BMJ Open

Representative estimates of COVID-19 infection fatality rates from four locations in India: cross-sectional study

Rebecca Cai,[1] Paul Novosad ![ORCID],[2] Vaidehi Tandel,[3] Sam Asher,[4] Anup Malani ![ORCID] [5]

¹Development Data Lab, Washington, District of Columbia, USA
²Economics, Dartmouth College, Hanover, New Hampshire, USA
³Real Estate and Planning Henley Business School, University of Reading, Reading, UK
⁴Economics, Johns Hopkins University School of Advanced International Studies, Washington, District of Columbia, USA
⁵University of Chicago Law School, Chicago, Illinois, USA

Correspondence to
Dr Anup Malani;
amalani@uchicago.edu

ABSTRACT

Objectives To estimate age-specific and sex-specific mortality risk among all SARS-CoV-2 infections in four settings in India, a major lower-middle-income country and to compare age trends in mortality with similar estimates in high-income countries.

Design Cross-sectional study.

Setting India, multiple regions representing combined population >150 million.

Participants Aggregate infection counts were drawn from four large population-representative prevalence/seroprevalence surveys. Data on corresponding number of deaths were drawn from official government reports of confirmed SARS-CoV-2 deaths.

Primary and secondary outcome measures The primary outcome was age-specific and sex-specific infection fatality rate (IFR), estimated as the number of confirmed deaths per infection. The secondary outcome was the slope of the IFR-by-age function, representing increased risk associated with age.

Results Among males aged 50–89, measured IFR was 0.12% in Karnataka (95% CI 0.09% to 0.15%), 0.42% in Tamil Nadu (95% CI 0.39% to 0.45%), 0.53% in Mumbai (95% CI 0.52% to 0.54%) and an imprecise 5.64% (95% CI 0% to 11.16%) among migrants returning to Bihar. Estimated IFR was approximately twice as high for males as for females, heterogeneous across contexts and rose less dramatically at older ages compared with similar studies in high-income countries.

Conclusions Estimated age-specific IFRs during the first wave varied substantially across India. While estimated IFRs in Mumbai, Karnataka and Tamil Nadu were considerably lower than comparable estimates from high-income countries, adjustment for under-reporting based on crude estimates of excess mortality puts them almost exactly equal with higher-income country benchmarks. In a marginalised migrant population, estimated IFRs were much higher than in other contexts around the world. Estimated IFRs suggest that the elderly in India are at an advantage relative to peers in high-income countries. Our findings suggest that the standard estimation approach may substantially underestimate IFR in low-income settings due to under-reporting of COVID-19 deaths, and that COVID-19 IFRs may be similar in low-income and high-income settings.

Strengths and limitations of this study

► This study provides representative estimates of the age-specific COVID-19 infection fatality rate (IFR) in four socioeconomically diverse regions of India, a major lower-middle-income country, using the standard method for estimating IFR.

► Due to high measurement cost, there are very few age-specific IFR estimates in low-income and middle-income countries (LMICs), despite concerns that LMICs are more vulnerable and plausibly have different mortality patterns.

► This study uses the primary method of estimating IFR in settings around the world, combining population-representative prevalence/seroprevalence surveys with official death reports, allowing direct methodological comparison with dozens of similar estimates from high-income countries.

► We provide population-representative estimates for over 150 million people using the largest sample to date in an LMIC, and the first documentation of IFR among the large, highly vulnerable population of migrant workers.

► The main limitation is our reliance on official reports of confirmed COVID-19 deaths, which, due to under-reporting and undertesting, likely underestimate the true number of deaths.

INTRODUCTION

Measuring the infection fatality rate (IFR) for SARS-CoV-2 has been a major research objective since the beginning of the global pandemic. Reliable IFR estimates are essential for policy decisions on non-pharmaceutical interventions and vaccine allocation,[1–3] and comparison of waves and variants. IFR estimates almost universally rely on large-scale seroprevalence samples drawn from the general population, matched to official death data. Because of these data requirements, the vast majority of age-specific IFR estimates are based on data from high-income countries (HICs)[2–6]; meta-analyses estimating age-specific IFR in low-income and middle-income countries (LMICs)[7 8] rely on untested assumptions that key epidemiological characteristics (eg, transmission dynamics, age-specific death rate) in HICs are generalisable to low-income settings. Studies measuring IFR

in LMICs mostly report age-aggregated IFR,[9–13] which are difficult to compare across contexts; the age pattern of infection may vary and aggregate IFRs skew higher where older people contract a larger share of infections. Estimates of age-specific IFR in LMICs have only been made from small or non-representative samples.[14 15]

Early modellers of lower-income settings warned that IFRs could be higher, due to worse baseline population health and under-resourced healthcare systems.[8 15 16] Other researchers observed low case fatality rates in sub-Saharan Africa and proposed that vaccination, infection history and effective mitigation strategies might have reduced mortality.[17 18] The age pattern of deaths in lower-income countries has skewed younger than in HICs, more so than can be explained by age distribution alone.[19–21]

We calculated age-specific IFRs from four samples in India representing a combined population exceeding 150 million. We used population-representative seroprevalence surveys in the city of Mumbai (N≅7000, population 12.5 million) and in the states of Karnataka (N≅1200, population 61 million) and Tamil Nadu (N≅26 000, population 71 million). By matching these surveys to age-specific administrative death data, we calculated IFR without relying on non-representative testing data. Additionally we drew on a survey of COVID-19 prevalence among randomly sampled short-term outmigrants (N≅4000 infections, population minimum 10 million), mostly working-age males, returning home to the state of Bihar with mortality follow-up. Because these migrants were randomly sampled and tracked until recovery or death, the death rate among those who tested positive is interpretable as an IFR.

Our objective was to calculate age-specific IFRs in four locations and compare them to international estimates, which are based mostly on HICs. We further examined heterogeneity of IFR within India and by age and sex.

Importantly, data collection took place during India's first wave of COVID-19 between March and December 2020. India has since undergone a second, more severe wave between March and June 2021, characterised by much higher case counts, new and potentially more transmissible variants and a health system crisis.[22] Excess mortality and reports suggest more severe infections and higher mortality in the second wave.[22] Our IFR estimates apply to the first wave, and should not be interpreted as representative for the second.

METHODS

We studied three states and one megacity with disparate demographic and health characteristics (table 1). Qualitatively, Tamil Nadu and Karnataka are large, relatively wealthy, southern Indian states. Mumbai is India's most populous city, and the capital of the western state Maharashtra. Tamil Nadu, Karnataka and Maharashtra have relatively robust healthcare infrastructure and vital registration.[23] In contrast, the northern state Bihar is one of the poorest in India, with the lowest stock of hospital beds per capita.[24]

The Bihar sample is limited to a subpopulation of returning migrants, primarily young male labourers who lost work opportunities during lockdown. The returning migrants to Bihar are part of a large population of internal labour migrants in India; a conservative estimate from the 2001 Census found that nearly 30 million workers migrated within India for employment.[25] Tens of millions of migrants exited cities immediately after lockdown, including 6.3 million travelling on specially designated trains ('Shramik Specials') between May and August, 2020.[26 27] Short-term migrants were on average very poor even before the pandemic.[28] India's sudden lockdown left them unemployed, and many experienced extreme physical and economic duress on the long journey home.[29 30]

India began its first nationwide lockdown on 24 March 2020, and by July 2021 had the second-highest number of country-wide confirmed COVID-19 cases in the world. The Indian government spends roughly 1.5% of gross domestic product on healthcare, one of the world's lowest rates.[31] Discussion of India's COVID-19 preparedness has focused on under-resourced public hospitals, a

Table 1 Health and demographic context of sample locations

	Median age	GDP/capita	Cumulative infections on July 31	Cumulative COVID-19 deaths on July 31	Hospital beds per 100 000 population
	Population census 2011	NSDP nominal (2018–2019 INT$)	JHU CSSE COVID-19 data[44]	JHU CSSE COVID-19 data[44]	Kapoor et al[24]
Bihar	19.9	640	51 233	296	25.55
Maharashtra*	28.2	2802	411 798	14 994	172.94
Karnataka	27.4	3082	124 115	2314	391.62
Tamil Nadu	29.9	2831	245 859	3935	174.83
India	24.0	1964	1 695 988	36 511	137.62

Row 2 indicates the data source.
*Mumbai is the capital city of Maharashtra.
GDP, Gross domestic product; JHU CSSE, Johns Hopkins University Center for Systems Science and Engineering; NSDP, Net state domestic product.

largely unregulated private healthcare sector and fear and stigma among the public surrounding infection.[31]

Data sources and study design

In Mumbai, Karnataka and Tamil Nadu, we matched representative seroprevalence surveys to administrative reports of confirmed COVID-19 deaths.

In Mumbai, seroprevalence surveys were conducted for 2 weeks in July 2020 with representative sampling of three wards, one from each of the city's three zones, stratified by age, sex and slum/non-slum dwellers.[10] Enumerators sought voluntary consent to sample one member per household, rotating through age-gender groups. Thus, the sample composition is representative for city-wide age and sex, subject to consent rates. The sample consisted of 6904 participants (4202 from slums and 2702 from non-slums), tested for IgG antibodies to the SARS-CoV-2 N-protein using the Abbott Diagnostics Architect test. Data on cumulative deaths were collected from daily reports from the municipal governing body.

In Karnataka, seroprevalence surveys were conducted from 15 June 2020 to 29 August 2020, in representative samples of urban and rural areas in 20 out of 30 districts, stratified to generalise to 5 regions spanning all districts.[32] We can, therefore, take the ELISA positive test rate as an unbiased measure of region-level positivity rate. The sampling frame was not age stratified or sex stratified, and older individuals were oversampled relative to population age composition. We assume that ELISA positive test rate is representative by age–sex–region group, because there was no evidence that the age of the consenting member of each household was associated with seropositivity in the home. A total of 1196 participants were tested with an ELISA for antibodies to the receptor binding domain of the SARS-CoV-2 virus, developed by Translational Health Science and Technology Institute in India. We collected district-level death data from the Government of Karnataka Department of Health and Family Welfare bulletins.

In Tamil Nadu, a representative seroprevalence survey was conducted between 19 October 2020 and 30 November 2020, of adults aged 18 and older, covering the state's 37 districts.[33] Collection times within districts were often significantly shorter. Enumerators divided districts into health unit districts, then randomly sampled urban and rural clusters. Within clusters, enumerators started at a randomly selected GPS starting point, sampling one person from households adjacent to the starting point (using the Kish method) to provide a biosample. Because household members were selected randomly, we similarly assume seropositivity is representative at the age–sex–district level. Seropositivity was tested using either the iFlash-SARS-CoV-2 IgG or the Vitros anti-SARS-CoV-2 IgG CLIA kit. The analytical subsample was 26 107 antibody tests that could be conclusively determined as positive or negative. Case-level data on 12 019 recorded statewide COVID-19 deaths, from March to December 2020 was collected from daily government reports.

In Bihar, the state government began COVID-19 testing among returning out-of-state migrants soon after the first positive case was identified in a migrant on 22 March 2020. On 4 May, Bihar began to randomly select migrants for testing. Random testing continued until 21 July, though for a brief window (22 May–31 May) only migrants returning from seven major cities were sampled. We isolated the subsample of randomly selected migrants, yielding 4362 individuals with positive tests.[29] Tests were conducted with TrueNat machines manufactured by MolBio Diagnostics in Goa, with positive tests confirmed by real-time PCR kits.[34] Bihar attempted to track all migrants who tested positive until they eventually recovered or died.

In all locations, population data came from the 2012 Socio-Economic and Caste Census.

Statistical analysis

In Mumbai, Karnataka and Tamil Nadu, we estimated infection counts from representative seroprevalence surveys. Methods for estimating infection counts are described in detail below. We matched infection counts to deaths assuming that the infection-seroconversion delay is on average 2 days shorter than the infection-death delay.[35 36] To implement this, we calculated IFR as the cumulative number of deaths reported as of 2 days after the end of seroprevalence testing, divided by the number of infections. Testing sensitivity to this assumption, we replicate results using deaths from 1 and 2 weeks after last day of seroprevalence testing, effectively generating upper bounds for the number of deaths (online supplemental figures 1–3 in the online supplemental file 1). Where multiple evaluations of the antibody tests' sensitivity/specificity existed, we tested robustness to assuming minimum sensitivity (online supplemental figures 4 and 5 in online supplemental file 1).

In Mumbai, we first adjusted for test sensitivity and specificity using the Rogan-Gladen correction,[37] then calculated aggregate seroprevalence for each sampled ward and multiplied by ward population to estimate infection count. We estimated infection counts in non-sampled wards by assuming a constant rate of government under-reporting in wards in the same zone. This approach was supported by very similar case-to-seroprevalence ratios in the three wards with seroprevalence data (online supplemental table 1). Age-specific and sex-specific infection shares were based on the seroprevalence survey (online supplemental figure 6).

In Karnataka, we adjusted for test inaccuracies,[37] then used census population counts to aggregate from regional to state-level infection counts, reweighting to match regional age–sex distributions. Methods for matching dates and deaths to infections is described in detail in (online supplemental figure 7. Because the seroprevalence survey period in Bangalore spanned 2 months (compared with less than 3 weeks in the other regions), we show results excluding Bangalore, where deaths may

have been overestimated due to the longer survey period (online supplemental figure 8).

In Tamil Nadu, we first calculated the population-representative seropositivity rate by district–age–sex group and type of test kit, then adjusted for test inaccuracies. We estimated the number of statewide infections per district–age–sex group by combining kit-specific seroprevalence estimates and multiplying by population, then summing across districts. In sensitivity checks, we re-estimated IFR limiting samples to districts where seroprevalence surveillance lasted less than 3 weeks (online supplemental table 2 and figure 9).

In Bihar, although enumerators attempted to track outcomes for all migrants, 1530 (35%) infected individuals could not be tracked. In main estimates, we assumed that their fatality rates were the same as successfully tracked individuals; in sensitivity checks, we considered the possibility that all survived. High attrition is common in studies of migrant workers,[29] with follow-up in this case complicated by the ongoing crisis. We limited our analytic sample to 3921 randomly sampled male migrants, for whom 2536 outcomes are known.

Information on underlying sample size, seroprevalence rate and number of deaths used to calculate IFRs in each location are in online supplemental tables 3–6 and online supplemental file 1.

Matching representative seroprevalence surveys to administrative death data is the primary method of IFR measurement everywhere in the world.[2 4 5] In Bihar, because migrants were randomly sampled, there was no selection on symptomatic or severe cases, and mortality rates among positive cases can be interpreted as IFRs. As noted above, short-term migrants from Bihar are economically marginalised; their IFRs can be understood as representative for migrants, but not necessarily the general population.

We calculated IFRs in 10-year age bins, plus bins 10–49 and 50–89, in all locations. We used two large-scale meta-analyses[1 7] of age-specific SARS-CoV-2 IFRs as reference groups. Both Levin et al[1] and O'Driscoll et al[7] draw almost exclusively from seroprevalence samples from Europe and the USA. The application of these samples to mortality in LMICs (as in O'Driscoll et al[7]) requires the as-yet untested assumption that multiple epidemiological factors (eg, transmission dynamics) are uniform between HIC and LMIC. Levin et al[1] do not report IFR by sex; we estimated sex-specific IFRs in Levin et al[1] by assuming the same sex ratio in IFR as reported in O'Driscoll et al[7]. For the larger age bins, we weighted age-specific IFR estimates from sample populations and meta-analyses by the Indian national population distribution, to ensure differences across contexts were driven by differences in age-specific IFRs, rather than population age distribution.

We calculated the slope of the natural log of IFR as a function of age by fitting a linear function to the most granular age-specific IFR data that could be obtained in each location. Additional details on the underlying

samples and the methodology are in online supplemental materials. All analyses were conducted in Stata V.16.0.

Patient and public involvement

No patients were directly involved in this study. Patients would not be able to identify themselves in the data.

There was no direct data collection for this study; all data were gathered secondhand from public or published sources. The data used for measuring seroprevalence, COVID-19 deaths, and population were all anonymised and aggregated before we accessed it. We retrieved seroprevalence rate data in all locations from public sources, aggregated by age and sex.[i] [10 29 33 38] Seroprevalence studies were designed and implemented in partnership with local city and state governments. Details of patient involvement, protocols and institutional ethics approval for each seroprevalence study have been published in separate papers, and in reports from the respective governments.[10 29 32 33]

RESULTS

We plotted age-specific IFR for each location on a log scale, to enable comparison at all ages despite exponential increases at higher ages found in all countries (figure 1A,B). For both males and females, there is substantial variation in IFR across the four locations in India. In Karnataka, age-specific IFRs are 10 times lower than those reported in the meta-analyses, and 25 times lower over age 70. In Tamil Nadu, estimates were 2–4 times lower than those in the meta-analyses. In Mumbai, estimates were close to the lower of the two meta-analyses at younger ages,[7] but were considerably lower than meta-analyses after age 60. For 60–69 year-old men, for example, we measured an IFR of 0.17% (95% CI 0.092% to 0.240%) in Karnataka, 0.45% (95% CI 0.397% to 0.0497%) in Tamil Nadu and 0.62% (95% CI 0.591% to 0.647%) in Mumbai (table 2); the two meta-analyses reported male IFR of 1.02%[7] and 1.86%[1] in this age group.

In contrast, mortality among male migrants returning to Bihar was an order of magnitude higher. Mortality among males aged 60–69 was extremely high but measured imprecisely due to the small sample of older males (4.26%, 95% CI 0.0% to 10.0%). The larger age bins allowed a more precise measure of IFR in Bihar (table 3). In both the 10–49 and 50–89 age bins, mortality in Bihar was an order of magnitude higher than in the other Indian locations and at least twice as high as rates in meta-analyses, after weighting to the Indian age distribution to ensure cross-context comparability. For the 50–89 age group, estimates were not precise enough to rule out equality between Bihar and the other locations. For the 10–49 age group, we can rule out equality (p<0.01).

To the extent that an IFR advantage exists in India, it appears more strongly among the elderly. In most cases,

---

[i]Details on public sources for seroprevalence data. Bihar migrant data may be requested from the Government of Bihar. Positive test rates by age, gender, ward, and slum in Mumbai can be found in the online supplement of.[10] The same rates by district in Tamil Nadu can be found in the online supplement of.[33] The same rates by region in Karnataka can be found in the supplement of.[38]
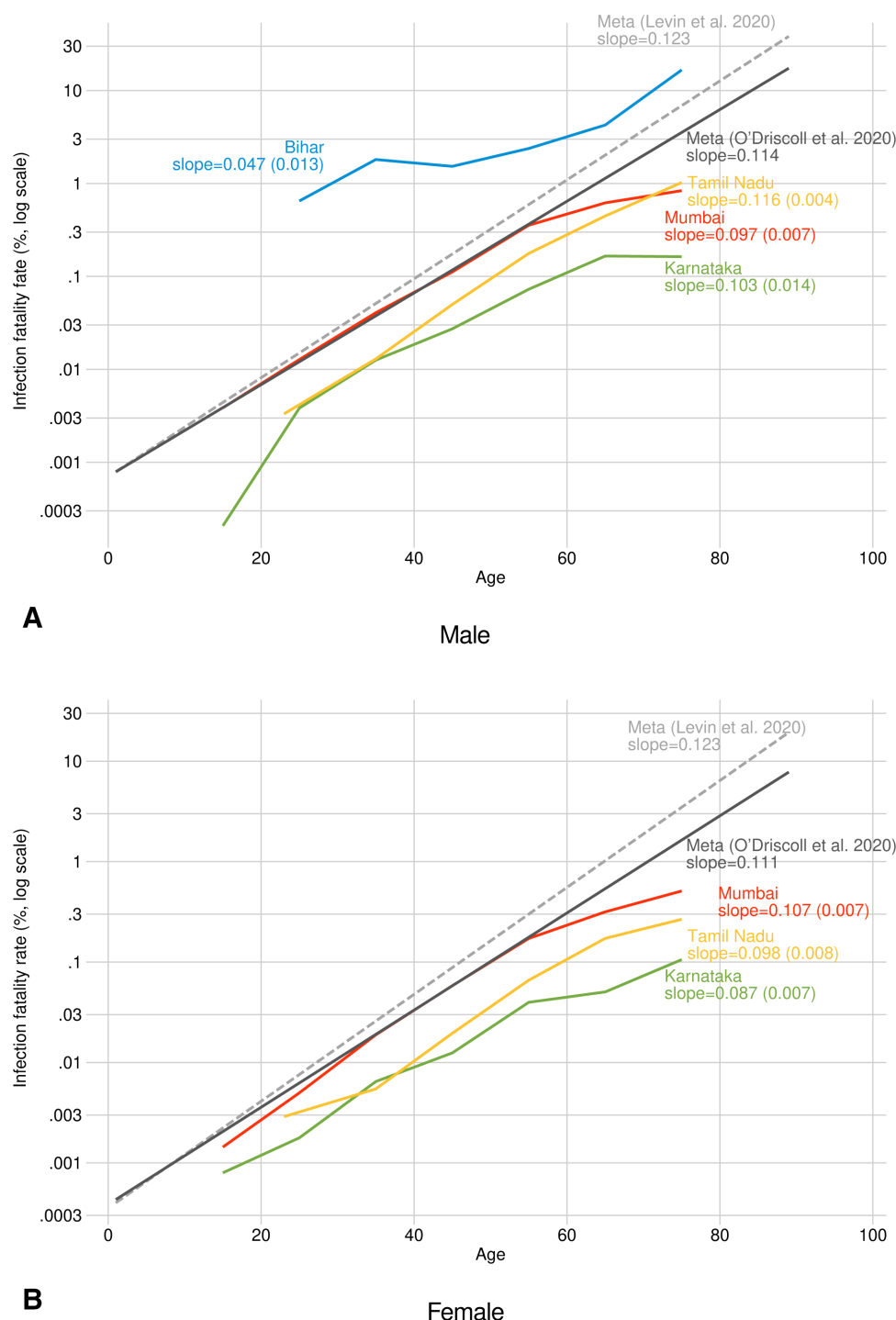
Figure 1 Age-specific infection fatality rate (IFR), comparing four locations in India with international estimates. Point estimates of age-specific IFR in (A) males and (B) females combining representative prevalence/seroprevalence studies and government-reported COVID-19 deaths. IFRs were estimated for age bins 10–19 (Mumbai and Karnataka only), 20–29,…,60–69 and 70+ in India. Slope of IFR age trends from the meta-analyses calculated by fitting a linear regression between age and natural log of IFR.

the overall increase in IFR with age was considerably less steep than in the reference meta-analyses (figure 1), particularly at older ages. The meta-analyses suggest that an 80-year-old has about 100× the IFR of a 40-year-old; in Mumbai, the increase in risk factor is 40× and in Bihar it is only 10×. Specifically, male IFR increased on average by 4.7%, 9.6%, 10.3% and 11.6% with each year of age in

Bihar, Mumbai, Karnataka and Tamil Nadu, respectively. We calculated comparable figures in the meta-analyses as 11.4%[7] and 12.3%.[1] Slopes for Indian females were uniformly flatter than those for the reference groups (figure 1B).

The main estimates are replicated in online supplemental materials under a range of different scenarios

Table 2 Age-specific infection fatality rates (%) from four locations in India

Age	Mumbai Male	Mumbai Female	Karnataka Male	Karnataka Female	Tamil Nadu Male	Tamil Nadu Female	Bihar migrants Male
10–19	0.004 (0.004 to 0.004)	0.001 (0.001 to 0.002)	0.000 (0.000 to 0.000)	0.001 (0.000 to 0.001)	NA	NA	0.000 (0.000 to 0.000)
20–29	0.013 (0.012 to 0.013)	0.005 (0.005 to 0.005)	0.004 (0.002 to 0.005)	0.002 (0.001 to 0.002)	0.003 (0.003 to 0.004)	0.003 (0.003 to 0.003)	0.649 (0.131 to 1.166)
30–39	0.041 (0.039 to 0.043)	0.019 (0.018 to 0.020)	0.013 (0.009 to 0.016)	0.006 (0.005 to 0.008)	0.013 (0.012 to 0.014)	0.005 (0.005 to 0.006)	1.810 (0.795 to 2.825)
40–49	0.112 (0.106 to 0.117)	0.058 (0.055 to 0.061)	0.027 (0.022 to 0.032)	0.012 (0.010 to 0.015)	0.050 (0.046 to 0.054)	0.020 (0.018 to 0.021)	1.529 (0.199 to 2.859)
50–59	0.355 (0.339 to 0.372)	0.172 (0.163 to 0.180)	0.073 (0.058 to 0.088)	0.040 (0.032 to 0.048)	0.177 (0.162 to 0.192)	0.066 (0.061 to 0.071)	2.381 (0.000 to 5.043)
60–69	0.619 (0.589 to 0.649)	0.317 (0.301 to 0.332)	0.166 (0.092 to 0.240)	0.050 (0.030 to 0.070)	0.447 (0.397 to 0.497)	0.172 (0.157 to 0.188)	4.255 (0.000 to 10.026)
70–89	0.837 (0.797 to 0.878)	0.511 (0.486 to 0.536)	0.163 (0.074 to 0.252)	0.106 (0.051 to 0.160)	1.024 (0.894 to 1.155)	0.267 (0.233 to 0.301)	12.500 (0.000 to 35.418)

Infection fatality rates as percentages. 95% CIs in parentheses.
NA, not available.

Table 3 Age-specific IFRs in India ages 10–49 and 50–89

Age	Mumbai Male	Mumbai Female	Karnataka Male	Karnataka Female	Tamil Nadu Male	Tamil Nadu Female	Bihar migrants Male
10–49	0.033 (0.032 to 0.034)	0.016 (0.016 to 0.017)	0.009 0.007 to 0.010	0.004 (0.004 to 0.005)	0.013* (0.013 to 0.014)	0.006* (0.006 to 0.007)	0.851 (0.467 to 1.235)
50–89	0.530 (0.517 to 0.543)	0.285 (0.278 to 0.292)	0.120 (0.090 to 0.150)	0.056 (0.043 to 0.069)	0.420 (0.390 to 0.459)	0.140 (0.132 to 0.149)	5.393 (0.000 to 11.156)

IFRs as percentages. 95% CIs in parentheses.
*In Tamil Nadu, seroprevalence collection and deaths were restricted to adults aged 18+. The 10–49 group assumes IFR in the 10–17 and 18–29 groups were equal, for weighting purposes.
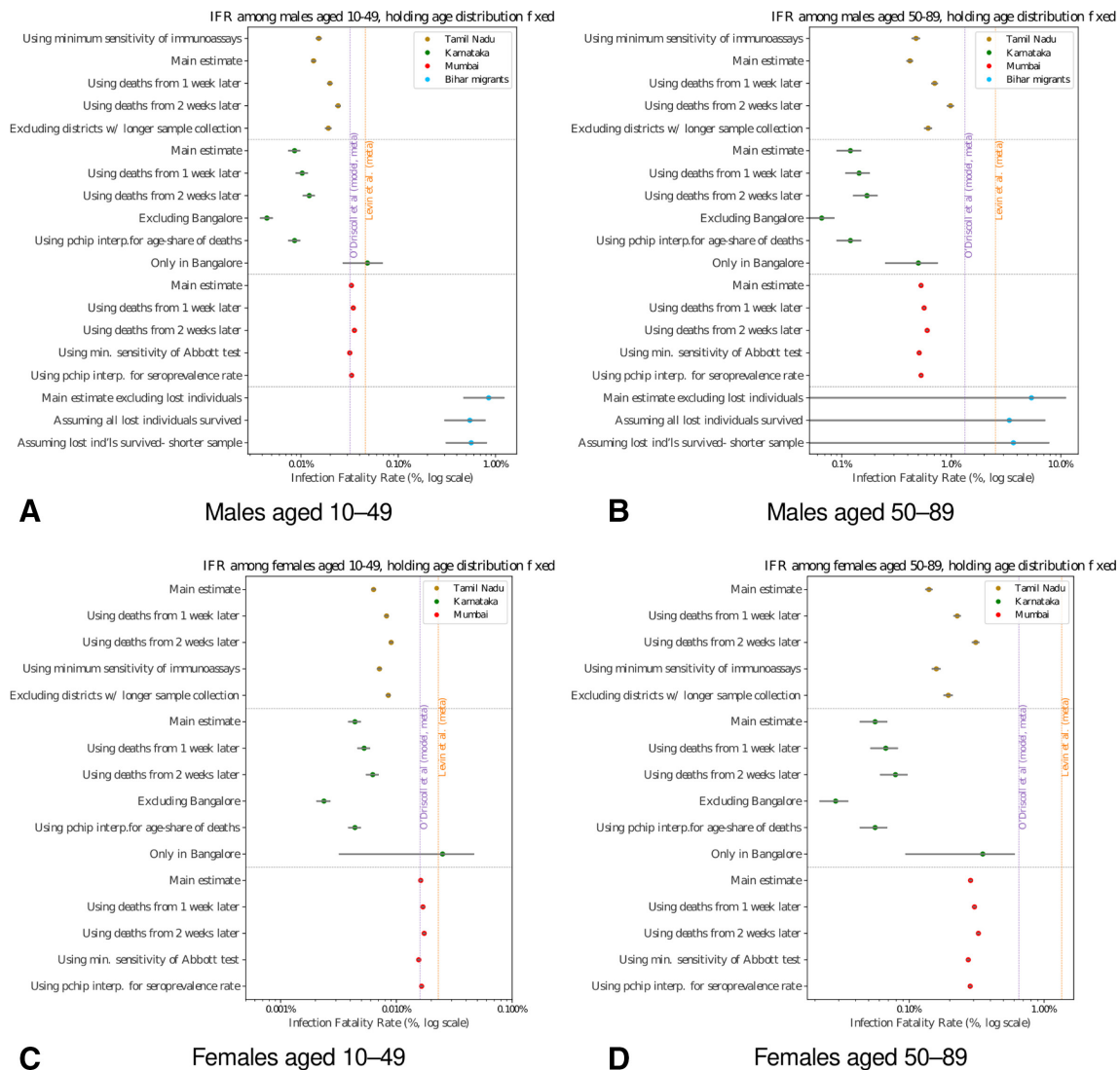IFR, infection fatality rate.

Figure 2 Age-specific infection fatality rates (IFRs) India: sensitivity checks. Main estimates and sensitivity checks of IFR of (A) males aged 10–49 years, (B) males aged 50–89, (C) females aged 10–49 and (D) females aged 50%–89%. 95% CIs shown in grey. In all locations, including meta-analyses, age-specific IFRs in smaller age bins have been weighted to India's national age distribution, controlling for cross-location differences in population age. See online supplemental file 1 for details of sensitivity checks.

and assumptions; the ordering of IFRs across regions and with respect to the reference groups is highly robust (figure 2A–D).

DISCUSSION
Principal findings
Using best-practice methods applied in many HICs, we found substantial heterogeneity in age-specific COVID-19 IFR in India. In all four locations, we found a weaker increase in IFR over age than seen in other countries.

In Mumbai, Karnataka and Tamil Nadu, estimated IFRs were considerably lower than those measured in richer countries. These results are qualified by the fact that COVID-19 deaths are known to be under-reported in these locations, as we discuss below. In a tracked sample of male migrants returning to Bihar,

IFR estimates were an order of magnitude higher than the other two locations and twice as high as the international reference groups.

Our Mumbai IFR estimates are representative for the city while Tamil Nadu and Karnataka estimates are representative for the state. IFR estimates for migrants returning to Bihar are plausibly generalisable to the tens of millions of migrant workers who exited cities, returning primarily to poorer rural areas, in the first months of the pandemic. Migrant workers differ from the general population, typically living in dense quarters that increase disease transmission,[25] with higher poverty rates,[28] lower baseline health and higher prevalence of malaria, respiratory infections and acute febrile illness.[25] In these aspects, our findings on migrants have some generalisability to other extremely disadvantaged populations. However, the

actual journey migrants undertook is a unique risk factor. Overpacked trains likely heightened transmission and long travel distances, often on foot, increased physical vulnerability.[27]

Strengths and weaknesses of the study

The strength of this study was the use of seroprevalence data representing over 150 million people, with a sufficiently large sample to calculate age-disaggregated IFR in a lower-middle-income country. The main weakness of the study is that, like all COVID-19 population estimates, our results depend on the quality of underlying mortality data. The largest potential source of bias was our use of official reports of COVID-19 deaths, which undercount the true number of deaths in all contexts.[23 39]

Though estimates of under-reporting are highly uncertain, accounting for misreporting of deaths brings IFRs in three of the study locations close to estimates from HICs. Focusing on the 50–89 age group, in Mumbai, a doubling of COVID-19 deaths is required to put estimated IFR in the range of the meta-analyses. It is plausible that deaths in Mumbai were undercounted by a factor of 2; between March and July, Mumbai recorded 6600 excess deaths in addition to the 6400 COVID-19 deaths used in this study.[39]

In Karnataka and Tamil Nadu, COVID-19 deaths would have to be under-reported by factors of 10 and 3 respectively to bring IFR in line with international estimates. Crude estimates from recently published data from India's Civil Registration System suggest excess mortality rates during the first COVID-19 wave were approximately six times higher than official COVID-19 deaths in both Karnataka and Tamil Nadu.[40] If this ratio between excess mortality and reported COVID-19 deaths is an accurate measure of the death under-reporting rate, then this puts IFRs in Mumbai and Tamil Nadu close to the range of the HIC results, and Karnataka only slightly lower.

While these IFR estimates remain subject to bias, we note that we calculated IFR with the standard methodology used in many cross-national settings, many of which are also characterised by under-reporting of COVID-19 deaths. As described in the online supplemental file 1, wherever possible we made conservative choices that would bias our IFR estimates upward rather than downward. In particular, antibodies may fade over time, so seroprevalence tests provide a lower bound on the cumulative infection rate.[41]

Official misreporting of COVID-19 deaths would not bias our IFR estimates in Bihar, due to the mortality follow-up methodology underlying these estimates. For our Bihar estimates to match the range of meta-analyses, deaths would need to have been overcounted by a factor of 2 for ages 50–89, and by 10 for ages 10–49. However, we do not know the base rate of migrant death. If migrant deaths would be high in absence of COVID-19, due to migrants' arduous return journeys, we may overstate the mortality attributable to COVID-19 in this group.

Comparison with other studies

Few other studies have used sufficiently large seroprevalence samples to estimate age-specific IFR for a large lower-income population. Seroprevalence-based IFR estimates for older individuals in a Brazilian city[14] were slightly lower than our estimate for Bihari migrants, and much higher than our seroprevalence-based estimates. However, seroprevalence samples of non-representative groups in sub-Saharan Africa implied high infection rates, suggesting either low overall mortality or substantial under-reporting of deaths, consistent with our findings in India.[11 17 42]

Studies have noted that the pattern of mortality in LMICs skews younger than would be predicted from the age distributions of death in HICs.[19 21] Our study suggests that a flatter age profile in IFRs in lower-income settings could be a major factor driving this difference.

CONCLUSION AND FURTHER RESEARCH

In large samples representing India's higher-income South, we found IFRs that broadly corresponded to those reported in richer countries, after adjusting for under-counting. Among a sample of economically distressed migrants, we found IFRs that were twice as high, plausibly due to severe economic and physical distress. Migrant workers have worse health than the general population at baseline[25 43]; the circumstances at the beginning of the pandemic may have made this group exceptionally vulnerable to adverse health events following viral infection.

At the time of writing, these estimates are among the best available in a lower-income setting. Improved surveillance and accounting of SARS-CoV-2 are critical investments that would improve our understanding of the fatality risk of the virus in lower-income settings. Further research is necessary to determine if IFRs are similar in high-income and low-income settings.

ORCID iDs
Paul Novosad http://orcid.org/0000-0001-8331-6630
Anup Malani http://orcid.org/0000-0002-2594-5778

REFERENCES

1 Levin AT, Hanage WP, Owusu-Boaitey N, *et al*. 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.

2 Pastor-Barriuso R, Pérez-Gómez B, Hernán MA, *et al*. Infection fatality risk for SARS-CoV-2 in community dwelling population of Spain: nationwide seroepidemiological study. *BMJ* 2020;371:m4509.

3 Yang W, Kandula S, Huynh M, *et al*. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *Lancet Infect Dis* 2021;21:203–12.

4 Perez-Saez J, Lauer SA, Kaiser L, *et al*. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. *Lancet Infect Dis* 2021;21:e69-e70.

5 Salje H, Tran Kiem C, Lefrancq N, *et al*. Estimating the burden of SARS-CoV-2 in France. *Science* 2020;369:208–11.

6 Ward H, Atchison CJ, Whitaker M, *et al*. Antibody prevalence for SARS-CoV-2 in England following first peak of the pandemic: REACT2 study in 100,000 adults. *MedRxiv* 2020.

7 O'Driscoll M, Ribeiro Dos Santos G, Wang L, *et al*. Age-Specific mortality and immunity patterns of SARS-CoV-2. *Nature* 2021;590:1–6.

8 Ghisolfi S, Almås I, Sandefur JC, *et al*. Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ Glob Health* 2020;5:e003094.

9 Hallal PC, Hartwig FP, Horta BL, *et al*. SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Glob Health* 2020;8:e1390–8.

10 Malani A, Shah D, Kang G, *et al*. Seroprevalence of SARS-CoV-2 in slums versus non-slums in Mumbai, India. *Lancet Glob Health* 2021;9:e110-e111.

11 Uyoga S, Adetifa IMO, Karanja HK, *et al*. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* 2021;371:79–82.

12 Ioannidis JPA. Infection fatality rate of COVID-19 inferred from seroprevalence data. *Bull World Health Organ* 2021;99:BLT.20.265892.

13 Xu X, Sun J, Nie S, *et al*. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. *Nat Med* 2020;26:1193–5.

14 Picon RV, Carreno I, da Silva AA, *et al*. Coronavirus disease 2019 population-based prevalence, risk factors, hospitalization, and fatality rates in southern Brazil. *Int J Infect Dis* 2020;100:402–10.

15 Verity R, Okell LC, Dorigatti I, *et al*. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 2020;20:669–77.

16 Walker PGT, Whittaker C, Watson OJ, *et al*. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science* 2020;369:413–22.

17 Nordling L. The pandemic appears to have spared Africa so far. scientists are struggling to explain why. *Sci AAAS* 2020.

18 Njenga MK, Dawa J, Nanyingi M, *et al*. Why is there low morbidity and mortality of COVID-19 in Africa? *Am J Trop Med Hyg* 2020;103:564–9.

19 Demombynes G. *COVID-19 age-mortality curves are Flatter in developing countries*. Washington, DC: The World Bank, 2020.

20 Philip M, Ray D, Subramanian S. Decoding India's low Covid-19 Case Fatality Rate. *J Human Dev Capabil* 2021;22:27–51.

21 Novosad P, Jain R, Campion A, *et al*. COVID-19 mortality effects of underlying health conditions in India: a modelling study. *BMJ Open* 2020;10:e043165.

22 Samarasekera U. India grapples with second wave of COVID-19. *Lancet Microbe* 2021;2:e238.

23 Chatterjee P. Is India missing COVID-19 deaths? *The Lancet* 2020;396:657.

24 Kapoor G, Hauck S, Sriram A, *et al*. State-wise estimates of current hospital beds, intensive care unit (ICU) beds and ventilators in India: are we prepared for a surge in COVID-19 hospitalizations? *MedRxiv* 2020.

25 Adsul BB, Laad PS, Howal PV, *et al*. Health problems among migrant construction workers: a unique public-private partnership project. *Indian J Occup Environ Med* 2011;15:29–32.

26 Dutta A. Parliament monsoon session: 97 people died on-board Shramik trains, govt tells Rajya Sabha. *Hindustan Times* 2020.

27 Gettleman J, Raj S, Yasir S, *et al*. The virus trains: how Lockdown chaos spread Covid-19 across India. *N Y Times* 2020.

28 Deshingkar P, Khandelwal R, Farrington J. Support for migrant workers: The missing link in India's development. Overseas Development Institute 2008.

29 Malani A, Mohanan M, Kumar C, *et al*. Prevalence of SARS-CoV-2 among workers returning to Bihar gives snapshot of COVID across India. *MedRxiv* 2020.

30 Abi-Habib M, Yasir S. India's Coronavirus Lockdown Leaves Vast Numbers Stranded and Hungry. *N Y Times* 2020.

31 Chetterje P. Gaps in India's preparedness for COVID-19 control. *Lancet Infect Dis* 2020;20:544.

32 Mohanan M, Malani A, Krishnan K, *et al*. Prevalence of SARS-CoV-2 in Karnataka, India. *JAMA* 2021;325:1001.

33 Malani A, Tandel V, Parasa R. Seroprevalence in Tamil Nadu in October-November 2020. *MedRxiv* 2021.

34 Chief Minister's Office Public Relations Cell. Bihar Government. Pre-release cm-243: directions for random testing 2020 https://prsindia.org/files/covid19/notifications/4896.BR_CM_Directions_Random_Testing_May_7.pdf

35 Linton NM, Kobayashi T, Yang Y, *et al*. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med* 2020;9. doi:10.3390/jcm9020538. [Epub ahead of print: 17 02 2020].

36 Stringhini S, Wisniak A, Piumatti G, *et al*. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* 2020;396:313–9.

37 Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71–6.

38 Mohanan M, Malani A, Krishnan K, *et al*. Prevalence of COVID-19 in rural versus urban areas in a low-income country: findings from a State-Wide study in Karnataka, India. *MedRxiv* 2020.

39 Banaji M. Mumbai had 13k "Excess" deaths in March-July. How does its COVID story change? Wire Sci, 2020. Available: https://science.thewire.in/the-sciences/covid-19-mumbai-all-cause-mortality-data-ifr-bmc-seroprevalence-survey/

40 Asher S, Lunt T, Matsuura P, *et al*. Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform. *World Bank Econ Rev* 2021;lhab003.

41 Ripperger TJ, Uhrlaub JL, Watanabe M, *et al*. Orthogonal SARS-CoV-2 serological assays enable surveillance of low-prevalence communities and reveal durable humoral immunity. *Immunity* 2020;53:925–33.

42 Mukherjee S. Why does the pandemic seem to be hitting some countries harder than others? *New Yorker n.d*.

43 Choudhari R. COVID 19 pandemic: mental health challenges of internal migrant workers of India. *Asian J Psychiatr* 2020;54:102254.

44 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–4.