BMC Research Notes

RESEARCH NOTE

Open Access

Estimating the wave 1 and wave 2 infection fatality rates from SARS-CoV-2 in India

Soumik Purkayastha[1], Ritoban Kundu[2], Ritwik Bhaduri[2], Daniel Barker[1], Michael Kleinsasser[1], Debashree Ray[3,4] and Bhramar Mukherjee[1,5,6*]

Abstract

Objective: There has been much discussion and debate around the underreporting of COVID-19 infections and deaths in India. In this short report we first estimate the underreporting factor for infections from publicly available data released by the Indian Council of Medical Research on reported number of cases and national seroprevalence surveys. We then use a compartmental epidemiologic model to estimate the undetected number of infections and deaths, yielding estimates of the corresponding underreporting factors. We compare the serosurvey based ad hoc estimate of the infection fatality rate (IFR) with the model-based estimate. Since the first and second waves in India are intrinsically different in nature, we carry out this exercise in two periods: the first wave (April 1, 2020–January 31, 2021) and part of the second wave (February 1, 2021–May 15, 2021). The latest national seroprevalence estimate is from January 2021, and thus only relevant to our wave 1 calculations.

Results: Both wave 1 and wave 2 estimates qualitatively show that there is a large degree of "covert infections" in India, with model-based estimated underreporting factor for infections as 11.11 (95% credible interval (CrI) 10.71–11.47) and for deaths as 3.56 (95% CrI 3.48–3.64) for wave 1. For wave 2, underreporting factor for infections escalate to 26.77 (95% CrI 24.26–28.81) and to 5.77 (95% CrI 5.34–6.15) for deaths. If we rely on only reported deaths, the IFR estimate is 0.13% for wave 1 and 0.03% for part of wave 2. Taking underreporting of deaths into account, the IFR estimate is 0.46% for wave 1 and 0.18% for wave 2 (till May 15). Combining waves 1 and 2, as of May 15, while India reported a total of nearly 25 million cases and 270 thousand deaths, the estimated number of infections and deaths stand at 491 million (36% of the population) and 1.21 million respectively, yielding an estimated (combined) infection fatality rate of 0.25%. There is considerable variation in these estimates across Indian states. Up to date seroprevalence studies and mortality data are needed to validate these model-based estimates.

Keywords: Case fatality rate, Excess deaths, False negative rates, India, RT-PCR test, SEIR model, Underreporting

Introduction

Main text

In late August 2020, India was predicted to surpass the United States in terms of reported case counts from SARS-CoV-2 infections. To the surprise of many modelers the curve turned corner in late September with the highest number (97,894) of daily new cases reported on 16 September 2020 [1]. After a steady decline for nearly five months, the curve started rising again, growing into an astronomic second wave. The highest number (414,280) of daily new cases in wave 2 was reported on May 6, 2021. As of May 15, 2021, India has reported 24.7 million cases, the second highest in the world, and nearly 270 thousand deaths, the third highest in the world. In this brief report, we reconcile estimates of the infection fatality rate (IFR) inferred from seroprevalence studies with

*Correspondence: bhramar@umich.edu
[1] Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA
Full list of author information is available at the end of the article

Purkayastha *et al. BMC Res Notes*   (2021) 14:262

Page 2 of 7

epidemiologic model-based estimates that account for underreporting of infections and deaths in India for wave 1. We then proceed to compute, compare and combine wave 1 with wave 2 IFR estimates.

Methods

Synthesizing evidence from seroprevalence studies

We review available seroprevalence results that vary across states and specifically across rural versus urban areas. Whereas in many major metros and slum areas the seroprevalences were reported to be more than 50%, in rural areas there is a wide variation (Table 1). The latest national serosurvey (from 17 December 2020 to 8 January 2021) reports 21.4% of all Indians above age 18 have antibodies present that indicate past SARS-CoV-2 infection [2]. Since approximately 59% [3] of India's 1.36 billion citizens are above age 18 and 10.45 million infections were reported as of 8 January 2021, this points to approximately 172.47 million infections, with an implied underreporting factor of 16.5 (172.47/10.45). In other words, only 6% of India's COVID-19 infections are reported, while 94% remained undetected or unreported. We use this estimated number of infections to calculate the IFR. Regional studies based on crematorium data and counting obituaries in India have suggested an underreporting factor in the range of 2 to 5 for COVID-deaths; this is at best ad hoc and anecdotal in nature and no rigorous quantification of missing death numbers is currently available [4].

Model-based estimates

Using a compartmental epidemiologic model (as explained in the *Supplementary Methods*) with a compartment for unascertained cases and deaths after accounting for the false negative rates of RT-PCR and rapid antigen tests used in India [5] we estimate the national and state-level IFR in India by inferring underreporting factors for cases and deaths. We assume that the estimated total infections (deaths) are comprised of reported and unreported infections (deaths). The model divides the population into ten disjoint compartments: S (Susceptible), E (Exposed), T (Tested), U (Untested), P (Tested positive), F (Tested False Negative), RR (Reported Recovered), RU (Unreported Recovered), DR (Reported Deaths) and DU (Unreported Deaths), as described in Additional file 1: Figure S1. A set of nine differential equations govern the transmission dynamics, which are approximated by means of discrete recurrence relations. For any compartment X, the instantaneous rate of change at time t (given by $\frac{dX}{dt}$) is approximated by the difference of counts in that specific compartment on the $(t+1)$

th day and the (t) th day, i.e., say $X(t+1) - X(t)$. Parameters are estimated using Bayesian techniques by generating samples from the posterior distribution using a Metropolis–Hastings algorithm with Gaussian proposal density, with 95% credible intervals (CrI) to quantify uncertainty of the estimates. Additional file 1: Table T4 presents an overview of the parameter descriptions and settings for this model.

Comparing and combining waves 1 and 2

Due to the stress on the healthcare and reporting infrastructure, the fatality and underreporting processes were very different across the two waves. Thus, we consider two separate phases of the pandemic, with wave 1 from April 1, 2020–January 31, 2021 and wave 2 starting on February 1, 2021. This definition is artificial and is guided by the fact that the national effective reproduction number (R_{eff}) crossed unity for the first time in 2021 on February 14 and we allow a two-week incubation period before that date. Using daily time series of case, death and recovery counts we compare fatality rates and underreporting factors associated with the two time periods using the compartmental models. Further, using observed data from the two waves and the model-based underreporting factor estimates, we compute cumulative case and death counts for the total duration of waves 1 and 2. We multiply the wave-specific cumulative counts with relevant underreporting factors and sum over both waves to get combined counts of cases and deaths. The estimated numbers of cumulative deaths and infections provide us with a combined IFR estimate for India as of May 15.

Results

IFR estimates for wave 1 using seroprevalence surveys

The observed case fatality rate (CFR) in India is low. With 154,428 deaths and 10.76 million cases reported as of January 31, 2021 the estimated CFR for wave 1 is 1.435% (95% confidence interval 1.428–1.442%) [1]. The estimated number of infections from the January seroprevalence survey imply an approximate infection fatality rate of 0.09% (i.e. 154,428/172.47 M). The anecdotal underreporting factor for deaths (in the range of 2–5) implies an ad hoc estimate of IFR in the range of 0.19–0.45%.

Estimates from epidemiological models

For wave 1 our estimate for the national IFR_1 (observed cumulative deaths/estimated cumulative total infections) is 0.129% (95% CrI 0.125–0.134%) and IFR_2 (estimated total cumulative deaths/estimated total cumulative infections) is 0.461% (95%

Purkayastha *et al. BMC Res Notes*   (2021) 14:262

Page 3 of 7

Table 1 Summary of results from various serological surveys conducted in India during 2020–21

Part A: State-level results from serological surveys conducted by the Indian Council of Medical Research in 2020–21

State[a]	Population (from 2011 Census)[b]	Serosurvey I[c] (May–June 2020)		Serosurvey II[d] (August–September 2020)		Observed cumulative cases February 7[e] (per million)	Observed cumulative deaths February 7[e] (per million)
		# of people tested	# of positive samples (%)	# of people tested	# of positive samples (%)		
Maharashtra	112,374,333	2385	19 (0.80)	2681	348 (12.98)	18,189.84	456.6
Kerala	33,406,061	1193	4 (0.34)	1282	11 (0.86)	28,989.92	115.79
Karnataka	61,095,297	1199	3 (0.25)	1287	186 (14.45)	15,427.01	200.28
Andhra Pradesh	49,577,103	1192	8 (0.67)	1245	352 (28.27)	17,920.03	144.4
Tamil Nadu	72,147,030	1200	16 (1.34)	1259	207 (16.44)	11,667.8	171.64
Uttar Pradesh	199,812,341	3616	15 (0.42)	3628	226 (6.23)	3009.75	43.48
West Bengal	91,276,115	2000	22 (1.10)	2097	219 (10.44)	6259.81	111.83
Odisha	41,974,219	1202	7 (0.58)	1223	294 (24.04)	7995.86	46.74
Rajasthan	68,548,437	1188	8 (0.67)	1212	27 (2.23)	4641.87	40.44
Chattisgarh	25,545,198	1210	4 (0.33)	1199	34 (2.84)	12,038.47	146.45
India	1,210,193,422	28,000	156 (0.56)	29,082	3135(10.8)	9027.74	128.77

Part B: Results from some other serological surveys conducted in India in 2020–21

Region	Study setting	Study period	# of people tested	% of positive samples
Delhi (Round 1)[f]	Urban	June–July, 2020	21,387	22.9
Delhi (Round 2)[g]	Urban	August 1–7, 2020	15,046	28.4
Delhi (Round 3)[g]	Urban	September 1–7, 2020	17,049	24.1
Delhi (Round 4)[g]	Urban	October 15–21, 2020	15,015	24.7
Delhi (Round 5)[h]	Urban	January, 2021	28,000	56.1
Tamil Nadu[i]	Rural and Urban	October–November, 2020	26,640	26.9 (rural areas) 36.9 (urban areas)
Mumbai (Round 1)[j]	Urban	First half of July, 2020	4234 (slum areas) 2702 (non-slum areas)	57.0 (slum areas) 16.0 (non-slum areas)
Mumbai (Round 2)[k]	Urban	Last half of August, 2020	3024 (slum areas) 2176 (non-slum areas)	45.2 (slum areas) 17.1 (non-slum areas)
Pune[l]	Urban	July 20–August 5, 2020	1659	51.3
Chennai (Round 1)[c]	Urban	July 17–28, 2020	12,405	18.4
Chennai (Round 2)[c]	Urban	October 8–15, 2020	6366	30.1
Indore[d]	Urban	August 11–23, 2020	7100	7.75
Karnataka[m]	Rural and urban	June 15–August 29, 2020	15,624	44.1 (rural areas) 53.8 (urban areas)
Jammu and Kashmir[c]	Rural and urban	October, 2020	6230	38.8
India[n]	Rural and urban	December 17, 2020–January 8, 2021 (Serosurvey III)	28,589 (general population) 7171 (healthcare workers)	21.4 (adults) 25.3 (children ≥ 10 years) 25.7 (healthcare workers)

[a] The first ten states with maximum cumulative COVID-19 cases (as of 31 January 2021) are included in this table

[b] Information sourced from Wikipedia. (https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population)

[c] The first national serosurvey conducted by the Indian Council of Medical Research (ICMR) began on May 11 and ended on June 4, 2020. A randomly sampled, community-based survey was conducted in 700 villages/wards, selected from the 70 districts of 21 chosen states of India, categorized into four strata based on the incidence of reported COVID-19 cases. Four hundred adults per district were enrolled from 10 clusters with one adult per household. A total of 28,000 adults were enrolled in the survey (Murhekar, Manoj V., et al. "Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May–June 2020." Indian Journal of Medical Research 152.1 (2020): 48. https://doi.org/10.4103/ijmr.IJMR_3290_20)

[d] The second national serosurvey conducted by the ICMR began on from August 18 and ended on September 20, 2020. A strata-based sampling design similar to that of the first serosurvey (see (b) above) was used. A total of 29,082 individuals aged 10 years or older were enrolled in the survey. (Murhekar, Manoj V., et al. "SARS-CoV-2 antibody seroprevalence in India, August–September, 2020: findings from the second nationwide household serosurvey." The Lancet Global Health (2021). https://doi.org/10.1016/S2214-109X(20)30544-1)

[e] As of February 7, 2021. Information sourced from Coronavirus Outbreak In India—COVID-19 tracker (www.covid19india.org)

[f] Data from media reports (The Hindu. Published online July 22,2020. https://www.thehindu.com/news/cities/Delhi/percentage-of-people-with-antibodies-high/article32156162.ece)

Table 1 (continued)

g Data from a preprint on repeated, cross-sectional, multi-stage sampling serosurvey conducted from all districts and wards of Delhi, with two-stage allocation proportional to population size. (Sharma, Nandini, et al. "The seroprevalence and trends of SARS-CoV-2 in Delhi, India: A repeated population-based seroepidemiological study". medRxiv (2021). https://doi.org/10.1101/2020.12.13.20248123)

h Data from media reports (Hindustan Times. Published online February 02, 2021. https://www.hindustantimes.com/cities/delhi-news/delhis-5th-sero-survey-over-56-people-have-antibodies-against-covid19-101612264534349.html)

i Data from a preprint on a population-representative serological survey conducted in all districts of Tamil Nadu, India in October–November 2020. (Malani, Anup, et al. "SARS-CoV-2 Seroprevalence in Tamil Nadu in October–November 2020." medRxiv (2021). https://doi.org/10.1101/2021.02.03.21250949)

j Data collected by a consortium of government organisations (NITI Aayog and Municipal Corporation of Greater Mumbai) and research institutes (Tata Institute of Fundamental Research and IDFC Institute) (https://www.tifr.res.in/TSN/article/Mumbai-Serosurvey%20Technical%20report-NITI.pdf)

k Data collected by a consortium of government organisations (NITI Aayog and Municipal Corporation of Greater Mumbai) and research institutes (Tata Institute of Fundamental Research and IDFC Institute) (https://www.tifr.res.in/TSN/article/Mumbai-Serosurvey%20Technical%20report-NITI_BMC-Round-2%20for%20TIFR%20website.pdf)

l Data from a preprint on multi-stage cluster random sampling of participants recruited from Pune sub-wards classified as high incidence settings for a serosurvey. (Ghose, Aurnab, et al. "Community prevalence of antibodies to SARS-CoV-2 and correlates of protective immunity in an Indian metropolitan city". medRxiv (2021). https://doi.org/10.1101/2020.11.17.20228155)

m Data from a research letter on a population-representative serological survey conducted in all districts of Karnataka, India in June 15–August 29 2020 (Mohanan M, Malani A, Krishnan K, Acharya A. Prevalence of SARS-CoV-2 in Karnataka, India. JAMA. Published online February 04, 2021. https://doi.org/10.1001/jama.2021.0332)

n Data from media reports (PTI. Published online February 04, 2021. https://www.ndtv.com/india-news/over-21-of-indias-population-may-have-had-covid-19-shows-sero-survey-2363166)

CrI 0.455–0.468%) with an underreporting factor for cases estimated at 11.11 (95% CrI 10.71–11.47) and for deaths at 3.56 (95% CrI 3.48–3.64). These model-based estimates in wave 1 are largely consistent with the estimates from the latest and third nationwide seroprevalence study.

In wave 2, using the same model we see a stark contrast with wave 1, with case and death underreporting factor estimates escalate to 26.73 (95% CrI 24.26–28.81) and 5.77 (95% CrI 5.34–6.15) respectively, leading to IFR_1 estimate of 0.032% (95% CrI 0.029–0.035%) and IFR_2 estimate of 0.183% (95% CrI 0.18–0.186%). This pattern is consistent with wave 2 CFR being estimated at 0.845% (95% CrI 0.840–0.849%), 59% of wave 1 estimate.

Figure 1 shows underreporting factors and estimated infections and deaths in waves 1 and 2 for India while Fig. 2 highlights state-level variations in IFR_1, IFR_2, CFR for waves 1 and 2 for 20 states in India with large case/death counts.

Combining waves 1 and 2

The composite CFR as of May 15 stands at 1.1%. The estimate for total (reported + unreported) cumulative case count for waves 1 and 2 combined is 491.73 (95% CrI 453.03–524.56) million, while the estimated number of total (reported + unreported) deaths is 1216.35 (95% CrI 1154.21–1272.70) thousand. This leads to a combined IFR_1 estimate of 0.06% and IFR_2 estimate of 0.24%. Detailed numerical estimates of underreporting factors across states for waves 1 and 2 are presented in Additional file 1: Tables T1, T2 and T3 and Additional file 1: Figures S2 and S3.

Discussion

Despite accounting for underreported deaths, the large number of asymptomatic/undetected infections (more than 90% by any calculation) indicate a lower IFR in India in comparison with other Western countries. A meta-analysis across the world places the pooled mean of IFRs at 0.68% (95% CI: 0.53–0.82%) [6], while another meta-analysis places the median at 0.27% [7] (with a range of 0–1.63%). Seroprevalence surveys and epidemiologic models qualitatively agree on the estimated IFR for India for wave 1. Up to date serosurvey and excess death/mortality data are needed to validate wave 2 and combined estimates. The estimated number of total infections as of May 15 suggests roughly 36% of Indians have an active or past infection, a number that will need to be verified with synchronous sero-surveys.

The current reduction in fatality rates in wave 2 that we notice could be primarily due to two reasons, one is that we do not have the same length of follow-up period and complete data on the decay phase of wave 2 curve. The second could be the different age composition of the infected populations in the two waves; it has been reported that the younger population got infected in larger numbers in wave 2 and they have lower risk of COVID-19 mortality. A fraction of the older population (aged 65 + years) also got vaccinated during wave 2. However, this hypothesis about reduced fatality rates in wave 2 cannot be verified without more granular, age-sex stratified nationwide time-series data on case and death counts, which is currently unavailable.

Purkayastha *et al. BMC Res Notes*   (2021) 14:262

Page 5 of 7

Fig. 1 Comparison of observed and estimated case and death counts and associated underreporting factors from waves 1, 2 and both waves combined
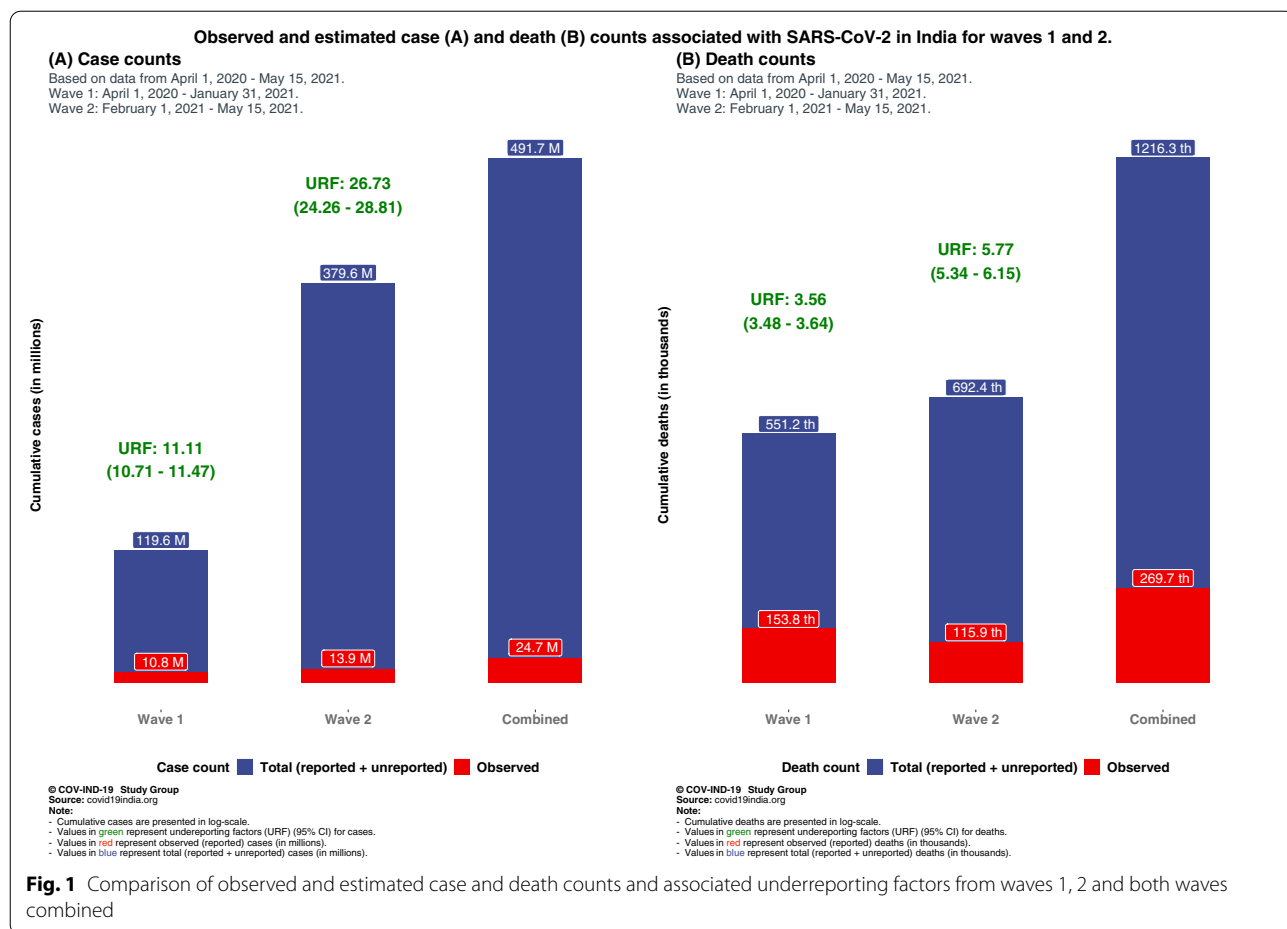
Limitations

We do not have a rigorous way to validate the extent of underreporting of deaths. An excess death calculation based on historical mortality data is infeasible at this point due to absence of all-cause-mortality data in the last three years from India. India has a very young population with only 6.4% in age group 65 + (compared to the US where this proportion is 16.5%) so a comparison of overall IFR between India and say the US is not fair, and only age-specific IFRs should be calculated and compared when more data become available. We do recognize that wave 2 information is appreciably incomplete, and the estimates will change as we have more complete information on deaths. For example, while our wave 2 analysis period ended on May 15, the highest daily number of deaths (4529 daily new deaths) were reported shortly after on May 18. Thus, our analysis presents an updated but incomplete picture of wave 2.

Abbreviations
CFR: Case fatality rate; CI: Confidence interval; CrI: Credible interval; IFR: Infection fatality rate; RT-PCR: Real time reverse transcript polymerase chain
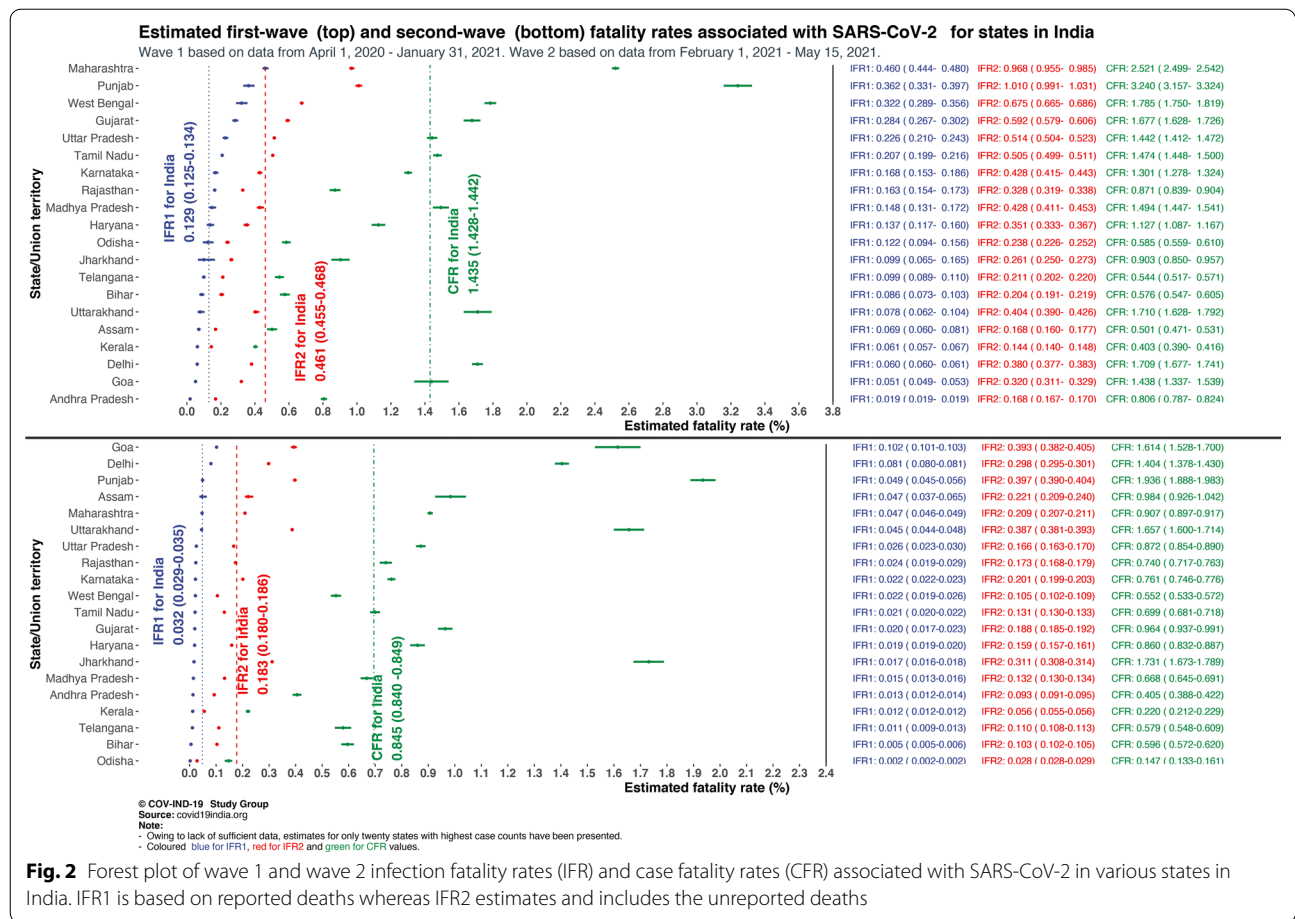
Estimated first-wave (top) and second-wave (bottom) fatality rates associated with SARS-CoV-2 for states in India

Wave 1 based on data from April 1, 2020 - January 31, 2021. Wave 2 based on data from February 1, 2021 - May 15, 2021.

Wave 1 (top panel), IFR1 for India 0.129 (0.125-0.134); IFR2 for India 0.461 (0.455-0.468); CFR for India 1.435 (1.428-1.442):

State/Union territory	IFR1	IFR2	CFR
Maharashtra	0.460 (0.444-0.480)	0.968 (0.955-0.985)	2.521 (2.499-2.542)
Punjab	0.362 (0.331-0.397)	1.010 (0.991-1.031)	3.240 (3.157-3.324)
West Bengal	0.322 (0.289-0.356)	0.675 (0.665-0.686)	1.785 (1.750-1.819)
Gujarat	0.284 (0.267-0.302)	0.592 (0.579-0.606)	1.677 (1.628-1.726)
Uttar Pradesh	0.226 (0.210-0.243)	0.514 (0.504-0.523)	1.442 (1.412-1.472)
Tamil Nadu	0.207 (0.199-0.216)	0.505 (0.499-0.511)	1.474 (1.448-1.500)
Karnataka	0.168 (0.153-0.186)	0.428 (0.415-0.443)	1.301 (1.278-1.324)
Rajasthan	0.163 (0.154-0.173)	0.328 (0.319-0.338)	0.871 (0.839-0.904)
Madhya Pradesh	0.148 (0.131-0.172)	0.428 (0.411-0.453)	1.494 (1.447-1.541)
Haryana	0.137 (0.117-0.160)	0.351 (0.333-0.367)	1.127 (1.087-1.167)
Odisha	0.122 (0.094-0.156)	0.238 (0.226-0.252)	0.585 (0.559-0.610)
Jharkhand	0.099 (0.065-0.165)	0.261 (0.250-0.273)	0.903 (0.850-0.957)
Telangana	0.099 (0.089-0.110)	0.211 (0.202-0.220)	0.544 (0.517-0.571)
Bihar	0.086 (0.073-0.103)	0.204 (0.191-0.219)	0.576 (0.547-0.605)
Uttarakhand	0.078 (0.062-0.104)	0.404 (0.390-0.426)	1.710 (1.628-1.792)
Assam	0.069 (0.060-0.081)	0.168 (0.160-0.177)	0.501 (0.471-0.531)
Kerala	0.061 (0.057-0.067)	0.144 (0.140-0.148)	0.403 (0.390-0.416)
Delhi	0.060 (0.060-0.061)	0.380 (0.377-0.383)	1.709 (1.677-1.741)
Goa	0.051 (0.049-0.053)	0.320 (0.311-0.329)	1.438 (1.337-1.539)
Andhra Pradesh	0.019 (0.019-0.019)	0.168 (0.167-0.170)	0.806 (0.787-0.824)

Estimated fatality rate (%)

Wave 2 (bottom panel), IFR1 for India 0.032 (0.029-0.035); IFR2 for India 0.183 (0.180-0.186); CFR for India 0.845 (0.840-0.849):

State/Union territory	IFR1	IFR2	CFR
Goa	0.102 (0.101-0.103)	0.393 (0.382-0.405)	1.614 (1.528-1.700)
Delhi	0.081 (0.080-0.081)	0.298 (0.295-0.301)	1.404 (1.378-1.430)
Punjab	0.049 (0.045-0.056)	0.397 (0.390-0.404)	1.936 (1.888-1.983)
Assam	0.047 (0.037-0.065)	0.221 (0.209-0.240)	0.984 (0.926-1.042)
Maharashtra	0.047 (0.046-0.049)	0.209 (0.207-0.211)	0.907 (0.897-0.917)
Uttarakhand	0.045 (0.044-0.048)	0.387 (0.381-0.393)	1.657 (1.600-1.714)
Uttar Pradesh	0.026 (0.023-0.030)	0.166 (0.163-0.170)	0.872 (0.854-0.890)
Rajasthan	0.024 (0.019-0.029)	0.173 (0.168-0.179)	0.740 (0.717-0.763)
Karnataka	0.022 (0.022-0.023)	0.201 (0.199-0.203)	0.761 (0.746-0.776)
West Bengal	0.022 (0.019-0.026)	0.105 (0.102-0.109)	0.552 (0.533-0.572)
Tamil Nadu	0.021 (0.020-0.022)	0.131 (0.130-0.133)	0.699 (0.681-0.718)
Gujarat	0.020 (0.017-0.023)	0.188 (0.185-0.192)	0.964 (0.937-0.991)
Haryana	0.019 (0.019-0.020)	0.159 (0.157-0.161)	0.860 (0.832-0.887)
Jharkhand	0.017 (0.016-0.018)	0.311 (0.308-0.314)	1.731 (1.673-1.789)
Madhya Pradesh	0.015 (0.013-0.016)	0.132 (0.130-0.134)	0.668 (0.645-0.691)
Andhra Pradesh	0.013 (0.012-0.014)	0.093 (0.091-0.095)	0.405 (0.388-0.422)
Kerala	0.012 (0.012-0.012)	0.056 (0.055-0.056)	0.220 (0.212-0.229)
Telangana	0.011 (0.009-0.013)	0.110 (0.108-0.113)	0.579 (0.548-0.609)
Bihar	0.005 (0.005-0.006)	0.103 (0.102-0.105)	0.596 (0.572-0.620)
Odisha	0.002 (0.002-0.002)	0.028 (0.028-0.029)	0.147 (0.133-0.161)

Estimated fatality rate (%)

© COV-IND-19 Study Group
Source: covid19india.org
Note:
- Owing to lack of sufficient data, estimates for only twenty states with highest case counts have been presented.
- Coloured blue for IFR1, red for IFR2 and green for CFR values.

Fig. 2 Forest plot of wave 1 and wave 2 infection fatality rates (IFR) and case fatality rates (CFR) associated with SARS-CoV-2 in various states in India. IFR1 is based on reported deaths whereas IFR2 estimates and includes the unreported deaths

reaction; SEIR: Susceptible-exposed-infected-recovered; URF: Underreporting Factor.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13104-021-05652-2.

Additional file 1: **Figure S1**. Schematic diagram for the SEIR-fansy model with imperfect testing and misclassification. **Figure S2.** Estimated first wave underreporting factors for cases and deaths associated with SARS-CoV-2 for states in India. **Figure S3.** Estimated first wave underreporting factors for cases and deaths associated with SARS-CoV-2 for states in India. **Table T1.** Summary of the different metrics for the states and the nation for wave 1, on 31st January, 2021. **Table T2.** Summary of the different metrics for the states and the nation for wave 2, on 15th May, 2021. **Table T3.** Summary of the different metrics for the states and the nation for waves 1 and 2 combined, on 15th May, 2021. **Table T4.** Parameter values and descriptions for the SEIRfansy model.

Authors' contributions
SP created the initial draft, conducted the literature review and collaborated on the analysis. RK and RB created the R package SEIR-fansy and implemented the epidemiologic models. DB and MK collaborated on analysing data from the second wave. DR helped create and modify the final draft and address issues raised by the review panel. BM conceived the project, planned the analysis and wrote the draft of the paper. All authors read, reviewed, edited and approved the manuscript for submission.

Availability of data and materials
All data are publicly available at covid19india.org. We used reported daily case, death and recovery counts for India and its states and union territories from April 1, 2020 to May 15, 2021. The statistical package SEIR-Fansy developed by the authors is available at covind19.org.

Declarations

Ethics approval and consent to participate
Not applicable. The analysis is based on publicly available completely de-identified aggregate counts. The study is exempt from IRB review as no patient participation or contact is involved.

Consent for publication
Not applicable.

Competing interests
There are no conflicts of interest perceived or declared by any of the authors.

Purkayastha *et al. BMC Res Notes*   (2021) 14:262

Page 7 of 7

Author details
[1]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA. [2]Indian Statistical Institute, Kolkata, West Bengal 700108, India. [3]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. [4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. [5]Center for Precision Health Data Science, University of Michigan, Ann Arbor, MI 48109, USA. [6]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.

References
1. COVIDIndia.org. COVID-19 Tracker Updates for India for State-wise and District-wise data. Published 2020. https://covidindia.org. Accessed 21 May 2020
2. Press Trust of India. Over 21% of India's population may have had COVID-19, shows sero survey. Published February 4, 2021. https://www.ndtv.com/india-news/over-21-of-indias-population-may-have-had-covid-19-shows-sero-survey-2363166
3. Wikipedia. Demographics of India: structure of the population. https://en.wikipedia.org/wiki/Demographics_of_India#Structure_of_the_population. Accessed 20 Feb 2021
4. The Wire. COVID-19: Cremation, burial records suggest Delhi's death toll is over twice the official figure. Published May 27, 2020. https://thewire.in/government/delhis-covid-19-deaths-data. Accessed 20 Feb 2021
5. 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. Epidemiology. 2020. https://doi.org/10.1101/2020.09.24.20200238.
6. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. Int J Infect Dis. 2020;101:138–48. https://doi.org/10.1016/j.ijid.2020.09.1464.
7. 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.