Variation in COVID-19 Data Reporting Across India: 6 Months into the Pandemic

Varun Vasudevan[1*], Abeynaya Gnanasekaran[1], Varsha Sankar[2],
Siddarth A. Vasudevan[3] and James Zou[4]

Abstract | India reported its first case of COVID-19 on January 30, 2020. Six months since then, COVID-19 continues to be a growing crisis in India with over 1.6 million reported cases. In this communication, we assess the quality of COVID-19 data reporting done by the state and union territory governments in India between July 12 and July 25, 2020. We compare our findings with those from an earlier assessment conducted in May 2020. We conclude that 6 months into the pandemic, the quality of COVID-19 data reporting across India continues to be highly disparate, which could hinder public health efforts.

REVIEW ARTICLE

1 Introduction

Two key components in containing the COVID-19 pandemic are public awareness and public trust in the government. These components critically depend on timely and accessible dissemination of COVID-19 data by the government[1]. While there are studies showing disparities in personal healthcare access in India, very little was known about the quality of access to public health data across India, especially during the early months of COVID-19 pandemic[2,3]. To address this problem, we developed a semi-quantitative framework to assess the quality of COVID-19 data reporting, and used it to calculate a COVID-19 Data Reporting Score (CDRS) for 29 state and union territory (UT) governments of India[4]. This assessment was done during the 2-week period from May 19 to June 1, 2020. The study showed a strong disparity in the quality of COVID-19 data reporting across India—CDRS varied from 0.61 (good) to 0.0 (poor) across the country, with a median value of 0.26.

In this communication, we present the findings from a second assessment of the quality of COVID-19 data reporting across India. This study was done during the 2-week period from July 12 to July 25, 2020, and includes 35 states[1] and UTs of India. Hereafter, this 2-week period is referred to as the scoring period. Lakshadweep

was excluded from the study as it did not have any COVID-19-positive cases as of July 12, 2020. Hereafter, the first assessment done during May is referred to as study-1 and the second assessment from July is referred to as study-2.

2 Methods

Our scoring framework consists of 45 indicators spanning four key dimensions of public health data reporting—availability, accessibility, granularity, and privacy[4,5]. These indicators capture the presence or absence of a piece of information in the reported data and the format in which it is reported. We would like to emphasize that our framework does not assess the "accuracy of the reported data."

In the availability dimension, we check the availability of basic data such as, daily and cumulative number of confirmed cases, deaths, and recoveries in the state[5]. To assess the accessibility of data, we check for the presence of trend graphics, availability of data in English, and the ease of getting to the web page where data are reported. Trend graphics are important because they make it easier to see patterns in the data. To evaluate the granularity of data, we check whether the state is reporting cumulative data stratified by age, gender, comorbidity, and districts. Granular data helps a layperson connect with the data at a personal level. To assess if a state is ensuring privacy while reporting data, we check if any personally

[1] Institute for Computational and Mathematical Engineering (ICME), Stanford University, 475 Via Ortega, Stanford, CA 94305, USA.
[2] Palo Alto, USA.
[3] Neuchatel, Switzerland.
[4] Department of Biomedical Data Science, Stanford University, Stanford, USA.
*devan@stanford.edu

---

[1] Hereafter, unless specified otherwise, the word state refers to both a state and union territory in India.

J. Indian Inst. Sci. | VOL 100:4 | 885–892 October 2020 | journal.iisc.ernet.in

⌀ Springer

Table 1: CDRS scoring table. Each "Metric-Report Item" pair is an indicator. Overall there are 45 indicators. The scores that an indicator can take are listed in the table. NA denotes not applicable. This table is filled for each state by inspecting the COVID-19 data reported by that state

Dimension	Metric	Report item				
		Confirmed	Deaths	Recovered	Quarantine	ICU
Availability	Total (cumulative)	0, 1	0, 1	0, 1	0, 1	0, 1
	Daily	0, 1	0, 1	0, 1	0, 1	0, 1
	Historical daily data	0, 1	0, 1	0, 1	0, 1	0, 1
Accessibility	Ease of access	0, 1				
	Availability in English	0, 1				
	Total trend graphics	0, 1	0, 1	0, 1	0, 1	0, 1
	Daily trend graphics	0, 1	0, 1	0, 1	0, 1	0, 1
Granularity	Total stratified by age	0, 1	0, 1	0, 1	NA	0, 1
	Total stratified by gender	0, 1	0, 1	0, 1	NA	0, 1
	Total stratified by comorbidity	0, 1	0, 1, 2	0, 1	NA	0, 1
	Total stratified by districts	0, 1	0, 1	0, 1	0, 1	0, 1
Privacy	Compromise in privacy	− 1, 1				

identifiable information of COVID-19 suspects or patients are made publicly available on the state's COVID-19 data reporting page. The report items shown as column headers in Table 1 represent five possible stages in which an individual can find themselves during the pandemic.

Each "Metric-Report Item" pair shown in Table 1 is an indicator. The entries in the table represent the possible scores an indicator can earn[4]. This table is filled for each state during the scoring period by checking the data reported by that state. For example, if a state is reporting total confirmed COVID-19 cases then a score of 1 is assigned to that indicator. The scores recorded in the table are collectively referred to as the scoring data.

Using the scoring data, four categorical scores, one for each dimension, and an overall score is calculated for each state. The categorical scores are obtained by summing the scores earned by the indicators in that dimension. The overall score is the normalized sum of the four categorical scores, and is referred to as the COVID-19 Data Reporting Score (CDRS). For further details on the scoring metrics, scoring process, and score calculation, refer to our article introducing the CDRS framework[4].

3 Results and Discussion

CDRS and the normalized categorical scores for the states in India are tabulated in Table 2. The categorical scores are normalized by the difference of maximum and minimum score possible in that category. The value of CDRS across states indicates a strong disparity in the quality of COVID-19 data reporting in India. The five number summary of CDRS is, min = 0.00, first quartile = 0.20, median = 0.30, third quartile = 0.35, and maximum = 0.63. The geographical disparity in CDRS is evident from the map[2] shown in Fig. 1.

Figure 2 lists states in the decreasing order of CDRS. As seen in the figure, Karnataka is at the top, Bihar and Uttar Pradesh are at the bottom. Bihar and Uttar Pradesh get a CDRS of 0 because they do not release any COVID-19 data on their government or health department website. Figure 2 also shows the incremental change in CDRS from its previous value calculated during study-1 conducted between May 19 and June 1, 2020. As seen in Fig. 2 CDRS has increased in 12 states and decreased in 5 states since the previous study. Figure 3 presents boxplots showing CDRS across India from study-1 and study-2. As seen in the figure the median value has increased slightly from 0.26 to 0.30.

Figure 4 shows the number of states that get a non-zero score on an indicator in our framework. Among the 35 states assessed in this study, 33 states report some data on the COVID-19 situation in the state. Bihar and Uttar Pradesh continue to not publish any data on their government or health department website. The remaining 33

---

[2] The map was generated using Tableau Desktop software version 2020.2.1 and the boundary information for regions in India was obtained as shapefiles from Datameet Org (http://projects.datameet.org/maps/).
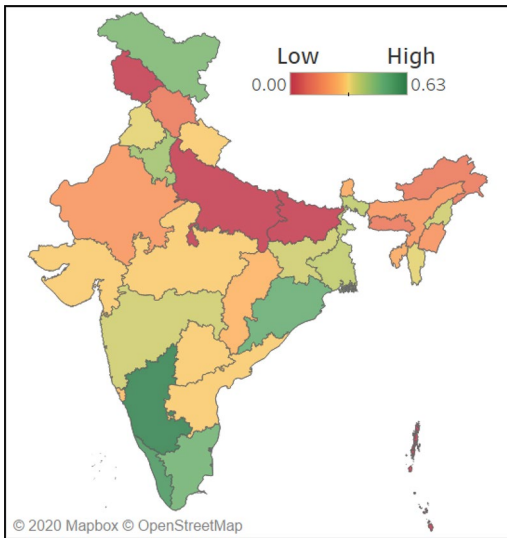
Figure 1: Filled map showing CDRS across India. The map represents the disparity in the quality of COVID-19 data reporting across India. Dark green (red) indicates states that have high (low) quality data reporting.

states report the total deaths and recovered cases, while only 32 of them report the total confirmed cases. Gujarat does not report the total confirmed cases but reports the number of active cases.

CDRS of 12 states have improved in study-2 as compared to study-1. Nine of the 12 states, namely, Andhra Pradesh, Chhattisgarh, Goa, Haryana, Karnataka, Kerala, Ladakh, Uttarakhand, and West Bengal have started reporting more granular data. This is encouraging and is definitely a step in the right direction.

In general, the states continue to score the lowest in the granularity dimension. Jharkhand, which had the highest granularity score in study-1 has stopped reporting age- and gender-stratified data for the total confirmed cases, deaths, and recoveries since June 8, 2020. Hence, its normalized granularity score dropped from 0.50 to 0.17 in this study. It might be worthwhile to investigate what led the Jharkhand government to stop reporting age- and gender-stratified data.

Punjab and Chandigarh compromised the privacy of individuals under quarantine by releasing personally identifiable information on their
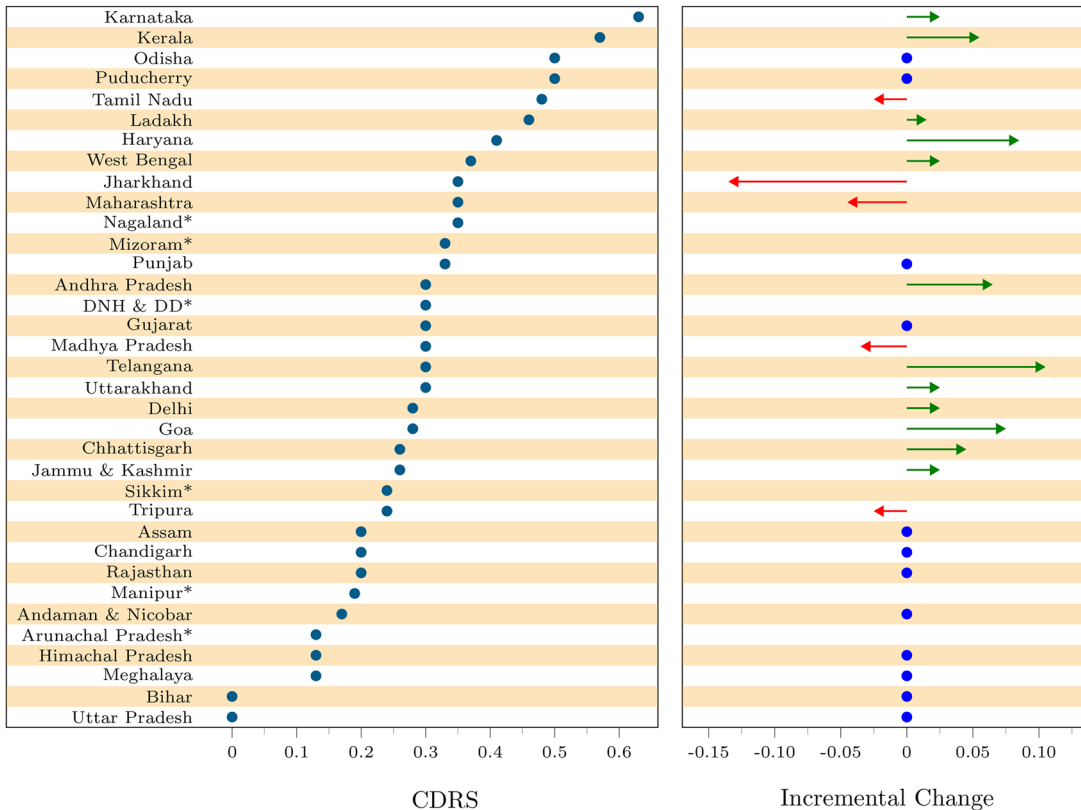

Figure 2: Left: A dot plot showing the spread of CDRS values. States are sorted in the decreasing order of CDRS. Right: The incremental change in CDRS since study-1. Incremental change is not shown for states (marked by an *) that were excluded in study-1.

J. Indian Inst. Sci. |VOL 100:4|885–892 October 2020|journal.iisc.ernet.in

887

Springer

Table 2: CDRS and the normalized categorical scores for the states in India. States are listed in the alphabetical order.

	State / Union Territory	Accessibility score	Availability score	Granularity score	Privacy score	CDRS
1	Andaman and Nicobar Islands	0.17	0.27	0.00	0.50	0.17
2	Andhra Pradesh	0.08	0.60	0.17	0.50	0.30
3	Arunachal Pradesh	0.17	0.20	0.00	0.50	0.13
4	Assam	0.17	0.20	0.17	0.50	0.20
5	Bihar	0.00	0.00	0.00	NA	0.00
6	Chandigarh	0.42	0.27	0.00	-0.50	0.20
7	Chhattisgarh	0.08	0.40	0.22	0.50	0.26
8	Dadra and Nagar Haveli and Daman and Diu	0.17	0.47	0.22	0.50	0.30
9	Delhi	0.17	0.67	0.00	0.50	0.28
10	Goa	0.17	0.60	0.06	0.50	0.28
11	Gujarat	0.08	0.60	0.17	0.50	0.30
12	Haryana	0.42	0.47	0.33	0.50	0.41
13	Himachal Pradesh	0.17	0.20	0.00	0.50	0.13
14	Jammu and Kashmir	0.08	0.47	0.17	0.50	0.26
15	Jharkhand	0.17	0.67	0.17	0.50	0.35
16	Karnataka	0.67	0.73	0.50	0.50	0.63
17	Kerala	0.75	0.67	0.33	0.50	0.57
18	Ladakh	0.42	0.73	0.22	0.50	0.46
19	Madhya Pradesh	0.08	0.60	0.17	0.50	0.30
20	Maharashtra	0.42	0.47	0.17	0.50	0.35
21	Manipur	0.17	0.33	0.00	0.50	0.19
22	Meghalaya	0.17	0.20	0.00	0.50	0.13
23	Mizoram	0.17	0.67	0.07	0.50	0.33
24	Nagaland	0.50	0.40	0.17	0.50	0.35
25	Odisha	0.67	0.60	0.28	0.50	0.50
26	Puducherry	0.67	0.67	0.22	0.50	0.50
27	Punjab	0.17	0.73	0.17	-0.50	0.33
28	Rajasthan	0.17	0.27	0.11	0.50	0.20
29	Sikkim	0.17	0.47	0.00	0.50	0.24
30	Tamil Nadu	0.50	0.60	0.33	0.50	0.48
31	Telangana	0.25	0.67	0.00	0.50	0.30
32	Tripura	0.17	0.27	0.22	0.50	0.24
33	Uttar Pradesh	0.00	0.00	0.00	NA	0.00
34	Uttarakhand	0.17	0.47	0.22	0.50	0.30
35	West Bengal	0.17	0.67	0.22	0.50	0.37

official websites. Chandigarh releases the name and address of people under home quarantine on a daily basis. Punjab released name, age, gender, and mobile number of persons inbound to the state from New Delhi on May 10, 2020[4]. As of July 25, 2020, the document is still present on the Punjab government's health department website.

3.1 Additional Comments

Testing: The strategy recommended by ICMR for COVID-19 testing in India has evolved over time[6–8]. The degree of relevance of testing data in understanding the spread of COVID-19 within a state depends on the testing strategy (e.g., how people are chosen for testing). Therefore, we did
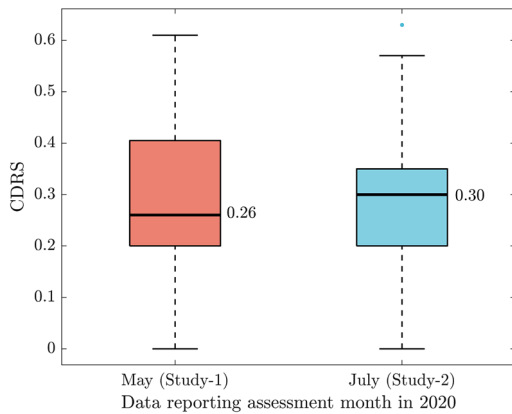
888

🗇 Springer

J. Indian Inst. Sci.|VOL 100:4|885–892 October 2020|journal.iisc.ernet.in

Figure 3: Boxplots showing CDRS across India from the assessments conducted during May (study-1) and July 2020 (study-2). In the boxplot for July the outlier denotes Karnataka.

not include an indicator in our framework to score the reporting of testing data. However, we note that all the states in India report some data on testing. But the reported testing data in most states do not distinguish total samples tested from total persons tested. In other words, most states are reporting total samples tested without specifying how many of them are unique. This is an important limitation to the data that is available to track the testing in a state[9]. For instance, in the case of Tamil Nadu which reports both total samples and total persons tested, the difference between those two numbers is more than a lakh as on August 7, 2020[10].

Age brackets: Karnataka, Odisha, and Tamil Nadu report total number of confirmed cases stratified by age. Karnataka and Kerala report the

total number of deaths stratified by age. However, the number of age brackets used by each of these states is different, making it difficult to compare the age distribution of confirmed and deceased individuals across states. For example, Karnataka, Odisha, and Tamil Nadu use eight, four, and three age brackets, respectively, to report the total number of confirmed cases stratified by age.

Aarogya Setu mobile app: On April 02, 2020, the Indian government launched Aarogya Setu mobile app with the objective of enabling Bluetooth-based contact tracing, mapping of likely hotspots, and dissemination of relevant COVID-19 information[11]. To use the app, one has to register with a mobile number, agree to its data sharing policy, and give it access to Bluetooth and location information. While access to phone number, Bluetooth, and location information might be necessary for contact tracing, we believe that expecting people to provide such information just to access critical COVID-19 data is unreasonable. Therefore, we did not consider data reported through the Aarogya Setu app while scoring the states. However, we would like to mention that the app reports cumulative and daily data for confirmed, deaths, and recoveries, both as text and trend graphics for all states.

Data aggregation platforms: covid19india.org is a volunteer-driven nationwide COVID-19 data aggregation initiative. They collect and report COVID-19 data from across the country. While the initiative is noteworthy, it does not replace the need for high-quality data reporting on official government websites for the following reason. The initiative can fill-in gaps in the accessibility dimension described in our framework. However, they cannot fill-in for the gaps along the availability and

Metric	Confirmed	Deaths	Recovered	Quarantine	ICU Cases
Ease of access	32				
Availability in English	29				
Trend graphics: Total	9	9	10	1	0
Trend graphics: Daily	7	5	5	0	0
Total	32	33	33	16	4
Daily (New)	26	20	24	4	3
Historical data (Date wise)	17	14	16	1	1
Stratified by age	3	2	0		0
Stratfied by gender	4	1	0		0
Stratified by comorbidities	0	8	0		0
Stratified by districts	21	20	22	9	1

Figure 4: Table shows the number of states that get a non-zero score on an indicator. For example, (1) total confirmed is 32 indicating that 32 states report total confirmed COVID-19 cases, (2) availability in English is 29 indicating that 29 states are reporting data in English. Privacy indicator is not shown in this table.

889

J. Indian Inst. Sci. |VOL 100:4|885–892 October 2020|journal.iisc.ernet.in ⚛ Springer

granularity dimensions resulting from the lack of corresponding data released by the government.

4 Conclusion

Our assessment informs the public health efforts in India about the disparity in the quality of COVID-19 data reporting across the country. The available evidence shows that an improvement in the quality of data reporting is required all across India. The disparity in CDRS shows the lack of a unified framework for reporting COVID-19 data in India, and highlights the need for a national agency like Indian Council of Medical Research (ICMR) to monitor or audit the quality of data reporting done by the states. The disparate reporting score also reflects inequality in individual access to public health information and privacy protection based on the state of residence[4].

Overall, there is an urgent need to fill the gaps in COVID-19 data reporting across the states. There has been only a marginal improvement in the quality of COVID-19 data reporting done by the states between May and July. With the pandemic being far from over, it is imperative that the states continue to learn from each other and improve their data reporting. We conclude this communication by quoting the following from the Economic Survey of India, "Given that sophisticated technologies already exist to protect privacy and share confidential information, governments can create data as a public good within the legal framework of data privacy. In the spirit of the Constitution of India, data should be 'of the people, by the people, for the people'."[12]

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Data availability

The curated scoring data used to calculate CDRS is publicly available at https://github.com/varun-vasudevan/CDRS-India/tree/master/study2_july. The states can use the scoring data to identify the limitations in their data reporting and improve upon them.

Appendix

Sources for scoring data		
	State/Union Territory	Data reporting websites
1	Andaman and Nicobar Islands	https://dhs.andaman.gov.in/
2	Andhra Pradesh	http://hmfw.ap.gov.in/covid_19_dailybulletins.aspx http://hmfw.ap.gov.in/covid_dashboard.aspx
3	Arunachal Pradesh	http://covid19.itanagarsmartcity.in/
4	Assam	https://covid19.assam.gov.in/
5	Bihar	No sources
6	Chandigarh	http://chdcovid19.in/
7	Chhattisgarh	http://cghealth.nic.in/ehealth/covid19/pages/index.html
8	Dadra and Nagar Havel iand Daman and Diu	https://dddcovid19.in/
9	Delhi	https://delhifightscorona.in/ http://web.delhi.gov.in/wps/wcm/connect/doit_health/Health/Home/Covid19/Bulletin+July+2020 https://coronabeds.jantasamvad.org/
10	Goa	https://www.goa.gov.in/covid-19/ https://nhm.goa.gov.in/corona-virus-important-links-iec/
11	Gujarat	https://gujcovid19.gujarat.gov.in/
12	Haryana	http://www.nhmharyana.gov.in/page.aspx?id=208 https://gisgmda.maps.arcgis.com/apps/dashboards/5cade394ece3496a9e0c4f168f9536a2
13	Himachal Pradesh	http://www.nrhmhp.gov.in/
14	Jammu and Kashmir	https://www.jkinfonews.com/index.aspx
15	Jharkhand	https://www.jharkhand.gov.in/Home/Covid19Dashboard http://jrhms.jharkhand.gov.in/Press-Release.aspx
16	Karnataka	https://covid19.karnataka.gov.in/english
17	Kerala	https://dashboard.kerala.gov.in/index.php
18	Ladakh	http://covid.ladakh.gov.in/
19	Madhya Pradesh	http://mphealthresponse.nhmmp.gov.in/covid/

890

Springer

J. Indian Inst. Sci.|VOL 100:4|885–892 October 2020|journal.iisc.ernet.in

	State/Union Territory	Data reporting websites
20	Maharashtra	https://www.covid19maharashtragov.in/mh-covid/dashboard https://arogya.maharashtra.gov.in/1175/Novel–Corona-Virus
21	Manipur	http://nrhmmanipur.org/?page_id=621
22	Meghalaya	http://meghalayaonline.gov.in/covid/login.htm
23	Mizoram	https://mcovid19.mizoram.gov.in/ https://health.mizoram.gov.in/posts https://dipr.mizoram.gov.in/posts
24	Nagaland	https://nagahealth.nagaland.gov.in/ https://covid19.nagaland.gov.in/
25	Odisha	https://statedashboard.odisha.gov.in/
26	Puducherry	https://covid19dashboard.py.gov.in/ https://covid19.py.gov.in/
27	Punjab	https://dronamaps.com/corona.html#/ http://pbhealth.gov.in/media-bulletin.htm https://corona.punjab.gov.in/
28	Rajasthan	http://www.rajswasthya.nic.in/
29	Sikkim	https://covid19sikkim.org/
30	Tamil Nadu	https://stopcorona.tn.gov.in/
31	Telangana	https://covid19.telangana.gov.in/
32	Tripura	https://tripura.gov.in/covid-test https://covid19.tripura.gov.in/ https://covid19.tripura.gov.in/Visitor/ViewStatus.aspx
33	Uttar Pradesh	No sources
34	Uttarakhand	http://health.uk.gov.in/pages/view/101-covid19-health-bulletin-for-uttarakhand
35	West Bengal	https://www.wbhealth.gov.in/

References

1. O'Malley P, Rainford J, Thompson A (2009) Transparency during public health emergencies: from rhetoric to reality. Bull World Health Organ 87:614–618

2. Fullman N, Yearwood J, Abay SM, Abbafati C, Abd-Allah F, Abdela J et al (2018) Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016. Lancet 391(10136):2236–2271

3. Barik D, Thorat A (2015) Issues of unequal access to public health in India. Front Public Health 3:245

4. Vasudevan V, Gnanasekaran A, Sankar V, Vasudevan SA, Zou J (2020) Disparity in the quality of COVID-19 data reporting across India. medRxiv (in submission). https://doi.org/10.1101/2020.07.19.20157248. https://www.medrxiv.org/content/early/2020/07/21/2020.07.19.20157248. Accessed 8 Aug 2020

5. WHO, Global surveillance for COVID-19 caused by human infection with COVID-19 virus: interim guidance. https://apps.who.int/iris/rest/bitstreams/1272502/retrieve. Accessed 14 July 2020

6. ICMR, Testing strategy for COVID-19 testing in India (March 09, 2020). https://www.mohfw.gov.in/pdf/ICMRstrategyforCOVID19testinginIndia.pdf. Accessed 08 Aug 2020

7. ICMR, Revised strategy of COVID-19 testing in India (Version 3, March 20, 2020). https://www.icmr.gov.in/pdf/covid/strategy/2020-03-20_covid19_test_v3.pdf. Accessed 08 Aug 2020

8. ICMR, Strategy for COVID-19 testing in India (May 18, 2020). https://www.icmr.gov.in/pdf/covid/strategy/Testing_Strategy_v5_18052020.pdf. Accessed 08 Aug 2020

9. JHU Coronavirus Resource Center, Which U.S. states meet WHO recommended testing criteria? https://coronavirus.jhu.edu/testing/testing-positivity. Accessed 08 Aug 2020

10. Health and Family Welfare Department, Tamil Nadu, Media Bulletin August 08, 2020. https://stopcorona.tn.gov.in/wp-content/uploads/2020/03/Media-Bulletin-07-08-20-COVID-19-6-PM.pdf. Accessed 07 Aug 2020

11. Ministry of Electronics and Information Technology, Aarogya Setu is now open source. https://static.mygov.in/rest/s3fs-public/mygov_159050700051307401.pdf. Accessed 08 August 2020

12. Ministry of Finance (2019) Economic Survey 2018–2019, vol 1. https://www.indiabudget.gov.in/budget2019-20/economicsurvey/index.php. Accessed 25 June 2020

891

J. Indian Inst. Sci. |VOL 100:4|885–892 October 2020|journal.iisc.ernet.in

Springer

Varun Vasudevan is a Ph.D. candidate in the Institute for Computational and Mathematical Engineering at Stanford University. Before joining Stanford, he obtained a Masters in Electrical and Computer Engineering from Purdue University. His research focuses on applications of machine learning to Radiation Treatment Planning and Medical Diagnosis with emphasis on explainability, computational efficiency and low-cost solutions.

Abeynaya Gnanasekaran is a Ph.D. candidate in the Institute for Computational and Mathematical Engineering at Stanford University. Her research interests are in Numerical Linear Algebra and High Performance Computing. Before joining Stanford, she obtained her Bachelors in Chemical Engineering from Indian Institute of Technology Madras, India.
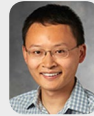
Varsha Sankar obtained a Master's in Electrical Engineering from Stanford University and is currently a Data Scientist in the industry. Before that, she completed Bachelors in Electronics and Communication Engineering from Anna University, Chennai, India.

Siddarth A. Vasudevan obtained a Ph.D. in Material Science from ETH Zürich and is currently working as a Postdoctoral Fellow in the industry. Prior to Ph.D., he obtained a Master's and Bachelors in Chemical Engineering from Delft University of Technology and National Institute of Technology, Karnataka, respectively.

James Zou is an Assistant Professor of Biomedical Data Science and, by courtesy, of Computer Science and Electrical Engineering at Stanford University. He received his Ph.D. from Harvard University in 2014.

892

Springer

J. Indian Inst. Sci.|VOL 100:4|885–892 October 2020|journal.iisc.ernet.in