

# Assessment of COVID-19 data reporting in 100+ websites and apps in India

Varun Vasudevan<sup>1¶</sup>, Abeynaya Gnanasekaran<sup>1¶</sup>, Bhavik Bansal<sup>2</sup>, Chandrakant Lahariya<sup>3</sup>,

Giridara Gopal Parameswaran<sup>4</sup>, James Zou<sup>5\*</sup>

<sup>1</sup>Institute for Computational & Mathematical Engineering, Stanford University, California, United States of America.

<sup>2</sup>All India Institute of Medical Sciences, New Delhi, India

<sup>3</sup>Public Policy and Health Systems Specialist, New Delhi, India

<sup>4</sup>Center for Disease Dynamics, Economics & Policy, New Delhi, India

<sup>5</sup>Department of Biomedical Data Science, Stanford University School of Medicine, California, United States of America

\*Corresponding author

Email: jamesz@stanford.edu

¶These authors contributed equally to this work.

# Abstract

**Introduction:** India is among the top three countries in the world both in COVID-19 case and death counts. With the pandemic far from over, timely, transparent, and accessible reporting of COVID-19 data continues to be critical for India's pandemic efforts.

**Methods:** We systematically analyze the quality of reporting of COVID-19 data in over one hundred government platforms (web and mobile) from India.

**Results:** Our analyses reveal a lack of granular data in the reporting of COVID-19 surveillance, vaccination, and vacant bed availability. As of 5 June 2021, age and gender distribution are available for less than 22% of cases and deaths, and comorbidity distribution is available for less than 30% of deaths. Amid rising concerns of undercounting cases and deaths in India, our results highlight a patchy reporting of granular data even among the reported cases and deaths. Furthermore, total vaccination stratified by healthcare workers, frontline workers, and age brackets is reported by only 14 out of India's 36 subnationals (states and union territories). There is no reporting of adverse events following immunization by vaccine and event type.

**Discussion:** By showing what, where, and how much data is missing, we highlight the need for a more responsible and transparent reporting of granular COVID-19 data in India.

## Introduction

How many men and women have died due to COVID-19 in each state in India? What adverse events have been reported following vaccination? Which states are reporting the number of vacant oxygen beds in their hospitals? Such questions have strong public health implications. Is data reporting from the national and subnational governments in India *granular* enough to answer such questions? We answer that in this paper by documenting and analyzing the reporting of surveillance data,[1] vaccination monitoring data,[2] and bed availability during the second wave of COVID-19, focusing on granular information. Age-gender distribution for cases and deaths, adverse events following immunization stratified by vaccine and event type, and number of vacant oxygen beds are few

examples of such granular information. Reporting granular COVID-19 data is important as it enables public health personnel to track disease spread, vaccination, and adverse events across different sub-populations.[2,3] Granular data is also more transparent and informative for the general public.

Assessment of *reporting data quality* is timely and important for the following reasons.

- Several articles continue to mention the lack of surveillance data from India.[3–5] Therefore, it is necessary to understand and document what, where, and how much data is missing.
- Assessing the reporting of vaccination monitoring data informs how India fares now and the improvements necessary to overcome future vaccine hesitancy challenges.
- The second COVID-19 wave is significantly larger than the first and led to a severe shortage of resources like oxygen beds.[6] Therefore, it is important to know if the surveillance reporting adapted to the worsening pandemic[7,8] and if there was reporting on the resources that were in shortage.

Recent studies have highlighted that the official reports in India could be undercounting the true number of COVID-19 cases and death, which raises substantial public health challenges.[3,9,10] This work focuses on the complementary question of among the data that is reported, whether useful granularity is provided.

## Methods

Between 22 May and 5 June 2021, we assessed digital platforms (web and mobile) hosted by the national and subnational (state and union territory) governments for reporting data on COVID-19 surveillance, vaccination monitoring, and bed availability. We first checked the MyGov mobile app and CoWIN dashboard hosted by the Indian national government to report surveillance and vaccination monitoring data.[11,12] For each subnational, we then checked their government and health department websites/apps and performed a google search. Overall, we assessed more than 100 digital platforms. See S1 Text for details.

**Surveillance reporting.** Using the indicators developed in [7], we check for availability, accessibility, granularity, and privacy violation in the reporting of confirmed, deceased, recovered, quarantine, and critical/ICU (intensive care unit) COVID-19 cases. Availability indicators check for total, daily, and historical data; accessibility indicators check for ease of access and reporting in English; granularity indicators check for total data stratified by age, gender, comorbidity, and districts; and privacy indicator checks if privacy is violated by including personally identifiable information in the reporting. Overall there are 45 indicators used in our assessment framework. We calculate two scores for each subnational as described in [7]. One, a surveillance reporting score (referred to as COVID-19 data reporting score in [7]) using all the above indicators, and two, a *granular* surveillance reporting score (referred to as granularity score in [7]) using just the granularity indicators. Both scores range between 0 (low) and 1 (high).

To get a handle on the scale of missing granular data, we narrow our focus on the reporting of age and gender for confirmed cases; and age, gender, and comorbidity for deaths. Among subnationals reporting these items, some disaggregate the cumulative numbers by the items; some disaggregate the daily numbers by the items, and the remaining report the items for each individual. Considering all subnationals that report one or more of age, gender, and comorbidity, in any of the three forms mentioned above, we calculate the percentage of cases and deaths for which age, gender, and comorbidity distribution is available as of 5 June, 2021.

**Vaccination reporting.** Disaggregated monitoring of vaccination is essential to measure the progress and effectiveness of India's vaccination campaign.[2] We developed a minimal set of indicators to assess the reporting quality of vaccination monitoring data. The indicators reflect recommendations from WHO and LANCET COVID-19 Commission India Task Force,[2,13] and the vaccine operational guidelines from the Ministry of Health and Family Welfare (MoHFW) of India.[14]

Indicators are grouped into three dimensions and are as follows. *Availability*: Daily and total vaccination. *Accessibility*: Daily vaccination trend graphic. *Granularity*: 1) Total vaccination stratified by districts and eligibility category (health care workers, front line workers, age 45+, age 18–44). 2) Total AEFI (adverse events following immunization) stratified by vaccine type (Covishield, Covaxin, Sputnik V); and event type (severe, serious).

For all indicators, except the AEFI ones, we check if data is reported separately for each dose (first and second). MoHFW classifies AEFI into three types: *minor* (e.g., pain and swelling at the injection site, fever), *severe* (e.g., non-hospitalized cases of anaphylaxis, sepsis), and *serious* (e.g., deaths, hospitalizations).[14] We check for reporting on severe and serious events. Trend graphics are used as an indicator because they are concise and make it easier to identify patterns. Eligibility category refers to the order of eligibility in which vaccines were rolled out in India.

**Vacant bed availability reporting.** When resources such as oxygen beds are in shortage,[6] it is important to report their vacancy to reduce panic among people in need. Therefore, we checked if subnationals report the number of vacant ICU/oxygen/ventilator beds disaggregated by districts/hospitals.

## Results

**Surveillance reporting.** The geographical variation in surveillance reporting scores is shown in Fig 1A. See S2 Table for each subnational's score. The median score is 0.46. MyGov provides seamless access to total and daily numbers and trend graphics for confirmed, recovered, and deaths for each subnational.[11] However, granular information such as cumulative numbers stratified by districts, age, gender, or comorbidity, is unavailable on MyGov, as shown in Fig 1B.

**Fig 1: (A)** Map showing the variation in surveillance reporting score across India. 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 (<http://projects.datameet.org/maps/>).

**(B)** Table indicating what surveillance data is being reported (or not) for each subnational on the MyGov app. **(C)** Table indicating the vaccination data reported (or not) for each subnational on the CoWIN dashboard. **(D)** Subnationals sorted in the decreasing order of granular surveillance reporting score from the current assessment (June 2021). The scores from previous assessments (2020) are shown for comparison.[7,8] The table also shows which subnationals are reporting (or not) vaccination coverage stratified by eligibility category; AEFI stratified by vaccine and event type; and vacant ICU/oxygen/ventilator bed availability disaggregated by districts/hospitals.

Fig 1D lists subnationals in the decreasing order of their *granular* surveillance reporting score.

Scores from previous assessments are shown for comparison. The northeastern state of Nagaland scored highest by reporting granular data through weekly bulletins. They report cumulative cases and deaths disaggregated by age and gender and cumulative deaths disaggregated by comorbidities, as shown in Fig 2A. In contrast, the lowest scoring subnationals report little or no granular data. As of 5 June 2021, age and gender distribution are available for less than 22% of cases and deaths, and comorbidity distribution is available for less than 30% of deaths. See Fig 2B, Fig 2C, and S3 Text for details. Even now, some subnationals do not report data stratified by districts. See the dataset released with this paper to know what each subnational is reporting.

**Fig 2: (A)** Age, gender, and comorbidity data for deaths provided in the weekly bulletin of Nagaland government on 5 June 2021 as examples of high-quality granular surveillance reporting. **(B)** Shows the % of cases for which age and gender distribution are available. Each subnational reporting that data is represented by a colored rectangle whose width denotes the % of total cases in India reported in that subnational. **(C)** Shows the % of deaths for which age, gender, and comorbidity data distribution are available. Each subnational reporting that data is represented by a colored rectangle whose width denotes the % of total deaths in India reported in that subnational. **(D)** Table showing the variation in the format of reporting of age, gender, and comorbidity among the subnationals that are reporting those data for the deaths. **(E)** Subnationals that stopped reporting total data stratified

by age, gender, comorbidity or district after either of the surveillance reporting assessments conducted in 2020.[7,8] \*See S4 Text for more details.

**Privacy violations.** Chandigarh and Haryana are violating privacy by including individually identifiable information in their reporting. Chandigarh continues to release a document ([http://chandigarh.gov.in/health\\_covid19.htm](http://chandigarh.gov.in/health_covid19.htm)) containing the name and address of people who have completed/under quarantine. The document is over a thousand pages with over 100,000 entries. Haryana is releasing a document (<https://haraadesh.nic.in/>) containing the name, age, gender, and address of cases from the Jhajjar district (see S6 Fig).

**Vaccination reporting.** CoWIN, launched in 2021, is a cloud-based information technology solution for planning, implementing, monitoring, and evaluating COVID-19 vaccination.[14] CoWIN dashboard reports the following for each subnational, district, and dose. Daily and total vaccination numbers, and daily vaccination trend graphics. CoWIN does not report total vaccination stratified by eligibility category for each dose. For AEFI, CoWIN reports daily AEFI numbers and the cumulative percentage. The number of severe and serious events disaggregated by vaccine type is missing (Fig 1C). Only 14 out of 36 subnationals shown in Fig 1D report on their digital platforms the total vaccination stratified by eligibility category for each dose. Karnataka is the only subnational that is reporting the number of severe and serious AEFI cases. AEFI reporting stratified by vaccine type is absent on all subnational platforms.

**Vacant bed availability reporting.** 20 out of 36 subnationals listed in Fig 1D report them by hospitals and frequently update them. It is a commendable effort to ensure the effective utilization of resources. Other subnationals are either not publishing any data on vacant bed availability or are reporting the total/vacant number of beds without classifying them. We encourage these subnationals to be more granular in reporting.

# Discussion

This is the largest study of its kind to assess the quality of COVID-19 data reporting in India. We did a comprehensive assessment of 100+ national and subnational government digital platforms (web and mobile) to identify what is present and what is missing in the reporting of surveillance data, bed availability, and vaccination monitoring data.

Overall, the quality of surveillance reporting has improved since 2020. Median surveillance reporting score has increased from 0.26 in May 2020 [7] and 0.30 in July 2020 [8]. This increase is primarily due to the consistent availability of high-level surveillance data through MyGov. However, the reporting of granular information such as age, gender, and comorbidity continues to be poor.

Age and gender distribution is available for about 1 in 5 cases and deaths in India. Similarly, comorbidity distribution is available for about 1 in 3 deaths. That is a staggeringly low number for a country with more than 344 thousand deaths. Essentially, we do not know even the basic information about who is getting infected and who is dying. The implications can be dire as it prohibits researchers from tracking age-gender specific trends, identifying high-risk subgroups, and validating hypotheses on infection fatality rates.[3,5] Going further, disaggregating cases and deaths by vaccination status (fully/partially/not vaccinated) is also essential to estimate the vaccine effectiveness in various sub-groups. Maharashtra, the state with the most deaths, does not report the age, gender, and comorbidity distribution. Even among subnationals reporting granular data for deaths, differences in the format of reporting make it difficult for comparison (Fig 2D).

A few subnationals have discontinued reporting certain granular items since the assessments in 2020 (Fig 2E). We highlight three specific instances. First, Karnataka, the state with the best surveillance reporting in 2020,[7,8] is no longer publishing war-room bulletins that had age and gender data for cases. Second, Kerala has stopped reporting comorbidity for deaths. There are claims that Kerala is undercounting deaths by attributing a portion of them as death due to



comorbidity.[15,16] The removal of comorbidity data could strengthen such claims. Third, Jharkhand, a model state for granular reporting in the initial months, stopped reporting age and gender data as the pandemic worsened. It is important to scrutinize these changes in reporting to understand the bottlenecks or motives that led to the changes.

On the one side, there is inadequate reporting of essential granular data like age and gender distribution. On the other side, even personally identifiable information is being published by subnationals like Chandigarh and Haryana. The public health benefits of the personally identifiable information released by these subnationals are unclear. Data reported by the government should include only the information necessary for public health activities.[17] Reporting personal data can discourage people from cooperating with the government or lead to discrimination against specific people.[18]

The quality of surveillance reporting in India has been analyzed extensively in three studies, including the current one. The first two studies were during the first wave of COVID-19 (roughly 3 and 6 months into the pandemic),[7,8], and the current study was during the second wave (after 15 months). Two crucial lessons that we learned collectively from these assessments are as follows. First, subnational governments are unlikely to make much progress in granular surveillance reporting without an official guideline from the central government on what data they have to report publicly. In fact, without someone to hold the subnational governments accountable, they can even switch from good to poor reporting practices (e.g., Karnataka and Jharkhand). Second, while official documents from the government, including a recent white paper from NITI Aayog (Vision 2035: Public Health Surveillance in India),[19] embrace the importance of privacy, there is an evident lack of awareness about privacy among officials releasing surveillance data.

We make three comments about the reporting of vaccination monitoring data. First, through the CoWIN dashboard, anyone can access vaccination coverage data for all subnationals and districts. It is a remarkable feat for such a large country. Second, governments should at least report

vaccination coverage disaggregated by eligibility category. In the coming months, more disaggregated reporting based on gender, pre-existing conditions (comorbidities, pregnancy), socioeconomic, rural-urban, and other equity factors are necessary to ensure no sub-groups are left behind.[2] Third, there is an urgent need for reporting AEFI by vaccine type, sub-population affected, gender, and severity. Detailed and transparent AEFI data can increase citizens' confidence in vaccines, especially as these vaccines are still in the emergency use authorization phase.[20] A large part of the success of polio elimination in India can be credited to disaggregated program data and a robust AEFI reporting system.

This study has several limitations that would be interesting to address in follow up research. One main analysis limitation is that we can not evaluate the effect of data reporting quality on the containment of the virus. Therefore, our results should not be interpreted as "good reporting means good control of the pandemic." While transparent and timely reporting of data is necessary, it is not sufficient. As discussed in the introduction, there could be a substantial number of COVID-19 cases and deaths that are under-reported, and we do not quantify these in our analysis. Our assessment is restricted to national and subnational (state and union territory) platforms and does not include district platforms.

## Conclusions

By not reporting granular details, governments are making a choice to make certain information invisible to the scientific community and the public. One interesting direction for future research is to explore what decisions shape how governments in India are reporting COVID-19 surveillance and vaccination monitoring data.[21,22]

As researchers and health professionals, our goal here is to advocate for change through measurement. Through a semi-quantitative approach, we showed the specifics and magnitude of missing granular data across India. Our findings provide the largest and most recent evidence for

lack of granularity in India's COVID-19 data reporting. Governments in India should recognize the importance of reporting granular data and make it a priority before the next wave of COVID-19. As a start, we recommend reporting the following. First, age and gender distribution for cases and deaths, and comorbidities for deaths. Second, details of serious/severe AEFI cases. Third, vaccination coverage for each dose stratified by eligibility category.

**Acknowledgements:** We want to thank healthcare workers across the globe for their efforts during the pandemic. We also thank the members of India COVID SOS and the Stanford community for their support and insightful feedback on a version of the draft.

**Contributors:** V.V. and A.G. contributed equally. V.V., A.G., and J.Z. conceived the study. V.V. and A.G. designed the methods with feedback from C.L., G.P. and J.Z. V.V., A.G., and B.B. collected the data. V.V. and A.G. performed the analysis. V.V., A.G., B.B., C.L. and G.P. wrote the first draft. J.Z. edited the drafts and supervised the project. All authors reviewed and approved the final manuscript.

**Funding:** J.Z. is supported by discretionary funding from Stanford University.

**Competing Interests:** The authors declare that there are no competing interests.

**Patient consent for publication:** Not Applicable

**Ethics Approval:** Not Applicable

**Data availability statement:** All the data used in this study are publicly available at [https://github.com/varun-vasudevan/CDRS-India/tree/master/study3\\_june\\_2021](https://github.com/varun-vasudevan/CDRS-India/tree/master/study3_june_2021).

# References

1. WHO. Global surveillance for COVID-19 caused by human infection with COVID-19 virus: Interim guidance [Internet]. [cited 2021 Jul 8]. Available from: <https://apps.who.int/iris/rest/bitstreams/1272502/retrieve>
2. WHO. Monitoring COVID-19 vaccination: Considerations for the collection and use of vaccination data [Internet]. [cited 2021 May 23]. Available from: <https://www.who.int/publications/i/item/monitoring-covid-19-vaccination-interim-guidance>
3. Zimmermann LV, Salvatore M, Babu GR, Mukherjee B. Estimating COVID-19– Related Mortality in India: An Epidemiological Challenge With Insufficient Data. *Am J Public Health*. 2021 Jul 1;111(S2):S59–62.
4. Pulla P. 'There are so many hurdles.' Indian scientists plead with government to unlock COVID-19 data | Science | AAAS [Internet]. [cited 2021 Jul 7]. Available from: <https://www.sciencemag.org/news/2021/05/there-are-so-many-hurdles-indian-scientists-plea-d-government-unlock-covid-19-data>
5. Purkayastha S, Kundu R, Bhaduri R, Barker D, Kleinsasser M, Ray D, et al. Estimating the wave 1 and wave 2 infection fatality rates from SARS-CoV-2 in India. *BMC Res Notes*. 2021 Jul 8;14(1):262.
6. Lancet T. India's COVID-19 emergency. *The Lancet*. 2021 May 8;397(10286):1683.
7. Vasudevan V, Gnanasekaran A, Sankar V, Vasudevan SA, Zou J. Disparity in the quality of COVID-19 data reporting across India. *BMC Public Health*. 2021 Jun 24;21(1):1211.
8. Vasudevan V, Gnanasekaran A, Sankar V, Vasudevan SA, Zou J. Variation in COVID-19 Data Reporting Across India: 6 Months into the Pandemic. *J Indian Inst Sci*. 2020 Oct 1;100(4):885–92.
9. Deshmukh Y, Suraweera W, Tumbe C, Bhowmick A, Sharma S, Novosad P, et al. Excess mortality in India from June 2020 to June 2021 during the COVID pandemic: death registration, health facility deaths, and survey data. *medRxiv*. 2021 Jul 23;2021.07.20.21260872.
10. Shewade HD, Parameswaran GG, Mazumder A, Gupta M. Adjusting Reported COVID-19 Deaths for the Prevailing Routine Death Surveillance in India. *Front Public Health* [Internet]. 2021 [cited 2021 Aug 5];0. Available from: <https://www.frontiersin.org/articles/10.3389/fpubh.2021.641991/full>
11. Government of India. MyGov App [Internet]. [cited 2021 Jul 8]. Available from: <https://play.google.com/store/apps/details?id=in.mygov.mobile&hl=en>
12. Government of India. CoWIN Dashboard [Internet]. [cited 2021 Jul 7]. Available from: <https://dashboard.cowin.gov.in/>
13. Lancet COVID-19 Commission, India Task Force. Managing India's Second COVID-19 Wave: Urgent Steps [Internet]. 2021 [cited 2021 May 23]. Available from: <https://static1.squarespace.com/static/5ef3652ab722df11fcb2ba5d/t/6076f57d3b43fb2db4a7c9c9/1618408831746/India+TF+Policy+Brief+April+2021.pdf>
14. Ministry of Health and Family Welfare, India. COVID-19 Vaccines: Operational Guidelines (Updated as on 28 December 2020) [Internet]. [cited 2021 Jun 11]. Available from: <https://www.mohfw.gov.in/pdf/COVID19VaccineOG111Chapter16.pdf>
15. Banaji M. The Importance of Knowing How Many Have Died of COVID-19 in India - The Wire Science [Internet]. [cited 2021 Jul 7]. Available from: <https://science.thewire.in/health/how-many-died-covid-19-india-undercounting-underreporting-prevent-next-surge/>
16. Biswas S. India coronavirus: How a group of volunteers "exposed" hidden Covid-19 deaths. *BBC News* [Internet]. 2020 Nov 20 [cited 2021 Jul 7]; Available from: <https://www.bbc.com/news/world-asia-india-54985981>
17. Lee LM, Gostin LO. Ethical collection, storage, and use of public health data: a proposal for a national privacy protection. *JAMA*. 2009 Jul 1;302(1):82–4.
18. Lancet T. India under COVID-19 lockdown. *The Lancet*. 2020 Apr 25;395(10233):1315.
19. Blanchard J, Washington R, Becker M, Namasivayam V, Gopal KM, Sarwal R. Vision 2035:

Public Health Surveillance in India - A White Paper [Internet]. figshare; 2021 [cited 2021 Aug 8]. Available from:

[https://figshare.com/articles/book/Vision\\_2035\\_Public\\_Health\\_Surveillance\\_in\\_India\\_-\\_A\\_White\\_Paper/14093323/4](https://figshare.com/articles/book/Vision_2035_Public_Health_Surveillance_in_India_-_A_White_Paper/14093323/4)

20. Joshi J, Das MK, Polpakara D, Aneja S, Agarwal M, Arora NK. Vaccine Safety and Surveillance for Adverse Events Following Immunization (AEFI) in India. *Indian J Pediatr*. 2018 Feb 1;85(2):139–48.
21. Pine KH, Liboiron M. The politics of measurement and action. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015. p. 3147–56.
22. Laxminarayan R, Jameel S, Sarkar S. India's Battle against COVID-19: Progress and Challenges. *Am J Trop Med Hyg*. 2020 Oct 7;103(4):1343–7.

## Supporting Information

**S1 Text. Shortlisting digital platforms.**

**S2 Table. Surveillance reporting score for each subnational.**

**S3 Text. Calculating the amount of missing granular surveillance data.**

**S4 Text. Additional notes on Figs 1 and 2.**

**S5 Text. Suggestions on granular reporting of testing data**

**S6 Fig. Privacy violation in the reporting from Haryana.**

# Supporting Information

## S1. Shortlisting digital platforms

If no relevant data was found on subnational government or health department websites and mobile applications, a google search was done using the phrases "<subnational> government covid website," "<subnational> government corona website," and "covid bed availability <subnational>." We parsed through the first two pages of search results and shortlisted all government sources. The government websites listed at [help.covidthoday.in](https://help.covidthoday.in) were also included.

The complete list of digital platforms that we shortlisted and checked are publicly available through the dataset released with this paper. At least two authors checked each platform independently and arrived at a consensus on what data is being reported. We excluded data reported on social media platforms due to their inherent limitations.[1]

## S2. Surveillance reporting score for each subnational

**Table S2.1:** Surveillance reporting score for each subnational calculated using all the indicators proposed in [1]. Subnationals are sorted in the decreasing order of score.

Subnational	Surveillance reporting score
Nagaland	0.61
Kerala	0.57
Ladakh	0.52
Odisha	0.52
Punjab	0.52
Tamil Nadu	0.52
Puducherry	0.50
Tripura	0.50
West Bengal	0.50
Chhattisgarh	0.48
Gujarat	0.48
Haryana	0.48
Karnataka	0.48
Maharashtra	0.48
Uttarakhand	0.48
Andhra Pradesh	0.46
Jammu and Kashmir	0.46

Subnational	Surveillance reporting score
Jharkhand	0.46
Meghalaya	0.46
Mizoram	0.46
Telangana	0.46
Madhya Pradesh	0.43
Rajasthan	0.43
Sikkim	0.43
Chandigarh	0.41
Andaman and Nicobar	0.39
Arunachal Pradesh	0.39
Assam	0.39
Delhi	0.39
Goa	0.39
Himachal Pradesh	0.39
Manipur	0.39
*Bihar	0.33
*Dadra and Nagar Haveli and Daman and Diu	0.33
*Lakshadweep	0.33
*Uttar Pradesh	0.33

\*Data from these subnationals are available only through MyGov.

### S3. Calculating the amount of missing granular surveillance data

To get a handle on the scale of missing granular data, we narrow our focus on the reporting of age and gender for confirmed cases; and age, gender, and comorbidity for deaths. Among subnationals reporting these items, some disaggregate the cumulative numbers by the items; some disaggregate the daily numbers by the items, and the remaining report the items for each individual.

Considering all subnationals that report one or more of age, gender, and comorbidity, in any of the three forms mentioned above, we calculated the percentage of cases and deaths for which age, gender, and comorbidity distribution is available. As of 5 June 2020, India had over 28.6 million cases and over 344 thousand deaths. Among these, we have age and gender distribution for 13.75% and 15.24% of the cases, as shown in Fig 2B. Similarly, as of 5 June 2020, we have age, gender, and comorbidity distribution for 19.46%, 21.96%, and 28.65% of the deaths, respectively, as shown in Fig 2C. The calculations are as shown in the tables below.

**Table S3.1:** Subnationals that report age and/or gender data for cases. Case numbers as of 5 June 2021 are taken from the MyGov app. Total cases in India = 28.6 million

Subnational	Total cases in subnational	Percentage of total cases in India = $100 \times (\text{Total cases in subnational} / \text{Total cases in India})$
Nagaland	22387	0.078
Odisha	798699	2.793
Tamil Nadu	2.19E+06	7.657
Telangana	587664	2.055
Haryana	760914	2.661
Uttarakhand	332959	1.164

**Table S3.2:** Calculation of percentage of cases for which age and gender distribution are available in India as of 5 June 2021.

Granular item	Subnationals reporting the granular item	Total %
Age	Nagaland (0.078) + Odisha (2.793) + Tamil Nadu (7.657) + Telangana (2.055) + Uttarakhand (1.164)	13.75
Gender	Nagaland (0.078) + Odisha (2.793) + Tamil Nadu (7.657) + Telangana (2.055) + Haryana (2.661)	15.24

**Table S3.3:** Subnationals that report age, gender, and/or comorbidity data for deaths. Death numbers as of 5 June 2021 are taken from the MyGov app. Total deaths in India = 344082

Subnational	Total deaths in subnational	Percentage of total deaths in India = $100 \times (\text{Total deaths in subnational} / \text{Total deaths in India})$
Karnataka	30895	8.979
Nagaland	416	0.121
Tamil Nadu	26128	7.594
Haryana	8605	2.501
Kerala	9510	2.764
Chhattisgarh	13162	3.825
Telangana	3346	0.972
West Bengal	16034	4.660



**Table S3.4:** Calculation of percentage of deaths for which age, gender, and comorbidity distribution are available in India as of 5 June 2021.

Granular item	Subnationals reporting the granular item	Total %
Age	Karnataka (8.979) + Nagaland (0.121) + Tamil Nadu (7.594) + Kerala (2.764)	19.46
Gender	Karnataka (8.979) + Nagaland (0.121) + Tamil Nadu (7.594) + Haryana (2.501) + Kerala (2.764)	21.96
Comorbidity	Karnataka (8.979) + Nagaland (0.121) + Tamil Nadu (7.594) + Haryana (2.501) + Chhattisgarh (3.825) + Telangana (0.972) + West Bengal (4.660)	28.65

## S4. Additional notes on Figs 1 and 2

- Fig 1D shows a decrease in the *granular* surveillance reporting score for Tripura from assessment 1 in May 2020 (0.29) to assessment 2 in July 2020 (0.22). It is not because they **stopped** reporting a granular item. The reason is as follows. Tripura did not have any COVID-19 deaths until the completion of assessment 1. So, we excluded relevant granular indicators for deaths during assessment 1. However, by the time of assessment 2, Tripura had COVID-19 deaths, and their reporting did not have all the granular information. They got a lower score because we included all indicators in this assessment.
- Fig 2E: Chhattisgarh stopped reporting the comorbidity distribution for *total* deaths after the assessments in 2020. Instead, they are now reporting comorbidity distribution for *daily* deaths.
- Fig 2E: Karnataka stopped reporting the age distribution for *total* deaths after the assessments in 2020. However, they are continuing to report the age of *each* deceased person.

## S5 Suggestions on granular reporting of testing data

The number of cases reported by a subnational is a function of its testing capacity and testing coverage. We make a few suggestions to subnationals aimed at granular test reporting. Apart from reporting the number of tests and test positivity, subnationals could report the following. First, the reporting lag (interval between symptom onset and confirmation), which is crucial in estimating the effective reproductive number and understanding the demand in the testing capacity.[2] Second, test numbers and results disaggregated by symptomatic versus asymptomatic patients; contacts traced versus others; and type of testing (RT-PCR, rapid antigen, etc). This data is representative of the real-time community spread of the virus and how much the health system is in control.

## S6 Privacy violation in the reporting from Haryana

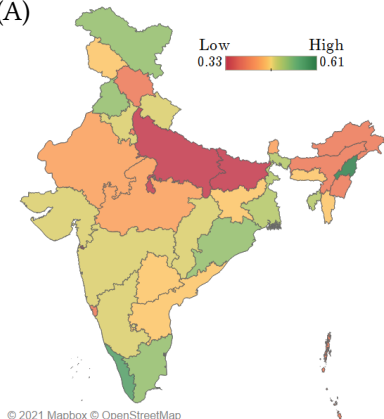
<b>ORDER</b>							
Whereas it has been reported that following persons of District Jhajjar found positive for COVID-19 infection and admitted in Institutional isolation/home isolation as mentioned below.							
S.No	Name of Patient (with age)	Address	Institutional Isolation or Home Isolation	Boundaries of Containment zone to be notified	Health facility centre to monitor the Patient	Houses and Population of Containment Zone	Remarks (New Containment and Extension)
1.			Home isolation		PHC Dujana	Houses: 08 Population : 30	Ext of order 392/MB Date 04.09.21
2.			Home isolation		PHC Nuna Majra	Houses: 15 Population :87	Ext of order 278/MB Date 04.07.21

**Fig S6.1:** Screenshot of the document published by Haryana containing the name, age, gender, and address of COVID-19 cases. We blacked out personal information in the screenshot.

## References for supporting information

1. Vasudevan V, Gnanasekaran A, Sankar V, Vasudevan SA, Zou J. Disparity in the quality of COVID-19 data reporting across India. BMC Public Health. 2021 Jun 24;21(1):1211.
2. Cori A, Kamvar Z, Stockwin J, Jombart T, Dahlquist E, FitzJohn R, et al. EpiEstim v2.2-3: A tool to estimate time varying instantaneous reproduction number during epidemics [Internet]. 2021. Available from: <https://github.com/mrc-ide/EpiEstim>

(A)



(B)

Data reported on MyGov app for each subnational

Daily & total data	Confirmed	✓
	Death	✓
	Recovered	✓
Data stratified by	Age	✗
	Gender	✗
	Comorbidity	✗
	Districts	✗

(C)

Data reported on CoWIN for each subnational

Vaccination stratified by district & dose	Daily	✓
	Daily trend	✓
	Total	✓
Vaccination stratified by eligibility category		✗
AEFI stratified by	Vaccine type	✗
	Event type	✗

(D)

Subnational (State / Union territory)	Granular surveillance reporting score			Vaccination monitoring			Vacant ICU/oxygen /ventilator bed availability
	May 2020	July 2020	June 2021	By eligibility category	AEFI by vaccine type	AEFI by event type	
Nagaland	–	0.17	0.50	✓	✗	✗	✗
Haryana	0.17	0.33	0.33	✗	✗	✗	✓
Tamil Nadu	0.33	0.33	0.33	✗	✗	✗	✓
Kerala	0.22	0.33	0.28	✓	✗	✗	✓
Odisha	0.28	0.28	0.28	✓	✗	✗	✗
Karnataka	0.39	0.50	0.22	✓	✗	✓	✓
Ladakh	0.21	0.22	0.22	✓	✗	✗	✗
Puducherry	0.22	0.22	0.22	✓	✗	✗	✓
Tripura	0.29	0.22	0.22	✗	✗	✗	✗
Uttarakhand	0.17	0.22	0.22	✓	✗	✗	✓
West Bengal	0.17	0.22	0.22	✗	✗	✗	✓
Andhra Pradesh	0.00	0.17	0.17	✗	✗	✗	✓
Chhattisgarh	0.17	0.22	0.17	✗	✗	✗	✓
Gujarat	0.17	0.17	0.17	✓	✗	✗	✓
Jammu & Kashmir	0.17	0.17	0.17	✗	✗	✗	✗
Jharkhand	0.50	0.17	0.17	✓	✗	✗	✗
Madhya Pradesh	0.17	0.17	0.17	✓	✗	✗	✓
Maharashtra	0.17	0.17	0.17	✗	✗	✗	✗
Meghalaya	0.00	0.00	0.17	✗	✗	✗	✗
Mizoram	–	0.07	0.17	✓	✗	✗	✗
Punjab	0.17	0.17	0.17	✓	✗	✗	✓
Telangana	0.00	0.00	0.17	✗	✗	✗	✓
Rajasthan	0.11	0.11	0.11	✗	✗	✗	✓
Sikkim	–	0.00	0.11	✗	✗	✗	✗
Andaman & Nicobar	0.00	0.00	0.00	✗	✗	✗	✗
Arunachal Pradesh	–	0.00	0.00	✗	✗	✗	✗
Assam	0.17	0.17	0.00	✗	✗	✗	✗
Bihar	0.00	0.00	0.00	✗	✗	✗	✓
Chandigarh	0.00	0.00	0.00	✗	✗	✗	✓
Delhi	0.00	0.00	0.00	✗	✗	✗	✓
DNH & DD	–	0.22	0.00	✗	✗	✗	✓
Goa	0.00	0.06	0.00	✗	✗	✗	✓
Himachal Pradesh	0.00	0.00	0.00	✓	✗	✗	✓
Lakshadweep	–	–	0.00	✗	✗	✗	✗
Manipur	–	0.00	0.00	✓	✗	✗	✗
Uttar Pradesh	0.00	0.00	0.00	✗	✗	✗	✗

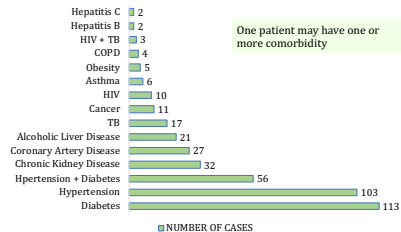
**Symbols:** – for score unavailable; — for no change; ▲ for increase; ▼ for decrease; ✓ for is reporting; and ✗ for is not reporting

(A)

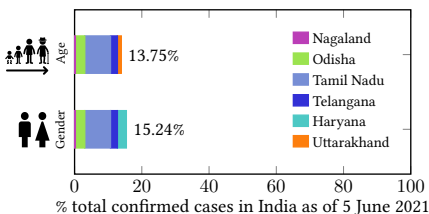
Distribution of deaths by age group		
Age group	Total 2020 (%)	Total 2021 (%)
<14	0	0
15-29 yrs	4	3
30-44 yrs	26	26
45-59 yrs	29	33
60-74 yrs	25	27
> 75	15	12
<b>TOTAL</b>	<b>100</b>	<b>100</b>

Distribution of deaths by gender		
Gender	Total 2020 (%)	Total 2021 (%)
Female	28	30
Male	72	70
<b>TOTAL</b>	<b>100</b>	<b>100</b>

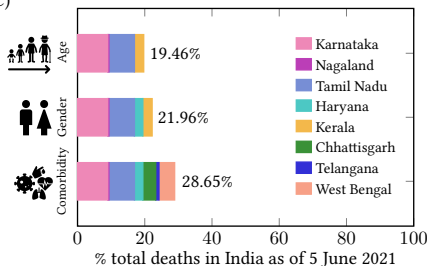
Comorbidities associated with COVID-19 deaths till date			
	Total Deaths	With Comorbidity	Without Comorbidity
Number	402	196	21
Percentage(%)	100	49	5



(B)



(C)



(D)

Subnational	Age	Gender	Comorbidity
Chhattisgarh	✗	✗	2
Haryana	✗	2	2
Karnataka	E	E	E
Kerala	E, 4	E	✗
Nagaland	6	2	15
Tamil Nadu	E	E	E
Telangana	✗	✗	2
West Bengal	✗	✗	2

✗ : Not reporting

E : reported for Each

number : Number of categories/brackets

(E)

Subnational	Age	Gender	Comorbidity	Districts
Assam	—	—	—	—
Chhattisgarh*	—	—	—	—
DNH&DD	—	—	—	—
Goa	—	—	—	—
Jharkhand	—	—	—	—
Karnataka*	—	—	—	—
Kerala	—	—	—	—

— : stopped reporting for confirmed

— : stopped reporting for deaths

— : stopped reporting for recovered

— : no change from earlier assessments