

CTIS Validation Analysis

17/1/2022

Abstract

Introduction

The COVID-19 pandemic and the ensuing crisis has been quantified like no previous crisis, not dissimilar to its societal impact. As the pandemic hit, Governments, business organizations and researchers from various fields across the world faced an urgent need to collect data on assessing the impact of the pandemic and the response at all levels - countries, states, households, individuals. However, interviewer administered, face-to-face surveys of general population which were commonplace prior to the crisis were no longer feasible due to non-pharmaceutical interventions such as physical distancing and lockdowns. At the same time, the widespread use of mobile phones provided alternate modes data collection for research to meet the demands generated by the pandemic. With the opportunity of conducting surveys remotely during the national lockdown when there was a need for scientifically collected data for decision making in the wake of the COVID-19 pandemic, several public and private organizations launched remote modes of surveys and the existing ones which were fielding face-to-face interviews also transitioned their surveys to remote modes of data collection. In the absence of face-to-face surveys, the remote modes of data collection filled this void and satisfied the need of a quick turnaround in the absence of travel time and helped in informing policy response in a timely manner.

However, the validity and reliability of the estimates generated by remote modes of data collection have not been established due to incomplete coverage of population having access to mobile phones and self selection resulting from response bias (Tsuboi et al. 2015; Bethlehem 2010). In a recent paper in Nature, Bradley et al. report that larger sample sizes do not protect against bias (Bradley et al. 2021). Using vaccination estimates from three different surveys in the United States by comparing them with benchmark CDC numbers, Bradely et al. demonstrate that two of the three surveys with large sample sizes significantly overestimated the vaccine uptake. Further, they provide evidence on two key issues related to vaccination estimates from surveys - 1) The difference in vaccination estimates of the two large surveys compared to the benchmark CDC numbers increased over time. This shows that that the magnitude of survey errors increased over time and 2) The relative estimates between subgroups of the population i.e., states in the US and 18-64 age group were no better than absolute estimates. Taken together, these findings suggest that both the absolute estimates and relative estimates over space and time for the two large surveys are biased and do not accurately represent the ground truths.

One of the surveys which Bradley et al. utilise is Facebook's COVID-19 Trends and Impact Survey (abbreviated to CTIS which was earlier referred to as COVID-19 Symptom Survey abbreviated to CSS) which is conducted in partnership with Carnegie Mellon University in the United States and University of Maryland (UMD) globally. CTIS is the largest public health survey till date conducted daily in more than 200 countries or territories globally and the survey estimates are used widely in research and to formulate public health policy (Salomon et al. 2021; Sukumaran et al. 2021; Adjodah et al. 2021; Babalola and Krenn 2021). Our earlier work, following the news about vaccine breakthrough to fight the Coronavirus, also leveraged data from the CTIS to present estimates of vaccine hesitancy in India across states and for the survey time period (Chowdhury, Motheram, and Pramanik, n.d.). Given the paucity of data that existed then, the CTIS has been a useful tool for us in understanding and tracking the changing nature of the pandemic and the public response to it.

In light of Bradley et. al's work, we examine the CTIS estimates for India and the compare them with available information in COVID-19 administrative data recorded in the CoWIN application portal. Governments at state and national level in India release aggregated data of CoWIN application data in the form of daily bulletins which are further consolidated by a crowdsourced volunteer group called COVID19Bharat. In the following section, we present results from two validation exercises comparing CTIS estimates in India to the widely used official public health metrics obtained from COVID19Bharat.

Results

Vaccine uptake estimates comparison

The CTIS questionnaire includes two questions on vaccination which can be compared to official estimates of COVID-19 Vaccination. Firstly, all respondents are asked whether they had a COVID-19 vaccination with yes, no and don't know as response choices. Bradley et al. use the estimates from responses to this question to estimate the vaccine uptake (of at least one dose) for CTIS-US. Further, if answered in positive to the earlier question, the respondent is further probed on how many COVID-19 vaccinations he/she has received with 1 dose, 2 doses and don't know as response choices. Strictly speaking, comparisons with respondents who have taken at least dose are more likely to align with official estimates of one dose of vaccination given the way the questions are probed. However, we also present results of percentage of respondents who have responded with two doses with official estimates of second dose of vaccination at the national level.

In Figure 1, we present the comparison of CTIS and Official Estimates of COVID-19 Vaccination by one and two doses of vaccination at the National Level. The official estimates are scaled to the eligible target group of people above 18 years old using the Census projections for the year 2021 published by the Government of India. Similar to Bradley et al., we observe the vaccine uptake in CTIS to be higher than in the administrative data. However, while the estimates at earlier time points diverge greatly due to coming from a low base and the CTIS sample characteristics, the divergence between CTIS decreases over time as more adults in the general population are vaccinated. The scaling factors of excess vaccination uptake for one dose and two doses at March 2021 stand at 8 and 120 respectively. By April-May 2021, the excess vaccine uptake in CTIS falls to 2 and 15 and continues the downward trend until October 2021.

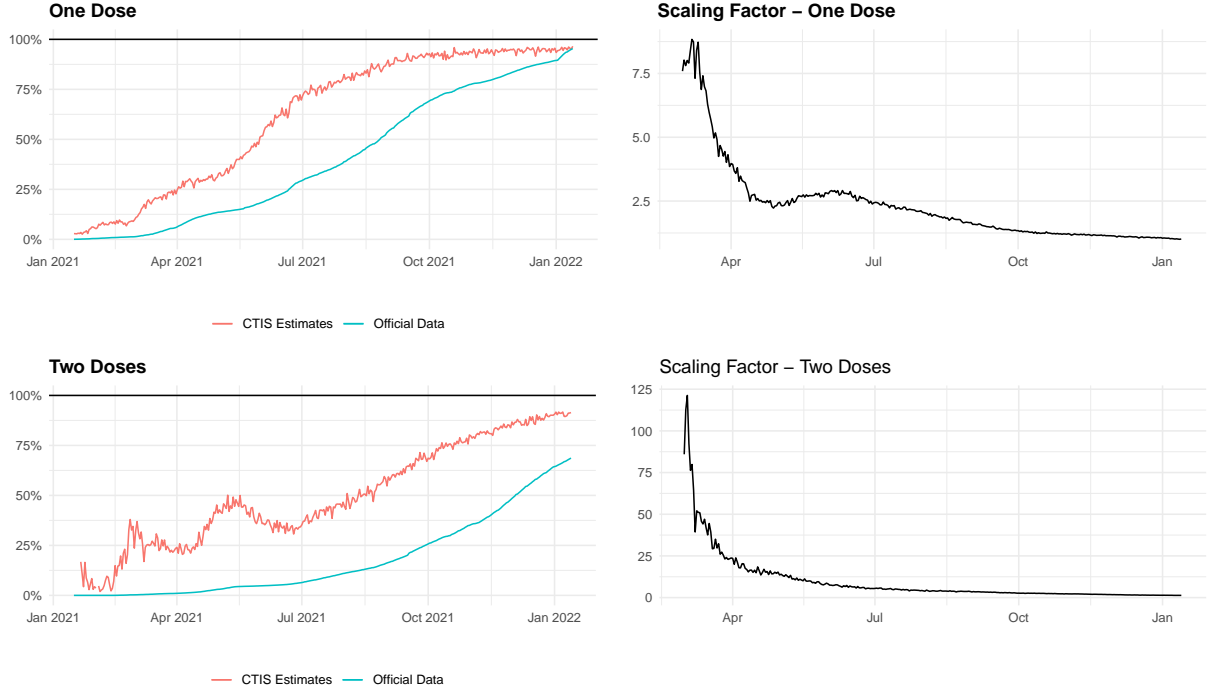


Figure 1: Vaccine Uptake in India - CTIS Estimates and Administrative Data Comparison

The convergence in estimates can be explained by a couple of factors. The COVID-19 vaccination rollout was initially restricted to self-registration on the CoWIN application. While the rollout was extremely useful in gathering information about the vaccinated, this perhaps had an unintended effect of increasing the divergence in vaccination rates between rural and urban areas due to the existing digital divide according to multiple news reports (Correspondent 2021; Sharma 2021). This is also observed in CTIS data which is shown in Figure 2. From the CTIS microdata, we plot the 7 day moving average vaccination rates by place of residence for respondents on whom there is complete data for both questions. Vaccination rates in rural areas catch up with the urban cities and towns in the second half of the year according to the CTIS microdata estimates. An important caveat here is that urbanicity is not used in the weighting procedure to correct for the sample bias which we discuss more in the section on study design and sample representativeness.

At the state level, we observe a similar convergence in trends over time with some exceptions. Due to small sample sizes, CTIS estimates have missing data for smaller states (by population size) on some days. Accounting for that, we plot the 7-day moving average of CTIS trends versus the official state level vaccination for the select states (determined by population sizes) in Figure 4. The CTIS picks up the trends in early movers such as Kerala and Madhya Pradesh which complete 90% first dose coverage by October 2021. In almost all the states, the CTIS estimates start higher and plateau earlier than the trends in the official vaccination numbers.

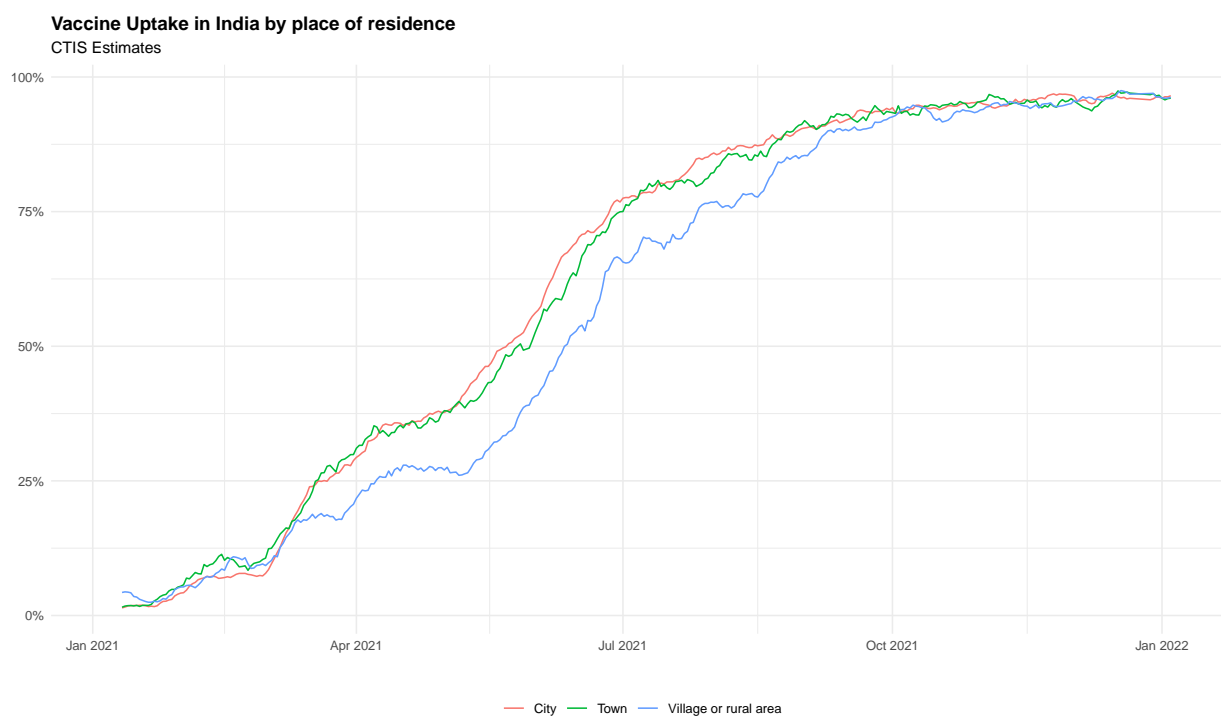


Figure 2: Vaccine Uptake in India by place of residence - CTIS Estimates

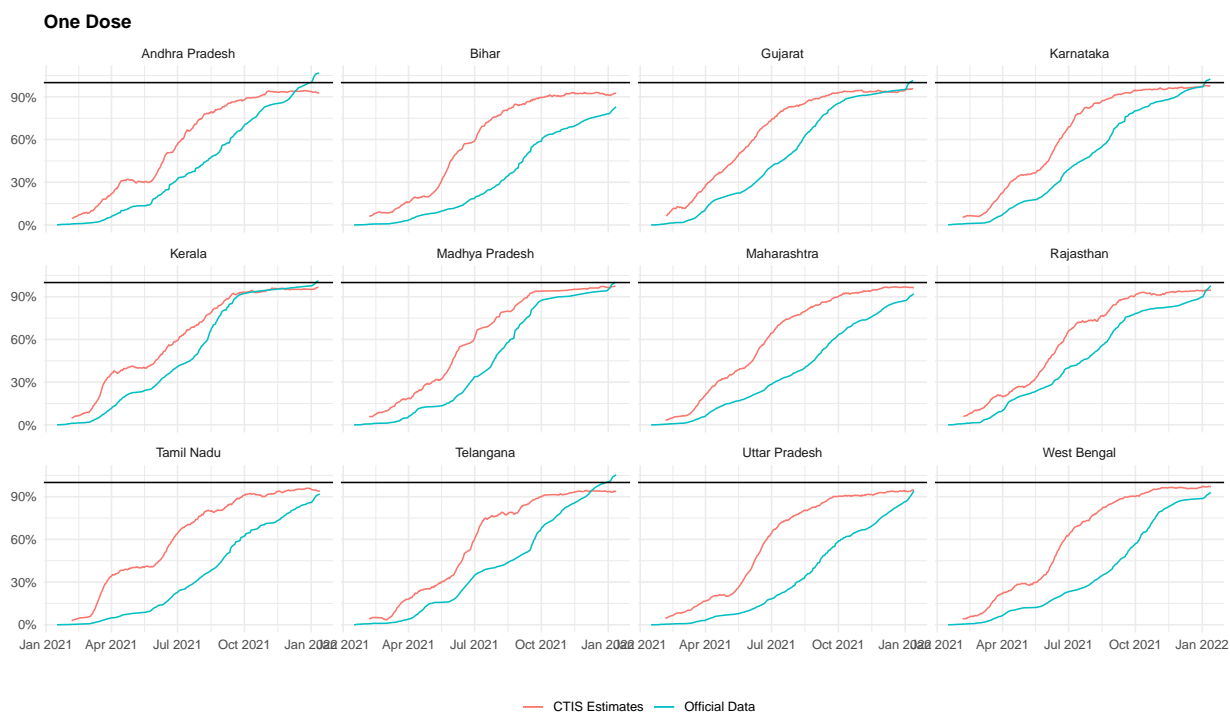


Figure 3: Vaccine Uptake in Indian States - CTIS Estimates and Administrative Data Comparison

As Bradley et al. note, one benefit of large sample, repeated cross-sectional surveys such as CTIS can be to compare estimates of spatial and demographic subgroups. To determine the validity of relative CTIS vaccination estimates versus the official numbers, we plot the daily Spearman correlation coefficient obtained from rank tests of states over time in Figure 5. To smoothen the noisy daily correlations, we also plot the 15-day moving average of the daily points. The 15-day moving average of Spearman rank correlation which starts from a low base improves to nearly 0.45 till wave 2 in April, 2021 and then declines to nearly 0.15 in the following three months as a result of rapid uptake in vaccination in certain states and subgroups during the panic of delta wave. In the “normal” months subsequent to the delta wave starting August 2021, the correlations range between 0.1-0.4. As the states converge to a full dose of first vaccination, we observe a wider dispersion of daily correlations suggesting that the CTIS estimates may not be fully or accurately capturing the final mile of vaccination among the remaining pockets of the population.

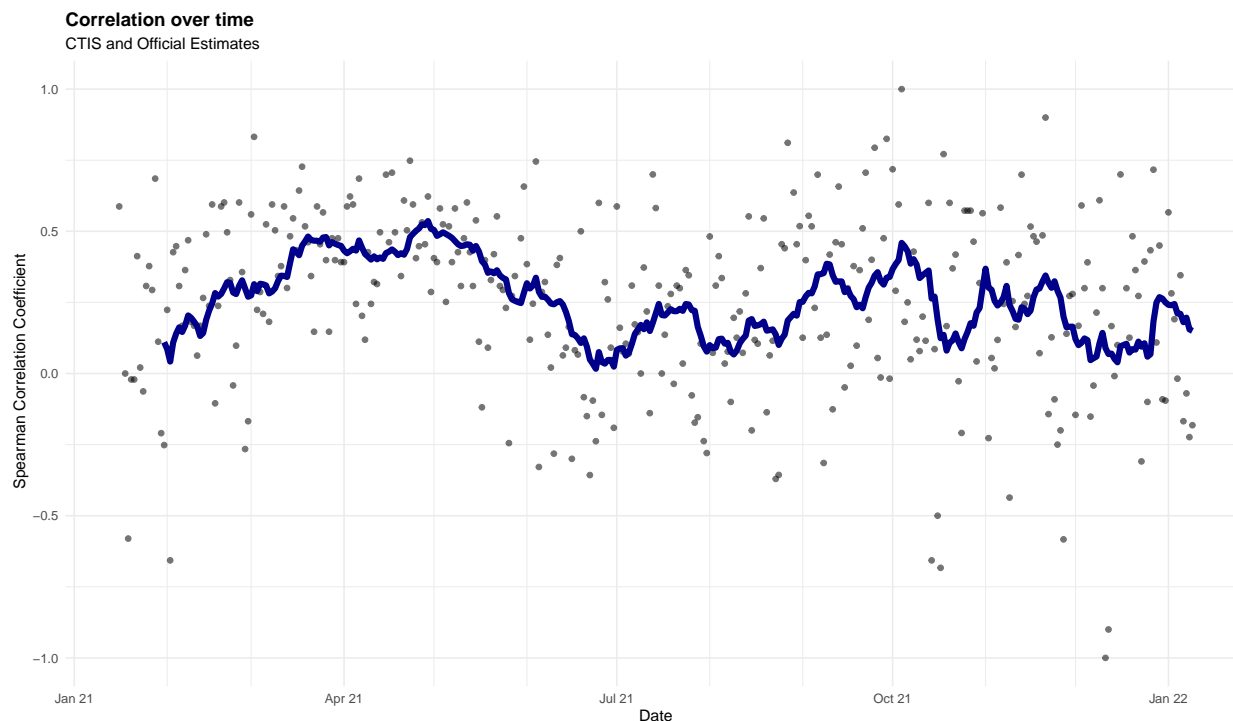


Figure 4: Daily state rank correlation over time

CLI and COVID-19 Cases

The original rationale behind CTIS as mentioned in the documentation is that since experiencing symptoms is often a precursor to going to the hospital or getting more seriously ill, if this works, it can produce a weekly forecast how many cases hospitals will see in the days ahead, as well as an early indicator of where the outbreak is growing and where the curve is being successfully flattened. Several studies on symptom tracking and forecasting COVID cases (Reinhart, n.d.; Salomon et al. 2021; Menni et al. 2020; Sukumaran et al. 2021).

In order to track the symptoms, the CTIS asks respondents 14 questions on various symptoms and if they have experienced any of them in the 24 hours prior to taking the survey. Facebook-UMD construct an indicator named COVID-like Illness (CLI) based on the percentage of positive responses to 3 of these 14 symptoms namely fever, along with cough, shortness of breath, or difficulty breathing. Facebook-UMD provide daily estimates of percentage of survey respondents that have reported COVID-like illness (CLI) publicly.

In Figure 6, we present the comparison of CLI with official COVID-19 diagnostic metrics. During Wave-2 in India amidst the emergence of delta variant between April-June 2021, both cases and deaths peak sequentially and plateau to a relatively low level after that period. We see that 7-day moving average of percentage of respondents with CLI trend also form a crest and a trough mirroring the cases. However, while reported cases and deaths have collapsed post wave 2 in June 2021, the percentage of respondents with CLI has not and moreover it keeps increasing past the peak seen during wave 2 reaching a peak of nearly 6% in November 2021. We also note that while testing has decreased post wave-2, the decrease is relatively small and not enough to entirely explain this divergence.

COVID-19 Diagnostic and CTIS-CLI trends in India

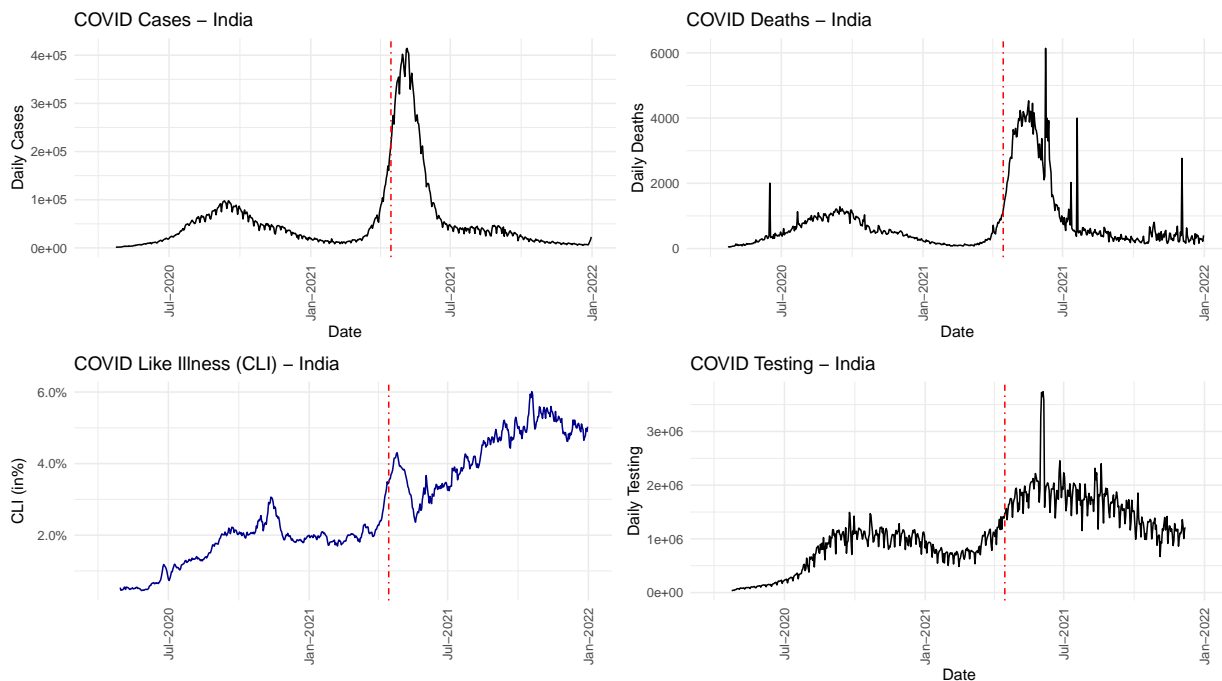


Figure 5: CLI - Comparison with COVID-19 Diagnostic Data

In Figure 7, we present at the CLI trend at state level by northern (Gujarat is included in the northern states although it lies in western India strictly speaking) and southern states. We observe that while the 7-day moving average CLI trend for southern states remains steady at relatively low levels, CLI trend in the northern states keeps increasing post wave 2 which contributes majorly to the overall national CLI. There are two questions at play here - 1) Why do we see an persistence of trend in the CLI trends at national level

when daily COVID-19 cases have fallen? 2) Secondly, irrespective of question 1, why do the northern and southern states exhibit different trends in CLI post wave 2, *ceteris paribus*.

COVID Like Illness (CLI) Symptoms

Trend over time in select states of India

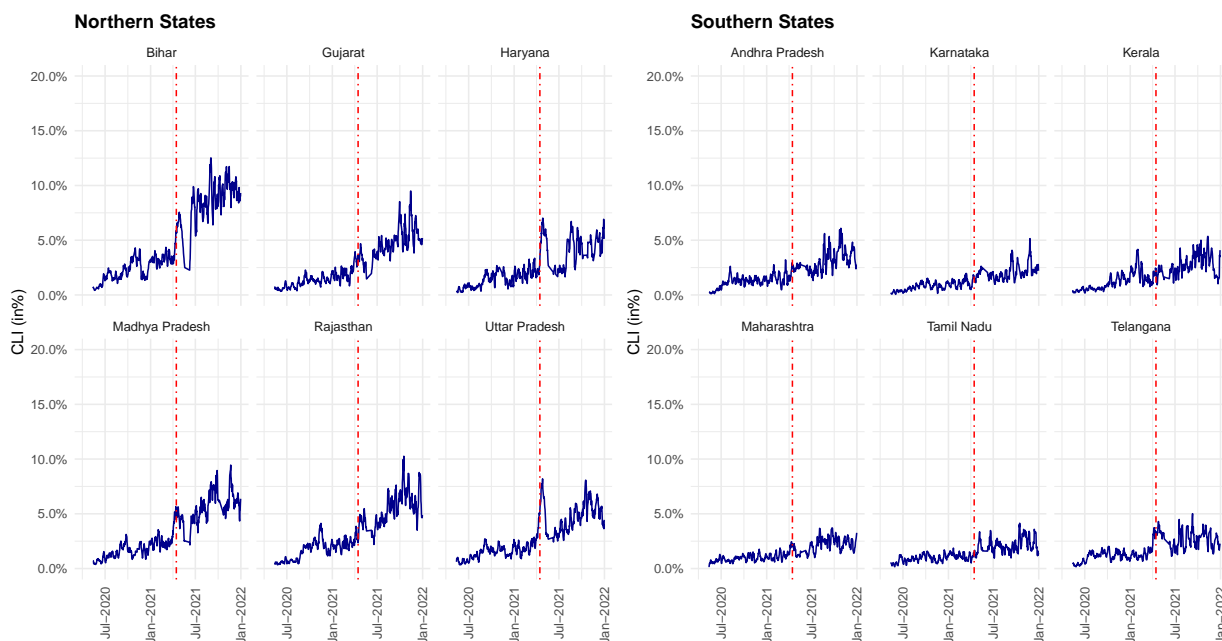


Figure 6: Reported CLI - Indian States

One potential hypothesis is the confounding of symptoms between COVID-19 and seasonal flu. Although the symptoms of COVID-19 are similar to those of the seasonal flu, including chest pains, fatigue, fever, and a persistent cough, the addition of anosmia, the loss of anosmia and dysgeusia, or smell and taste has been shown to be an important indicator of whether an individual has been affected with COVID-19 [menni2020]. Whether respondent has had Anosmia in 24 hours prior to the interview is one of the 14 questions on symptoms asked in CTIS. In Figure 8, we plot the 7-day moving average of percentage of respondents in the selected states who have reported Anosmia as a symptom. Similar to CLI trends, we see a major divergence in the trends by northern and southern states.

Anosmia Reported Symptom
Trend over time in select states of India

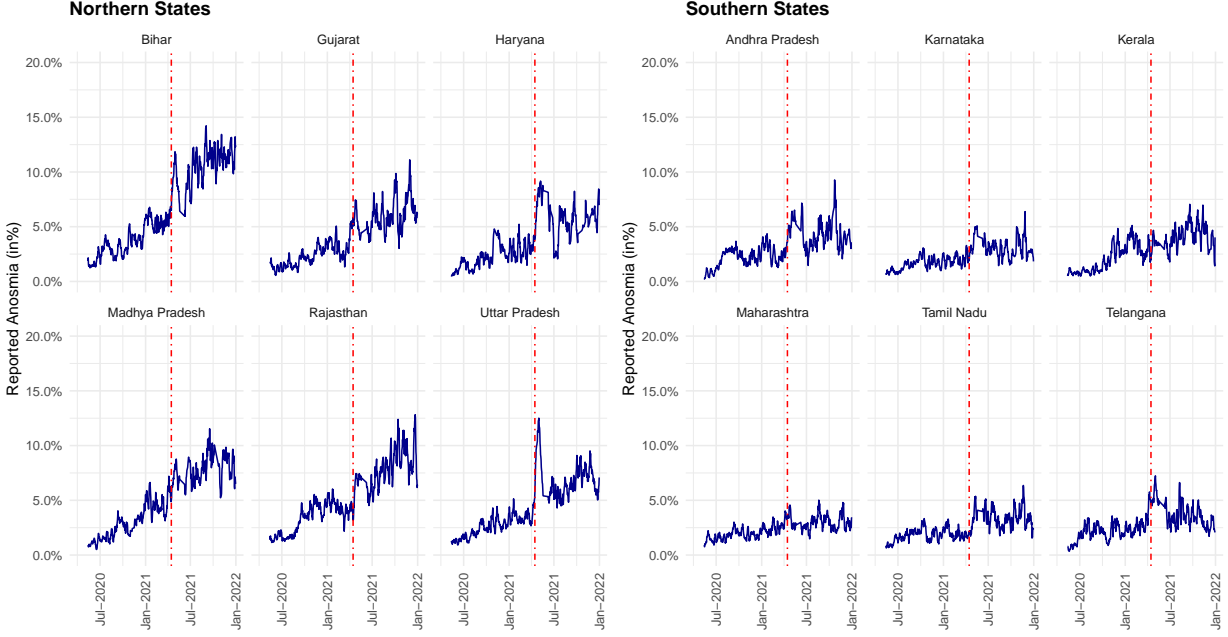


Figure 7: Reported Anosmia - Indian States

Discussion

In this section, we discuss the study design and sample representativeness of CTIS and conclude by listing some possible learnings for future survey research.

Study Design & Sample Representativeness

CTIS is an opt-in survey of Facebook users of over 2 billion monthly users over 18 years of age sampled from Facebook’s Active User Base (FAUB). CTIS employs a repeated cross section design to draw new random sample of users daily. There could be repetition of survey request to the same users (particularly in small states), but should be treated as a new sample. To select daily sample, stratified random sampling is considered (strata being the states) in order to provide representative sample at the national level. While the proportion of females and older subgroups is lower in the FAUB, FB-UMD use a two-step weighting procedure to account for this. The objective of constructing weights is to provide a weight per respondent so that respondents of CSS better represent the target population.

Table 1 presents the unweighted and weighted sample characteristics by interview completion status for a randomly selected daily sample from December 1, 2021. We see that the two-step weighting procedure partially corrects the bias in gender and age groups (to a lesser extent) represented in the CTIS samples. However, urbanicity a key demographic variable is not used in the weighting procedure by FB-UMD. Most of the respondents in FAUB are from cities and as a result of not accounting for this, the estimates are skewed towards urban population. The education variable also cannot be used in analysis for the way it is probed and coded.

Characteristic	Unweighted Estimates		Weighted Estimates	
	Incomplete, N = 1,063 ¹	Completed, N = 1,528 ¹	Incomplete ²	Completed ²

Respondent Gender				
Male	82 (7.7%)	1,240 (81%)	6.1%	55%
Female	17 (1.6%)	263 (17%)	7.6%	43%
Prefer not to answer	2 (0.2%)	16 (1.0%)	0.2%	1.6%
Seen but unanswered	2 (0.2%)	9 (0.6%)	<0.1%	0.7%
Missing	960 (90%)	0 (0%)	86%	0%
Respondent Age Group				
18-24	21 (2.0%)	287 (19%)	2.4%	16%
25-34	37 (3.5%)	594 (39%)	3.6%	35%
35-44	17 (1.6%)	314 (21%)	1.0%	16%
45-54	12 (1.1%)	167 (11%)	4.6%	15%
55-64	4 (0.4%)	104 (6.8%)	0.6%	10%
65-74	5 (0.5%)	43 (2.8%)	0.8%	6.3%
>=75 years	0 (0%)	17 (1.1%)	0%	1.3%
Seen but unanswered	1 (<0.1%)	2 (0.1%)	<0.1%	0.1%
Missing	966 (91%)	0 (0%)	87%	0%
Area of Residence				
City	44 (4.1%)	833 (55%)	7.1%	62%
Town	11 (1.0%)	250 (16%)	1.0%	15%
Village or rural area	26 (2.4%)	441 (29%)	2.4%	23%
Seen but unanswered	1 (<0.1%)	4 (0.3%)	<0.1%	0.3%
Missing	981 (92%)	0 (0%)	89%	0%

¹n (%)

²%

Sample characteristics represent the latest daily sample of December 01, 2021

Another major limitation of CTIS is that the demographic details of the respondents are probed at the end of the interview (this is corrected in version 12 of the survey). Owing to this structure of CTIS survey design, the sample characteristics for most non-respondents are missing both at item and unit level. Facebook's original sample frame (FAUB) from which the respondents are selected does have this information and more however, this information is not shared with researchers. This poses a major limitation on any analysis of CTIS data in its current form as it is impossible to get any sense of the non-respondents in the survey.

Surveys and Gold Standards

Results from validation of CTIS estimates indicate that the vaccination estimates from CTIS certainly inflated compared to the actual numbers. Further, the CTIS samples is also biased towards more educated, internet savvy, urban respondents even after correction for survey non-response and individuals who are not on the Facebook platform. However, this does not imply that the CTIS data is no longer useful and or that it cannot be taken seriously. We think otherwise and present our reasoning with an example where the symptom trends observed in the CTIS diverge from the official estimates and the administrative estimates cannot be considered as goldstandard. It is unclear at this moment as to why the percentage of respondent with CLI keeps increasing post wave 2. There are several hypothesis including the persistence of symptoms in long covid patients (Carfi et al. 2020; Sudre et al. 2021), role of absolute humidity and seasonality in the COVID-19 transmission dynamics (Shaman and Galanti 2020; Liu et al. 2021; Baker et al. 2021; Phillips 2021) . Unlike in the case of vaccination, the official COVID-19 diagnostic numbers cannot be substituted for a gold standard.

While the COVID official diagnostic numbers such as cases, deaths, tested are much less of a “gold standard” compared to the vaccination numbers ((Karlinsky and Kobak 2021)), they are used widely used to track the progress of the pandemic despite their short comings (Serikbayeva, Abdulla, and Oskembayev 2021). For instance, test positivity rate is a widely used metric which suffers from selection bias in testing and test

positivity rates has been documented (Wälde, n.d.; Díaz-Pachón and Rao 2021). In India, even the peak of about 3 million daily tests falls well short of the total population at 1.3 billion and an adult population of more than 900 million. Further, state’s action or the lack of it in mounting an appropriate response has a considerable effect on outcomes (Serikbayeva, Abdulla, and Oskenbayev 2021; Shah 2021). In Figure 8, we show the wide range of inter-state responses using total number of tests per adult population as a crude proxy of the state response in testing.

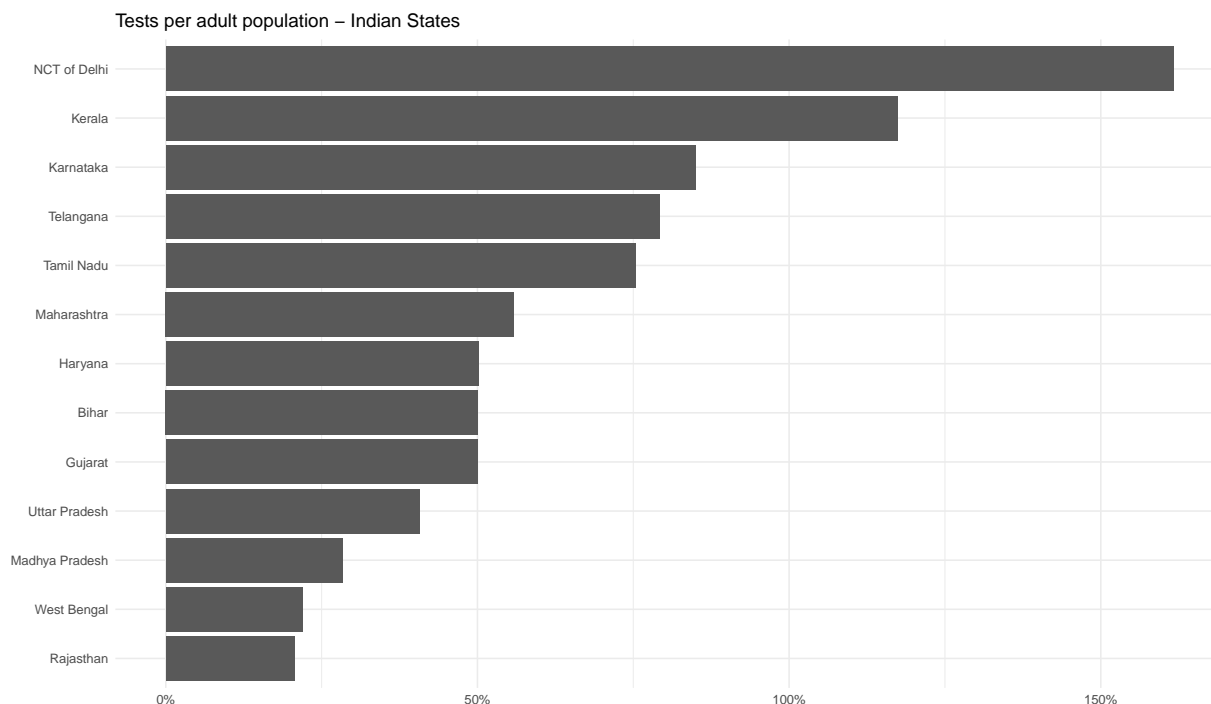


Figure 8: Testing per adult - Select Indian States

In response to Bradley et al., Professor Frauke Kreuter who co-leads the Global CTIS writes “The (survey) quality is very difficult to assess, because there is usually no independently verified ‘ground truth’ or ‘gold standard’ with which to compare survey data.” (Kreuter 2021). This rings even truer in the context of India and other developing economies where the gold standard estimates are deficient collecting the required information in some cases, defunct in others or not at all present in some others. Even in situations where the benchmark estimates are available, they have been called into question and validation is difficult due to lack of all the available data such as metadata and paradata (process data) of the surveys (Meyer, Mok, and Sullivan 2015). The vaccine uptake example is a rare outlier and the official numbers are certainly as close to a “gold standard” as one can get. However, this is not true of majority if not all of the variables on which information is collected in sample surveys. Historically, official surveys such as the National Sample Survey (NSS) or National Family Health Survey (NFHS) have been considered to be the “gold standard” surveys in India. However, the lack of data on household incomes in NSS has led to researchers using data from a panel study called Indian Human Development Study (IHDS). More recently, several academic papers have been published (Malani and Ramachandran 2021) using survey data from a private entity called Centre for Monitoring Indian Economy (CMIE) primarily due to the lack of clarity on frequency and data quality of NSS’s employment-unemployment surveys. This adoption has taken place despite issues of sample representativeness and methodology of surveys such as CMIE conducted by private sector (Somanchi 2021; Bhalla and Das, n.d.). Going forwards, it is clear that any research in social sciences has to make use of data coming in all shapes and forms. For surveys with biased samples, novel techniques such as Multilevel Regression and Poststratification (also known as Mister P or MRP) have shown promise in correcting the survey biases (Wang et al. 2015).

As Professor Kreuter writes, “For certain inferential tasks, surveys with deficiencies can be useful. The usefulness of a data set can be evaluated only in the context of a specific research question.” We agree with one caveat that the survey methods and data including metadata and process data be made available transparently. This means embracing more openness in all forms - accessibility, availability and less transaction costs in general to acquire the needed information.

Data Availability

Data used in the analysis come from various sources. CTIS aggregate estimates at state and national levels are obtained from UMD’s Global CTIS Open Data API. CTIS Microdata is made available through UMD’s Global CTIS Microdata Repository to researchers upon request. Separately, an R interface to access UMD’s Open Data and Microdata APIs is made available as an R package (<https://github.com/am0505/ctisglobal>). While CTIS open data estimates are made publicly available for other countries, we restrict our analysis to India. Administrative data for states and national level in India on COVID-19 is obtained from COVID19Bharat (<https://covid19bharat.org/>). Data on state and national level population by subgroups is obtained from Population Projections Report by Census of India 2011 (https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf).

Code Availability

Code used to produce these analyses is available on GitHub at https://github.com/am0505/ctis_paper.

References

- Adjodah, Dhaval, Karthik Dinakar, Matteo Chinazzi, Samuel P. Fraiberger, Alex Pentland, Samantha Bates, Kyle Staller, Alessandro Vespignani, and Deepak L. Bhatt. 2021. “Association Between COVID-19 Outcomes and Mask Mandates, Adherence, and Attitudes.” *PLOS ONE* 16 (6): e0252315. <https://doi.org/10.1371/journal.pone.0252315>.
- Babalola, S, and S Krenn. 2021. “COVID Behaviors Dashboard. Johns Hopkins Center for Communication Programs in Collaboration with Facebook Data for Good, Delphi Group at Carnegie Mellon University, University of Maryland Social Data Science Center, Global Outbreak Alert and Response Network.” <https://covidbehaviors.org/>.
- Baker, Rachel E., Wenchang Yang, Gabriel A. Vecchi, C. Jessica E. Metcalf, and Bryan T. Grenfell. 2021. “Assessing the Influence of Climate on Wintertime SARS-CoV-2 Outbreaks.” *Nature Communications* 12 (1): 846. <https://doi.org/10.1038/s41467-021-20991-1>.
- Bethlehem, Jelke. 2010. “Selection Bias in Web Surveys.” *International Statistical Review / Revue Internationale de Statistique* 78 (2): 161–88. <https://www.jstor.org/stable/27919830>.
- Bhalla, Surjit, and Tirthatanmoy Das. n.d. “?population, Education, and Employment in India: 1983-2018?” https://scholar.google.com/citations?view_op=view_citation&hl=en&user=J-EOoXsAAAAJ&citation_for_view=J-EOoXsAAAAJ:u5HHmVD_uO8C.
- Bradley, Valerie C., Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng, and Seth Flaxman. 2021. “Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake.” *Nature*, December, 1–6. <https://doi.org/10.1038/s41586-021-04198-4>.
- Carfi, Angelo, Roberto Bernabei, Francesco Landi, and Gemelli Against COVID-19 Post-Acute Care Study Group. 2020. “Persistent Symptoms in Patients After Acute COVID-19.” *JAMA* 324 (6): 603–5. <https://doi.org/10.1001/jama.2020.12603>.
- Chowdhury, Soumi Roy, Abhinav Motheram, and Santanu Pramanik. n.d. “Covid-19 Vaccine Hesitancy: Trends Across States, over Time.” <http://www.ideasforindia.in/topics/human-development/covid-19-vaccine-hesitancy-trends-across-states-over-time.html>.

- Correspondent, Legal. 2021. “On CoWIN, Supreme Court Flags Digital Divide.” *The Hindu*, June. <https://www.thehindu.com/news/national/on-cowin-supreme-court-flags-digital-divide/article34711169.ece>.
- Díaz-Pachón, Daniel Andrés, and J. Sunil Rao. 2021. “A simple correction for COVID-19 sampling bias.” *Journal of Theoretical Biology* 512 (March): 110556. <https://doi.org/10.1016/j.jtbi.2020.110556>.
- Karlinsky, Ariel, and Dmitry Kobak. 2021. “Tracking Excess Mortality Across Countries During the COVID-19 Pandemic with the World Mortality Dataset.” Edited by Miles P Davenport, Marc Lipsitch, Marc Lipsitch, Lone Simonsen, and Ayesha Mahmud. *eLife* 10 (June): e69336. <https://doi.org/10.7554/eLife.69336>.
- Kreuter, Frauke. 2021. “What Surveys Really Say.” *Nature*, December. <https://doi.org/10.1038/d41586-021-03604-1>.
- Liu, Xiaoyue, Jianping Huang, Changyu Li, Yingjie Zhao, Danfeng Wang, Zhongwei Huang, and Kehu Yang. 2021. “The Role of Seasonality in the Spread of COVID-19 Pandemic.” *Environmental Research* 195 (April): 110874. <https://doi.org/10.1016/j.envres.2021.110874>.
- Malani, Anup, and Sabareesh Ramachandran. 2021. “Using Household Rosters from Survey Data to Estimate All-Cause Mortality During COVID in India.” <https://www.nber.org/papers/w29192>.
- Menni, Cristina, Ana M. Valdes, Maxim B. Freidin, Carole H. Sudre, Long H. Nguyen, David A. Drew, Sajaysurya Ganesh, et al. 2020. “Real-Time Tracking of Self-Reported Symptoms to Predict Potential COVID-19.” *Nature Medicine* 26 (7): 1037–40. <https://doi.org/10.1038/s41591-020-0916-2>.
- Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. “Household Surveys in Crisis.” <https://www.nber.org/papers/w21399>.
- Phillips, Nicky. 2021. “The Coronavirus Is Here to Stay — Here’s What That Means.” *Nature* 590 (7846): 382–84. <https://doi.org/10.1038/d41586-021-00396-2>.
- Reinhart, Alex. n.d. “CTIS: The COVID-19 Trends and Impact Survey,” 40.
- Salomon, Joshua A., Alex Reinhart, Alyssa Bilinski, Eu Jing Chua, Wichada La Motte-Kerr, Minttu M. Rönn, Marissa B. Reitsma, et al. 2021. “The US COVID-19 Trends and Impact Survey: Continuous Real-Time Measurement of COVID-19 Symptoms, Risks, Protective Behaviors, Testing, and Vaccination.” *Proceedings of the National Academy of Sciences* 118 (51). <https://doi.org/10.1073/pnas.2111454118>.
- Serikbayeva, Balzhan, Kanat Abdulla, and Yessengali Oskembayev. 2021. “State Capacity in Responding to COVID-19.” *International Journal of Public Administration* 44 (11-12): 920–30. <https://doi.org/10.1080/01900692.2020.1850778>.
- Shah, Ajay. 2021. “Indian Health Policy in the Light of Covid-19: The Puzzles of State Capacity and Institutional Design.” <https://ideas.repec.org/p/anf/wpaper/4.html>.
- Shaman, Jeffrey, and Marta Galanti. 2020. “Will SARS-CoV-2 Become Endemic?” *Science*, October. <https://doi.org/10.1126/science.abe5960>.
- Sharma, Rohit. 2021. “India: Digital Divide and the Promise of Vaccination for All.” <https://blogs.lse.ac.uk/southasia/2021/06/28/india-digital-divide-and-the-promise-of-vaccination-for-all/>.
- Somanchi, Anmol. 2021. “Missing the Poor, Big Time: A Critical Assessment of the Consumer Pyramids Household Survey.” <https://ideas.repec.org/p/osf/socarx/qmce9.html>.
- Sudre, Carole H., Benjamin Murray, Thomas Varsavsky, Mark S. Graham, Rose S. Penfold, Ruth C. Bowyer, Joan Capdevila Pujol, et al. 2021. “Attributes and Predictors of Long COVID.” *Nature Medicine* 27 (4): 626–31. <https://doi.org/10.1038/s41591-021-01292-y>.
- Sukumaran, Rohan, Parth Patwa, Sethuraman T. V, Sheshank Shankar, Rishank Kanaparti, Joseph Bae, Yash Mathur, et al. 2021. “COVID-19 Outbreak Prediction and Analysis Using Self Reported Symptoms.” *Journal of Behavioral Data Science* 1 (1): 154–69. <https://doi.org/10.35566/jbds/v1n1/p8>.
- Tsuboi, Satoshi, Honami Yoshida, Ryusuke Ae, Takao Kojo, Yosikazu Nakamura, and Kunio Kitamura. 2015. “Selection bias of Internet panel surveys: a comparison with a paper-based survey and national governmental statistics in Japan.” *Asia-Pacific Journal of Public Health* 27 (2): NP2390–2399. <https://doi.org/10.1177/1010539512450610>.
- Wälde, Klaus. n.d. “How to Remove the Testing Bias in CoV-2 Statistics.” <https://www.iza.org/publications/dp/13785/how-to-remove-the-testing-bias-in-cov-2-statistics>.
- Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. “Forecasting Elections with Non-Representative Polls.” *International Journal of Forecasting* 31 (3): 980–91. <https://doi.org/10.1016/j.ijforecast.2014.06.001>.