sciendo
Proceedings of the 5th International Conference on Economics and Social Sciences (2022), ISSN 2704-6524, pp. 845-856

The 5th International Conference on Economics and Social Sciences
Fostering recovery through metaverse business modelling
June 16-17, 2022
Bucharest University of Economic Studies, Romania

On the Potential of Google's "Popular times" Data in Epidemiology

Eduard C. MILEA[1*], Marian NECULA[2], Maria-Magdalena ROȘU[3]

DOI: 10.24789788367405072-079

Abstract

Human interaction and mobility patterns are one of the key factors in modeling and controlling epidemiological outbreaks, seen with the ongoing SARS-COV2 pandemic. Most public health policies, adopted over the past two years, were designed to contain virus transmission by imposing restrictions on human mobility and interactions.

We provide a test case scenario of using crowdsourced data from Google's "Popular times" graph available on Google Maps for various points of interest present there. We posit that the data available in this graph can be used as a proxy for both human mobility and human interactions, and at a more granular level, could be used to evaluate the current epidemiological situation and assess the impact of the ongoing public health restrictions. The data collected during our study period was then plotted as a heatmap overlaid atop OpenStreetMaps for ease of visualization.

The dataset is comprised of a randomly stratified sample of restaurants, bars, and clubs in Bucharest, for which we collected the daily rates "busyness" for each of the points of interest in the sample. For hypothesis testing, we employed several classification algorithms/models to assess Google's "Popular times" statistical strength in predicting current epidemiological status and public health policy effects.

Keywords: COVID-19 transmission, crowd-generated data, epidemiology, human mobility, social interactions.

JEL Classification: C22, C53, I10, I12.

1. Introduction

The COVID-19 pandemic brought forth by SARS-CoV-2 came with many disruptions to our modern life style, and while the pandemic potential of

[1] Bucharest University of Economic Studies, Bucharest, Romania, mileaeduard12@stud.ase.ro.
[2] Bucharest University of Economic Studies, Bucharest, Romania, neculamarian18@stud.ase.ro.
[3] Bucharest University of Economic Studies, Bucharest, Romania, rosumaria13@stud.ase.ro.
* Corresponding author.

coronaviruses coupled with habitat disruptions were discussed by scientists for many years, the world as a whole was still taken by surprise.

The main public health policies adopted by governments around the world, consisted of lockdowns and severe mobility restrictions for various periods across the early phases of the pandemic. These policies were estimated to have saved millions of lives by curbing the viral spread and keeping the hospitals from being overloaded. However, with more knowledge about the virus being obtained, nations moved away from lockdowns to other non-pharmaceutical interventions (NPI), and with vaccines and better treatment protocols becoming available, blanket restrictions were mostly lifted.

In Romania, the initial lockdown and early policies proved to be highly successful, with the moving average 7 days (MA7) cases before the lockdown was lifted on May 14th, 2020, being 250, and death MA7 at around 23. However, as the lockdown was lifted, cases MA7 jumped to ~1200 by the end of August, 2020, with the death MA7 being at around 45, continuing to increase at an accelerated rate once more restrictions were lifted globally and people interacted more.

Previous research, on other diseases with a respiratory component to them, has shown that mobility and human interaction patterns correlate well with disease transmission (Findlater, Bogoch, 2018), thus it was expected that NPIs will help reduce the spread; however, exact causality of NPIs is hard to establish within a population due to the myriad of host intrinsic and extrinsic factors affecting disease transmission, especially respiratory ones.

2. Problem Statement

Existing literature describes the role human social interactions have in viral transmission dynamics, especially for respiratory diseases (Leung, 2021; Buckee, Noor, Sattenspiel, 2021; Cauchemez et al., 2011). Typical models used in epidemiology are based on SIR (Hethcote, 1989), and other derivatives, wherein they typically forecast growth based on certain disease/pathogen characteristics, such as R0 (i.e. how infectious a disease is), how large the susceptible (S) population is, how many people were already infected (I), and how many recovered (R) and assumed to be completely immune afterwards.

These models were developed to be used in small isolated populations with the specific assumptions of a constant population size and homogeneity in interactions between the individuals of said populations.

Our approach, while also focusing on local populations, uses Google's Popular Times data to explain and potentially forecast the local daily cases. Popular Times is a Google Maps feature which reports historical averages for each hour of each day of the week, in addition to this, it also reports, where available, the "live busyness" value of venues. This live value is determined based on Android's built-in location sharing (Google, 2020). Because the data is aggregated on a temporal basis from local devices, it is able to act as a proxy for local social interaction patterns.

3. Research Questions

We know from the existing literature that the incubation period of SARS-CoV-2 fluctuates to some degree between populations (due to intrinsic and extrinsic host factors), as well as between circulating variants, with the median value ranging between 3-7 days after contact until symptom onset, for both Omicron and Delta variants (Del Águila-Mejía et al., 2022; Mefsin et al., 2022; Zhang et al., 2021). In addition to this lag interval, there exists also a human and bureaucratic lag interval.

The human one stems from the time the symptomatic person decides to get tested, be it because they recognized their symptoms as being related to COVID-19, or because they were required to get tested. The other lag stems from how long the test result takes to come back to the person, and be officially counted in the national statistics. This second, bureaucratic, lag interval will fluctuate depending on the local epidemiological situation and the backlog of testing sites and authorities.

Thus, in order to assess the predictive power of our indicator, we first had to approximate the local bureaucratic lag, and use it to estimate approximate interval between a potential exposure event and being registered in the official statistics.

4. Research Methods

4.1 Data Sources

Venue lists

We compiled a list of popular venues in Bucharest from various websites (e.g., TripAdvisor, Metropotam), from which we kept only ones that had any sort of rating or review. The resulting list was then fed through a Python script to gather and append to each of the venues their peak popular hour for each day of the week, while dropping venues which had no available Popular Times data for the full week. Our final list contained a total of 63 venues, which was then fed through a Python script to gather and add to each of the venues their peak popular hour for each day of the week.

Table 1. Venue type distribution per Google's data

Venue type	
Beer	27
Restaurant	25
Bar	10
Clubs	1

Source: Authors' own work.

COVID-19 data

We used the reports made available on the official coronavirus-related government websites "datelazi.ro" and "stirioficiale.ro". We built 2 datasets from these sources, one containing daily SARS-CoV-2 infections in Bucharest, and one containing total daily tests, accounting for our study period, from February 10th, 2022 to February 26th.

Popular Times

Since this data is not publicly exposed in Google's Maps API, nor any other public API made available by Google, we employed two different methods for redundancy.

The first method was to build a Selenium scraper that would run daily and visit each of the venues on the list at their peak hour to scrape the "busyness" value, and two hours after the curfew started. This method, however, requires supervision as Google sometimes detects unusual actions and requires a captcha puzzle to be solved.

The second method involved using a third-party platform (i.e., besttime.app) which makes available via API their services to obtain the required popular times data without the issues of the previous method. The dataset obtained from both methods was congruent.

4.2 Data Analysis

To analyse the correlation between human mobility and new COVID-19 cases, we need to account for the lag between first coming into contact with the virus, and when the case is officially recorded in the public data. This lag will consist of the incubation period and the bureaucratic lag.

In order to approximate what the bureaucratic lag was during February, the month of our study period, we fitted the number of daily tests and cases from Bucharest, to a linear regression, and selected the lag which best fit our data, which was 4 days.

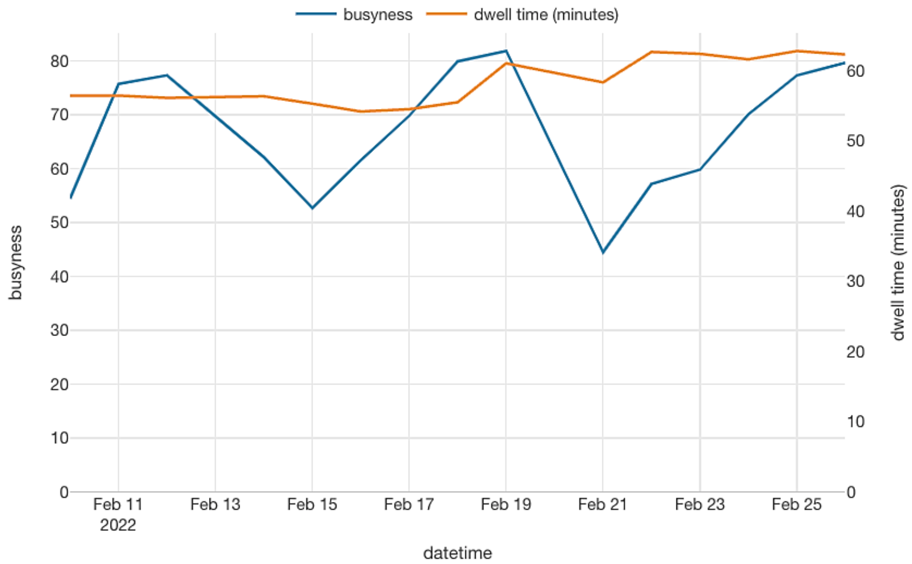
For analysing the correlation between our mobility indicators and new daily cases in Bucharest, we offset the data from our dates by the median incubation period (5 days median of 3-7 range) and the bureaucratic lag (4 days), thus 9 days.

5. Findings

5.1 Busyness and dwell time values

Figure 1 presents the aggregated busyness and dwell time values across our study period. Dwell time represents how much time people spend in a location. The first dip is on Tuesday, February 15th, while the second is one is on Monday, February 21th.

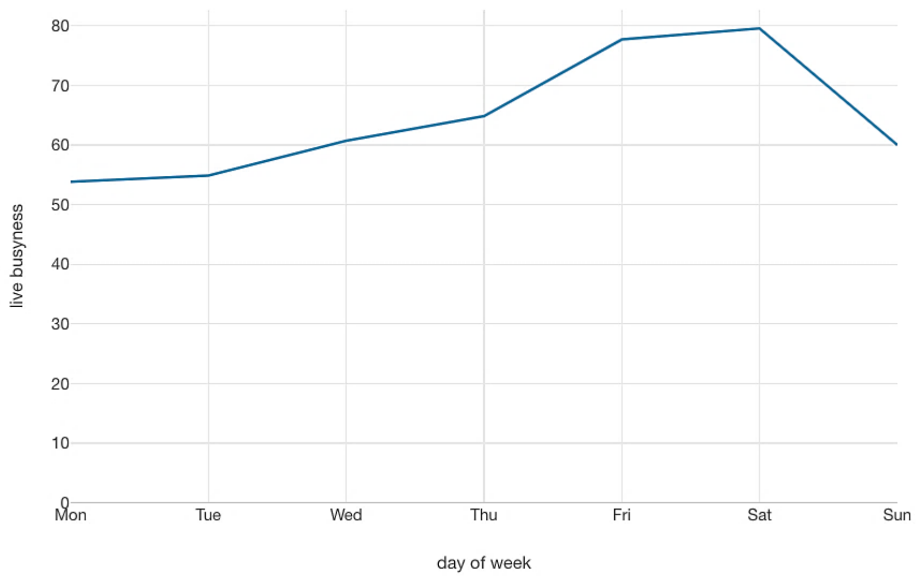
Figure 1. Busyness and dwell times



Source: Authors' own work.

It is reasonable to assume a cyclicity in busyness trends based on days of the week, with Monday being the lowest, barring any significant events such as February 14th which seems to have pushed the busyness higher than on February 21th in.

Figure 2. Busyness across day of the week



Source: Authors' own work.

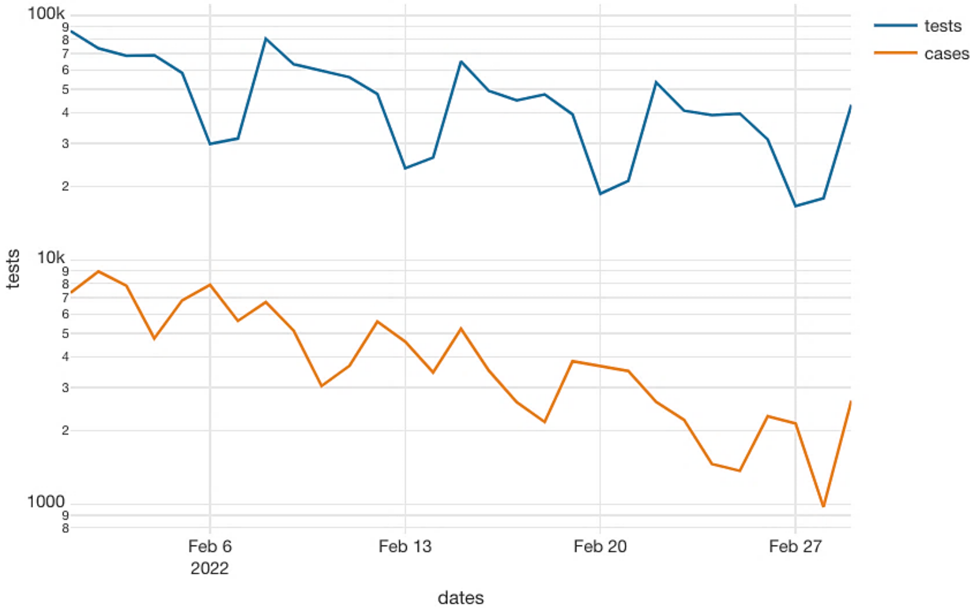
In Figure 2, we look at the aggregated data for days of the week, where we see that Monday is indeed the lowest. As expected, Thursday to Sunday shows a peak, reflecting social interaction patterns.

It is also possible that the increased number of observations in this timeframe is from a number of venues which are not open for the full week.

5.2 Daily tests versus new cases in Bucharest

Figure 3 shows the test-case relationship during the month of our study, whereas Table 2 presents the result of the regression where 4 days offset between test date and the official case count was used. With an adjusted R-squared value of 0.8876, the model has a good fit within the time period analysed and with the offset used.

Figure 3. Tests versus cases



Source: Authors' own work.

Table 2. Test vs cases regression statistics

R Squared	Adj. R Squared	RMSE	F Ratio	P Value
0.915714260	0.887619014	4,061.26678129	32.5932098161	0.0106

Source: Authors' own work.

5.3 Effect of busyness and dwell time on cases

We fitted the busyness value and dwell times on a linear regression versus cases. Table 3 presents the power of the model.

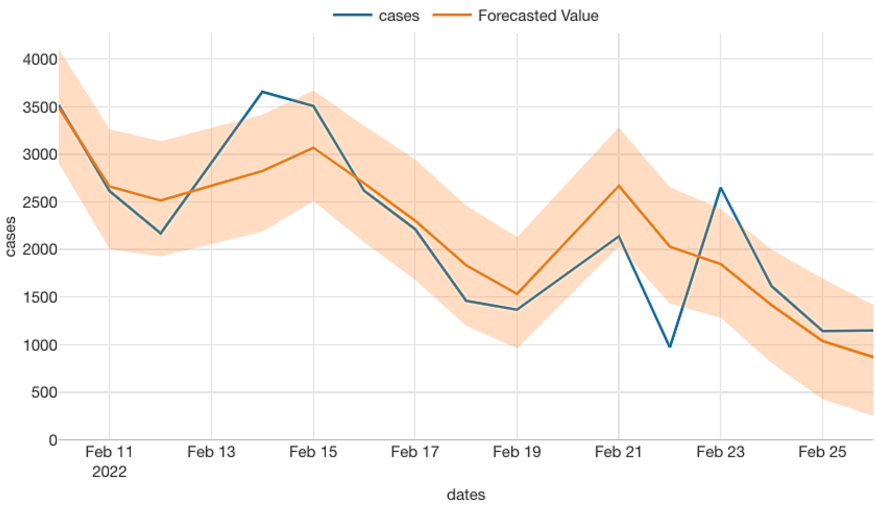
Table 3. Busyness and dwell time vs cases regression statistics

R Squared	Adj. R Squared	RMSE	F Ratio	P Value
0.584231011	0.514936179	562.94886967	8.4310907369	0.0051

Source: Authors' own work

Figure 4 shows how the same variables, when used in Facebook's time series Prophet forecasting procedure, is able to predict accurately the daily cases in Bucharest. Table 4 presents the summary of the model, with a mean absolute percentage error of 20%, the model is considered good.

Figure 4. Prophet forecasting model



Source: Authors' own work

Table 4. Prophet summary

RMSE	MAE	MAPE
473.6198	358.4501	0.2055

Source: Authors' own work.

From an epidemiological standpoint, the more time someone spends in a location, the bigger the increase in risk of acquiring an infection. Thus, if someone spends more time in an already busy location, we expect that fitting busyness x dwell time (named abs_index in our data) as a factor to increase the fit of the model, which is confirmed in Table 5.

Table 5. abs_index regression statistics

R Squared	Adj. R Squared	RMSE	F Ratio	P Value
0.631954504	0.603643312	529.65570998	22.321720127	0.0003

Source: Authors' own work.

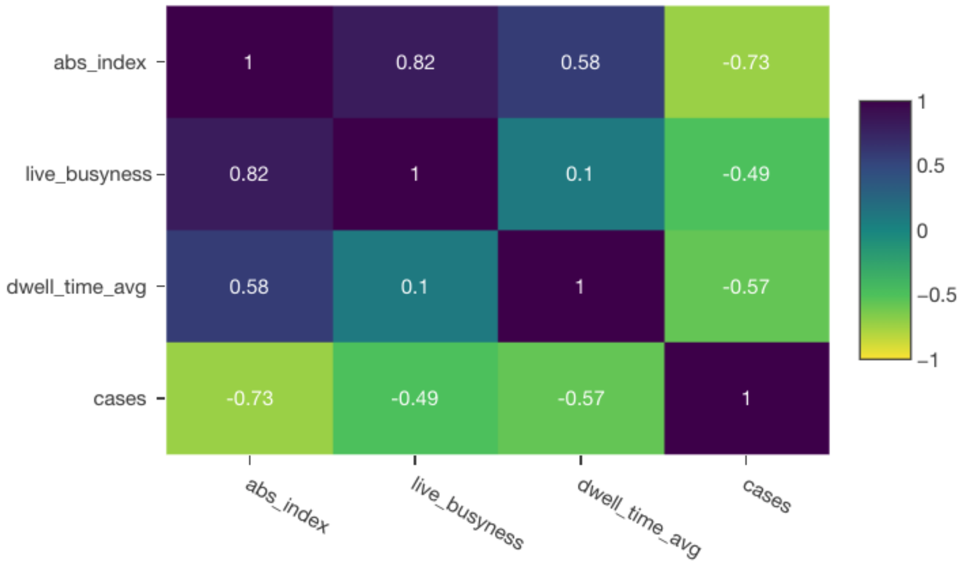
Testing the abs_index, along with the standalone values, using Spearman's rank correlation, we obtain the results presented in Table 6 and Figure 5. These results show how intertwined human behaviour is in cases, as they are in a perpetual feedback loop. As cases go up, people self-correct their behaviour and start avoiding busy places, thus causing venue busyness and dwell times go down. But as they go down and cases peak and start falling, people start to go out more.

Table 6. abs_index correlation statistics

		Correlation	P Value	ρ (rho) Value
abs_index	live_busyness	0.81785714	0.0002971	102
abs_index	dwell_time_avg	0.57857143	0.02636913	236
live_busyness	dwell_time_avg	0.09642857	0.73373753	506
live_busyness	cases	-0.4892857	0.06659231	834
dwell_time_avg	cases	-0.5678571	0.02980447	878
abs_index	cases	-0.7285714	0.00292671	968

Source: Authors' own work.

Figure 5. abs_index correlation graphic

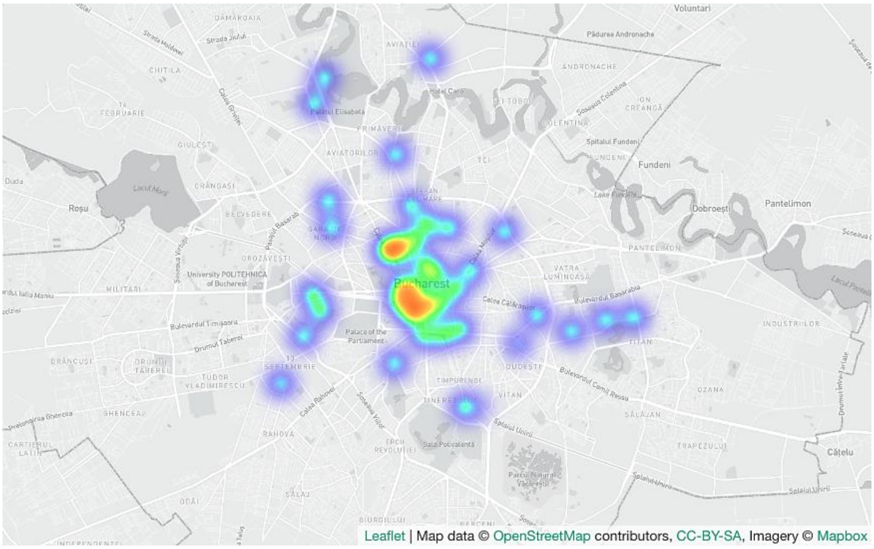


Source: Authors' own work.

5.4 Venue activity before and after curfew

During the period of our study, in Romania several restrictions were in place, one of them being a 10 PM curfew time for public places.

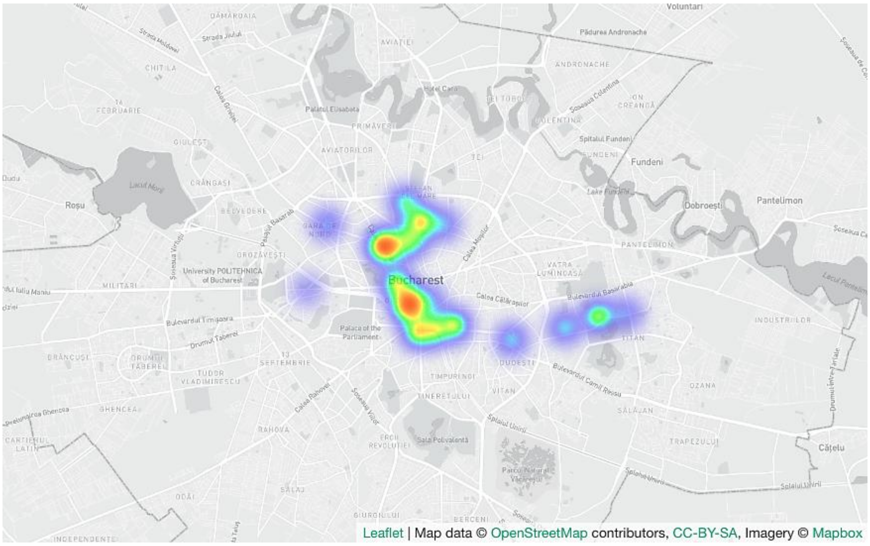
Figure 6a. Activity before curfew



Source: Authors' own work.

As per our collected data, Figure 6a shows the aggregated activity of 63 venues, before the closing time imposed by the 10 PM curfew.

Figure 6b. Activity after curfew



Source: Authors' own work.

After curfew, 23 venues still present activity at 23:59 when the snapshot was taken.

6. Conclusions

Our approach shows the potential to explain, at least in the short term, the evolution of local daily cases. An advantage to this approach is that, in essence, the indicators are based on crowd-generated data, which is publicly available for anyone to scrape and use.

With only 2 indicators, that of venue busyness and time spent by patrons, our model is able to explain the change in cases, and because the indicators dynamically reflect human interaction and mobility patterns, a live forecast model can be deployed to make predictions based on them with decent accuracy and few resources.

An unexpected result of our approach comes from also being able to track adhesion to public health policies, specifically the one imposing a curfew for venues. With a decent number of venues still having live traffic after the curfew hour, it could potentially reflect the mistrust people have in the measures implemented by the authorities, which in a pandemic situation proposes both health and governmental challenges.

Moreover, the two challenges are strongly connected with a governmental crisis leading to inefficacy to manage public health. As such, mistrust in government authority is associated with misperceptions concerning the threat proposed by COVID-19 (Jennings et al., 2021). Extant literature underlines mistrust in authority as the main factor leading to vaccine hesitancy (Hudson, Montelpare, 2021), low engagement in safety behaviours (Blair et al., 2021), and refusing assumed responsibilities such as lockdown measures (Hilhorst, Mena, 2021). As health crises such as COVID-19 surprise in various ways, the public needs to combat the generalized sense of uncertainty with the trust that ministers adequately respond to ambiguous situations (Boin, 2008). In turn, policymakers need to trust the public to implement the measures targeted at reducing the virus spread (Ahern, Loh, 2021). With problems arising in rapid succession, this type of reciprocal trust can only be guaranteed by consistent evidence. Studies suggest that, indeed, controversies should be addressed through thought through crisis communication, responsible journalism, and rumours dispelled (Yu, Lasco, and David, 2021; Jacob et al., 2021). Finally, effective crisis management based on public trust is crucial not only for society as a whole but also for individual well-being and mental comfort with stress, depression, and anxiety steadily increasing across the globe due to the outbreak of COVID-19 (Bawankule et al., 2020).

References

[1] Ahern, S., Loh, E. (2021). Leadership during the COVID-19 pandemic: building and sustaining trust in times of uncertainty, *BMJ Leader*, 5(4), doi:10.1136/leader-2020-000271.

[2] Bawankule, S., Khatib, M.N., Quazi Syed, Z., Gaidhane, A., Telrandhe, S., Godhiwala, P. (2020). Depression, Anxiety and Stress among the general population in the time of COVID-19 lockdown: A cross-sectional study protocol, *International Journal of Research in Pharmaceutical Sciences*, 11, pp. 360-364, doi:10.26452/ijrps.v11iSPL1.2726.

[3] Blair, R.I., Kaur, N.A., Oriaku, I., Levsky, S., Gans, K., Patterson, F., Crowley, J.P., Satti, A. (2021). Is Trust in Medical and Public Authorities and the Media Associated with COVID-19 Related Stress, Perceptions, Safety Behaviors, Inflammatory Biomarkers and Sleep Quality Among Black Adults? In: TP11. TP011 use of health services and medical education research to evaluate the effect of COVID-19, American Thoracic Society International Conference Abstracts, *American Thoracic Society*, pp. A1495-A1495, doi:10.1164/ajrccm-conference.2021.203.1_MeetingAbstracts.A1495.

[4] Boin, A. (2008). *Governing after Crisis: The Politics of Investigation, Accountability and Learning*. Cambridge, Cambridge University Press, doi:10.1017/CBO9780511756122.

[5] Buckee, C., Noor, A., Sattenspiel, L. (2021) Thinking clearly about social aspects of infectious disease transmission, *Nature*, 595(7866), pp. 205-213, doi:10.1038/s41586-021-03694-x.

[6] Cauchemez, S., Bhattarai, A., Marchbanks, T.L., Fagan, R.P., Ostroff, S., Ferguson, N.M., Swerdlow, D., the Pennsylvania H1N1 working group, Sodha, S.V., Moll, M.E., Angulo, F.J., Palekar, R., Archer, W.R., Finelli, L. (2011). Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza, *Proceedings of the National Academy of Sciences*, 108(7), pp. 2825-2830, doi:10.1073/pnas.1008895108.

[7] Del Águila-Mejía, J., Wallmann, R., Calvo-Montes, J., Rodríguez-Lozano, J., Valle-Madrazo, T., Aginagalde-Llorente, A. (2022). Secondary Attack Rate, Transmission and Incubation Periods, and Serial Interval of SARS-CoV-2 Omicron Variant, Spain. *Emerging Infectious Diseases*, 28(6), pp. 1224-1228, doi:10.3201/eid2806.220158.

[8] Findlater, A., Bogoch, I.I. (2018). Human Mobility and the Global Spread of Infectious Diseases: A Focus on Air Travel, *Trends in Parasitology*, 34(9), pp. 772-783, doi:10.1016/j.pt.2018.07.004.

[9] Google (2020). *Behind the scenes: popular times and live busyness information*, 15 October 2020, Google, https://blog.google/products/maps/maps101-popular-times-and-live-busyness-information/, accessed: 7 July 2022.

[10] Hethcote, H.W. (1989). Three Basic Epidemiological Models. In: S.A. Levin, T.G. Hallam, L.J. Gross (eds.). *Applied Mathematical Ecology*. Biomathematics. Berlin, Heidelberg, Springer, pp. 119-144, doi:10.1007/978-3-642-61317-3_5.

[11] Hilhorst, D., Mena, R. (2021). When Covid-19 meets conflict: politics of the pandemic response in fragile and conflict-affected states, *Disasters*, 45(S1), pp. S174-S194, doi:10.1111/disa.12514.

[12] Hudson, A., Montelpare, W.J. (2021). Predictors of Vaccine Hesitancy: Implications for COVID-19 Public Health Messaging, *International Journal of Environmental Research and Public Health*, 18(15), p. 8054, doi:10.3390/ijerph18158054.

[13] Jacob, J., Stephen, S., Issac, A., Krishnan, N., Vadakkethil Radhakrishnan, R., R, V.V., Dhandapani, M., Jose, S., Sm, A., Nair, A.S. (2021). Determinants of Willingness for COVID-19 Vaccine: Implications for Enhancing the Proportion of Vaccination Among Indians, *Cureus*, 13(5), p. e15271, doi:10.7759/cureus.15271.

[14] Jennings, W., Stoker, G., Valgarðsson, V., Devine, D., Gaskell, J. (2021). How trust, mistrust and distrust shape the governance of the COVID-19 crisis, *Journal of European Public Policy*, 28(8), pp. 1174-1196, doi:10.1080/13501763.2021.1942151.

[15] Leung, N.H.L. (2021). Transmissibility and transmission of respiratory viruses, *Nature Reviews Microbiology*, 19(8), pp. 528-545, doi:10.1038/s41579-021-00535-6.

[16] Mefsin, Y., Chen, D., Bond, H.S., Lin, Y., Cheung, J.K., Wong, J.Y., Ali, S.T., Lau, E.H.Y., Wu, P., Leung, G.M., Cowling, B.J. (2022). *Epidemiology of infections with SARS-CoV-2 Omicron BA.2 variant in Hong Kong, January-March 2022*.p.2022.04.07.22273595, doi:10.1101/2022.04.07.22273595.

[17] Yu, V.G., Lasco, G., David, C.C. (2021). Fear, mistrust, and vaccine hesitancy: Narratives of the dengue vaccine controversy in the Philippines, *Vaccine*, 39(35), pp. 4964-4972, doi:10.1016/j.vaccine.2021.07.051.

[18] Zhang, M., Xiao, J., Deng, A., Zhang, Y., Zhuang, Y. et al. (2021). Transmission Dynamics of an Outbreak of the COVID-19 Delta Variant B.1.617.2 – Guangdong Province, China, May-June 2021, *China CDC Weekly*, 3(27), pp. 584-586, doi:10.46234/ccdcw2021.148.