Reassessing Reported Deaths and Estimated Infection Attack Rate during the First 6 Months of the COVID-19 Epidemic, Delhi, India

Margarita Pons-Salort, Jacob John, Oliver J. Watson,
Nicholas F. Brazeau, Robert Verity, Gagandeep Kang, Nicholas C. Grassly

India reported >10 million coronavirus disease (COVID-19) cases and 149,000 deaths in 2020. To reassess reported deaths and estimate incidence rates during the first 6 months of the epidemic, we used a severe acute respiratory syndrome coronavirus 2 transmission model fit to data from 3 serosurveys in Delhi and time-series documentation of reported deaths. We estimated 48.7% (95% credible interval 22.1%–76.8%) cumulative infection in the population through the end of September 2020. Using an age-adjusted overall infection fatality ratio based on age-specific estimates from mostly high-income countries, we estimated that just 15.0% (95% credible interval 9.3%–34.0%) of COVID-19 deaths had been reported, indicating either substantial underreporting or lower age-specific infection-fatality ratios in India than in high-income countries. Despite the estimated high attack rate, additional epidemic waves occurred in late 2020 and April–May 2021. Future dynamics will depend on the duration of natural and vaccine-induced immunity and their effectiveness against new variants.

India had just under 150,000 reported coronavirus disease (COVID-19) deaths in 2020, fewer per 1 million persons than many other countries, such as Spain, France, the United Kingdom, and the United States (https://www.ourworldindata.org). This discrepancy could in part be because of a younger population but also because of incomplete documentation of overall deaths and of deaths with COVID-19 as a cause (*1*,*2*). Assessing the extent of underreporting of COVID-19 cases and deaths is essential for estimating actual disease burden and likely future trends in transmission.

Multiple severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) seroprevalence surveys conducted during 2020 in Delhi, one of India's largest metropolitan areas (20 million residents), offered us an opportunity to assess the completeness of reported COVID-19 deaths and estimate the actual infection attack rate. SARS-CoV-2 transmission in Delhi has led to several waves of infection and death (Figure 1). At the beginning of the epidemic, all SARS-CoV-2 testing relied on reverse transcription PCR (RT-PCR), but after mid-June 2020, use of antigen-based rapid diagnostic tests (Ag-RDTs), which have lower sensitivity, quickly exceeded use of RT-PCR tests (Appendix Figure 1, https://wwwnc.cdc.gov/EID/article/28/4/21-0879-App1.pdf). Three serosurveys conducted in Delhi during 2020 that sampled participants >4 years of age found age- and sex-adjusted seropositivity rates (uncorrected for test sensitivity and specificity) of 22.8% in July, 28.7% in August, and 25.1% in September (Appendix Table 1) (*3*). The July survey found a difference in seropositivity between residents living inside or outside of slum areas (25.3% vs. 19.2%; $p<0.001$), but the August survey did not (28.9% vs. 28.8%; $p = 0.94$), and the September survey did not report this information.

Materials and Methods

Study Overview

We developed a SARS-CoV-2 transmission model to estimate the incidence of infection and changes in the reproduction number (R) after the start of nonpharmaceutical interventions, including lockdowns

Author affiliations: Imperial College London School of Public Health, London, UK (M. Pons-Salort, O.J. Watson, N.F. Brazeau, R. Verity, N.C. Grassly); Christian Medical College, Vellore, India (J. John, G. Kang)

DOI: https://doi.org/10.3201/eid2804.210879

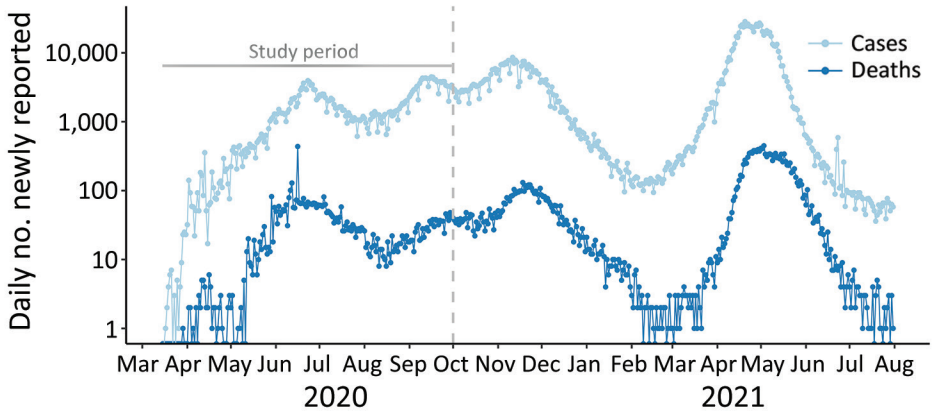


Figure 1. Daily number of coronavirus disease reported cases (light blue) and deaths (dark blue) on a logarithmic scale, Delhi, India, March 15, 2020–August 31, 2021. The gray vertical dashed line indicates the end of the study period, September 30, 2020.

(Appendix Table 2, Figure 2). We used Bayesian Markov chain Monte Carlo to fit the model to the 3 seroprevalence surveys and the time-series of reported deaths. We estimated the proportion of COVID-19 deaths reported by comparing reported deaths to the number expected based on the age-adjusted infection-fatality ratio (IFR) we used in the model. We used age-specific IFR estimates based on data from 7 countries in Europe; New York, USA; and Brazil (*4*) to estimate a median age-adjusted 0.39% IFR (95% prediction interval 0.21%–0.85%) for Delhi; median age-adjusted IFR in high-income countries with older populations, such as the United Kingdom, was ≈1%, based on data through June 2020 (*5*,*6*). The age-adjusted IFR for Delhi that we used was very similar to the 0.39% obtained using early data from China (*6*) and 0.40% from a meta-analysis based on data from advanced economies (as defined by membership in the Organization for Economic Cooperation and Development [https://www.oecd.org]) (*7*).

Epidemiologic and Demographic Data

We obtained data on the number of confirmed SARS-CoV-2 cases and deaths reported daily in Delhi beginning March 14, 2020, from COVID19India (*8*), a volunteer-driven, crowdsourced initiative that collates data from several sources, including the Ministry of Health and Family Welfare. Cases and deaths that occurred before March 14 were reported as cumulative numbers. Because we did not know specifically when these pre–March 14 cases and deaths occurred, we did not use these data for parameter inference. For our model, we used data from the 3 serosurveys conducted in Delhi (*3*) on dates of sample collection, number of samples tested, seropositivity rate found, and reported estimates of sensitivity and specificity of the assay used in each of the three serosurveys (Appendix Table 1). We used projections of the 2021 population in Delhi from the National Commission on Population (*9*) and stratified the population by 10-year age groups.

Transmission Model

To model SARS-CoV-2 transmission, we used a susceptible-exposed-infected-recovered (SEIR) deterministic transmission model (Appendix Figure 4).We did not stratify the population by age for the transmission parameters, assuming random mixing by

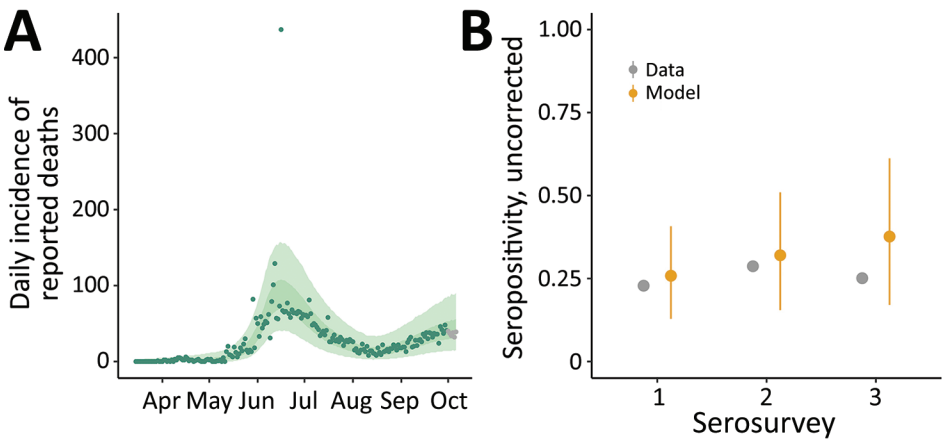


Figure 2. Model fit for reported coronavirus disease deaths and severe acute respiratory syndrome coronavirus 2 seroprevalence, Delhi, India, March 15–September 30, 2020. A) Model fit to the time-series of reported deaths (dots), showing 50% (darker green shading) and 95% (lighter green shading) credible intervals (CrIs). The last 6 points (shown in gray) were not used for parameter inference. B) Model fit (orange dots) to seroprevalence data (gray dots) from 3 serosurveys, conducted in July, August, and September 2020, showing medians and 95% CrIs (error bars).

age, meaning that epidemic growth was equivalent in all age groups in the model. We did not account for births or deaths from causes other than COVID-19 because of the model's short timeframe.

Because epidemic growth rate is determined by the reproductive number and the generation time, T_c (i.e., time interval between infection times of an infector-infectee pair) (*10*), we fixed the generation time to T_c = 6.5 days on the basis of previous observations (*11*) and estimated the reproduction number. We split the generation time into the mean durations of the preinfectious ($d_E = 1/\omega$) and infectious ($d_I = 1/\gamma$) periods, so that $T_c = d_E + d_I$ (*10*); we fixed d_E and d_I using information on the duration of the incubation period (i.e., time between infection and onset of symptoms) and the fact that infectiousness starts ≈1 day before symptoms start (*12*–*14*). Given an ≈5.5-day incubation (i.e., presymptomatic) period (*15*,*16*), to give the correct generation time, we assumed a mean duration of the preinfectious period of d_E = 5.5 – 1.0 = 4.5 days and a mean duration of the infectious period of d_I = 6.5 – 4.5 = 2 days.

Disease Progression and Death Model

We modeled disease progression and death after infection independent of the transmission process (Appendix Figure 4). Because the model has been used for other purposes, it also included transitions to hospitalizations, but these were not relevant for our work and did not affect the results. We used the 5.5-day mean incubation period and a peaked distribution modeled with an Erlang distribution with shape parameter 6 (*15*). We assumed that one third of infections were asymptomatic, although there is high variability in the observed proportion of asymptomatic infections across studies (*17*–*19*).

We separately tracked the proportion of total infections leading to hospitalization and those leading to death; those hospitalized who eventually died were represented in both groups. We age-adjusted the proportion of infections leading to hospitalization with versus without critical care using demographics from Delhi and age-stratified estimates from China (*6*). That is, we computed a weighted average of the age-stratified estimates, assigning weights by the share of the corresponding age classes. We based the proportion of infections leading to death on estimates of age-stratified IFR (*4*) applied to the population of Delhi.

We set average time from symptom onset to hospitalization as 5.8 days, consistent with observations in China (*20*). For hospitalization without critical care, we assumed a mean 9.8-day stay; if critical care was required, we assumed 9.8 days in critical care, followed by 3.3 days recovery outside of critical care, based on early estimates from the United Kingdom (*21*). The average time from symptom onset to death was ≈16 days (*6*). Using these estimates, we assumed a 10-day mean for time between hospitalization and death. These values might differ for India, but no domestic data were available at that time.

Parameter Inference

We fitted the transmission model to both the seroprevalence data and reported daily COVID-19 deaths (Appendix). We allowed the reproduction number to change at 5 different time points corresponding to changes in interventions (Appendix Table 2). Denoting the basic reproduction number during the first infection period (i.e., before any changes) as R_0, the reproduction number after i number of changes as R_i (i in periods 1–5), we conducted parameterization of R_i as $R_i = R_0 \times (1 + r_1) \times ... \times (1 + r_i)$, where $r_1,\ldots,r_5$ measured the relative change in the reproduction number from one period to the previous one.

We estimated R_0 and the subsequent changes at each time point, $r_1,\ldots,r_5$, the initial number of infected [$E(0) + I(0)$], the reporting, θ, and overdispersion of deaths, k. We assumed February 19, 2020 (28 days before the first 10 cases were reported), as the starting time (t_0) for the simulations and estimated the number of infected persons at that time point, [$E(0) + I(0)$]. To prevent parameter estimates being biased by the earliest phase of the epidemic, when underreporting of deaths might have been greatest, we computed likelihood using data collected from March 29, 2020, when the first COVID-19 death was reported, through September 30, 2020, the end of the 6-month study period.

We could not estimate a change in transmission at the first time point (r_1), corresponding to the start of the lockdown on March 25, because no deaths were reported during March 15–28; we therefore assumed $r_1 = 0$. We assumed May 4, when the first lockdown relaxations were introduced, as the next time point for change in the reproduction number (r_2). Therefore, estimates of the reproduction number during February 19–May 4, 2020, from the beginning of the simulations through r_2, implicitly accounted for any effects of the lockdown during that time. Because R_0 was highly correlated with the initial number of infected, we estimated the total number of infections just before r_2 and back-calculated the initial number of infected persons using a simple exponential growth model to define the relationship between R_0 and the epidemic growth rate for a SEIR model (*9*,*22*).

We performed 100,000 iterations using Markov chain Monte Carlo in the lazymcmc software package (*23*) and uniform prior distributions to estimate model parameters; we ran 4 chains with different starting values to check convergence. We performed all analyses using R version 4.0.2 (*24*).

Results

Our model fit the data well for both the death time-series (Figure 2, panel A) and seroprevalence survey data (Figure 2, panel B), except for the last serosurvey, in which we estimated an increase in seropositivity from the previous survey, instead of a slight decrease. This difference might have been because the observation model did not account for waning antibodies and the possibility of seroreversion. However, the third serosurvey used a different testing kit, which might also have contributed to this difference. We estimated that the first peak in infection incidence was reached on May 31, at a median of 294,930 (95% credible interval [CrI] 143,271–440,702) new infections per day (Appendix Figure 5). Incidence at the second peak, reached on September 17, was lower, at a median of 79,032 (95% CrI 40,484–109,140) new infections per day. Assuming that changes in transmission occurred beginning at the times of each change in interventions and accounting for the reduction in susceptible persons, we estimated that the effective reproduction number, R_{eff}, increased with the first relaxation of the lockdown introduced May 4 (beginning of phase 3); in June and July, during the first 2 reopening phases, R_{eff} was <1; in August, R_{eff} then increased again to >1 (Figure 3, panel A), resulting in a median infection attack rate of 48.7% (95% CrI 22.1%–76.8%) by the end of September. After that, Delhi experienced a large third wave of cases and deaths (Figure 1), suggesting that even with approximately half the population having been infected, the herd immunity threshold had not yet been reached at that time. Of interest, a serosurvey conducted in January 2021 found a sex- and age-adjusted seroprevalence of 56.1%, probably indicating a steep increase in the cumulative number of infections, reflecting the effects of this third wave of transmission.

Using a 0.39% age-adjusted IFR, we estimated reported deaths to be 15.0% (95% CrI 9.3%–34.0%) of actual deaths (Figure 4; Appendix Figure 6). Repeating the analysis using an age-adjusted IFR of 0.21%, corresponding to the lower bound of the 95% prediction interval for IFR based mostly on age-specific high-income country (HIC) data (*4*), increased the proportion of reported deaths to 28% (95% CrI 18–59%) of actual deaths (Figure 4).

On the basis of infection incidence determined using our model, we also estimated the probability of detecting COVID-19 cases over time by comparing the number of reported cases to the estimated incidence of symptomatic infections (Figure 5, panel A). The probability of detecting infection quickly increased over the last weeks of March, fluctuated until mid-June, then remained relatively consistent through the end of September; a median of 7.1% of all symptomatic infections was detected during July 1–September 30, 2020 (Figure 5, panel B).

Discussion

The low proportion of reported deaths relative to actual deaths we found is consistent with findings from other cities in India, where seroprevalence surveys suggested substantially greater exposure to infection than predicted on the basis of reported COVID-19 deaths. For example, comparing seroprevalence during the first half of July 2020 in Mumbai (*25*) with cumulative deaths at that time suggested that only 21% of deaths were reported (Appendix Table 3). Similarly, a large-scale prospective, active-surveillance study conducted in the district of Madurai, Tamil Nadu, India, during the first wave of COVID-19 in summer

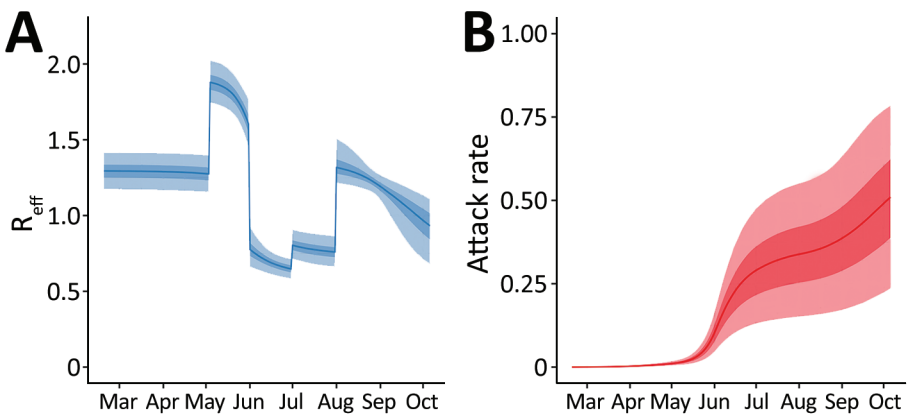


Figure 3. Effective reproduction number, R_{eff}, for model of coronavirus disease deaths and severe acute respiratory syndrome coronavirus 2 attack rate, Delhi, India, March 15–September 30, 2020. A) Median and 50% (dark blue shading) and 95% (light blue shading) credible intervals (CrIs) of the estimated R_{eff} from the model. Changes were assumed to occur beginning when changes in the interventions were introduced. B) Median and 50% (dark red shading) and 95% (light red shading) CrIs of the estimated infection attack rate from the model.

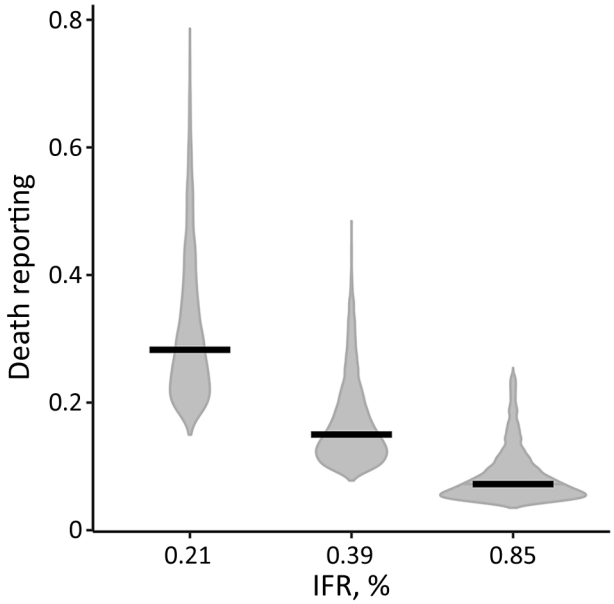


Figure 4. Estimated reporting of coronavirus disease deaths, Delhi, India, March 15–September 30, 2020. Violin plots show the posterior distribution of the estimate of death reporting for 3 different values for the assumed age-adjusted IFR, using age-stratified estimates of IFR based on data from mostly high-income countries; 0.21% corresponds to lower bound, 0.39% to the median, and 0.85% to the upper bounds of the IFR based on data documented elsewhere (*4*). Horizontal black lines indicate the median values of the posterior distributions. IFR, infection-fatality ratio.

2020 found that only 11.0% of deaths were reported, compared with expected deaths based on IFR estimates from other settings (*26*). This high level of underreporting might reflect incomplete or delayed reporting of deaths and a failure to report COVID-19 as a suspected or confirmed cause of death, particularly in the absence of a SARS-CoV-2 test result.

The extent of underreporting might also reflect our use of an age-specific IFR for India derived from mostly HIC data. Age-specific IFR may be lower in India for several reasons. First, the prevalence of underlying medical conditions that increase the risk for severe COVID-19 after infection is somewhat lower in India than in the countries that informed the age-specific IFR estimates for our model (Appendix Figure 7) (*27*). However, correcting the Delhi IFR to account for the lower prevalence of underlying conditions only marginally reduced the age-adjusted IFR (≤0.02%). Second, a recent study that analyzed COVID-19 deaths from Mumbai and Karnataka by age found that IFR rose less steeply with age than in HICs (R. Cai et al., unpub. data, https://doi.org/10.1101/2021.01.05.21249264). Third, differences in immunity reflecting exposure to a greater number of pathogens, including related coronaviruses, or simply lower frailty among those surviving to older ages in India compared with HICs could theoretically reduce the IFR in older groups, although data supporting these hypotheses are lacking (*28*; B. Chatterjee et al., unpub. data, https://doi.org/10.1101/2020.07.31.20165696). If the IFR in India was actually higher than in HICs, the proportion of deaths reported would be even lower. For example, using a 0.85% age-adjusted IFR, corresponding to the upper bound of the 95% prediction interval for IFR based on age-specific HIC data (*4*), would decrease the reported deaths to only 7% (95% CrI 4%–21%) of actual deaths (Figure 4).

The first limitation of our study is that we did not structure the transmission model by age, and therefore, did not account for differences in attack rates between age groups. However, age-structured models have predicted relatively homogeneous infection attack rates across age for India (*29*), consistent with age-stratified seroprevalence estimates (*3*), suggesting that any bias in our results from age-specific patterns of mixing and potentially lower attack rates in more susceptible older age groups is likely to be limited. Second, we assumed that the proportion of deaths

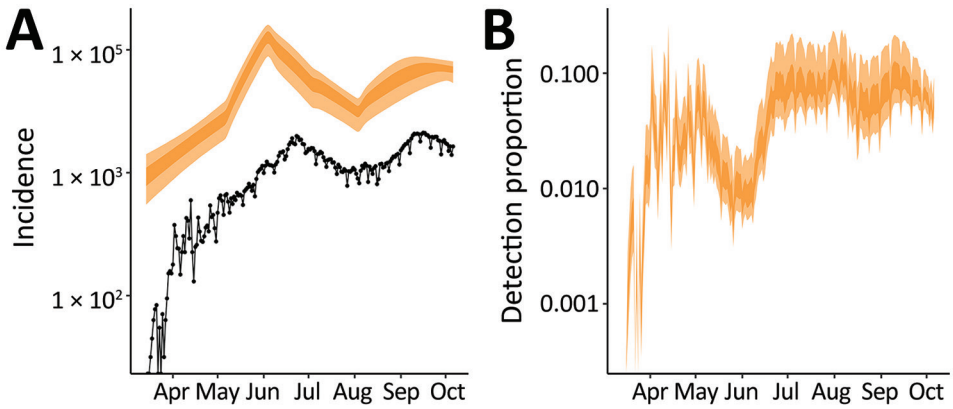


Figure 5. Reported cases of symptomatic severe acute respiratory syndrome coronavirus 2 infections and estimated actual number of cases, Delhi, India, March 15–September 30, 2020. A) Daily number of newly reported cases (black dots) and 50% (dark orange shading) and 95% (light orange shading) credible intervals (CrIs) for the estimated actual incidence of symptomatic infections, assuming that 2/3 infections are symptomatic. B) Estimated detection probability per symptomatic infection per day: 50% (dark orange) and 95% (light orange) CrIs.

reported was constant over the study period, but it might have changed over time. Therefore, our estimate of reported deaths represents an average over the study period. Finally, we used an age-specific IFR based on estimates mostly from HICs and explored sensitivity based on this assumption, including using data on underlying conditions in India. Further analyses using data from cohort studies or demographic surveillance specific to India will help to refine these estimates of IFR and the exact degree of underreporting of death.

The total number of new COVID-19 cases declined in India between mid-September 2020 and mid-February 2021 but started increasing again after that, and in April–May 2021, India experienced a devastating nationwide second epidemic wave bigger than the first one. How much of the country's population had already been infected before the second nationwide wave and whether the herd immunity threshold had been reached were unclear (*30*). Seroprevalence surveys conducted in major cities, such as Mumbai, reported seroprevalence rates >50% in slum areas for the first half of July 2020 (*25*), suggesting that infection spread very quickly over the first few months of the epidemic in certain pockets. However, seroprevalence rates <20% in non–slum areas showed that the epidemic was spatially highly heterogeneous. Understanding what brought the number of cases down after the first wave in different parts of India and how to interpret the serosurvey results related to building population immunity are key to understanding and predicting the dynamics of subsequent waves of COVID-19.

The SARS-CoV-2 Delta variant emerged in Maharashtra in late 2020 and spread across India during the first few months of 2021, replacing other variants. In vitro data characterizing the Delta variant found that it was less sensitive to serum neutralizing antibodies from persons previously infected with other variants and that it also had higher replication efficiency (*31*). These findings suggest that the predominance of the Delta variant in the upsurge of SARS-CoV-2 cases seen in India during April and May 2021 resulted from either immune escape in previously infected persons, increased transmissibility, or both. These mechanisms, together with possible waning of population immunity over time, likely explain the increase in SARS-CoV-2 cases in Delhi, despite the high attack rate that we estimated in September 2020 and the high reported seroprevalence (≈56% for both) in the round 5 (January 2021) and 6 (April 2021) cross-sectional serosurveys. Analysis of epidemiologic data is needed to disentangle how these mechanisms contributed to the second nationwide epidemic wave.

In conclusion, our analysis found reported COVID-19 deaths in Delhi during the first 6 months of the pandemic were well below the number of actual deaths. Our estimate of underreporting of deaths might reflect incomplete or delayed documentation or failure to report COVID-19 as a cause of death but may also reflect our use of an age-specific IFR for India, dertived from mostly HIC data.

Acknowledgments

We thank Nimalan Arinaminpathy for insightful comments on the manuscript, Marc Baguelin for helpful discussions on parameter inference, and James A. Hay for help using the lazymcmc R package.

M.P.-S. is a Sir Henry Dale fellow, a program jointly funded by the Wellcome Trust and the Royal Society (grant number 216427/Z/19/Z). M.P.-S., O.J.W., N.F.B., R.V., and N.C.G. acknowledge funding from the MRC Centre for Global Infectious Disease Analysis (MR/R015600/1), which is jointly funded by the U.K. Medical Research Council (MRC) and U.K. Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement, and is also part of the EDCTP2 programme supported by the European Union. M.P.-S., O.J.W., N.F.B., R.V., and N.C.G. also acknowledge funding from Community Jameel.

About the Author

Dr. Pons-Salort is a Sir Henry Dale fellow at Imperial College London. She uses statistical and mathematical models to study the dynamics of infectious diseases. Her recent work has focused on the epidemiology of COVID-19 and enteroviruses.

References

1. Pulla P. 'The epidemic is growing very rapidly': Indian government adviser fears coronavirus crisis will worsen. Nature. 2020;583:180. https://doi.org/10.1038/d41586-020-01865-w
2. Chatterjee P. Is India missing COVID-19 deaths? Lancet. 2020;396:657. https://doi.org/10.1016/S0140-6736(20)31857-2
3. Sharma N, Sharma P, Basu S, Saxena S, Chawla R, Dushyant K, et al. The seroprevalence of severe acute respiratory syndrome coronavirus 2 in Delhi, India: a repeated population-based seroepidemiological study. Trans R Soc Trop Med Hyg. 2021 Aug 2 [Epub ahead of print]. https://doi.org/10.1093/trstmh/trab109
4. Brazeau NF, Verity R, Jenks S, Fu H, Whittaker C, Winskill P, et al. COVID-19 infection fatality ratio: estimates from seroprevalence. London: Imperial College London; 2020 [cited 2021 Mar 23]. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-34-ifr
5. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and

immunity patterns of SARS-CoV-2. Nature. 2021;590:140–5. https://doi.org/10.1038/s41586-020-2918-0
6. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. [Erratum in Lancet Infect Dis. 2020;20:e116.]. Lancet Infect Dis. 2020;20:669–77. https://doi.org/10.1016/S1473-3099(20)30243-7
7. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. Eur J Epidemiol. 2020;35:1123–38. https://doi.org/10.1007/s10654-020-00698-1
8. COVID19India [cited 2021 Mar 23]. https://www.covid19india.org
9. Census of India. Population projections for India and states 2011–2036: report of the technical group on population projections. November 2019 [cited 2021 Mar 23]. https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf
10. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci. 2007;274:599–604. https://doi.org/10.1098/rspb.2006.3754
11. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiologyand transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis. 2020;20:911–9. https://doi.org/10.1016/S1473-3099(20)30287-5
12. Lau YC, Tsang TK, Kennedy-Shaffer L, Kahn R, Lau EHY, Chen D, et al. Joint estimation of generation time and incubation period for coronavirus disease (Covid-19). J Infect Dis. 2021 Aug 23 [Epub ahead of print]. https://doi.org/10.1093/infdis/jiab424
13. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6. https://doi.org/10.1016/j.ijid.2020.02.060
14. Xin H, Li Y, Wu P, Li Z, Lau EHY, Qin Y, et al. Estimating the latent period of coronavirus disease 2019 (COVID-19). Clin Infect Dis. 2021 Aug 28. https://doi.org/10.1093/cid/ciab746
15. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;172:577–82. https://doi.org/10.7326/M20-0504
16. Xin H, Wong JY, Murphy C, Yeung A, Taslim Ali S, Wu P, et al. The incubation period distribution of coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. Clin Infect Dis. 2021;73:2344–52. https://doi.org/10.1093/cid/ciab501
17. Oran DP, Topol EJ. The proportion of SARS-CoV-2 infections that are asymptomatic : a systematic review. Ann Intern Med. 2021;174:655–62. https://doi.org/10.7326/M20-6976
18. Beale S, Hayward A, Shallcross L, Aldridge RW, Fragaszy E. A rapid review and meta-analysis of the asymptomatic proportion of PCR-confirmed SARS-CoV-2 infections in community settings. Wellcome Open Res. 2020;5:266. https://doi.org/10.12688/wellcomeopenres.16387.1
19. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. PLoS Med. 2020;17:e1003346. https://doi.org/10.1371/journal.pmed.1003346
20. Gaythorpe K, Imai N, Cuomo-Dannenburg G, Baguelin M, Bhatia S, Boonyasiri A, et al. Symptom progression of COVID-19. Imperial College London. 2020 [cited 2021 Mar 23]. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-8-symptom-progression-covid-19
21. Knock ES, Whittles LK, Lees JA, Perez-Guzman PN, Verity R, FitzJohn RG, et al. Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. Sci Transl Med. 2021;13:eabg4262. https://doi.org/10.1126/scitranslmed.abg4262
22. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. [Erratum in Science. 2020;368:eabd4246]. Science. 2020;369:208–11. https://doi.org/10.1126/science.abc3517
23. Hay JA. lazymcmc R package [cited 2021 Mar 23]. https://github.com/jameshay218/lazymcmc
24. R: a language and environment for statistical computing [cited 2021 Mar 23]. https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing
25. Malani A, Shah D, Kang G, Lobo GN, Shastri J, Mohanan M, et al. Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India. Lancet Glob Health. 2021;9:e110–1. https://doi.org/10.1016/S2214-109X(20)30467-8
26. Laxminarayan R, B CM, G VT, Arjun Kumar KV, Wahl B, Lewnard JA. SARS-CoV-2 infection and mortality during the first epidemic wave in Madurai, south India: a prospective, active surveillance study. Lancet Infect Dis. 2021;21:1665–76. https://doi.org/10.1016/S1473-3099(21)00393-5
27. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, et al.; Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health. 2020;8:e1003–17. https://doi.org/10.1016/S2214-109X(20)30264-3
28. Kumar P, Chander B. COVID 19 mortality: Probable role of microbiome to explain disparity. Med Hypotheses. 2020;144:110209. https://doi.org/10.1016/j.mehy.2020.110209
29. Walker PGT, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science. 2020;369:413–22. https://doi.org/10.1126/science.abc0035
30. Chandrashekhar V. Herd immunity? India still has a long way to go, scientists say. Science. 2020;370:513. https://doi.org/10.1126/science.370.6516.513
31. Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira IATM, et al.; Indian SARS-CoV-2 Genomics Consortium (INSACOG); Genotype to Phenotype Japan (G2P-Japan) Consortium; CITIID-NIHR BioResource COVID-19 Collaboration. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. Nature. 2021;599:114–9. https://doi.org/10.1038/s41586-021-03944-y

Address for correspondence. Margarita Pons-Salort, Imperial College London, Norfolk Place, W2 1PG London, UK; email: m.pons-salort@imperial.ac.uk

Article DOI: https://doi.org/10.3201/eid2804.210879

Reassessing Reported Deaths and Estimated Infection Attack Rate during the First 6 Months of the COVID-19 Epidemic, Delhi, India

Appendix

Parameter Inference: Observation Models

To account for underreporting and overdispersion of the death data, we modeled the number of deaths with a negative binomial distribution:

$$deaths^{obs}(t) \sim NegBin(\mu = \theta \times deaths(t), k)$$

where the mean is $\theta \times deaths(t)$ and the variance is $\theta \times deaths(t) + (\theta \times deaths(t))^2/(\theta \times deaths(t))^k$.

We modeled the number of persons who would test seropositive each day with a given serological assay of sensitivity Se_j and specificity Sp_j as

$$seropositives_j(t) = \sum_{i=1}^{t-1} e_{inc}(i) * p_{sero,j}(t-i) + (1 - Sp_j) * s_{prev}(t)$$

where e_{inc} was the incidence of infection, $p_{sero,j}(\tau)$ was the probability of testing positive τ days after infection and s_{prev} was the number of susceptible persons, as in (J. Ojal, unpub. data). We assumed that $p_{sero,j}(\tau)$ increases linearly from day 0 of infection to Se_j 26 days after infection and remains constant after that (we did not consider seroreversion):

$$p_{sero,j}(\tau) = \begin{cases} Sej \times \frac{\tau}{26}, \text{ if } 0 < \tau < 26 \\ Sej, \text{ if } \tau \geq 26 \end{cases}$$

Finally, we linked the modelled number of seropositives at the mid–time point of each serosurvey, denoted here by ts_1 for survey 1, ts_2 for survey 2, and ts_3 for survey 3, to the data

from the 3 seroprevalence surveys using Beta binomial distributions to account for overdispersion:

$$seropositives(t_{sj})^{obs} \sim$$

$$BetaBinom(N = N_{samples,j}, p = {seropositives_j(t_{sj})}/{population}, \sigma = 1/0.7)$$

with the overdispersion parameter fixed to $\sigma = 1.0/0.7$, because of the small number of observations available to accurately estimate it.

Equations for the Transmission Model

See notation and description of variables in Appendix Table 4 and parameters in Appendix Table 5.

$$\frac{dS}{dt} = -\lambda S$$

$$\frac{dE}{dt} = \lambda S - \omega E$$

$$\frac{dI}{dt} = \omega E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$\lambda = R_i \gamma \frac{I}{N}$$

$$N = S + E + I + R$$

Equations for the Disease-Progression Model

See notation and description of variables in Appendix Table 4 and parameters in Appendix Table 5.

$$\frac{dPSY_1}{dt} = \lambda S - \frac{6}{D_{PSY}} PSY_1$$

$$\frac{dPSY_2}{dt} = \frac{6}{D_{PSY}} PSY_1 - \frac{6}{D_{PSY}} PSY_2$$

…

$$\frac{dPSY_6}{dt} = \frac{6}{D_{PSY}} PSY_5 - \frac{6}{D_{PSY}} PSY_6$$

$$\frac{d(SY + ASY)}{dt} = \frac{6}{D_{PSY}} PSY_6 - \frac{1}{D_{SY}} (SY + ASY)$$

$$\frac{dHNC}{dt} = p_{inf,hnc} \frac{1}{D_{SY}} (SY + ASY) - \frac{1}{D_{HNC}} HNC$$

$$\frac{dHC}{dt} = p_{inf,hc} \frac{1}{D_{SY}} (SY + ASY) - \frac{1}{D_{HC}} HC$$

$$\frac{dHSD}{dt} = p_{hc,hsd} \frac{1}{D_{HC}} HC - \frac{1}{D_{HSD}} HSD$$

$$\frac{dF}{dt} = p_{inf,death} \frac{1}{D_{SY}} (SY + ASY) - \frac{1}{D_F} F$$

$$\frac{dD}{dt} = \frac{1}{D_F} F$$

References

1. Sharma N, Sharma P, Basu S, Saxena S, Chawla R, Dushyant K, et al. The seroprevalence of severe acute respiratory syndrome coronavirus 2 in Delhi, India: a repeated population-based seroepidemiological study. Trans R Soc Trop Med Hyg. 2021;trab109. PubMed https://doi.org/10.1093/trstmh/trab109

2. Malani A, Shah D, Kang G, Lobo GN, Shastri J, Mohanan M, et al. Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India. Lancet Glob Health. 2020;9:e110–1.

3. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis. 2020;20:911–9. PubMed https://doi.org/10.1016/S1473-3099(20)30287-5

4. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;172:577–82. PubMed https://doi.org/10.7326/M20-0504

5. Lau YC, Tsang TK, Kennedy-Shaffer L, Kahn R, Lau EHY, Chen D, et al. Joint estimation of generation time and incubation period for coronavirus disease (Covid-19). J Infect Dis. 2021;jiab424. PubMed https://doi.org/10.1093/infdis/jiab424

6. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6. PubMed https://doi.org/10.1016/j.ijid.2020.02.060

7. Xin H, Li Y, Wu P, Li Z, Lau EHY, Qin Y, et al. Estimating the latent period of coronavirus disease 2019 (COVID-19). Clin Infect Dis. 2021;ciab746. PubMed https://doi.org/10.1093/cid/ciab746

8. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020;20:669–77. PubMed https://doi.org/10.1016/S1473-3099(20)30243-7

9. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, et al.; Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health. 2020;8:e1003–17. PubMed https://doi.org/10.1016/S2214-109X(20)30264-3

Appendix Table 1. Information on the 3 serosurveys conducted in Delhi (*1*) during the study period.*,†

Serosurvey no. (agency)	Midtime point	Samples, no.	Uncorrected seropositivity rate	Test sensitivity	Test specificity
1 (NCDC)	07/01/2020	19,041	0.2283	0.921	0.977
2 (MAMC)	08/04/2020	15,046	0.287	0.921	0.977
3 (MAMC)	09/04/2020	17,409	0.251	0.9912	0.9933

NCDC, National Center for Disease Control; MAMC: Maulana Azad Medical College
†Serosurveys 1 and 2 used the testing kit ELISA COVID-Kawach kit, serosurvey 3 used the ERBALISA COVID-19 IgG.

Appendix Table 2. National interventions to control COVID-19 in India.

Intervention	Start date	End date	Modeled change in R?	Observations
Janata curfew	03/22/2020	03/22/2020	No	14-hour curfew
Phase 1 lockdown	03/25/2020	04/14/2020	No*	Beginning of lockdown
Phase 2 lockdown	04/15/2020	05/03/2020	No	Lockdown extended
Phase 3 lockdown	05/04/2020	05/17/2020	Yes	Lockdown extended but with relaxations
Phase 4 lockdown	05/18/2020	05/31/2020	No	Lockdown extended
Unlock 1.0	06/01/2020	06/30/2020	Yes	Reopening phase
Unlock 2.0	07/01/2020	07/31/2020	Yes	Reopening phase
Unlock 3.0	08/01/2020	08/31/2020	Yes	Reopening phase
Unlock 4.0	09/01/2020	09/30/2020	No	Reopening phase

*Insufficient data available to estimate a change in transmission at that point.

Appendix Table 3. Estimate of death reporting in Delhi and Mumbai based on a simple back calculation based on data on the cumulative number of deaths up until the first serosurvey. For Mumbai, we used the seroprevalence data reported elsewhere (*2*), we assume that 53% of the population lives in slums, and account for different age distribution in slums and non-slums, based on data from the TIFR Covid-19 City-Scale Simulation Team (P. Harsha, unpub.data).

Area	Date†	Cumulative reported deaths	Seroprevalence, %‡	Population size, M	IFR, %¶	Deaths reported, %
Delhi	07/01/2020	2,803	22.86	20.86	0.39	15
Mumbai	07/07/2020	3,317	NA	12.8	NA	21
Slums	NA	NA	55.7	6.784	0.29	NA
Non-slums	NA	NA	16.2	6.016	0.51	NA

*IFR, infection fatality ratio; M, millions; NA, not available
†Midtime point of sample collection during the serosurvey.
‡Corrected for test sensitivity and specificity.
¶Based on model described elsewhere (N.F. Brazeau, unpub. data).

Appendix Table 4. Model variables.

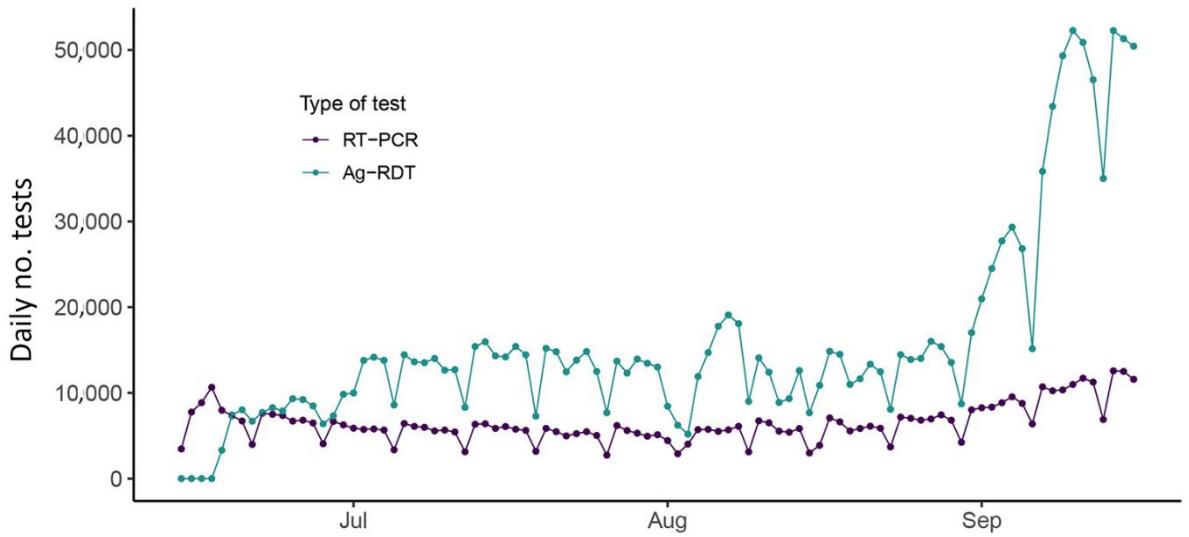
Notation	Description
S	Susceptible
E	Exposed
I	Infected
R	Recovered
N	Total population size
PSY1…PSY6	Presymptomatic 1…6*
SY	Symptomatic
ASY	Asymptomatic
HNC	Hospitalized, no critical care
HC	Hospitalized, critical care
HSD	Hospitalized, stepdown
F	Fatal infections
D	Deaths

*The 6 compartments were necessary to model the Erlang distribution with shape value 6)

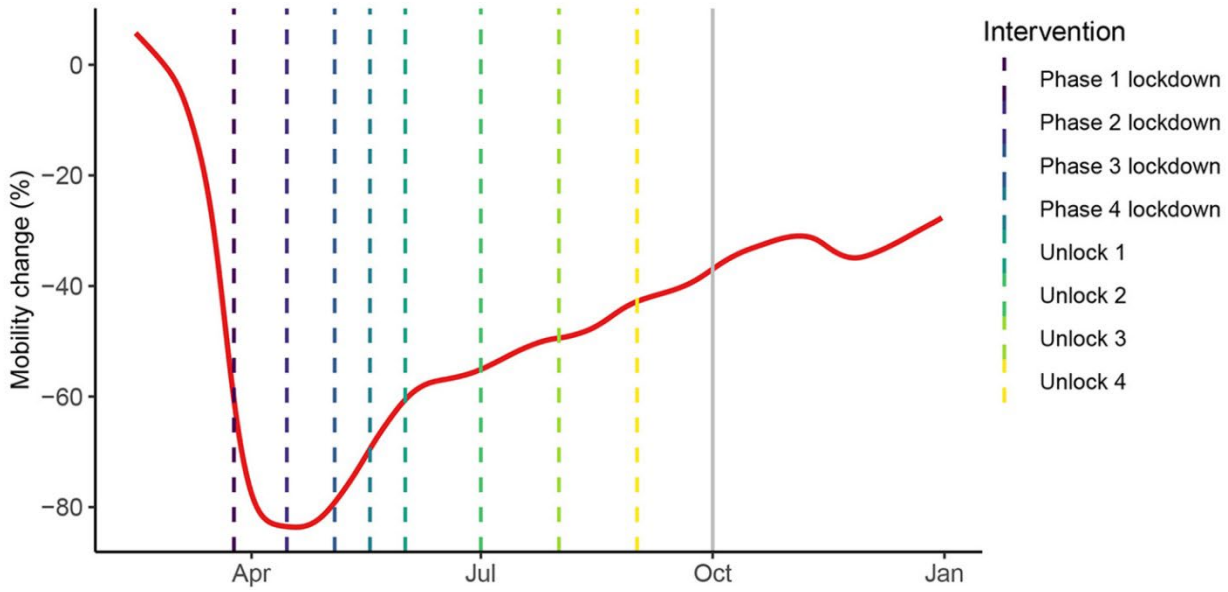
Appendix Table 5. Model parameters

Notation	Description	Value and units	Reference
R_0	Basic reproduction number	Estimated	NA
r1…r5	Coefficients modifying the reproduction number	Estimated	NA
$1/\omega$	Mean duration preinfectious period	4.5 d	Set using information on the generation time (*3*), the duration of the pre-symptomatic period (*4*) and the fact that infectiousness starts about one day before symptoms start (*5*–*7*) (see Methods for details)
$1/\gamma$	Mean duration infectious period	2 d	As above
D_{PSY}	Mean duration presymptomatic	5.5 d	(*4*)
D_{SY}	Mean duration symptomatic	5.8 d	(K. Gaythorpe, unpub. data)
D_{HNC}	Mean duration hospitalized non–critical care	9.8 d	Based on early, unpublished data from the UK
D_{HC}	Mean duration hospitalized critical care	9.8 d	Based on early, unpublished data from the UK
D_{HSD}	Mean duration hospitalized step-down	3.3 d	Based on early, unpublished data from the UK.
D_F	Mean duration infections leading to fatal outcome	10 d	Based on a mean time from symptom onset to hospitalization of 5.8 days (K. Gaythorpe, unpub. data) and an average time from symptom onset to death of ≈16 days (*8*)
$p_{INF,HNC}$	Proportion of infections leading to hospitalization for noncritical care (age-adjusted)	2.77×10^{-2}	(*8*)
$p_{INF,HC}$	Proportion of infections leading to hospitalization for critical care (age-adjusted)	6.64×10^{-3}	(*8*)
$p_{INF,D}$	Proportion of infections leading to death (age-adjusted IFR)	3.94×10^{-3} †	(N.F. Brazeau, unpub. data)
$p_{HC,HSD}$	Proportion of hospitalizations in critical care recovering (in stepdown compartment)	0.6	Working assumption

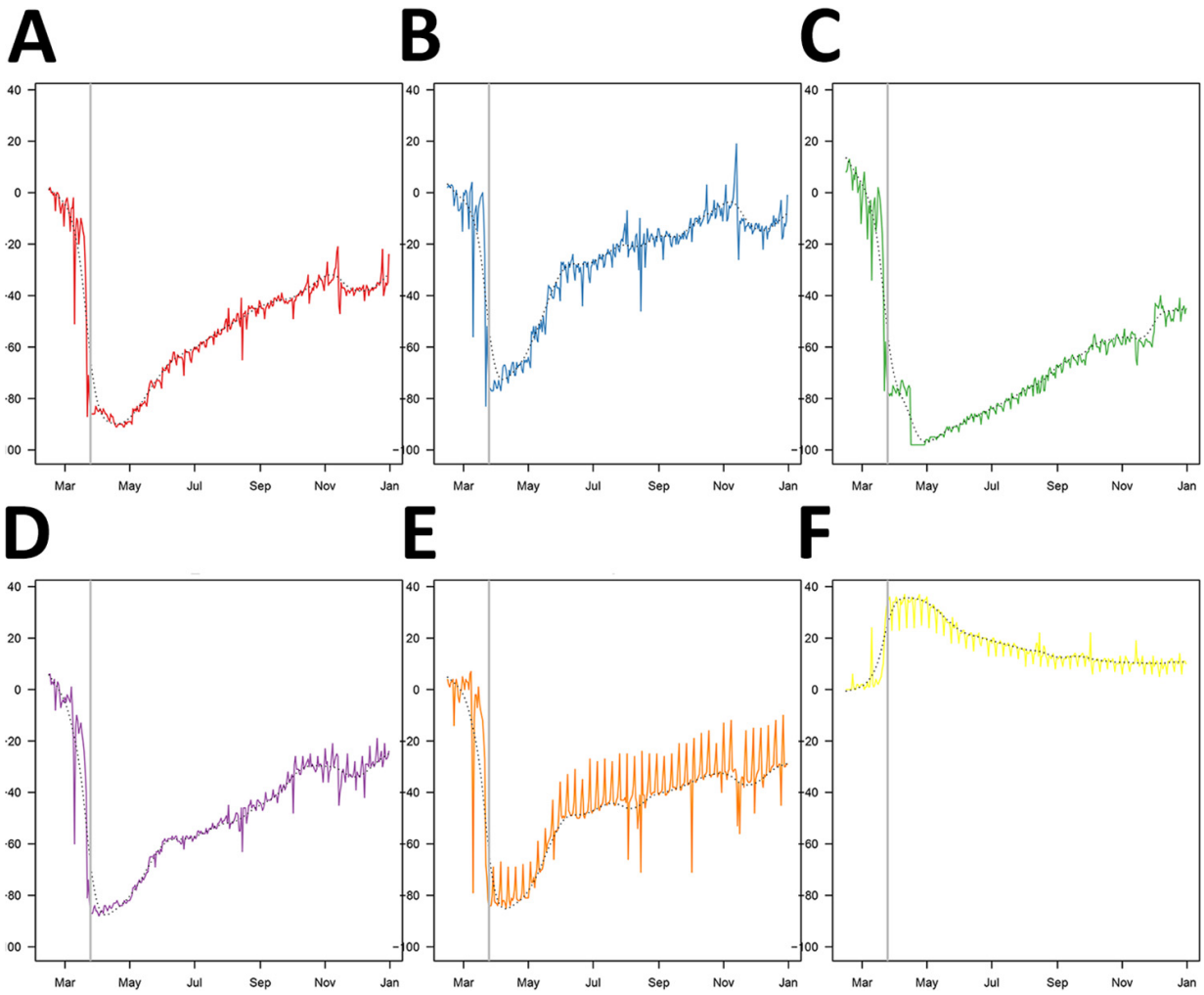
***IFR, infection fatality ratio; NA, not applicable**
†Value used in the main analysis. Sensitivity analysis performed based on this value.



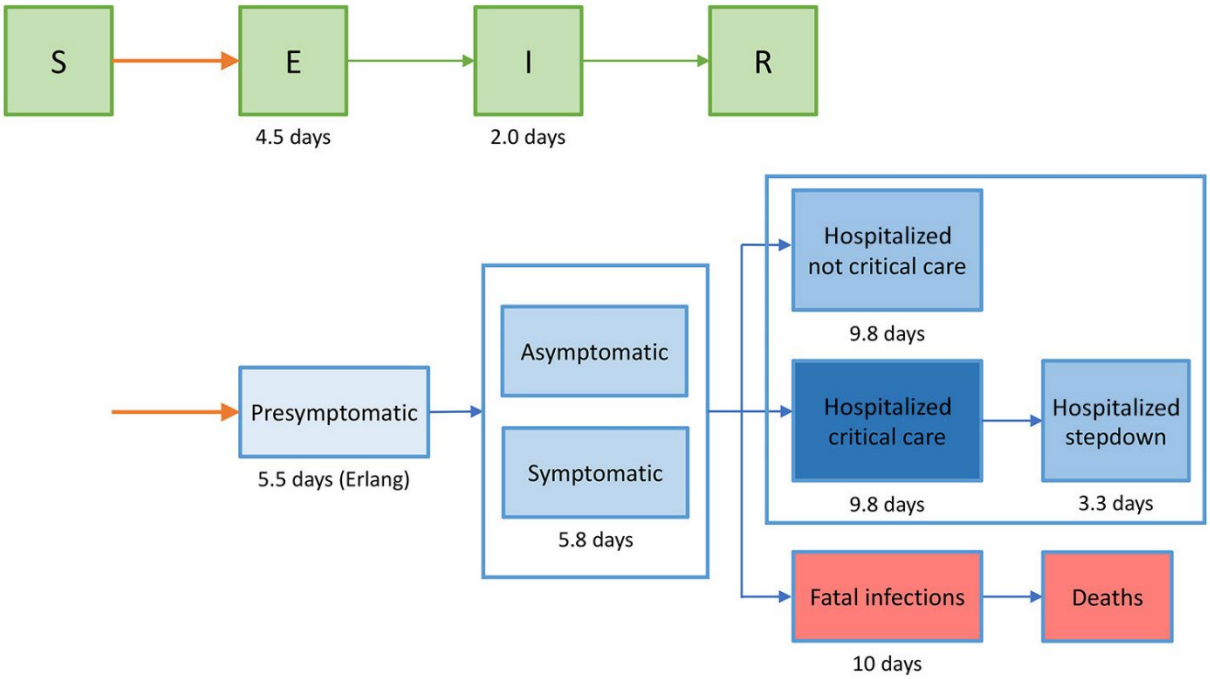
Appendix Figure 1. Daily number of reverse transcription PCR and antigen-detecting rapid diagnostic tests performed, June 14–September 16, 2020. Ag-RDT, antigen-detecting rapid diagnostic tests; RT-PCR, reverse transcription PCR



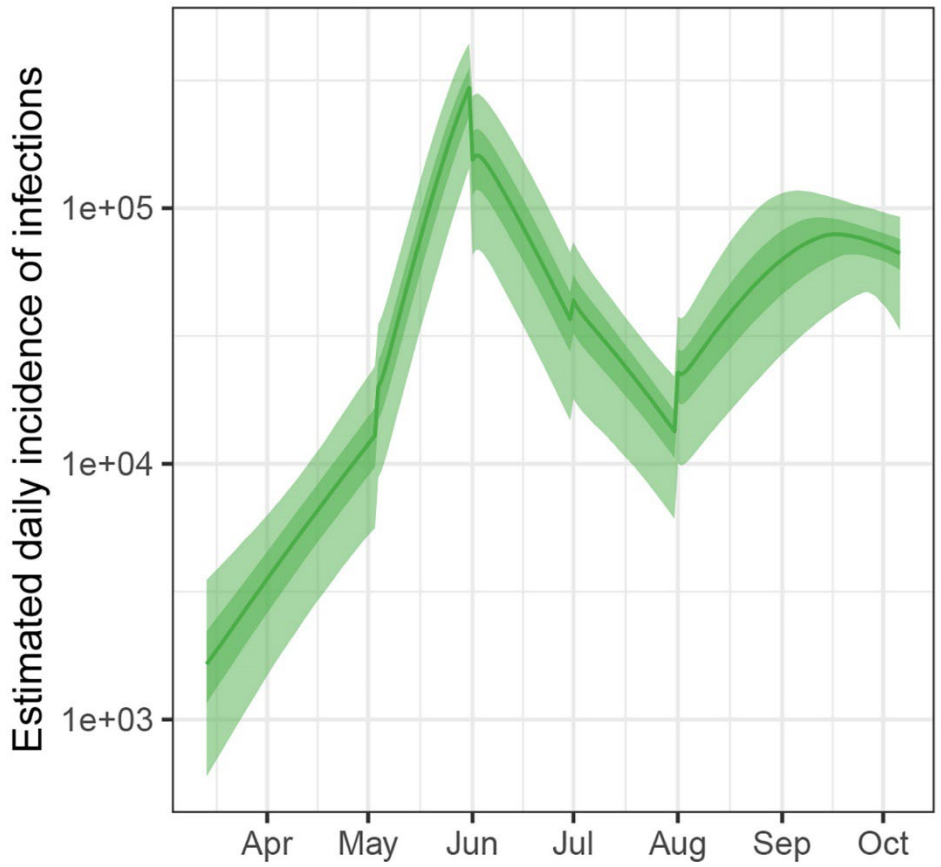
Appendix Figure 2. Changes in mobility over time (based on summarized Google data, February 15–December 31, 2020). The data were smoothed to remove weekend effect and averaged across 5 of the 6 data streams available (all except residential; Appendix Figure 3). Unlock 1–4 refers to gradual reopenings. Colored lines indicate the start of interventions (Appendix Table 2). Grey vertical line indicates the end of the study period (September 30, 2020).



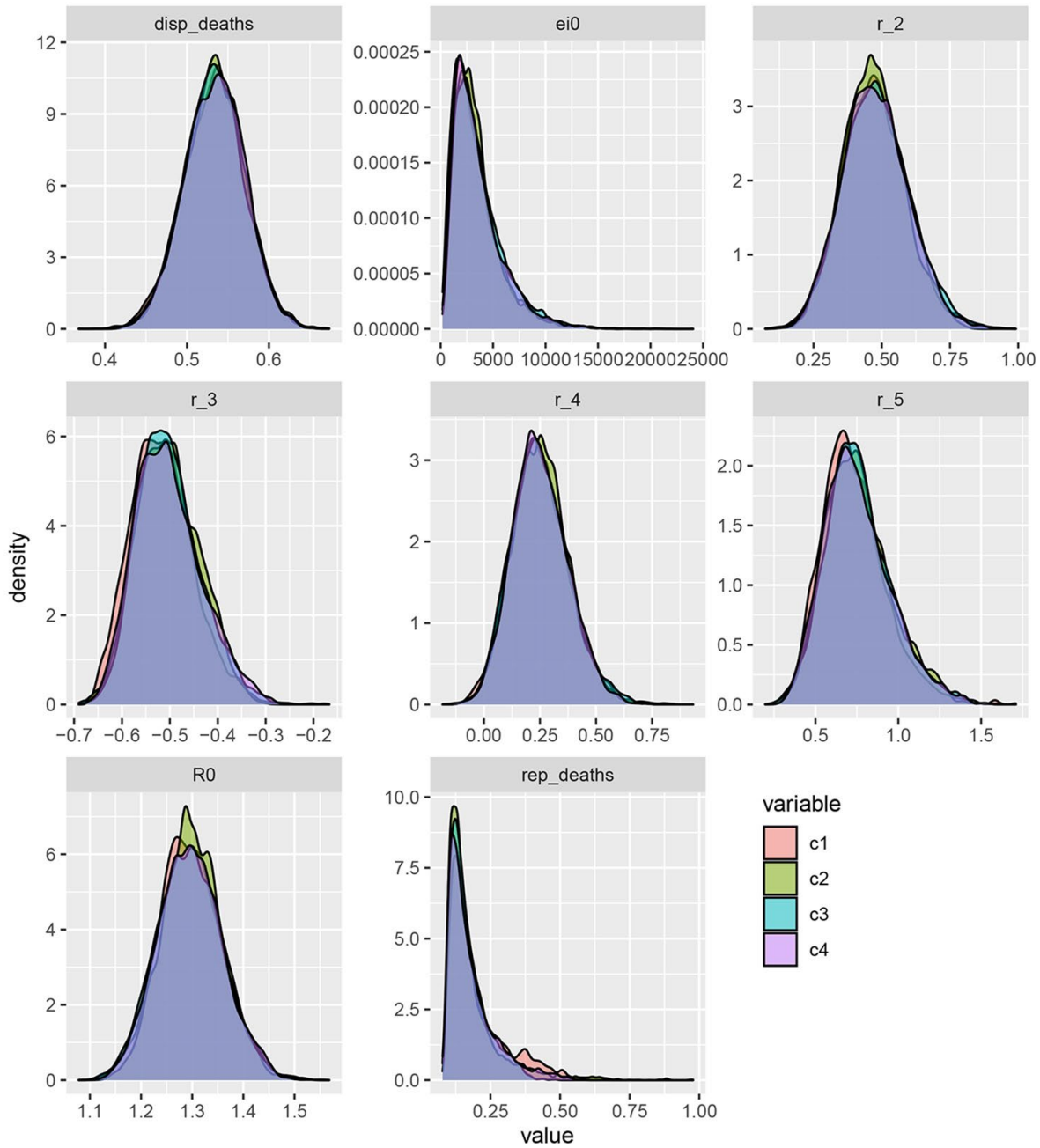
Appendix Figure 3. Changes in mobility over time for Delhi (based on raw data from Google). Raw data for the 6 available data streams are shown in red for retail and recreation (A), blue for grocery and pharmacy (B), green for parks (C), purple for transit stations (D), orange for workplaces (E), and in yellow for residential (F). Dashed black lines indicated smoothed raw data obtained with a spline after removing the points for Thursday to Sunday to remove the weekend effect. These smoothed lines for all the streams, except residential (F), were averaged to produce Appendix Figure 2. Vertical line in each panel indicates the beginning of the lockdown on March 25, 2020.



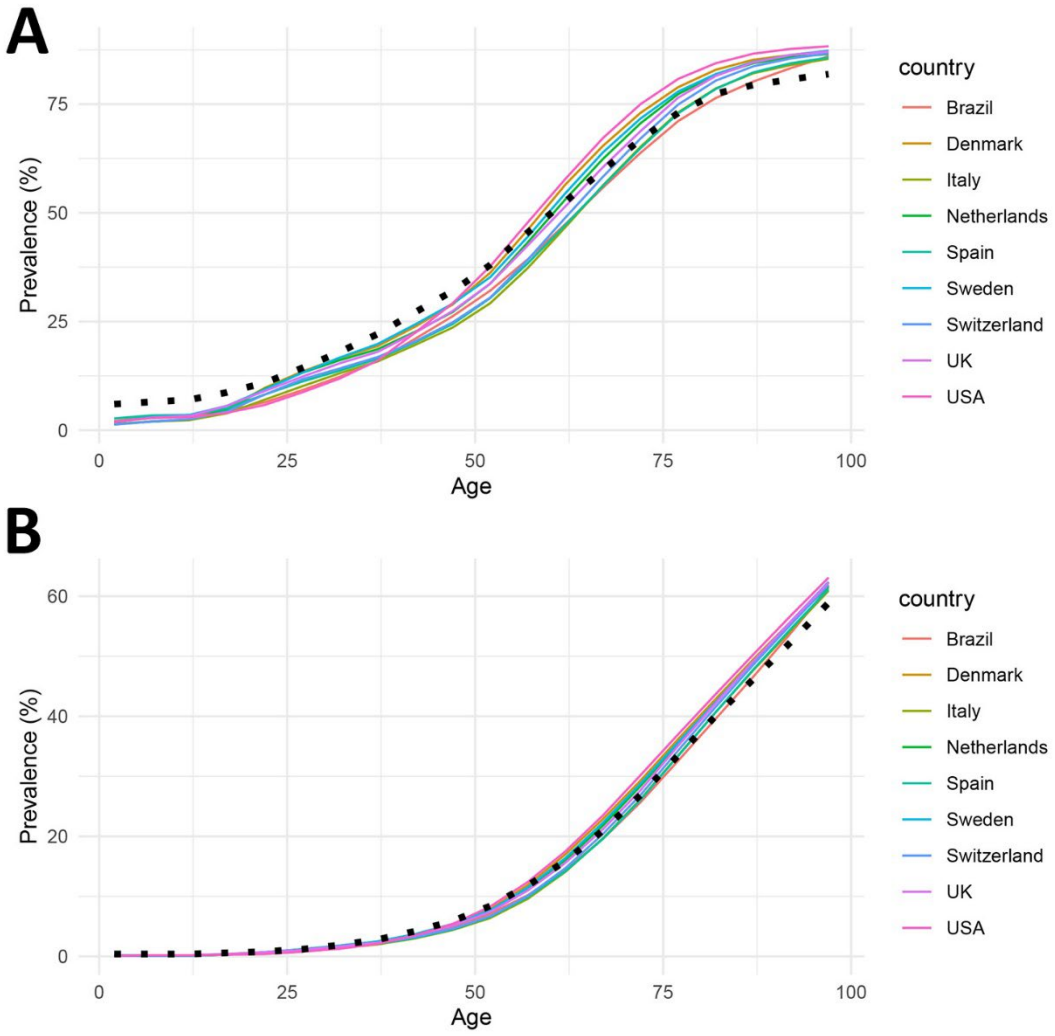
Appendix Figure 4. Model diagram. In green, the susceptible-exposed-infectious-removed (SEIR) transmission model. Blue boxes: the disease-progression model; red boxes: infections leading to death. S, susceptible; E, exposed; I, infectious; R, removed



Appendix Figure 5. Reconstructed estimated incidence of infections, with 50% (light green) and 95% (dark green) credible intervals.



Appendix Figure 6. Posterior densities of estimated parameters for 4 different Markov chain Monte Carlo chains initialised at different starting values.



Appendix Figure 7. Prevalence of A) ≥1 or B) ≥2 concurrent conditions for severe COVID-19 by age for India (black dotted line) and the 9 countries that informed the age-specific estimates of the infection fatality ratio used in the model (*9*).