). Hence, the application of Google Trends to track disease progression has a far-reaching influence for countries across the globe, regardless of their location or language. It is also

cost-effective, timesaving and does not carry any substantial economical or organizational burdens the healthcare system. Therefore, we ideally conceive that Google Trends may have the potential to assess the prevalence of COVID-19 in Pakistan and anticipate new waves of infection.

A survey conducted in Pakistan showed that nearly half (45%) of patients search about their health-related concerns on the Internet (11). Hence, using epidemiological data from Google trends could significantly represent the Pakistani population. Google search terms regarding key symptoms, such as loss of taste and loss of smell, may help in predicting the epidemiological trajectory of COVID-19 (12,13). Olfactory and gustatory dysfunctions have a strong association with COVID-19 patients, with anosmia showing the highest correlation (14). Tracking these symptoms can be a feasible and viable means for assessing the prevalence of COVID-19 and effectively target the government response.

This study was hence designed to assess the potential utility of Google search trends focused on COVID-19 symptoms (including anosmia and dysgeusia), in projecting the trajectory of the local pandemic outbreak in Pakistan through correlation with ongoing COVID-19 statistics.

MATERIAL AND METHODS

In order to avoid daily variations in the positivity rate, the number of weekly COVID-19 confirmed cases in Pakistan were retrieved from OurWorldInData.org, powered by the Johns Hopkins Coronavirus Resource Center (15). The data was further confirmed from the official website of the Ministry of National Health, Pakistan (16,17). This information was retrieved for a period between March 15, 2020 and June 15, 2021. Data was acquired from Google Trends (Google Inc., Mountain View, CA), using the following search terms, encompassing the most representative symptoms in COVID-19: fever, cough, headache, shortness of breath, taste loss and hearing loss, along with other virus-related keywords such as 'COVID-19', 'coronavirus', 'virus' and 'COVID. A weekly Google Trends score was obtained for each keyword on a scale of 100 points, reflecting the cumulative number of Google searches during the previous week. The maximum attainable score of 100 was defined as the highest search volume during the study period for a particular search.

The study was conducted in accordance with the Declaration of Helsinki, under the terms of relevant local legislation. This analysis was based on electronic searches in unrestricted, publicly available repositories, such that no informed consent or ethical committee approvals were needed.

STATISTICAL ANALYSIS

Statistical analysis was carried out using Microsoft Excel for Windows (2016) and R Software (version 4.0.2; R Foundation for Statistical Computing). Correlation analysis for individual search terms was used for assessing the time lags which generated the maximum achievable correlations between the weekly positive cases and Google trends timeline. The corresponding P values and 95% confidence intervals (CIs) were also calculated. A P value <0.05 was considered as statistically significant.

To calculate the quantitative effect of Google Trend score increment on subsequent rise in weekly cases, time series linear regression analysis was performed, and the time lag with maximum predictive value was computed. Adjusted R2 values and graphic analysis was undertaken to assess the combined model performance of positivity rate forecasting compared against national surveillance data.

RESULTS

The highest overall trend value for the study duration was achieved for fever (n=3036), followed by cough (n=2120), Coronavirus (n=1669), COVID (n=1417), headache (n=1284), COVID-19 (n=333), virus (n=329), shortness of breath (n=257), loss of smell (n=129) and loss of taste (n=106) respectively. From all searched terms, fever and cough during second week of June and last week of May 2021 attained the highest Google trend value of 100.

Time-series linear regression analysis is provided in Figure 1(a-e) and 2, summarizing the effects of the Google Trends search series when adjusted for the monthly trend of an increase in positive cases.

The linear regression model revealed loss of smell (adjusted R^2 of 0.7) with the significant 1-week, 2-week and 3-week– lagged time series, as the best predictor of weekly positive case counts, as further elaborated in Table 3 and Figure 1(a-e). The combined model yielded an excellent performance for forecasting positive cases with adjusted R^2 value of 0.83 as shown in Figure 2 and Table 2.

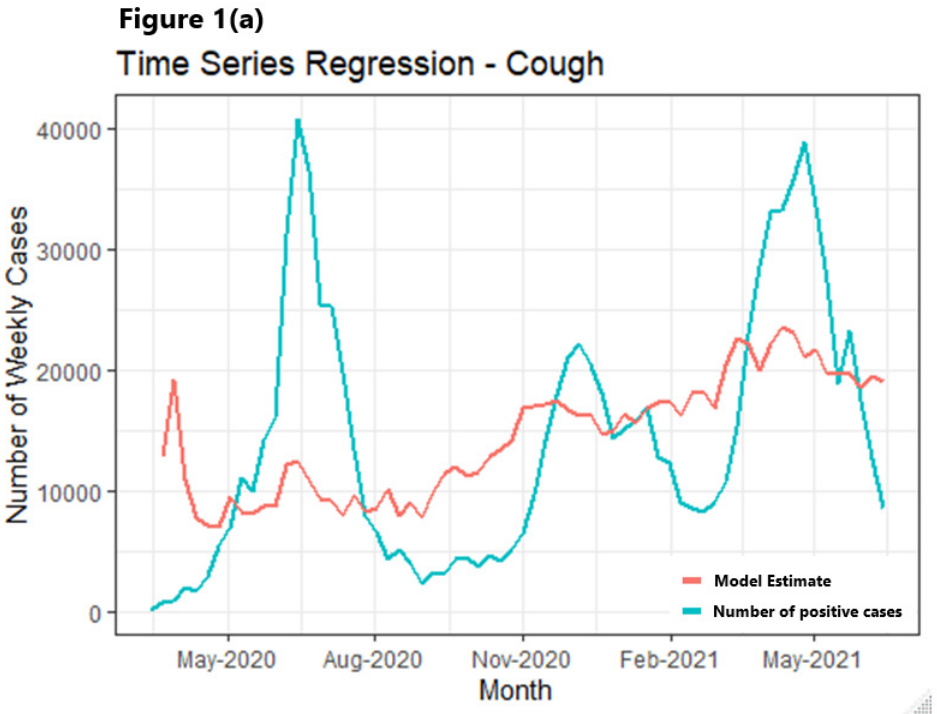
DISCUSSION

The results of our study demonstrate the existence of a statistically significant positive correlation between Google search terms and overall COVID-19 positivity rate in Pakistan. This was especially evident for search terms such as 'fever', 'smell loss', 'taste loss' and 'shortness of breath', with a time lag of 2 weeks, while for 'cough' and

Table 1 Cross-correlation analysis between weekly number of COVID-19 cases in Pakistan with Google Trends scores for suggestive symptoms

Search term	Optimal lag	Correlation	p-value
Fever	-2	0.437	<0.001
Headache	-8	0.349	0.005
Smell loss	-2	0.561	<0.001
Cough	-3	0.260	0.035
Taste loss	-2	0.618	<0.001
Shortness of breath	-2	0.289	0.019
Coronavirus	-3	-0.326	0.008
COVID	-1	0.501	<0.001
COVID-19	-7	-0.353	0.004
Virus	-4	0.322	0.009

Figure 1 (a-e) **Time-series linear regression analysis for weekly number of COVID-19 in Pakistan with Google Trends scores for suggestive symptoms**



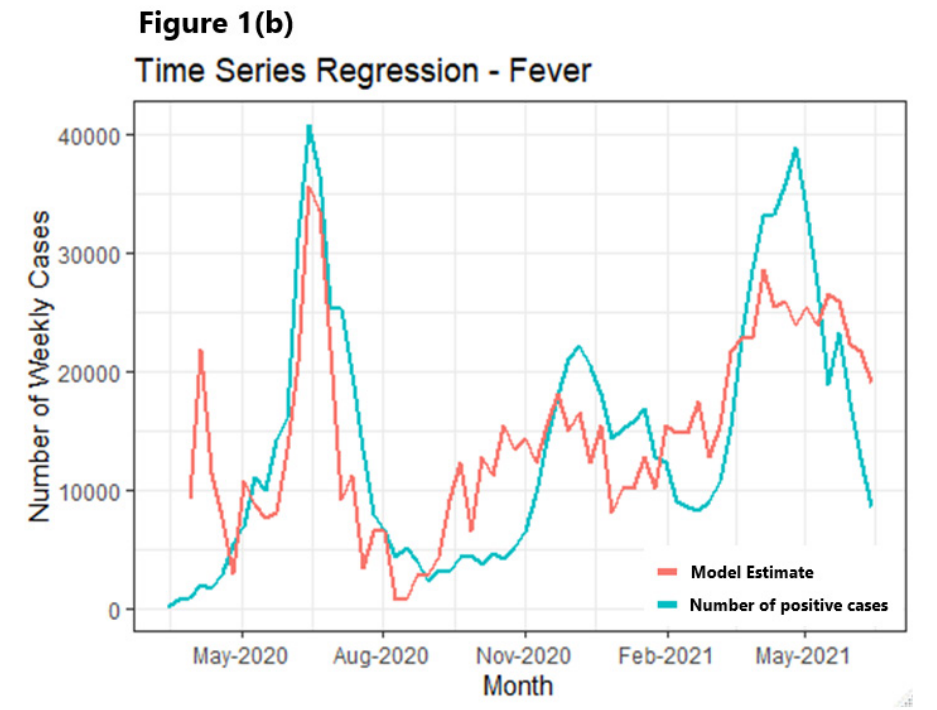


Figure 1(c)

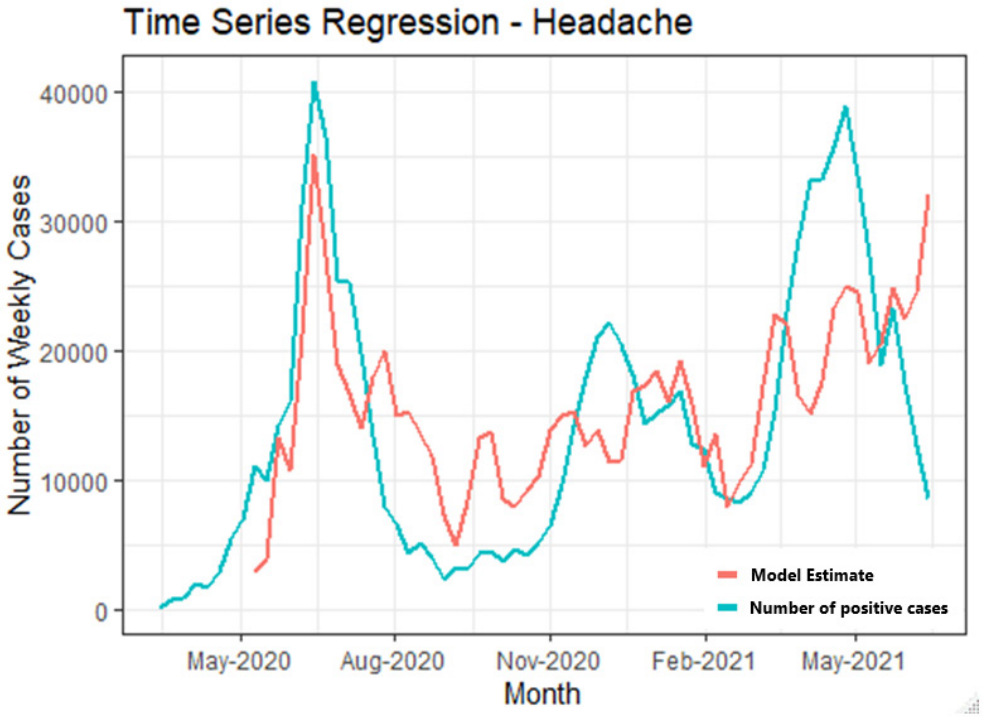


Figure 1(d)

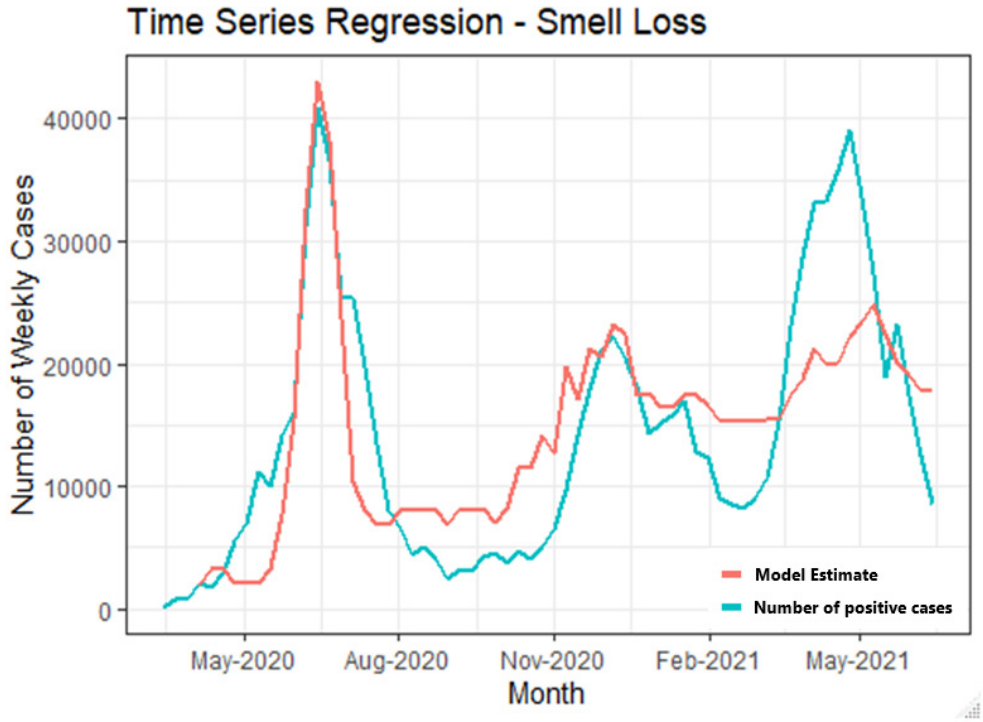


Figure 1(e)

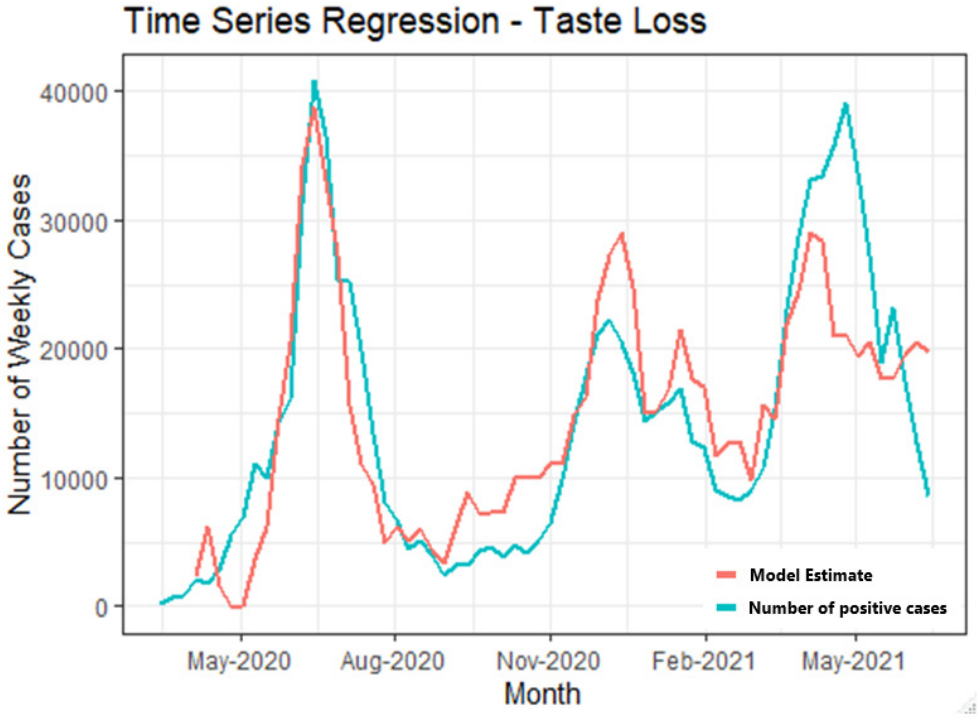


Table 2 Adjusted R^2 based on time-series linear regression analysis for the combined model for weekly number of COVID-19 in Pakistan with Google Trends scores

	Estimate	Std. Error	p-value
Month	798.4	204.1	<0.001
Fever (Lagged -2)	219.1	66.5	0.002
Taste Loss (Lagged -1)	1478.9	480.2	0.003
Taste Loss (Lagged -3)	1717.6	507.2	0.001
Covid (Lagged -1)	237.8	50.0	<0.001
Covid-19 (Lagged -7)	-1725.8	420.5	<0.001
Adjusted R^2 0.83			

Figure 2 Time-series linear regression analysis for the combined model for weekly number of covid-19 in Pakistan with Google Trends

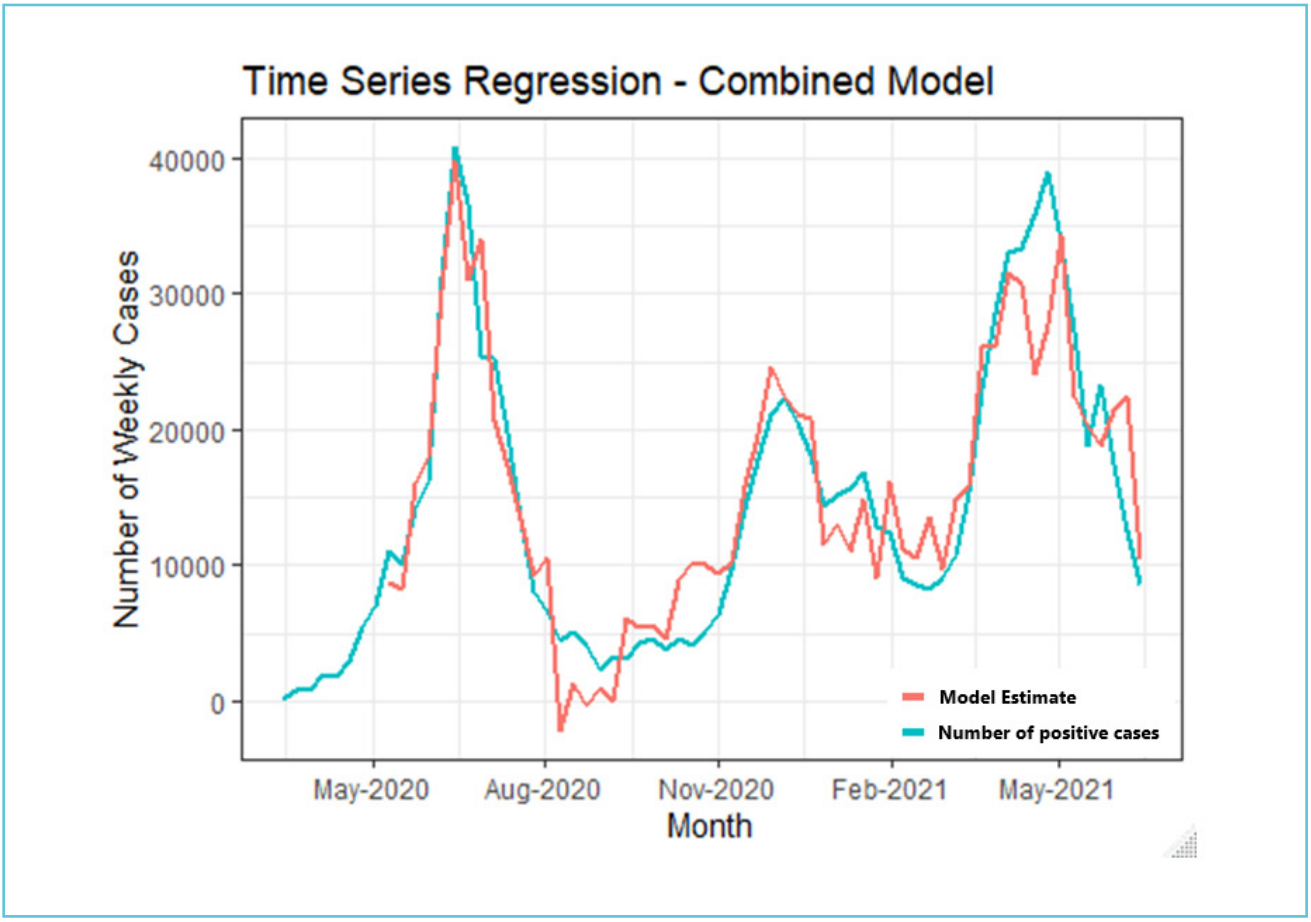


Table 3 Adjusted R^2 based on time-series linear regression analysis for weekly number of covid-19 in Pakistan with Google Trends scores for suggestive symptoms

Symptom	Adjusted R^2
Fever	0.36
Cough	0.44
Headache	0.60
Loss of smell	0.70
Loss of taste	0.70

'coronavirus' the time lag was 3 weeks. This model can hence successfully predict an increase in COVID-19 cases 2-3 weeks ahead of official diagnosis, thus allowing government and healthcare system to adapt and be prepared for the oncoming burden.

Google Trends is a useful tool for forecasting both healthcare and non-healthcare related epidemiological trends. The freely available data can help health authorities to anticipate increases in demands of testing capacity, as well as treatment facility, including availability of hospital beds, oxygen supply, access to ventilators and availability of adequate number of physicians and ancillary staff.

Although this study is a first-of-its-kind based in Pakistan, other countries have successfully used Google Trends to predict changes in the ongoing COVID-19 pandemic outbreak. Cherry et al. studied Google Trends data for 137 regions from 5 different countries, reporting that pathognomonic symptoms such as anosmia and dysgeusia can accurately predict the future incidence patterns of COVID-19 (12). Henry et al. reported similar results in Poland (18). In another study, Lippi et al. also describe significant associations of fever, fatigue and dyspnea with the COVID-19 outbreaks in Italy (19). The same group reported that the correlation between Google searches and COVID-19 cases became stronger with a lag of 2 weeks, as compared to the same week (13). The results of these studies are hence in keeping with the findings of our Pakistan-based analysis. The use of Google Trends in health policy making and management of pandemic could be especially useful for low-middle income countries (LMIC) like Pakistan, where resources are limited and strict and timely management of these resources can help curb the increasing pandemic.

The model we developed could be hence used in other LMIC, to direct resources where most required. Although our study utilized data from the whole country, region specific data can also be used to focus resources to regions which require them the most in near future.

The limitations of our study include the limited use of internet decrease literacy rate in developing countries. Also, the internet use behavior can be influenced by media communications, and possibly serve as a cofounder. Increased knowledge about self-reported symptoms can also decrease the use of internet for searching COVID-related information.

CONCLUSION

Google Trends is an effective tool for forecasting trends of the ongoing COVID-19 pandemic outbreak. We found a high correlation between Google searches for COVID-19 symptoms and diagnoses of SARS-CoV-2 infection, which can be used to direct resources where required or needed. Such data can help government authorities and health policy-making agencies to make well-informed decisions related to imposition of lockdown and provision of resources. Utilizing such data can help developing countries like Pakistan streamline their efforts against the pandemic and possibly prepare of outbreaks before the actually happening to minimize morbidity and mortality as well financial losses that may pursue.

❖❖❖❖❖

Data availability statement

The data that support the findings of this study are openly available by accessing https://trends.google.com/trends/?geo=PK and the links in references [15, 16].

Ethics statements

This analysis was based on electronic searches in unrestricted, publicly available repositories,

so that no informed consent or ethical committee approvals were needed.

List of authors

Dr. Sibtain Ahmed (SA), Assistant Professor
Dr. Muhammad Abbas Abid (MAA), Resident Medical Officer
Maria Helena Santos de Oliveira (MSO), Student, Biostatistics Master's Program
Dr. Zeeshan Ansar Ahmed (ZAA), Assistant Professor
Ayra Siddiqui (AS), MD Student
Dr. Imran Siddiqui (IS), Professor
Dr. Lena Jafri (LJ), Associate Professor
Prof. Giuseppe Lippi (GL)

Author contribution

SA performed the literature search, data analysis and write-up of the work in the first draft. MAA was involved in the write up, literature search and data collection. MSO did the data analysis and prepared graphs and tables. AS assisted in the writing of the first draft. ZAA and IS were involved in the critical revision of the article for the intellectual content. LJ conceived the idea, coordinated the writing of the paper and reviewed the final draft. GL provided supervision of the project, contributed to discussion of the results along with review and amelioration of the draft. All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

❖❖❖❖❖

REFERENCES

1. The United Nations Development Programme. Human Development Report 2020 The Next Frontier: Human Development and the Anthropocene: The United Nations Development Programme, New York, USA; 2020.

2. Khan, NU, Danish, L, Khan, HU, et al. Prevalence of dengue virus serotypes in the 2017 outbreak in Peshawar, KP, Pakistan. J Clin Lab Anal. 2020; 34:e23371. https://doi.org/10.1002/jcla.23371

3. Abbas Z, Abbas M. The cost of eliminating hepatitis C in Pakistan. Lancet Glob Health. 2020 Mar;8(3):e323-e324. doi: 10.1016/S2214-109X(20)30036-X. PMID: 32087161.

4. Ahmed A, Hashmi FK, Khan GM. HIV outbreaks in Pakistan. Lancet HIV. 2019 Jul;6(7):e418. doi: 10.1016/S2352-3018(19)30179-1. Epub 2019 Jun 13. PMID: 31204244.

5. Atif M, Malik I. Why is Pakistan vulnerable to COVID-19 associated morbidity and mortality? A scoping review. Int J Health Plann Manage. 2020;35(5):1041-1054. doi:10.1002/hpm.3016

6. Nishtar S, Boerma T, Amjad S, Alam AY, Khalid F, ul Haq I, et al. Pakistan's health system: performance and prospects after the 18th Constitutional Amendment. Lancet. 2013 Jun 22;381(9884):2193-206. doi: 10.1016/S0140-6736(13)60019-7. Epub 2013 May 17. PMID: 23684254.

7. Coronavirus Testing Survey. American Association for Clinical Chemis try. https://www.aacc.org/science-and-research/covid-19-resources/aacc-covid-19-testing-survey Accessed July 27, 2021.

8. Ginsberg, J., Mohebbi, M., Patel, R. et al. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014 (2009). https://doi.org/10.1038/nature07634

9. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google TrendsTM Analysis. Int J Infect Dis. 2020;95:192-197. doi:10.1016/j.ijid.2020.04.033

10. Chang Y, Chiang W, Wang W, et al Google Trends-based non-English language query data and epidemic diseases: a cross-sectional study of the popular search behaviour in Taiwan BMJ Open 2020;10:e034156. doi: 10.1136/bmjopen-2019-034156.

11. Nazir, Mariam & Soroya, Saira. (2021). Health Informatics: Use of Internet for Health Information Seeking by Pakistani Chronic Patients. Journal of Library Administration. 61. 134-146. 10.1080/01930826.2020.1845552.

12. Cherry G, Rocke J, Chu M, Liu J, Lechner M, Lund V, et al. Loss of smell and taste: a new marker of COVID-19? Tracking reduced sense of smell during the coronavirus pandemic using search trends. Expert Rev Anti Infect Ther. 2020 Nov;18(11):1165-1170. doi: 10.1080/14787210.2020.1792289. Epub 2020 Jul 16. PMID: 32673122; PMCID: PMC7441792.

13. Lippi G, Henry BM, Mattiuzzi C, Sanchis-Gomar F. Google searches for taste and smell loss anticipate Covid-19 epidemiology. medRxiv 2020.

14. Hoang MP, Kanjanaumporn J, Aeumjaturapat S, Chusakul S, Seresirikachorn K, Snidvongs K. Olfactory and

gustatory dysfunctions in COVID-19 patients: A systematic review and meta-analysis. Asian Pac J Allergy Immunol. 2020 Sep;38(3):162-169. doi: 10.12932/AP-210520-0853. PMID: 32563232.

15. Coronavirus Pandemic (COVID-19). OurWorldInData.org. https://ourworldindata.org/coronavirus (Accessed: 15/07/2021).

16. COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination. Governement of Pakistan. https://covid.gov.pk (Accessed: 15/07/2021).

17. Johns Hopkins University & Medicine Coronavirus Resource Center. https://coronavirus.jhu.edu (Accessed: 15/07/2021).

18. Henry BM, Szergyuk I, De Oliveira MHS, Lippi G, Juszczyk G, Mikos M. Utility of Google Trends in anticipating COVID-19 outbreaks in Poland. Polish archives of internal medicine 2021; 131:389-392.

19. Lippi G, Mattiuzzi C, Cervellin G. Google search volume predicts the emergence of COVID-19 outbreaks. Acta Bio Medica: Atenei Parmensis 2020; 91: e2020006.