Estimating COVID-19 related mortality in India: An epidemiological challenge with insufficient data

Lauren V. Zimmermann[1,2], Maxwell Salvatore [2,3], Giridhara R. Babu[4], and Bhramar Mukherjee[1,2,3]

1. Center for Precision Health Data Science, University of Michigan, Ann Arbor, MI

2. Department of Biostatistics, University of Michigan, Ann Arbor, MI

3. Department of Epidemiology, University of Michigan, Ann Arbor, MI

4. Life course Epidemiology Unit, Indian Institute of Public Health – Bengaluru, Public Health Foundation of India, Bengaluru, Karnataka

Abstract: The harrowing second wave of COVID-19 in India has led to much discussion over the quality and timeliness of reporting of deaths attributed to the pandemic. In this brief report, we aim to present the existing evidence, as well as the broader complexities surrounding the mortality burden of COVID-19 in India. This article sheds light on the following epidemiological issues: (1) general and India-specific challenges to COVID-19 death reporting, (2) latest COVID-19 mortality estimates in India as of May 16, 2021, (3) the apparent scale of uncaptured COVID-19 deaths, and (4) the role of disaggregated historic mortality trends in quantification of excess deaths attributed to COVID-19. We conclude with a set of high-level policy recommendations for improving the vital surveillance system and tracking of causes of death in India. We encourage direct efforts to integrate health data and indirect strategies for cross-validation of registered deaths. Such system-wide advances would drastically aid epidemiological research efforts and strengthen India's position to overcome future public health crises.

Background: As of May 16, 2021, India, a country with a population of 1.38 billion—was second only to the United States in the total number of reported SARS-CoV-2 cases (nearly 25 million) and third following the U.S. and Brazil in total reported deaths (over 270 thousand) [1]. Data from seroprevalence studies and limited excess mortality calculations offer evidence that the actual number of infections and deaths are likely much larger than the ones reported [2, 3].

We recognize that multiple challenges lead to underreporting of COVID-19 fatalities including: (1) deaths that occur outside of hospitals either are not captured or incur a lag, (2) deaths that are classified under comorbid illnesses, (3) deaths that are due to low access to quality healthcare and/or a shortage of healthcare resources, (4) deaths that are undetected as a result of an inadequate COVID-19 testing program. Our review of the existing evidence suggests that the problem is particularly acute for India, where a large number of deaths (especially ones happening outside a healthcare facility and/or in rural areas) routinely remain medically unreported [4].

Current Estimates: We report these numbers as of May 16, 2021. The overall case fatality rate (CFR) in India has remained low (1.09%) relative to estimates from other countries (1.77% in the United States, 2.78% in Brazil, 2.07% globally) [1]. However, India has a young population (e.g.,

proportion of the population 65+ is 6.4% in India versus 9.3% in Brazil and 16.5% in the U.S.) and as such age-specific mortality comparisons are more meaningful. The first and second waves of the pandemic in India are characteristically different in terms of both infections and deaths. The CFR for Wave 1 is 1.4%, while the CFR for Wave 2 is currently 0.8%. Some state-specific numerical estimates are presented in Table 1. It is hypothesized that the reduced CFR in Wave 2 is due to underreporting, pending data reconciliation from diverse sources, and a large number of infections in younger age groups with a lower risk for severe clinical presentation of SARS-CoV-2, an assertion yet to be verified [5].

Underreporting of Infections and Deaths: To estimate infection fatality rates (IFR), we, among other researchers, have used epidemiological models and seroprevalence surveys [6]. Such models [7] indicate the underreporting factor is around 10-20 for cases and around 2-5 for deaths, based on data from Wave 1 in India. According to these studies, the IFR for India is roughly 0.1% using observed death counts and 0.4% after incorporating underreporting of deaths (Table 1). The former resembles early estimates for Mumbai, Srinagar, and Karnataka using observed fatalities (0.09%, 0.06%, and 0.05%, respectively) [8, 9].

We note that anecdotal and media reports corroborate model estimates. For example, during Wave 1, a group of volunteers collected reported deaths from obituaries in newspapers and found the death count to be almost twice that officially reported [10]. Likewise, during this recent surge, a New York Times article noted that authorities in Gujarat reported between 73-121 daily COVID-19 related deaths in mid-April, contradicting a leading newspaper in Gujarat that cited the number was several times higher (around 610 daily deaths) [11]. Recently, an excess death calculation based on comparing death certificates issued in the state of Gujarat [12] showed that while the state reported 4,218 COVID-19 deaths during March 1-May 10, 2021 an estimated 61,000 excess deaths remained uncounted, indicating an underreporting factor of nearly 15. Moreover, comparisons to past years of satellite images revealing fires emitting from burial pyres has imprinted the sheer scale of additional lives lost to the pandemic in April 2021.

Unique Features: Due to delayed detection, the proportion of COVID-19 deaths with a narrow time to death window (from the date of confirmed diagnosis) is higher in select regions compared to the global findings. For example, a study found a considerable 18% of deaths across the states of Tamil Nadu and Andhra Pradesh occurred within 24 hours of diagnosis [13], suggesting a substantial lag in the initial diagnosis of COVID-19 compared to other countries. In Wave 2, a strained health system, a deficit of ICU beds, and inadequate oxygen monitoring for at-home isolation has collectively exacerbated this issue. The CFRs in India vary considerably across states (e.g., among large states, Kerala has the lowest and Punjab has the highest case fatality rates). This geographical heterogeneity is also reflected in the (albeit limited) regional excess death calculations available for 2020 [14].

Data Paucity: India, unlike other countries, does not have robust data that can be used for analysis [15]. The Ministry of Health and Family Welfare shared age and sex-disaggregated COVID-19 related data at the start of the pandemic, but the officials stopped reporting this information. We only have access to sporadic release of charts and tables in briefings and media reports. We join the research community in calling for these data as well as information on comorbidities, which are necessary to track age-sex specific trends, to identify high-risk subpopulations, and to validate hypotheses regarding rates of infections, severe cases, and deaths within subgroups of interest.

In terms of longevity and cause of death, India's most recent reporting of life expectancy and all-cause mortality estimates are from 2014-2018 and 2010-2013, respectively, precluding any meaningful, timely study of all-cause or excess mortality. According to the latest global excess mortality study (January 2021), 77 countries report data on all-cause mortality, enabling experts to compute country-specific excess mortality, which is largely considered the gold standard for estimating the burden of COVID-19 [15]. India is a notable exception [15]; in our opinion, the release of these figures is sorely needed.

Impact of Insufficient Data: Deficiency in the COVID-19 death reporting has harmful ramifications. It limits modelers' ability to predict the course of the pandemic, gauge its impact, and estimate healthcare resource needs—including oxygen supplies and hospital beds. This data deficient environment stunts overall policy efforts to improve public health outcomes and healthcare infrastructure. Without disaggregated epidemiological data, linked with genomic sequencing, assessing the lethality of virus strains and evaluating vaccine effectiveness becomes nearly impossible.

Recommendations Moving Forward: We offer general recommendations herein for systematizing the collection and advancing the quality of all-cause and disease-specific mortality in India. The Indian government recently announced a pilot trial of a personal digital health identifier, which would ultimately serve as an electronic key to a health data repository for each individual nationwide [16]. Integrating data across health systems offers a solution to capturing all-cause mortality in a more nationally representative way. With successful implementation, and multi-platform linkages, a digital health ID would enable comprehensive analysis of healthcare outcomes via continuous reporting and a breadth of available data.

Heterogenous data linkage holds promise for approximating unreported deaths, such as through tracking inactive Aadhaar cards (akin to social security cards in the U.S.), bank accounts, phone numbers, and social media accounts. Inspection of life insurance claims may also complement indirect validation efforts. Innovative strategies for surveillance using community healthcare and Accredited Social Health Activist (ASHA) workers are needed in rural India, where a proper reporting system is largely absent. We recommend strengthening the Civil Registration System (CRS) by leveraging community engagement and partnerships as well as collaborating with community and religious leaders to encourage prompt reporting by family members of the deceased. We need continued attention to medical certification of deaths, and mandatory linking to the CRS for India to meet international standards. Death not being reported reflect dishonor to the entire life of a person. When not captured and analyzed, the existing health inequities are further exacerbated. A fortified nationwide vital surveillance system, as well as timely and comprehensive data reporting, is at the heart of fighting this pandemic. An investment in a robust data ecosystem now will help safeguard India against future health crises.

Table 1. Summary of COVID-19 attributed mortality in India during 2020-21.

Part A: State-level results from the latest issued mortality estimates

State[a]	As of May 16, 2021[b]			As of January 31, 2021 [c,2]		Excess Deaths[d-g,3]
	Number of Cases Reported	Number of Deaths Reported	Case Fatality Ratio (CFR)[1]	Infection Fatality Rate (IFR)	Adjusted-Infection Fatality Rate (IFR)	
India	24,965,079	274,417	1.09%	0.13%	0.46%	*N/A*
Maharashtra	5,378,452	81,486	1.51%	0.46%	0.97%	*N/A*−Mumbai: *approx.* 21,000
Punjab	497,705	11,895	2.38%	0.36%	1.01%	*N/A*
Gujarat	752,619	9,121	1.21%	0.28%	0.59%	*approx.* 61,000
Karnataka	2,203,462	21,837	0.99%	0.17%	0.43%	*N/A*−Bengaluru: 10,248
Kerala	2,147,968	6,429	0.29%	0.06%	0.14%	*approx.* (-16,000)
Delhi	1,393,867	21,506	1.54%	0.06%	0.38%	*N/A*−New Delhi: *approx.* (-5,800)

Part B: State-level comparison of first and second waves[h]

State[a]	Wave 1: March 24, 2020 – February 14, 2021			Wave 2: February 15, 2021 – May 15, 2021		
	Number of Cases Reported	Number of Deaths Reported	Case Fatality Ratio (CFR)[1]	Number of Cases Reported	Number of Deaths Reported	Case Fatality Ratio (CFR)[1]
India	10,915,905	155,169	1.42%	13,766,623	114,550	0.83%
Maharashtra	2,064,181	51,526	2.50%	3,279,785	28,983	0.88%
Punjab	176,275	5,696	3.23%	314,457	5,996	1.91%
Gujarat	265,213	4,399	1.66%	479,165	4,638	0.97%
Karnataka	945,237	12,271	1.30%	1,226,661	9,169	0.75%
Kerala	1,004,041	3,986	0.40%	1,114,128	2,354	0.21%
Delhi	636,916	10,890	1.71%	750,465	10,353	1.38%

(a) The states that have issued excess deaths calculations as of May 16, 2021 (and Punjab) are included in this table.

(b) India COVID-19 Tracker. Available at: https://www.covid19india.org/. Accessed May 16, 2021.

(c) Purkayashtha, S., Kundu, R., & Bhaduri, R., Mukherjee B. (2021). Estimating the Infection Fatality Rate from SARS-CoV-2 in India. doi:10.2139/ssrn.3798552

(d) Annual excess deaths for the City of Mumbai and the State of Kerala are for 2020, respectively (S, R. (2021, April 11). Can New Mortality Data Explain India's Low COVID Death Numbers? *Indiaspend*. Available at: https://www.indiaspend.com/covid-19/mortality-data-kerala-mumbai-too-soon-to-say-india-covid19-less-deadly-second-wave-737270. Accessed May 16, 2021).

(e) Excess deaths for the State of Gujarat are from March 1 to May 10, 2021 (Scroll Staff (2021, May 14). Gujarat is undercounting Covid-19 deaths, shows 'Divya Bhaskar' report. *Scroll.In*. Available at: https://scroll.in/latest/994906/gujarat-is-undercounting-covid-19-deaths-shows-divya-bhaskar-report. Accessed May 16, 2021).

(f) Excess deaths for the City of Bengaluru are from January to July 2020 (Correspondent, S. (2020, September 2). 32% increase in mortality rate in Bengaluru: Patil. *The Hindu*. Available at: https://www.thehindu.com/news/national/karnataka/32-increase-in-mortality-rate-in-bengaluru-patil/article32509298.ece. Accessed May 16, 2021).

(g) Excess deaths for the City of New Delhi are from April to June 2020 (Joshi, M. (2020, July 18). Over 21,000 deaths registered in capital in April-June quarter, 5,800 less than in 2019. *The Indian Express*. Available at: https://indianexpress.com/article/cities/delhi/over-21000-deaths-registered-in-capital-in-april-june-quarter-5800-less-than-in-2019-6511345/lite/?__twitter_impression=true. Accessed May 16, 2021).

(h) Wave 1 is defined as starting from March 24, 2020, when the first nationwide lockdown was implemented in India. Wave 2 is defined as starting from February 15, 2021, when the national effective reproduction number for COVID-19 in India crossed unity. Estimates for Wave 2 are computed through May 15, 2021, which is the latest available data at the time of this report (© COV–IND–19 Study Group. Source: covid19india.org).

(1) CFR is the number of reported deaths divided by the number of reported infected cases.

(2) IFR and Adjusted-IFR are estimates from an extended SEIR model, where Adjusted-IFR accounts for under-reporting of COVID-19 deaths.

(3) Excess death calculations vary across regions, as approaches depend on underlying assumptions regarding the number of expected deaths. The general framework includes obtaining the difference between the observed death count and the average expected death count, as derived from previous years.

1. About the Authors: Lauren V. Zimmermann is a MS student with the Department of Biostatistics, Center for Precision Health Data Science, University of Michigan, Ann Arbor, MI. Maxwell Salvatore is a PhD student with the Departments of Epidemiology and Biostatistics, University of Michigan, Ann Arbor. Giridhara R. Babu is with the Life course Epidemiology Unit, Indian Institute of Public Health – Bengaluru, Public Health Foundation of India, Bengaluru, Karnataka. Bhramar Mukherjee is with the Departments of Epidemiology and Biostatistics, University of Michigan, Ann Arbor. She is also affiliated with the Center for Precision Health Data Science, University of Michigan, Ann Arbor.

2. Corresponding Author Contact Information: Bhramar Mukherjee, PhD, M4208 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109 (e-mail: bhramar@umich.edu).

3. Acceptance Date: May 21, 2021 12:44PM.

4. Contributor Statement: All authors contributed to the research, writing, editing, and review of this commentary. M. Salvatore and L.V. Zimmermann performed the numerical analysis. B. Mukherjee guided overall execution.

5. Acknowledgments: None.

6. Human Participant Protection: Not applicable; no IRB approval was required for this report.

References

1. Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, 20(5), 533-534.
2. Bhattacharyya, R., Kundu, R., Bhaduri, R., Ray, D., Beesley, L. J., Salvatore, M., & Mukherjee, B. (2021). Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seroprevalence estimates. *Scientific Reports, 11(1)*, 1-14. doi:10.1038/s41598-021-89127-1
3. Chatterjee, P. (2020). Is India missing COVID-19 deaths? *The Lancet, 396*(10252), 657. doi:10.1016/s0140-6736(20)31857-2
4. Mahapatra, P., & Chalapati Rao, P. V. (2001). Cause of death reporting systems in India: a performance analysis. *National Medical Journal of India, 14(3),* 154-162.
5. Joshi, S. (2021, May 7). More than 18,000 Covid deaths in Maharashtra during second wave, fatality rate lower than first wave. India Today. Available at: https://www.indiatoday.in/coronavirus-outbreak/story/covid-deaths-maharashtra-second-wave-fatality-rate-1799717-2021-05-07. Accessed May 16, 2021.
6. Cai, R., Novosad, P., Tandel, V., Asher, S., & Malani, A. (2021). Representative Estimates of COVID-19 Infection Fatality Rates from Three Locations in India. medRxiv. doi:10.2139/ssrn.3783787
7. Bhaduri, R., Kundu, R., Purkayastha, S., Kleinsasser, M., Beesley, L. J., & Mukherjee, B. (2020). Extending the Susceptible-Exposed-Infected-Removed (SEIR) model to handle the high false negative rate and symptom-based administration of COVID-19 diagnostic tests: SEIR-fansy. *Medrxiv*. doi:10.1101/2020.09.24.20200238

8. Ioannidis, J. P. A. (2020). Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bulletin of the World Health Organization*, *99*(1), 19–33F. doi:10.2471/blt.20.265892

9. Babu GR, Sundaresan R, Athreya S, Akhtar J, Pandey PK, Maroor SP, et al. (2020). The burden of active infection and anti-SARS-CoV-2 IgG antibodies in the general population: Results from a statewide survey in Karnataka, India. *medRxiv*. doi:10.1101/2020.12.04.20243949

10. Biswas, S. (2020, November 20). India coronavirus: How a group of volunteers "exposed" hidden Covid-19 deaths. *BBC News*. Available at: https://www.bbc.com/news/world-asia-india-54985981. Accessed May 16, 2021.

11. Gettleman, J., Yasir, S., Kumar, H., Raj, S., & Loke, A. (2021, May 13). As Covid-19 Devastates India, Deaths Go Undercounted. *The New York Times*. Available at: https://www.nytimes.com/2021/04/24/world/asia/india-coronavirus-deaths.html. Accessed May 16, 2021.

12. Scroll Staff (2021, May 16). Gujarat is undercounting Covid-19 deaths, shows 'Divya Bhaskar' report. *Scroll.In.* Available at: https://scroll.in/latest/994906/gujarat-is-undercounting-covid-19-deaths-shows-divya-bhaskar-report. Accessed May 16, 2021.

13. Laxminarayan, R., Wahl, B., Dudala, S. R., Gopal, K., Neelima, S., Reddy, K. J., ... & Lewnard, J. A. (2020). Epidemiology and transmission dynamics of COVID-19 in two Indian states. Science, 370(6517), 691-697. doi:10.1126/science.abd7672

14. S, R. (2021, April 11). Can New Mortality Data Explain India's Low COVID Death Numbers? *Indiaspend*. Available at: https://www.indiaspend.com/covid-19/mortality-data-kerala-mumbai-too-soon-to-say-india-covid19-less-deadly-second-wave-737270. Accessed May 16, 2021.

15. Karlinsky, A., & Kobak, D. (2021). The World Mortality Dataset: Tracking excess mortality across countries during the COVID-19 pandemic. medRxiv: the preprint server for health sciences. doi:10.1101/2021.01.27.21250604

16. TIMESOFINDIA.COM. (2020, August 15). PM Modi announces digital health ID: All you need to know. *The Times of India*. Available at: https://timesofindia.indiatimes.com/india/pm-modi-announces-digital-health-id-all-you-need-to-know/articleshow/77565257.cms. Accessed May 16, 2021.