

## COVID-19 VARIANTS OF CONCERNS TRACKING: HOW WE EASED OUT THE WHOLE PROCESS THROUGH OPEN-SOURCE SOFTWARE IN MADHYA PRADESH, INDIA

VEENA SINHA<sup>1</sup>, DIVYA SWAMI<sup>1</sup>, AANCHAL BIJLWAN<sup>1</sup>, SAURAV KUMAR<sup>1</sup>, YOGESH SINGH KAURAV<sup>1\*</sup>,  
VINEET KUMAR TIWARI<sup>1</sup>, SHAILENDRA KUMAR SINGH<sup>1</sup>, SHAIWYA SALAM<sup>1</sup>, VANDANA BHATT<sup>1</sup>,  
ASHISH VERMA<sup>1</sup>, SANJAY GOYAL<sup>2</sup>

<sup>1</sup>Directorate of Health Services, IDSP, DHS, Satpura Bhavan, Bhopal, Madhya Pradesh, India. <sup>2</sup>Revenue, Principal Revenue Commissioner Office, Vallabh Bhavan, Madhya Pradesh, India. Email: yskaurav@gmail.com

Received: 09 April 2022, Revised and Accepted: 15 May 2022

### ABSTRACT

**Objective:** To comprehend the evolution and spread of the severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) virus and also to prevent the future spread of the same, sequencing and analyzing the genomic data of SARS CoV-2 are essential. The objective of the present study is to describe the scope of improvement identified by the state of Madhya Pradesh in the data flow chain and the methodology designed to address the identified shortcomings.

**Methods:** The number of sources of sample data collection was altered as well as a series of Google Sheets were formulated as an open-source tool, to implement an efficient sample data-sharing platform. The application of the proposed tool (Google Sheets as a source of data collection and information sharing) was within the state of Madhya Pradesh, India.

**Result:** After utilizing this mechanism, the state was able to trace more than 80% VOCs and 3341 primary contacts and was also able to communicate this result to all stakeholders without much delay.

**Conclusion:** Based on successful implementation and results, the authors suggest widening the domain of the proposed tool to other states.

**Keywords:** Covid-19, Severe acute respiratory syndrome coronavirus 2, Whole-genome sequencing, Variants of concern, INSACOG.

© 2022 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2022v15i7.44895>. Journal homepage: <https://innovareacademics.in/journals/index.php/ajpcr>

### INTRODUCTION

Coronaviruses (CoVs) are enveloped single-stranded RNA viruses which belong to the order Nidovirales and the family Coronaviridae. CoVs are capable of transmitting/causing diseases in various birds and mammals. Moreover, the occurrence of CoVs in humans is seen since 1960. To the current date, seven human coronaviruses have been identified, the last being the novel virus that was first detected in December 2019 in China, and has been spreading rapidly since then. It is now officially called as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the name of the disease it causes is known as Covid-19. The SARS-CoV-2 is very transmissible and can easily spread from human to human. SARS-CoV-2 was immediately declared an emergency of international concern [1,2].

Globally, the SARS-CoV-2 virus has presented a vast public health threat of this century. However, with the help of effective diagnosis, appropriate treatment measures, and contact tracing, India has been successful in keeping a low mortality rate. With a total of 914,195 cases as of January 26, 2022, Madhya Pradesh has 70,870 active cases with a current positivity rate close to 12%. The cumulative fatality rate in MP is at 1.16% and the cumulative death count is 10,583. MP has currently been contributing an average (past 7 days) of 3.16% cases to the national daily caseload (as of January 26, 2021) [3].

Mutations are common in nature. Similarly, if two viruses infect the same cell simultaneously, coronaviruses can recombine genetically. With the increase in the number of COVID-19 cases, it became important to study the various mutations in the viral genome. Such a study enables us to carry out a comparison between the virus sample and virus lineage. The comparison aids in detecting whether the local outbreaks are a result of a single transmission of the virus or through multiple viral lineages.

While all mutations may not pose a comparatively higher risk, a few viral lineages are identified as a serious threat due to their transmission speeds and ability to cause higher fatality rates. A list of such currently designated variants of concern (VOCs) is presented in Table 1 [4-8].

The early identification of variants of concern is necessary to curb its spread through various restriction measures such as lockdowns, travel bans, and stopping group gatherings. Early identification, however, is possible only through a robust monitoring system. In the context of variants of concerns found in Britain, South Africa, and Brazil, regular monitoring, tracing, and sampling of the foreign passengers were needed and will be needed in the future for unprecedented challenges. The main challenge posed by the virus is continuously presenting itself in the form of new variants. Travelers are required to be traced, tested, and followed up which, in turn, requires an effective yet simple monitoring system.

During the second wave of COVID-19, health professionals, treatment facilities, and other infrastructures were overwhelmed. There were many notifications/letters from the Government of India for whole-genome sequencing and monitoring of foreign travelers was communicated to all the states but states were unable to deploy extra workforce or bear the extra cost of hiring human resources and procuring new infrastructure. Thus, effective management of resources is always the need of the hour. This requirement forms the basis of the need to implement an improved, more effective tool aimed at bringing various labs and offices on a common platform for higher synergy and to enable smooth information flow.

Whole-genome sequencing (WGS) enables the determination of the order of bases in the organism's genome in one process through

laboratory procedure [9]. The genetic material of an organism is termed a genome. A genome is considered an information manual that has all the information regarding making and maintaining the organism. For instance, human genomes constitute double-stranded DNA and have a total of four nucleotide base letters. Human genomes can have a length of over 3 billion base letters. Contrary to the human genomes, the genome of a virus can either contain DNA or RNA and is very small. The latest SARS-CoV-2 virus comprises a single RNA strand that has only 30,000 letters. These letters can be read one by one by the procedure known as sequencing. To confirm that the patient has contracted the latest coronavirus, the sequence of samples from his/her nose or mouth needs to match with the sequence of the coronavirus. Therefore, the symptoms are called of COVID-19 in the patient. The genomes of the virus are capable of mutating constantly, wherein a few letters are changed at a time as the virus divides and spreads. To track the virus spread, sequencing, recording, and analyzing the letters in the genome are used. To keep the vaccine up-to-date, tracking and sequencing of the virus are of utmost priority so that the vaccine is effective for all the strains [10].

Ascertaining the importance of whole-genome sequencing, the Indian SARS-CoV-2 genomics consortium (INSACOG), was initiated jointly by the Union Ministry of Health and Department of Biotechnology (DBT) along with the Council for Scientific and Industrial Research (CSIR) and the Indian Council of Medical Research (ICMR). The initiation was done in association with 38 laboratories to keep track of various variations detected in the SARS-CoV-2 genomes. INSACOG in itself is a multi-agency, multi-laboratory, and pan-India network to track variations in genomes in the SARS-CoV-2 with the aid of sentinel sequencing. This work is facilitated by the National Centre for Disease Control (NCDC), Delhi by involving the central surveillance unit (CSU) under the domain of the Integrated Disease Surveillance Programme (IDSP). CSU does the job of collecting data in a decentralized manner regularly from various states/districts [11,12]. For proper functioning, sentinel sites in each district/state were determined for collecting positive samples (as governed by CSU and SSU).

## MATERIALS AND METHODS

As per the national guidelines, the states were expected to have early identification of samples, prompt effective tracking, and accurate mapping for WGS as per the guidelines. The primary tasks embedded in the state for WGS activities were –

- Identification of district-wise samples as per the guidelines.
- Tracing of those samples by the sentinel labs.
- Transportation of these identified samples to the INSACOG laboratory by the laboratories.
- Collaboration between designated sentinel laboratories, districts, state, and the CSU.
- Early reporting and surveillance activities as per WGS results.
- Formation of a rapid response team in each state/UT by the health department [13].

After the identification of VOCs, the activities assigned to the district surveillance units (DSUs) were

- Contact tracing for the cases where a mutation is detected.
- Epidemiological detailing of the detected mutant in accordance with the number of cases, deaths in the community, etc.
- To detect a change in the severity or mortality using a clinical spectrum of the positive case.
- Collecting and sending samples for WGS to the mapped IGSL of all members in the family in which a variant has been detected. The same procedure is to be carried out for their contacts. This step should be overseen by the rapid response team.
- Taking mandatory containment measures in conjunction with the district administration.
- Reporting daily to NCDC.

To address these issues, the government of Madhya Pradesh has adopted a few strategies to effectively rollout tracking, treatment, and containment of patients affected with VOCs was

- Initially, the instruction from the Government of India was to select 10 sentinel sites in each state. This caused a problem in the complete geographical coverage of the state, particularly those with huge geographical boundaries. In Madhya Pradesh, it was decided by the state to increase the number of sentinel laboratories to 21 for effective coverage of all districts.
- A fixed target of 300 samples was set for each district (15 samples from each sentinel site would be sent on the first and third fortnight). The fixed target could not resolve the complexity of varying populations and positive samples over various districts. Hence, the ratio of positive samples to population per sentinel laboratories was decided clearly in Madhya Pradesh. This led to an effective district-wise distribution of samples.
- To collect the samples and to communicate the results of WGS, effective communication channels were required and there was a need for a single system that can facilitate the communication and coordination between districts, sentinel laboratories, integrated disease surveillance program (IDSP) teams, and bridge the gaps considering timely outcome measurement. Hence, an open-source software – Google Sheets was utilized. Google Sheets is an open-source cloud-based online application that allows working with other people in unison [14,15]. The application helps in creating and formatting spreadsheet documents in cloud space using the internet [16,17]. The Google Sheets applications make the data storage and dashboards creation easier [18-21].

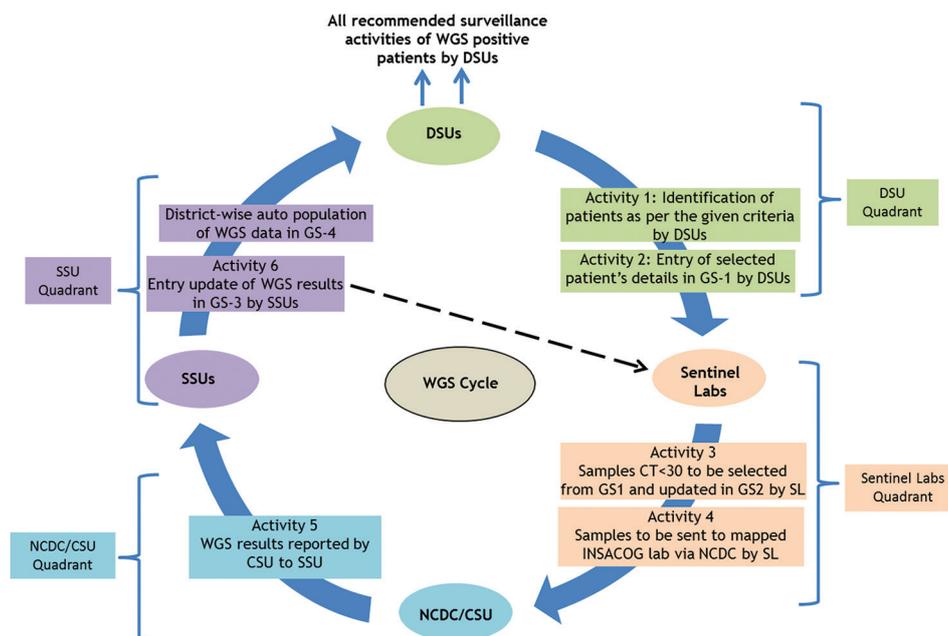
The state streamlined the whole process flow of sample identification to final tracking and containment activities by formulating Google Sheets for the same. Four separate Google Sheets were created for this task. The coordination and flow of data between DSUs, State Surveillance Units (SSUs), the National Centre for Disease Control/Central Surveillance Unit (NCDC/CSU), and the sentinel laboratories are as follows (Fig. 1).

### Google Sheet-1 (district sample line list)

Protected Google Sheet-1 was designed for the district IDSP teams, as shown in Fig. 1. All 52 districts were mapped to 21 sentinel laboratories in this Google Sheets to make them aware of their sentinel laboratories and vice-versa. Districts were tasked to identify the minimum number of samples designated to them twice a month and have to update sample details into Google Sheet-1 in the 1<sup>st</sup> and 3<sup>rd</sup> week of every month including buffer samples (twice the target of sentinel laboratory targets). Sample details contain COVID-19-positive patient basic details, sentinel site name, vaccination status, the reason for sending samples, category of the isolation status, hospitalization status, travel details, and reinfections status. A real-time dashboard was created on the same sheet to identify the gap in data entry and based on that the state IDSP cell and sentinel laboratories can review the details filled by districts and can communicate with them to fill in on time in case the district missed the timeline or mandatory details in Google Sheet-1. Effective communication can be initiated in this regard by coordinators from the state.

### Google Sheet-2 (Sentinel Site Lab Sample sheet)

All 21 sentinel laboratories were roped into Google Sheet-2 providing access to them along with district IDSP teams and others, as shown in Fig. 1. Sentinel laboratories have to choose eligible samples from sample details suggested by district IDSP teams in Google Sheet-1 and also consider the buffer samples in case of samples were found to be ineligible for WGS. Sentinel laboratories have to enter the exact sample count and details which are supposed to be sent to NCDC for whole-genome sequencing in the 2<sup>nd</sup> week and 4<sup>th</sup> week of every month. Details contain positive patient basic details entered by districts, CT value, courier service details, date of dispatch of the sample, etc. Again, with the help of a real-time dashboard, the state and district IDSP teams can review the sample details in case the sentinel laboratory missed the timeline or mandatory details in Google Sheet-2.



**Fig. 1: The WGS Cycle – coordination between DSUs, State Surveillance Units (SSUs), National Centre for Disease Control/Central surveillance unit (NCDC/CSU), and the sentinel laboratories (SLs). (GS-1: Google Sheet-1, GS-2: Google Sheet-2, GS-3: Google sheet-3, and GS-4: Google Sheet-4)**

**Table 1: Currently designated variants of concerns**

| WHO label | Pango lineage | Earliest documented               |
|-----------|---------------|-----------------------------------|
| Alpha     | B.1.1.7       | UK, September 2020                |
| Beta      | B.1.351       | South Africa, May 2020            |
| Gamma     | P.1           | Brazil, November 2020             |
| Delta     | B.1.617.2     | India, October 2020               |
| Omicron   | B.1.1.529     | Multiple countries, November 2021 |

#### Google Sheet-3 (WGS sample report)

The state team has to fetch the result from CSU and enter them into the Google Sheet-3 (WGS report sheet) at regular intervals, as displayed in Fig. 1. This sheet works as a data source of WGS results and VOC contact tracing sheet-3.

#### Google Sheet-4 (VOCs contact tracing)

This result sheet contains the sample results data, district-wise VOCs summary, tracing status of VOCs, and first contact details. Results of whole-genome sequencing get updated into Google Sheet-3 (WGS report sheet) at regular intervals, which work as a data source for WGS results and VOC contact tracing. Segregation of VOCs and WGS results summary can be seen by state IDSP officials, district IDSP officials, and concerned sentinel laboratories in VOC contact tracing sheet-4. States can review the tracing and sampling of primary contacts and initiate corrective actions if required.

#### Human resources/partners involved in implementations

Existing IDSP district teams, state IDSP team, human resources from sentinel laboratories, and coordinators from the NISHTHA project, Jhpiego was involved in WGS and UK passengers tracing activities. Online zoom-based briefing/training on the data flow of Google Sheets was continually provided by the state IDSP cell to all-district teams and sentinel laboratories.

## RESULTS

#### Evidence of effectiveness

Before implementing a robust network of Google Sheet-based communication; there were very little data available to review progress

in the initial tracing activities of VOCs in the initial month of November–December 2020. There were huge delays and issues in communication and coordination with districts, sentinel laboratories, and others.

After utilizing this mechanism, the state was able to trace more than 80% VOCs and 3341 primary contacts and was also able to communicate this result to all stakeholders without much delay. State officials, district officials, and concerned sentinel laboratories can instantly review the summary of VOCs, once the result from NCDC is updated and districts can initiate their tracing without formal communication. The state can present the status of results to the higher authority of the state and IDSP NPO referring to the WGS result summary and sample status.

## DISCUSSION

In the present study, Google Sheets were used as a simple yet effective tool to collect district sample lists, sentinel laboratory lists, and WGS report status as well as variant of concern (VOC) contact tracing. Similarly, a study conducted by Otoom *et al.* (2020) used an IoT-based framework, for monitoring and identifying (or predicting) potential coronaviruses cases, in real time. Equally important, this framework could be used to predict the treatment response of confirmed cases, as well as to better understand the nature of the COVID-19 disease. It consists of five main components: Symptom data collection and uploading, a quarantine/isolation center, a data analysis center, and an interface to health physicians, all of which are interconnected through a Cloud Infrastructure [22].

Google Sheet 1 was filled with sample details containing COVID-19-positive patient basic details, sentinel site name, vaccination status, the reason for sending samples, category of the isolation status, hospitalization status, travel details, and reinfections status. It helped to identify the gap in data entry and based on that the state IDSP cell and sentinel laboratories can review the details filled by districts and can communicate with them to fill them in on time. Likewise, Inbaraj *et al.* (2021) study to estimate the seroprevalence of COVID-19 in Bangalore Rural District of Karnataka, 6 months post the index case found 7 (95% CI 1:4.5–1:9) undetected infected individuals for every RT-PCR confirmed case. The data were collected through a questionnaire containing questions about demographic information (age, gender,

education, comorbidities such as diabetes, hypertension, lung disease, and cancer), history of exposure to COVID-19 infection (history of being diagnosed as a COVID-19 case, interaction, and household contacts with persons with confirmed COVID-19), any history of COVID-19-related symptoms a month before the survey. Infection fatality rate (IFR) was calculated as 12.38 per 10,000 infections as of October 22, 2020. History of self-reported symptoms and education were significantly associated with positive status ( $p < 0.05$ ) [23].

The coordination and flow of data between DSUs, State Surveillance Units (SSUs), the National Centre for Disease Control/Central Surveillance Unit (NCDC/CSU), and the sentinel laboratories were the key to tracking and containing COVID spread. In a study in Colombia, Turner *et al.* (2021) found that intersectoral collaboration has been critical in the response to COVID-19. Substantive and situational facilitators and impediments affected coordination. Policymakers should focus on addressing substantive barriers to coordination, such as pre-existing tensions and mistrust among national and local health-care actors, strict regulations, and limited financial and human resources, while also supporting situational enablers, such as aligning public and private actors' interests, intersectoral government support, and establishing frequent communication channels [24].

## CONCLUSION

Potential of scale: A robust network of linked Google sheets can accommodate millions of records without much significant cost and additional human resources. The system is easy to use and has low maintenance required. Human resources can be trained on data flow within a span of a few days. The system can be used as a tool to coordinate all WGS activities and effective communication with all stakeholders.

Cost: Madhya Pradesh is using open-source software such as Google Sheets and Zoom (free subscription) and no additional requirement of workforce. The state is effectively leveraging all available resources with zero additional cost requirements.

Summary of lessons and challenges: Regular follow-ups are required with districts to enter sample details in Google Sheet-1, and tracing of VOCs in Sheet-4. Regular follow-ups are required with laboratories entering final sample details in Google Sheet-2. Regular data correction and review of the network of linked Google Sheets are needed in case of technical errors and human errors.

## ACKNOWLEDGMENTS

We would like to acknowledge the support from higher state officials, the Additional Chief Secretary of Health, the Commissioner of Health, and the Director of Health. We also wish to acknowledge the immense support from the district's IDSP team of Madhya Pradesh and the development partner Jhpiego.

## AUTHORS' CONTRIBUTIONS

The authors' contributions are as follows: Dr. Veena Sinha: Final approval of the version to be published. Dr. Divya swami: Data interpretation and analysis, paper writing, and formatting. Dr. Aanchal Bijlwan: Paper writing, revision, and formatting. Dr. Saurav Kumar: Paper writing and formatting. Dr. Yogesh Singh Kaurav: Data collection and critical revision of the article. Vineet Kumar Tiwari: Paper writing and formatting. Shailendra Kumar Singh: Paper writing and formatting. Dr. Shaiwya Salam: Paper writing and formatting. Dr. Vandana Bhatt: Paper writing and formatting. Dr. Ashish Verma: Paper writing and formatting. Dr. Sanjay Goyal: Paper writing and formatting.

## CONFLICTS OF INTEREST

None.

## AUTHORS FUNDING

None.

## REFERENCES

- Bchetnia M, Girard C, Duchaine C, Laprise C. The outbreak of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): A review of the current global status. *J Infect Public Health* 2020;13:1601-10. doi:10.1016/j.jiph.2020.07.011, PMID 32778421.
- Kahn JS, McIntosh K. History and recent advances in coronavirus discovery. *Pediatr Infect Dis J* 2005;24:S223-7. doi:10.1097/01.inf.0000188166.17324.60, PMID 16378050.
- Mohfw. Available from: <https://www.mohfw.gov.in> [Last accessed on 2022 Jan 02].
- World Health Organization. Tracking SARS-CoV-2 Variants. Geneva: World Health Organization; 2021. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants> [Last accessed on 2021 Dec 21].
- Centers for Disease Control and Prevention. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/variants/understanding-variants.html> [Last accessed on 2021 Dec 10].
- Thye AY, Law JW, Pusparajah P, Letchumanan V, Chan KG, Lee LH. Emerging SARS-CoV-2 variants of concern (VOCs): An impending global crisis. *Biomedicines* 2021;9:1303. doi:10.3390/biomedicines9101303, PMID 34680420.
- Chakraborty C, Bhattacharya M, Sharma AR. Present variants of concern and variants of interest of severe acute respiratory syndrome coronavirus 2: Their significant mutations in S-glycoprotein, infectivity, immune escape, and vaccine activity. *Rev Med Virol* 2021;32:e2270.
- Parums V. Editorial: Revised World Health Organization terminology for variants of concern and variants of interest of SARS-CoV-2. *Med Sci Monit* 2021;27:e933622. doi:10.12659/MSM.933622, PMID 34149046.
- Gautam SS, Kc R, Leong KW, Mac Aogáin M, O'Toole RF. A step-by-step beginner's protocol for whole genome sequencing of human bacterial pathogens. *J Biol Methods* 2019;6:e110. doi:10.14440/jbm.2019.276
- Coronavirus Explained. Available from: <https://coronavirusexplained.ukri.org/en/article/und0001> [Last accessed on 2021 Dec 24].
- Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, *et al.* Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat Med* 2020;26:1405-10. doi:10.1038/s41591-020-0997-y, PMID 32678356.
- Available from: <https://dbtindia.gov.in/insacog> [Last accessed on 2022 Jan 02].
- Mohfw. Available from: <https://www.mohfw.gov.in/pdf/IndianSARSCoV2PDFGenomicsConsortiumGuidanceDocument.pdf> [Last accessed on 2022 Jan 02].
- Rahmah A, Sukmasetya P, Romadhon MS, Adriansyah AR. Developing Distance Learning Monitoring Dashboard with Google Sheet: An Approach for Flexible and Low-Price Solution In pandemic era. In 2020: International Conference on ICT for Smart Society (ICISS). IEEE Publications; 2020. p. 1-6.
- Google. How to Use Google Sheets. Available from: [https://support.google.com/docs/answer/6000292?hl=en&ref\\_topic=9\\_055343](https://support.google.com/docs/answer/6000292?hl=en&ref_topic=9_055343) [Last accessed on 2021 Dec 18].
- Voelker R, Boman C. Between the Sheets: A Library-wide Inventory with Google; 2017.
- Arnold T, Earl JH, Marshall CD. Using google sheets to determine mortgage information. *The J Wealth Manag* 2017;19:128-31. doi:10.3905/jwm.2017.19.4.128.
- Karl M, Kevin M. Google Sheets v Microsoft Excel: A comparison of the behavior and performance of spreadsheet users. In: *Psychology of Programming Annual Conference*. University of Sussex; 2014.
- Google. Overview: Differences between Sheets and Excel Google; 2020.
- Google. Available from: <https://support.google.com/a/users/answer/9331278?hl=en> [Last accessed on 2021 Sep 19].
- Acquaviva KD, Mugele J, Abadilla N, Adamson T, Bernstein SL, Bhayani RK, *et al.* Documenting social media engagement as scholarship: A new model for assessing academic accomplishment for the health professions. *J Med Internet Res* 2020;22:e25070. doi:10.2196/25070, PMID 33263554.
- Otoom M, Otoom N, Alzubaidi MA, Etoom Y, Banihani R. An IoT-based framework for early identification and monitoring of COVID-19 cases. *Biomed Signal Process Control* 2020;62:102149. doi:10.1016/j.bspc.2020.102149, PMID 32834831
- Inbaraj LR, George CE, Chandrasingh S. Seroprevalence of COVID-19 infection in a rural district of South India: A population-based seroepidemiological study. *PLoS One* 2021;16:e0249247. doi:10.1371/journal.pone.0249247, PMID 33788873.
- Turner S, Ulloa AM, Niño N, Godoy VV. The role of intersectoral action in response to COVID-19: A qualitative study of the roles of academia and the private sector in Colombia. *Int J Health Policy Manag* 2021. Epub ahead of print. doi:10.34172/ijhpm.2021.100, PMID 34523858