# Estimating COVID-19–Related Mortality in India: An Epidemiological Challenge With Insufficient Data

*Lauren V. Zimmermann, BS, Maxwell Salvatore, MS, Giridhara R. Babu, PhD, MPH, and Bhramar Mukherjee, PhD*

**ABOUT THE AUTHORS**

*Lauren V. Zimmermann is an MS student with the Department of Biostatistics, Center for Precision Health Data Science, University of Michigan, Ann Arbor. Maxwell Salvatore is a PhD student with the Departments of Epidemiology and Biostatistics, University of Michigan. Giridhara R. Babu is with the Life Course Epidemiology Unit, Indian Institute of Public Health—Bengaluru, Public Health Foundation of India, Bengaluru, Karnataka, India. Bhramar Mukherjee is with the Departments of Epidemiology and Biostatistics and the Center for Precision Health Data Science, University of Michigan.*

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 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) cases (nearly 25 million) and third following the United States and Brazil in total reported deaths (> 270 000).[1] Data from seroprevalence studies and limited excess mortality calculations offer evidence that the actual numbers 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 that either are not captured or incur a lag, (2) deaths that are classified under comorbid illnesses, (3) deaths that are attributable to low access to quality health care or a shortage of health care resources, and (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 health care facility 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 aged ≥ 65 years is 6.4% in India vs 9.3% in Brazil and 16.5% in the United States) 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%, and 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 attributable 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 (IFRs), we, among other researchers, have used epidemiological models and seroprevalence surveys.[6] Such models[7] indicate that the underreporting factor is around 10 to 20 for cases and around 2 to 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 2). The former resembles early estimates for Mumbai, Srinagar, and Karnataka using observed fatalities (0.09%, 0.06%, and 0.05%, respectively).[11,12]

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.[13] Likewise, during this recent surge, a *New York Times* article noted that authorities in Gujarat reported between 73 and 121 daily COVID-19–related deaths in mid-April, contradicting a leading newspaper in Gujarat that cited the number as several times higher (around 610 daily deaths).[10] Recently, an excess death

**TABLE 1— State-Level Comparison of COVID-19 Attributed Mortality for the First and Second Waves in India: March 24, 2020–February 14, 2021, and February 15, 2021–May 15, 2021**

| State | Wave 1: Mar 24, 2020–Feb 14, 2021[a] | | Wave 2: Feb 15, 2021–May 15, 2021[a] | |
|---|---|---|---|---|
| | No. of Cases Reported | No. of Deaths Reported (CFR)[b] | No. of Cases Reported | No. of Deaths Reported (CFR)[b] |
| 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) |

*Note.* CFR = case fatality rate.
*Source*. COVID19INDIA (https://bit.ly/3waJWzc).
[a]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.
[b]CFR is the number of reported deaths divided by the number of reported infected cases.

calculation based on comparing death certificates issued in the state of Gujarat[14] showed that while the state reported 4218 COVID-19–related deaths during March 1 to 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

As a result of delayed detection, the proportion of COVID-19–related deaths within a narrow time-to-death window (from the date of confirmed diagnosis) was higher in select regions compared with 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,[9] suggesting a substantial lag in the initial diagnosis of COVID-19 compared with other countries. In wave 2, a strained health system, a deficit of intensive care unit beds, and inadequate oxygen monitoring for at-home isolation have 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 CFRs). This geographical heterogeneity is also reflected in the (albeit limited) regional excess death calculations available for 2020.[15]

## DATA PAUCITY

India, unlike other countries, does not have robust and readily available mortality data that can be used for analysis.[16] The Ministry of Health and Family Welfare shared age- and sex-disaggregated COVID-19–related data at the start of the pandemic, but the official reporting of this information quickly stopped. 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- and 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 to 2018 and 2010 to 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.[16] India is a notable exception[16]; 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 health care resource needs—including oxygen supplies and hospital beds. Without disaggregated epidemiological data, linked with genomic sequencing, assessing the lethality of virus strains and evaluating vaccine effectiveness becomes nearly impossible. This data-deficient environment stunts overall policy efforts to improve public health outcomes and

**TABLE 2— State-Level Summary of the Latest Issued COVID-19 Attributed Mortality in India During 2020–2021**

| State[a] | As of May 16, 2021[b] | | As of Jan 31, 2021[c,d] | | Excess Deaths[e–i] |
|---|---|---|---|---|---|
| | No. of Cases Reported | No. of Deaths Reported (CFR)[j] | IFR | Adjusted IFR | |
| India | 24 965 079 | 274 417 (1.09) | 0.13 | 0.46 | NA |
| Maharashtra | 5 378 452 | 81 486 (1.51) | 0.46 | 0.97 | NA—Mumbai: ~21 000 |
| Punjab | 497 705 | 11 895 (2.38) | 0.36 | 1.01 | NA |
| Gujarat | 752 619 | 9 121 (1.21) | 0.28 | 0.59 | ~61 000 |
| Karnataka | 2 203 462 | 21 837 (0.99) | 0.17 | 0.43 | NA—Bengaluru: 10 248 |
| Kerala | 2 147 968 | 6 429 (0.29) | 0.06 | 0.14 | ~–16 000 |
| Delhi | 1 393 867 | 21 506 (1.54) | 0.06 | 0.38 | NA—New Delhi: ~–5 800 |

*Note.* CFR = case fatality rate; IFR = infection fatality rate; NA = not available.
[a]The states that have issued excess deaths calculations as of May 16, 2021 (and Punjab) are included in this table.
[b]COVID19INDIA (https://bit.ly/3waJWzc).
[c]Purkayashtha et al.[8]
[d]IFR and adjusted IFR are estimates from an extended susceptible–exposed–infected–removed (SEIR) model, where adjusted IFR accounts for underreporting of COVID-19 deaths.
[e]Annual excess deaths for the city of Mumbai and the state of Kerala are both for 2020.[9]
[f]Excess deaths for the state of Gujarat are from March 1 to May 10, 2021.[10]
[g]Excess deaths for the city of Bengaluru are from January to July 2020 (https://bit.ly/351BV3V).
[h]Excess deaths for the city of New Delhi are from April to June 2020 (https://bit.ly/3gqcSNj).
[i]Excess deaths 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.
[j]CFR is the number of reported deaths divided by the number of reported infected cases.

health care infrastructure for the future. For example, without knowing who are dying and the magnitude of COVID-19–related fatalities, one cannot design social and economic policies to protect the vulnerable and support the families left behind.

## RECOMMENDATIONS MOVING FORWARD

We offer general recommendations herein for systematizing the collection and advancing the quality of all-cause and disease-specific mortality data 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.[17] Integrating data across health systems offers a solution to capturing all-cause mortality in a more nationally representative way. With successful implementation, and multiplatform linkages, a digital health identifier would enable comprehensive analysis of health care outcomes via continuous reporting and a breadth of available individual-level data.

Heterogenous data linkage holds promise for approximating unreported deaths, such as through tracking inactive Aadhaar cards (akin to Social Security cards in the United States), bank accounts, phone numbers, and social media accounts. Inspection of life insurance claims may also complement indirect validation efforts in urban areas. Innovative strategies for surveillance using community health care and Accredited Social Health Activist workers are needed in rural India, where a proper reporting system is largely absent. We recommend strengthening the Civil Registration System 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 Civil Registration System for India to meet international standards. Death not being reported reflects dishonor to the entire life of a person. When not captured and analyzed in a timely manner, the existing health inequities are further exacerbated. A fortified nationwide vital surveillance system, as well as timely and comprehensive data reporting and cogent analysis, is at the heart of fighting this pandemic. An investment in a robust data ecosystem now will help safeguard India against future health crises. AJPH


## CORRESPONDENCE

Correspondence should be sent to Bhramar Mukherjee, PhD, M4208 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109 (e-mail: bhramar@umich.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.



## PUBLICATION INFORMATION

Full Citation: 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;111(S2):S59–S62.

Acceptance Date: May 21, 2021.

DOI: https://doi.org/10.2105/AJPH.2021.306419


## CONTRIBUTORS

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.

## CONFLICTS OF INTEREST

The authors report no conflicts of interest.

## REFERENCES


1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20(5):533–534. https://doi.org/10.1016/S1473-3099(20)30120-1
2. Bhattacharyya R, Kundu R, Bhaduri R, et al. Incorporating false negative tests in epidemiological models for SARS-CoV-2 transmission and reconciling with seroprevalence estimates. *Sci Rep.* 2021;11(1):9748. https://doi.org/10.1038/s41598-021-89127-1
3. Chatterjee P. Is India missing COVID-19 deaths? *Lancet.* 2020;396(10252):657. https://doi.org/10.1016/S0140-6736(20)31857-2
4. Mahapatra P, Chalapati Rao PV. Cause of death reporting systems in India: a performance analysis. *Natl Med J India.* 2001;14(3):154–162.
5. Joshi S. More than 18,000 Covid deaths in Maharashtra during second wave, fatality rate lower than first wave. *India Today*. May 7, 2021. 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. Representative estimates of COVID-19 infection fatality rates from three locations in India. *medRxiv*. 2021; Epub ahead of print. https://doi.org/10.2139/ssrn.3783787
7. Bhaduri R, Kundu R, Purkayastha S, Kleinsasser M, Beesley LJ, Mukherjee B. 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.* 2020; Epub ahead of print. https://doi.org/10.1101/2020.09.24.20200238
8. Purkayashtha S, Kundu R, Bhaduri R, Mukherjee B. Estimating the infection fatality rate from SARS-CoV-2 in India. Preprint. Posted online March 7, 2021. Available at: papers.ssrn.com/sol3/papers.cfm?abstract_id=3798552. Accessed June 11, 2021.
9. Laxminarayan R, Wahl B, Dudala SR, et al. Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science.* 2020;370(6517):691–697. https://doi.org/10.1126/science.abd7672
10. Gettleman J, Yasir S, Kumar H, Raj S, Loke A. As Covid-19 devastates India, deaths go undercounted. *New York Times*. May 13, 2021. Available at: https://www.nytimes.com/2021/04/24/world/asia/india-coronavirus-deaths.html. Accessed May 16, 2021.
11. Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bull World Health Organ.* 2021;99(1):19–33F. https://doi.org/10.2471/BLT.20.265892
12. Babu GR, Sundaresan R, Athreya S, et al. 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.* 2020; Epub ahead of print. https://doi.org/10.1101/2020.12.04.20243949
13. Biswas S. India coronavirus: how a group of volunteers "exposed" hidden Covid-19 deaths. *BBC News*. November 20, 2020. Available at: https://www.bbc.com/news/world-asia-india-54985981. Accessed May 16, 2021.
14. Scroll Staff. Gujarat is undercounting Covid-19 deaths, shows "Divya Bhaskar" report. *Scroll.In.* May 16, 2021. Available at: https://scroll.in/latest/994906/gujarat-is-undercounting-covid-19-deaths-shows-divya-bhaskar-report. Accessed May 16, 2021.
15. Rukmini S. Can new mortality data explain India's low COVID death numbers? *Indiaspend*. April 11, 2021. 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.
16. Karlinsky A, Kobak D. The world mortality dataset: tracking excess mortality across countries during the COVID-19 pandemic. *medRxiv*. 2021; Epub ahead of print. https://doi.org/10.1101/2021.01.27.21250604
17. PM Modi announces digital health ID: all you need to know. *Times of India*. August 15, 2020. 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.