medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Estimates of pandemic excess mortality in India based on civil registration data

Murad Banaji and Aashish Gupta[1]

Abstract

Background: The COVID-19 pandemic has had large impacts on population health. These impacts are less well understood in low-and middle-income countries, where mortality surveillance before the pandemic was patchy. Although limited all-cause mortality data are available in India, interpreting this data remains a challenge.

Objective: We use existing data on all-cause mortality from civil registration systems of twelve Indian states comprising around 60% of the national population to understand the scale and timing of excess deaths in India during the COVID-19 pandemic.

Methods: We characterize the available data, discuss the various reasons why these data are incomplete, and estimate the extent of coverage in the data. Comparing the pandemic period to 2019, we estimate excess mortality in twelve Indian states, and extrapolate our estimates to the rest of India. We explore sensitivity of the estimates to various assumptions, and present optimistic and pessimistic scenarios along with our central estimates.

Results: For the 12 states with available all-cause mortality data, we document an increase of 28% in deaths during April 2020-May 2021 relative to expectations from 2019. This level of increase in mortality, if it applies nationally, would imply 2.8-2.9 million excess deaths. More limited data from June 2021 increases national estimates of excess deaths during April 2020-June 2021 to 3.8 million. With more optimistic or pessimistic assumptions, excess deaths during this period could credibly lie between 2.8 million and 5.2 million. We find that the scale of estimated excess deaths is broadly consistent with expectations based on seroprevalence data and international data on COVID-19 fatality rates. Moreover, there is a strong association between the timing of excess deaths, and of recorded COVID-19 deaths.

Contribution: We show that the surveillance of pandemic mortality in India has been extremely poor, with around 8-10 times as many excess deaths as officially recorded COVID-19 deaths. Our findings highlight the utility of all-cause mortality data, as well as the significant challenges in interpreting such data from LMICs. These data reveal that India is among the countries most severely impacted by the pandemic. It is likely that in absolute terms India has seen the highest number of pandemic excess deaths of any country in the world.

1 Murad Banaji (m.banaji@mdx.ac.uk) is a mathematician at Middlesex University London and Aashish Gupta is a demographer at Harvard University (aashishgupta@hsph.harvard.edu). Additional methodological details and state factsheets are available at https://www.indiacovidmapping.org/. Data and Code are available at GitHub: https://github.com/muradbanaji/IndiaACMdata/tree/master/NationalEstimates. We thank Chris Leffler, Rukmini S and Srinivasan Ramani for helpful clarifications.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

1. Introduction

The COVID-19 pandemic has had large impacts on population health across the world. These impacts are less well understood in low- and middle-income countries, where routine mortality surveillance before the pandemic was patchy (World Health Organization 2014). Daily case counts and confirmed COVID-19 deaths have been widely used in policy and public discussions, including in India. These disease surveillance systems are innovative in the context of India. However, it is widely recognized that reported cases generally capture a small fraction of total infections, and that deaths in these systems are undercounts (Acosta et al. 2021; Banaji 2021b; Murhekar et al. 2021). All-cause mortality data, where available, has been used to understand the overall mortality impact of the pandemic (Aburto et al. 2021). In India, several recent media reports have reported large increases in registered deaths (Ramani 2021; Rukmini S 2021; The Hindu Data Team 2021). Interpreting these data and estimating the scale of excess deaths in India is challenging, however.

There are multiple reasons for this. First, the data are available only for some of India's 36 states and union territories. Second, vital registration varies substantially within the states from which all-cause mortality data are available (Government of India 2021b). States differ in terms of level of mortality registration before the pandemic, recording systems (for instance for online and offline registration), and the way mortality data are organized (e.g. by date of death or date of registration). These details have been documented at India COVID Mapping (2021). Available data may miss not only the deaths that were not registered, but also sometimes deaths that were registered in an offline system (Ramani 2021). Third, trends in registration and mortality need to be understood. Often, registration was slowly increasing prior to the COVID-19 pandemic (Rao and Gupta 2020), but was disrupted during the early phase of the pandemic. It is likely that some of this disruption was a consequence of national lockdown. It is not clear to what extent registration recovered as lockdown eased.

In this article, we use reported all-cause mortality data from India's civil registration system (CRS) to understand the scale and evolution of excess mortality in India during the COVID-19 pandemic. To do so, we compile contextual information on registration, baseline mortality, the extent to which available death records are complete, and trends/disruptions in death registration. We transparently lay out the many assumptions needed to estimate excess deaths on the basis of these data. Our results complement those by Anand, Sandefur, and Subramanian (2021); Deshmukh et al. (2021); and Leffler and Yang (2021).

We use data from twelve states comprising around 60% of the national population to evaluate the scale of pandemic excess mortality in these states up to May 2021. We then use this data to make estimates of excess deaths nationally, and use more limited data for June 2021 to extend these estimates to June. We assess the degree to which our estimates are sensitive to the various uncertainties, for example fluctuations in levels of registration, or possible differences in the mortality impact of the pandemic in the states for which we have data and the rest of the country.

Our findings indicate that the mortality impact of the pandemic in India has been severe. We estimate that the twelve states whose data we use saw around around 28% more deaths than expected from historical data between April 2020 and May 2021. The limited data from June 2021 shows considerable further rises, perhaps partly as a result of delays in registration.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Nationally, we estimate around 34% more deaths over a 15 month period from April 2020-June 2021 than expected from historical data. Uncertainties and incompleteness of the data, most importantly the possibility of changes in levels of death registration during the pandemic, mean that this surge could plausibly lie between 24% and 45%. These estimates place India amongst the harder hit countries in the world during the pandemic (Giattino et al. 2020; Karlinsky and Kobak 2021). In absolute terms, our central estimate amounts to around 3.8M excess deaths during April 2020-June 2021 with optimistic and pessimistic estimates of 2.8M and 5.2M excess deaths respectively. Although data from many countries is limited, these estimates make it likely that India is the country with the highest number of pandemic excess deaths in the world (see also The Economist 2021).

In order to contextualise the numbers, we examine how our estimates of excess deaths compare with official COVID-19 deaths, and how they align with expectations of COVID-19 deaths based on international data. Our estimates of excess deaths are broadly consistent with expectations based on COVID-19 fatality rates given India's age structure and the levels of spread estimated in seroprevalence surveys. On the other hand, by June 2021, we estimate a ratio of excess deaths to official COVID-19 deaths of over 9, indicating that official death counts have underestimated the scale of pandemic mortality by an order of magnitude.

2. The context of mortality and civil registration in India

In order to use registered deaths to estimate excess deaths in any given region, we need estimates of registration completion, namely the fraction of deaths which are registered, both before and during the pandemic. But there are uncertainties around registration completion prior to the pandemic, and trends in completion.

Pre-pandemic registration completion. Government estimates of completion rely on comparing registered deaths with expected deaths. The latter are derived using population estimates and survey-based estimates of the crude death rate (CDR). According to the 2018 Sample Registration System annual statistical report (ORGI 2020), henceforth, "the 2018 SRS report", India's CDR stood at 6.2 per 1K in 2018. Based on this estimate, registration completion in 2019 stood at 92%, as reported in the 2019 report on Vital Statistics of India based on the Civil Registration System (Government of India 2021), henceforth, the "2019 CRS report". There are, however, several reasons to believe that the estimated CDR of 6.2 is too low, and that completion in 2019 was less than 92%.

A variety of data sources and approaches detailed in Appendix 1, including from the UN population division and Wave 1 of the National Family Health Survey-5, give estimates of the national CDR in 2019 ranging from 6.2 to 7.5 per 1K, corresponding to registration completion in 2019 from 76% to 92%. The median of these estimates is around 82%. Most relevant from our point of view here, are sub-national estimates of registration completion in the 2019 CRS report based on sub-national estimates of CDR. We can derive from these a national CDR in 2019 of 6.6 per 1K, implying completion of around 86%. It is these sub-national estimates that we use here; but we bear in mind that the estimates of completion may be somewhat too high in some states.

Trends in total deaths. While there is uncertainty around the national CDR prior to the pandemic, there is also uncertainty about how it was changing from year to year. According to the SRS bulletins (ORGI 2020a), the national CDR saw a steady decline of around 1.5% per year from 6.5 in 2015 to 6.2 in 2018. There is no available estimate for 2019. According to UN estimates (World

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Bank 2021), however, India's estimated CDR was falling prior to 2015 and but saw a 1% increase during 2015-2019. Meanwhile, population projections suggest that the national population has been growing by around 1% per year (National Commission on Population 2019). This is also the estimated population growth rate in states whose data we use below. Thus the SRS estimates imply that up to 2018 year-on-year deaths were steady or falling slightly; while UN estimates suggest they could have been rising by a little over 1% per year.

Registration completion during 2018 and 2019. While the estimated changes in yearly deaths are fairly small, there was a larger increase in estimated registration completion between 2018 and 2019. According to the 2019 CRS report (Government of India 2021b), completion nationally increased from 84.6% to 92% between 2018 and 2019. Using the sub-national data, we find an increase in estimated registration completion from 81.1% to 86.5% nationally. The twelve states whose data we use saw estimated completion rise from 86.9% to 92.1% between 2018 and 2019. It is noteworthy that incomplete but increasing death registration in Bihar was a key factor in this increase: if we remove Bihar from the picture, the remaining eleven states together saw estimated registration completion rise from 95.7% to 98.4% between 2018 and 2019.

Since no SRS bulletin is available for 2019, all of these estimates assume no change in national or sub-national values of the CDR between 2018 and 2019. Note that according to these estimates registration was approaching 100% in many states.

Predicting expected deaths. What we learn from scrutinising civil registration data is that using historical data to predict expected deaths, key to estimating excess deaths, is challenging in the Indian context. Pre-pandemic trends in registered deaths mainly reflect trends in registration completion rather than in mortality. On the other hand, with registration completion at or close to completeness in many states, and with considerable disruption to registration during the pandemic, it could be misleading to extrapolate past trends of increasing registration completion into the pandemic period. For this reason, we choose a very simple approach below. For our central estimates, expected deaths are set at estimated 2019 levels; but we explore how changing levels of registration would affect the resulting estimates of excess mortality. We see this approach as sufficient for the purposes of making aggregate estimates; but making credible state-level estimates would require a more careful case-by-case analysis. Preliminary state-factsheets which attempt such state-level analysis are available at India COVID Mapping (2021).

3. The available data from twelve states

Henceforth, unless stated otherwise, all estimates of registration completion during 2019 are based on the sub-national data from the 2019 CRS report, as discussed above.

We use civil registration data from the following twelve states to arrive at estimates of excess mortality in India: Andhra Pradesh, Bihar, Haryana, Himachal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Punjab, Rajasthan, Tamil Nadu and West Bengal. We refer to these states collectively as STAR12 - a rough mnemonic for "12 States with Available Registration Statistics" In these states, partial or complete death registration data are available for at least January 2018 to May 2021. Data are additionally available for June 2021 from Andhra Pradesh, Karnataka and Punjab. The full data and code are available on GitHub (Banaji 2021a)

Overall, STAR12 accounted for 59% of the estimated 2019 national population and also 59% of estimated total deaths occurring during 2019. The estimated CDR in STAR12 was thus close to the national value. On the other hand, registration completion in STAR12 was somewhat higher than

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

the national average: around 92% of deaths occurring during 2019 in STAR12 were registered, as against 78% in the remainder of the country.

Often, publicly available data do not include all registered deaths because they are from online systems which may miss deaths registered using offline systems. The available data includes 85% of all the deaths registered in STAR12 during 2019. This amounts to 79% of all deaths estimated to have occurred in these states during 2019. We reference this fact by saying that "coverage in the data" was 79% in 2019, i.e., the fraction of estimated total deaths in 2019 in STAR12 which appear in the data was 79%. In individual states, coverage in the data during 2019 varies from 48% to 100% (see Table 1).

Table 1: Estimated deaths, registration, and coverage in STAR12, 2019.

	Deaths, registration, and coverage in 2019				
	Registered deaths	**Estimated registration completion (%)**	**Estimated deaths**	**Deaths recorded in the available data**	**Estimated coverage in the available data (%)**
Andhra Pradesh	401,472	100.0	401,472	363,649	90.6
Bihar	359,349	51.6	696,413	351,248	50.4
Haryana	188,910	100.0	188,910	183,795	97.3
Himachal Pradesh	43,633	86.4	50,501	40,970	81.1
Karnataka	508,584	100.0	508,584	508,584	100
Kerala	270,567	100.0	270,567	264,071	97.6
Madhya Pradesh	493,328	89.1	553,679	449,819	81.2
Maharashtra	693,800	100.0	693,800	462,028	66.6
Punjab	215,045	100.0	215,045	213,122	99.1
Rajasthan	451,315	98.6	457,723	219,814	48.0
Tamil Nadu	633,897	100.0	633,897	588,221	92.8
West Bengal	551,695	100.0	551,695	457,561	82.9
Total	**4,811,595**	**92.1**	**5,222,286**	**4,102,882**	**78.6**

Note: Registered deaths and estimates of completion from the 2019 CRS report are used to estimate total deaths during 2019 in STAR12 states. These can be compared to the number of deaths in the available data to get estimates of coverage in the data during 2019.

Coverage in the data in STAR12 rose from 73% in 2018 to 79% in 2019. Most of this increase reflects an increase in registration completion rather than, say, increasing use of online systems: in fact, the fraction of *registered* deaths appearing in the available data rose only slightly, from 83.6% in 2018 to 85.3% in 2019. If we were to omit Bihar, the data appears even more stable: the remaining eleven states saw coverage in the data rise only modestly from 80% in 2018 to 83% in 2019.

We proceed roughly as follows. Based on the number of registered deaths in 2019 and then during the pandemic period we compute excess registrations in each state of STAR12. We then use estimates of coverage in the data during 2019 to estimate total excess deaths in each state of

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

STAR12. We ask what the excess mortality estimated in STAR12, if applied nationally, would imply about total excess deaths in India during the pandemic period. We extend these estimates to June using the limited data available for June.

We bear in mind that there may have been shifts in expected mortality, and in coverage in the data during the pandemic, and carry out some analysis of how these might affect the estimates later.

4. Results: central estimates of excess deaths

4.1 Estimates of excess mortality in STAR12 during April 2020-May 2021

We begin by considering the 14 month period April 2020-May 2021 for which we have data from all states in STAR12. For each state, we set baseline expectations for registrations in a given month during the pandemic period to be registrations during this month in 2019. By summing the difference between registered deaths during each month of April 2020-May 2021 and the corresponding month of 2019, we obtain excess registrations over this period. In this way we find 1.3M excess registrations in STAR12 during April 2020-May 2021, amounting to 27% more registrations than expected from 2019 data.

If we scale the excess registrations in each state based on estimated coverage in the data during 2019, we find 1.7M excess deaths during April 2020-May 2021 in STAR12, an increase of 28% in deaths in these states relative to expectations from 2019. We refer to this percentage surge as a p-score (Aron and Muellbauer 2020; Giattino et al. 2020). This amounts to 2.1 excess deaths per thousand population in STAR12 during April 2020-May 2021. These estimates assume there were no variations in expected mortality, or in coverage in the data between 2019 and the pandemic period.

4.2 National estimates of excess mortality in STAR12 during April 2020-May 2021

Based on estimates for STAR12, we can compute national estimates of excess deaths via one of two approaches:

1. **mortality-based extrapolation**: we assume that per capita pandemic excess deaths nationwide were similar to in STAR12. For the denominator in these calculations we use estimated 2020 populations.
2. **P-score-based extrapolation**: we assume that the p-score nationwide during the pandemic period was similar to that in STAR12. As we do not have national monthly data from the pre-pandemic period, we take monthly deaths nationally during 2019 to be one twelfth of the estimated yearly total.

Both extrapolations are open to question, and we discuss later the effects on our estimates if the mortality impact in regions outside STAR12 was higher or lower than in STAR12. But the fact that the estimated crude death rate in STAR12 matches the national estimate means that these two approaches give similar results: 27-28% more deaths nationally during April 2020-May 2021 than expected from 2019 data, which amounts to 2.8-2.9M excess deaths.

Other approaches to extrapolation are possible: for example, we might attempt to use recorded COVID-19 deaths as a basis for extrapolation, i.e., to assume the ratio of excess deaths to COVID-19 deaths seen in STAR12 holds nationally. We reject this approach for reasons discussed in Section 6.1 below.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

4.3 Extending the national estimates to June 2021

Data for June 2021 is currently only available for three states of STAR12: Andhra Pradesh, Karnataka, and Punjab, which hold around 11% of the national population. During June 2021, registrations in these states were 123% above registrations during June 2019. Moreover, the ratio of excess deaths to recorded COVID-19 deaths in these three states jumped from 8.0 in May 2021 to 11.7 in June 2021. This suggests that delays in registration following the enormous mortality surge in May could be at least partly responsible for the high June excess deaths.

We can use either mortality-based or P-score-based extrapolations from these states to obtain national estimates of excess deaths for June 2021. When we add these to the estimates for April 2020-May 2021, we find estimates of national excess deaths during April 2020-June 2021 of 3.8M. This amounts to an increase of 34% over deaths expected during a 15 month period from 2019 data, or around 2.8 excess deaths per 1K. Clearly, these figures carry greater uncertainty, given the limited data from June.

5. Sensitivity of the results to assumptions: optimistic and pessimistic scenarios

How do the results change if we assume there were some changes in expected mortality and/or registration completion during the pandemic? In particular, there was evidence of increasing coverage in the data prior to the pandemic as discussed in Section 2. Further, in most of the states considered, there was evidence of disruption to death registration during the pandemic, especially in the early days (India COVID Mapping 2021). Overall, the data for STAR12 for March-April 2020 includes 10% fewer registrations than data for the same period in 2019. Finally, COVID-19 may have hit marginalised communities where death registration is weaker hardest (Bamezai et al. 2021). This would result in a reduction in registration completion.

The reality is that different states likely saw different levels of fluctuation in registration during the pandemic. The overall picture could well be that registration dropped below 2019 levels at some points during the pandemic, but returned to or above 2019 levels at other points. The question is of how such variations might have summed up over the course of the pandemic, and how much this matters to the estimates. Ultimately, survey estimates will be needed to shed light on this question.

Another question to consider is: how do the national estimates change if STAR12 saw a higher or lower mortality impact than the remainder of country?

Overall, we consider the effects on estimated excess mortality of four factors: pre-pandemic registration completion; changes in coverage in the data between 2019 and the pandemic period; changes in expected deaths between 2019 and the pandemic period; and possible differences in mortality impact of the pandemic in STAR12 and the rest of the country. A basic sensitivity analysis in Appendix 2 reveals that of these four factors, it is changes in coverage in the data which can potentially cause the greatest errors in estimated excess mortality. Based on this analysis, we consider two scenarios which we consider plausible best- and worst-case scenarios.

Optimistic scenario. We suppose that:

- There was a 2% increase in expected deaths in each state (associated with population growth and natural increases in CDR).
- Baseline coverage in the data was correctly estimated.
- There was a 5% relative increase in coverage in the data. In other words, in each state, the fraction of total deaths which appear in the available data increased by 5% during the

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

pandemic period relative to 2019 levels; but we cap registration completion in any given state at 100%.

- The remaining states and territories saw a 20% lower p-score/excess mortality than STAR12.

In this optimistic scenario, in STAR12 we obtain 1.6 excess deaths per 1K by May 2021. Nationally, we find 1.4-1.5 excess deaths per 1K by May 2021, rising to 2.0-2.1 by June. This equates to a total of 1.9-2.0M excess deaths up to May 2021, rising to 2.8M by June.

Pessimistic scenario. In this case, we suppose that:

- There was no change in expected deaths relative to 2019 levels.
- Baseline registration completion in the data was overestimated by 5% in each state.
- There was a 5% relative decrease in registration completion in the data in each state during the pandemic period.
- The remaining states and territories saw a 20% higher p-score/excess mortality than STAR12.

In this pessimistic scenario, in STAR12 we obtain 2.7 excess deaths per 1K by May 2021. Nationally, we find 2.9-3.0 excess deaths per 1K by May 2021, rising to 3.9 by June. This equates to a total of 4.0-4.1M excess deaths up to May 2021, rising to 5.2M by June. The results are summarised in Table 2 below.

Table 2: Central, optimistic, pessimistic estimates of excess deaths

	mortality-based extrapolation		**P-score-based extrapolation**	
	up to May 2021	**up to June 2021**	**up to May 2021**	**up to June 2021**
excess deaths (M)	2.8 [1.9, 4.0]	3.8 [2.8, 5.2]	2.9 [2.0, 4.1]	3.8 [2.8, 5.2]
excess deaths per 1K population	2.1 [1.4, 2.9]	2.8 [2.1, 3.9]	2.1 [1.5, 3.0]	2.8 [2.0, 3.9]
p-score (%)	27 [18, 37]	34 [25, 45]	28 [19, 37]	34 [24, 45]
excess deaths as a fraction of annual deaths (%)	32 [21, 43]	43 [31, 56]	32 [22, 44]	43 [31, 56]

Note: National estimates of excess deaths based on the data from STAR12. In brackets after each central estimate are estimates based on the optimistic and pessimistic scenarios respectively. These estimates should *not* be considered as confidence intervals, but rather plausible best and worst case scenarios for pandemic mortality.

6. Excess mortality, recorded COVID-19 deaths and comparison with COVID-19 fatality rates

In this section we assume no changes in expected deaths or registration completion, and use mortality based extrapolations to get national estimates. The goal is to examine how excess mortality estimates align with official COVID-19 deaths over time, and to find out if the overall scale of excess mortality is roughly consistent with what we might expect given the level of spread of disease in India, and the country's age structure.

6.1 Comparison with recorded COVID-19 deaths

We begin by noting the very high ratio of estimated excess deaths to official COVID-19 deaths. Overall, up to May 2021, estimated excess deaths relative to a 2019 baseline in STAR12 were 7.2

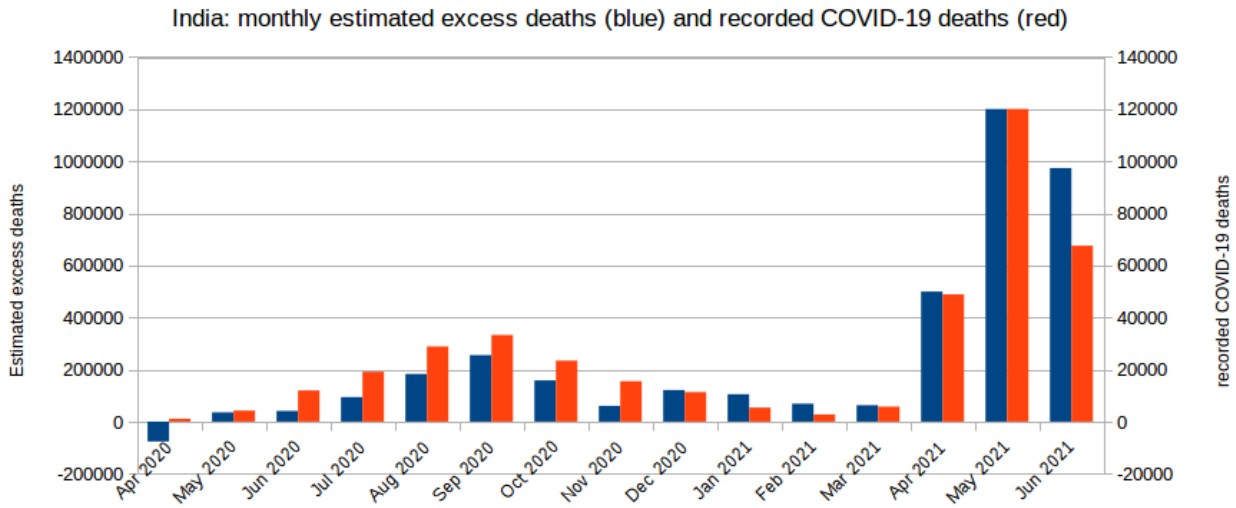
medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

times recorded COVID-19 deaths in these states. The ratio was even higher nationally: during April 2020-May 2021 the estimated national excess death toll was 8.5 times the official COVID-19 toll. If we include estimates up to June 2021, this ratio rises to 9.5. In the optimistic scenario detailed in the previous Section this ratio is 6.9, while in the pessimistic scenario it is around 13.0.

The fact that the estimated ratio of excess deaths to recorded COVID-19 deaths nationally is higher than in STAR12 could reflect the fact that COVID-19 disease and death surveillance in the remaining states and territories was weaker than in STAR12. Recall that pre-pandemic registration completion was considerably higher in STAR12 than in remaining states. And, in addition, there is evidence that COVID-19 disease surveillance was weaker outside STAR12: by June 2021, per capita COVID-19 cases from STAR12 were about 70% higher than in the remainder of the country, even though seroprevalence data from the fourth national serosurvey (PIB India 2021b, 2021a) indicates that seroprevalence was similar across the strata. It is for this reason that we considered extrapolation based on official COVID-19 data to carry a higher risk of bias than extrapolation based on the level of excess mortality, or the p-score.

Qualitatively, national estimates of monthly excess deaths align well with COVID-19 deaths, as seen in the plot of the two data-sets in Figure 1. Note the very different scales.

Figure 1: Estimated monthly excess deaths in India relative to a 2019 baseline, alongside recorded COVID-19 deaths.



Note: The excess deaths are estimated from data in STAR12 according to methods described in the text, and elaborated further in Appendix 2. Data on official COVID-19 deaths is from covid19india.org (COVID19India 2021)

During April 2020-Feb 2021, there is a strong linear association between estimated monthly excess deaths relative to a 2019 baseline and recorded COVID-19 deaths: the correlation coefficient is 0.84. During April 2020-May 2021, this rises to 0.98, and drops slightly to 0.96 over April 2020-June 2021.

Comparisons between the time-course of excess deaths and COVID-19 deaths provide some clues as to how coverage in the data in STAR12 may have changed during the pandemic. The data is consistent with initial registration disruption, followed by recovery to 2019 levels or above. The question is how these effects summed over the duration of time for which we have data. March-April 2020 saw 10% fewer registrations in STAR12 than March-April 2019. On the other hand, January-March 2021 saw around 5% more death registrations than expected based on the ratio of

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

excess to COVID-19 deaths in STAR12 during the remaining months of April 2020-May 2021. This could reflect improved coverage in the data by early 2021.

Given delays in registration, we should be cautious about interpreting changes in the ratio of excess deaths to COVID-19 deaths as indicative of changes in coverage. Karnataka, for example, saw 15% more death registrations during January-March 2021 than during the same period in 2019; this was despite the fact that coverage in the data was estimated to be 100% during 2019, and that during January-March 2021 official COVID-19 deaths in Karnataka were low. It is possible that delays in registration could account for the somewhat higher than expected registrations in states like Karnataka in the period between the country's two COVID-19 waves.

Increasing under-ascertainment of COVID-19 deaths? According to these estimates, the ratio of excess deaths to recorded COVID-19 deaths increased from around 6.7 during April 2020-February 2021, to around 11.3 during March-May 2021. This could reflect an increase in under-ascertainment of COVID-19 deaths during the huge 2021 surge or, possibly, an increase in non-COVID-19 excess deaths.

The scale of the increase in this ratio may be overestimated as a consequence of early disruption to registration and subsequent recovery. For example, during April 2020-February 2021 a modest 2% decrease in coverage in the data relative to 2019 levels leads to a 17% increase in estimated excess deaths during this period; in the other direction, a 5% increase in coverage in the data during March-May 2021 relative to 2019 levels, would reduce excess deaths estimates during this period by around 9%. These changes would have little effect on the overall estimate of pandemic excess deaths, but would raise the first wave ratio of excess deaths to official COVID-19 deaths to 7.8, and lower the second-wave ratio to 10.3. Note that even given such a shift in registration coverage, the ratio would be higher during the second wave.

6.2 Comparison of excess deaths and expected COVID-19 deaths based on international data

How do the estimates of excess mortality align with expectations of COVID-19 mortality given the scale of India's epidemic?

Based on India's estimated 2021 age-structure (National Commission on Population 2019), the meta-analysis of O'Driscoll et al. 2021 predicts a COVID-19 infection fatality rate (IFR) of 0.25%, while the meta-analysis of Levin et al. 2020 predicts COVID-19 IFRs of 0.42%-0.50% depending on the assumed age-distribution of the over-80s. There are some important caveats to such estimates: the meta-analyses differ, and both are based on 2020 COVID-19 fatality data and so, presumably, reflect the lethality of the original variants of SARS-CoV-2 circulating at the time. Moreover, the predictions assume even spread of disease across different age groups. Having noted these caveats, what would IFR estimates of 0.25%-0.50% imply about COVID-19 deaths nationally, and how do these expectations align with our estimates of excess deaths?

Wave 1. The third national serosurvey carried out during December 2020-January 2021 estimated seroprevalence of 24.1% nationally, corresponding to approximately 325M infections nationwide (Murhekar et al. 2021). Allowing a variation in prevalence of 10% in either direction, IFR estimates of 0.25%-0.50% would imply 0.73-1.46M COVID-19 deaths. During April 2020-February 2021, our central estimate of 1.05M excess deaths lies comfortably within this range.

The whole pandemic period upto June 2021. Preliminary results from the fourth national serosurvey (PIB India 2021b) found antibodies to SARS-CoV-2 in 62.3% of unvaccinated individuals sampled,

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

corresponding to an estimated 839M COVID-19 infections by June 2021. Bearing in mind that these are unadjusted figures, and that some prior infections – especially older ones – may be missed, let us suppose that 750-1000M infections (equivalent to infection rates of 56% to 74%) had occurred by this point. IFR estimates of 0.25%-0.50% would then imply 1.9-5.0M COVID-19 deaths. Again, our central estimate of 3.8M excess deaths is comfortably within this range.

Thus India's estimated excess mortality is broadly consistent with expected COVID-19 mortality based on meta-analyses using international data, and estimates of infection levels from seroprevalence surveys. In fact, the data are consistent with the following assertions, although none can be made definitively:

- The majority of excess deaths were likely COVID-19 deaths
- A significant minority of excess deaths may have been either avoidable COVID-19 deaths (caused, for example, by unavailability of medical care or oxygen); or non-COVID-19 deaths caused, for example, by disruptions to health-care.
- During the second wave, more lethal variants and/or overwhelmed health systems may have driven up COVID-19 IFR despite increasing vaccination coverage. We need to treat this conclusion with some caution: registration disruption may have been particularly acute during national lockdown and the early part of the pandemic, leading to underestimation of first wave excess mortality. On the other hand, recovery of registration coverage to 2019 levels or higher during the second wave could lead to some overestimation of second wave excess mortality.

7. Reflections on uncertainties and potential biases

There are several sources of uncertainty, which can lead to under- or over-estimation of excess mortality in STAR12 and nationally. These have already been discussed above, but we summarise and comment further on these below.

- *The data used is not up to date*. Delays in registration, and continued spread of disease, mean that we should expect further increases in estimates of excess deaths as more data becomes available. As some of the data we use is recorded by date of death (rather than date of registration), even totals for months where we already have data are likely to rise.

- *The data for June 2021 is from only three states.* Estimates using this data carry greater uncertainty. It is possible that Andhra Pradesh, Karnataka and Punjab saw greater or later spread than the national average. Recall, however, that in these states June saw a sharp rise in the ratio of excess to COVID-19 deaths, indicative of registrations delayed during the massive surge in mortality in May. We could see a similar pattern repeated in other parts of the country if June data becomes available.

- *Death registration completion prior to the pandemic may have been overestimated*. We discussed earlier how death registration completion at the national level is likely overestimated in the 2019 CRS report, and even the sub-national estimates may overestimate completion to some degree. Overestimating pre-pandemic completion is equivalent to underestimating pre-pandemic mortality. Taking into account some overestimation of pre-pandemic registration completion would push up estimates of total excess deaths in an easily quantifiable way. Excess deaths per million would go up, and so would the ratio of excess-to-COVID-19 deaths; but p-scores would not necessarily rise. Clearly, having accurate data on mortality is of importance even in non-pandemic times; the pandemic has, however highlighted how critical it is.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

- *Death registration levels could have changed during the pandemic*. As we saw in computations above, relatively small shifts in coverage in the data significantly impact the estimates. For example, a 5% drop in coverage across STAR12 causes a 20-25% increase in excess mortality estimates. Looking at the data all together, the strong association between excess deaths and official COVID-19 deaths over time provides supporting evidence that the bulk of the fluctuations in registered deaths reflect fluctuations in mortality driven by the pandemic, rather than trends in registration.

- *There may be natural changes in yearly deaths on account of population growth and a changing CDR.* The likely scale of such changes was discussed in Section 2, and found to be fairly modest. SRS and UN estimates suggest stable CDR, and population growth at around 1% a year. An increase in expected deaths of 2% relative to 2019 levels is plausible, and would cause a reduction of about 7% in estimates of excess mortality.

- *National surge/excess mortality may not match the estimates from STAR12*. It is currently unclear in what direction data from more states and territories might push the estimates. But there is no convincing reason to believe the rest of the country was less hard hit. Indeed, partial data (Acosta et al. 2021; Das 2021) suggests that some of the absent states, such as Uttar Pradesh and Gujarat, have been very badly hit during the pandemic. The effects of these uncertainties are fairly easily quantified. For example, 20% lower or higher excess mortality in the remainder of the country relative to STAR12 changes national excess mortality by around 8%.

Ultimately, as more data becomes available it will be possible to unravel the potential biases with more confidence. But even the most optimistic and pessimistic views of the data do not change the story qualitatively. Consistent with the wide spread of disease, excess mortality has been high, and surveillance of COVID-19 deaths has been very weak.

8. Summary of findings

The data from STAR12 allow us to infer with high confidence that STAR12 saw a major surge in mortality during the pandemic. The calculations here give estimates of 2.1 excess deaths per 1K population in these states during April 2020-May 2021. Repeated nationally, this level of excess mortality equates to around 2.8M excess deaths. Extending these estimates to June using more limited data we estimate around 3.8M excess deaths during April 2020-June 2021. Optimistic or pessimistic assumptions could shift these estimates by 20%-30% in either direction, while more up-to-date data are likely to push them up.

Given different age-structures and levels of development, comparing pandemic excess mortality in different countries is best done by considering excess deaths as a fraction of annual deaths. By this measure, India's estimated excess death toll up to June 2021 was around 43% of its normal annual toll. Uncertainties in the data mean that this figure could plausibly lie between 31% and 56%. Using comparisons with international data on pandemic excess deaths as a percentage of annual deaths (Karlinsky and Kobak 2021), even the lowest of these estimates places India amongst some of the hardest hit countries in the world.

The time-course of monthly excess deaths estimated using the process here displays a surprisingly strong association with recorded COVID-19 fatalities, rising and falling with the two waves of the epidemic. This suggests that the majority of these deaths reflect consequences of the pandemic,

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

rather than underlying trends in mortality or death registration. This is not to say that these were all deaths from COVID-19: there may, indeed, have been some non-COVID-19 excess deaths. However, comparing the excess deaths estimates to expectations of COVID-19 mortality based on disease spread suggests that the majority of excess deaths were likely from COVID-19.

However we look at it, official COVID-19 deaths have entirely failed to capture the scale of pandemic excess mortality in India. If most excess deaths were, indeed, from COVID-19 then under-ascertainment of COVID-19 deaths has been high, with around 8-10 excess deaths for every recorded COVID-19 death. There is also evidence that under-ascertainment increased during the huge second wave.

Our findings highlight the extent to which populations are vulnerable to mortality crises in LMIC contexts and the extent to which disease surveillance systems under-estimate mortality. This points to the urgent need for robust systems to monitor all-cause mortality, as well as to improve the availability of real time data from these systems. Our analysis reveals that data from India's civil registration system can be challenging to analyse. Even so, this data provides key insights into India's pandemic mortality crisis.

References

Aburto, J.M., Schöley, J., Zhang, L., Kashnitsky, I., Rahal, C., Missov, T.I., Mills, M.C., Dowd, J.B., and Kashyap, R. (2021). Recent Gains in Life Expectancy Reversed by the COVID-19 Pandemic. *medRxiv*. https://doi.org/10.1101/2021.03.02.21252772.

Acosta, R.J., Patnaik, B., Buckee, C., Balsari, S., and Mahmud, A. (2021). All-cause excess mortality in the State of Gujarat, India, during the COVID-19 pandemic (March 2020-April 2021). *medRxiv*. https://doi.org/10.1101/2021.08.22.21262432.

Anand, A., Sandefur, J., and Subramanian, A. (2021). Three new estimates of India's all-cause excess mortality during the COVID-19 pandemic. *Cent Glob Dev Work Pap* 416.

Aron, J. and Muellbauer, J. (2020). Transatlantic excess mortality comparisons in the pandemic. *INET Oxford Working Paper No. 2020-18.*

Bamezai, A., Banaji, M., Gupta, A., Pandey, S., Sharan, M.R., Sharma, K., and Singh, C.K. (2021). Survey evidence of excess mortality in Bihar in the second COVID-19 surge. *SocArxiv*.

Banaji, M. (2021a). IndiaACMdata/National Estimates/V1 [electronic resource]. https://github.com/muradbanaji/IndiaACMdata/tree/master/NationalEstimates.

Banaji, M. (2021b). What does mortality data tell us about Bihar's first Covid-19 wave? *Scroll.in*. https://scroll.in/article/999865/what-does-mortality-data-tell-us-about-bihars-first-covid-19-wave.

Banaji, M. and Gupta, A. (2021). *India: Notes on Estimates of Crude Death Rate and Registration Coverage*. IndiaCOVIDMapping.org. https://www.indiacovidmapping.org/reports/mortality/IndiaRegistrationNotes.pdf.

COVID19India (2021). Coronavirus in India: Latest Map and Case Count [electronic resource]. https://www.covid19india.org.

Das, S. (2021). Death Count In 24 UP Districts 43 Times More Than Official Covid-19 Toll — Article 14. *Article 14*. https://www.article-14.com/post/untitled-60cf605395758.

Deshmukh, Y., Suraweera, W., Tumbe, C., Bhowmick, A., Sharma, S., Novosad, P., Fu, S.H., Newcombe, L., Gelband, H., and Brown, P. (2021). Excess mortality in India from June 2020 to June 2021 during the COVID pandemic: death registration, health facility deaths, and survey data. *medRxiv*. https://doi.org/10.1101/2021.07.20.21260872.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Gerland, P. (2014). UN Population Division's methodology in preparing base population for projections: case study for India. *Asian Population Studies* 10(3):274–303.

Giattino, C., Ritchie, H., Roser, M., Ortiz-Ospina, E., and Hasell, J. (2020). Excess mortality during the Coronavirus pandemic (COVID-19). *Our World in Data*.

Government of India (2021a). National Family Health Survey - 5 State Report Bihar [electronic resource]. http://rchiips.org/nfhs/NFHS-5Report_BR.shtml.

Government of India (2021b). Vital Statistics of India Based on the Civil Registration System 2019. *New Delhi: Office of the Registrar General of India, Ministry of Home Affairs, Government of India*.

India COVID Mapping (2021). Mapping the impact of the COVID-19 pandemic in India [electronic resource]. https://www.indiacovidmapping.org/.

Karlinsky, A. and Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *Elife* 10:e69336.

Leffler, C.T., Yang, E., and Lykins, J.D. (2021). Preliminary Analysis of Excess Mortality in India During the Covid-19 Pandemic (Update September 26, 2021). *medRxiv*. https://doi.org/10.1101/2021.08.04.21261604.

Levin, A.T., Hanage, W.P., Owusu-Boaitey, N., Cochran, K.B., Walsh, S.P., and Meyerowitz-Katz, G. (2020). Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European journal of epidemiology*:1–16.

Murhekar, M.V., Bhatnagar, T., Thangaraj, J.W.V., Saravanakumar, V., Kumar, M.S., Selvaraju, S., Rade, K., Kumar, C.G., Sabarinathan, R., and Turuk, A. (2021). SARS-CoV-2 seroprevalence among the general population and healthcare workers in India, December 2020–January 2021. *International Journal of Infectious Diseases* 108:145–155.

National Commission on Population (2019). Population Projections for India and States 2011-2036. *Report of the Technical Group on Population Projections Constituted by the National Commission on Population*.

O'Driscoll, M., Dos Santos, G.R., Wang, L., Cummings, D.A., Azman, A.S., Paireau, J., Fontanet, A., Cauchemez, S., and Salje, H. (2021). Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature* 590(7844):140–145.

ORGI (2020a). *Sample Registration System Bulletin (Various Years)*. Office of the Registrar General and Census Commissioner of India, New Delhi: Government of India. https://censusindia.gov.in/vital_statistics/SRS_Bulletins/.

ORGI (2020b). *Sample Registration System Statistical Report 2018*. Office of the Registrar General and Census Commissioner of India, New Delhi: Government of India.

PIB India (2021a). Centre advises States to conduct State-specific Sero Surveys in consultation with ICMR to generate district-level data on sero-prevalence [electronic resource]. https://pib.gov.in/pib.gov.in/Pressreleaseshare.aspx?PRID=1739902.

PIB India (2021b). Implications of 4th Round of National Sero-Survey show that there is a ray of hope but there is no room for complacency. Non-essential travel must be discouraged and travel only if fully vaccinated - DG @ICMRDELHI #IndiaFightsCorona https://t.co/IwOs72LBAR. *@PIB_India*. https://twitter.com/PIB_India/status/1417443573948043271.

Ramani, S. (2021). On 'excess deaths' in Rajasthan. *The Hindu*. https://www.thehindu.com/news/national/on-excess-deaths-in-rajasthan/article35136631.ece.

Rao, C. and Gupta, M. (2020). The civil registration system is a potentially viable data source for reliable subnational mortality measurement in India. *BMJ Global Health* 5(8):e002586.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Rukmini S (2021). Madhya Pradesh saw nearly three times more deaths than normal after second wave of Covid-19 struck. *Scroll.in*. https://scroll.in/article/996772/madhya-pradesh-saw-nearly-three-times-more-deaths-than-normal-after-second-wave-of-covid-19-struck.

The Economist (2021). The pandemic's true death toll. *The Economist*. https://www.economist.com/graphic-detail/2021/09/27/americas-pandemic-is-now-an-outlier-in-the-rich-world.

The Hindu Data Team (2021). Data | India's excess deaths could be highest among nations with the most recorded COVID-19 fatalities. *The Hindu*. https://www.thehindu.com/data/data-indias-excess-deaths-could-be-highest-among-nations-with-most-recorded-covid-19-fatalities/article35636438.ece.

United Nations, Department of Economic and Social Affairs, Population Division (n.d.). *World Population Prospects 2019 Data Booklet*. (ST/ESA/SER.A/424).

World Bank (2021). Death rate, crude (per 1,000 people) - India | Data [electronic resource]. https://data.worldbank.org/indicator/SP.DYN.CDRT.IN?locations=IN.

World Health Organization (2014). *Global Civil Registration and Vital Statistics: A Scaling Up Investment Plan 2015-2024*. World Health Organization.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Appendix 1. Estimates of crude death rate and registration completion in 2019

The 2019 report on Vital Statistics of India based on the Civil Registration System (Government of India 2021b), provides both national and sub-national data on registered deaths, along with estimates of completion. The latter draw on estimates of the crude death rate (CDR) from the 2018 Sample Registration System annual statistical report (ORGI 2020b). Completion nationally or sub-nationally can be estimated by comparing total registered deaths to total expected deaths based on CDR and population estimates.

The 2018 SRS report estimates India's CDR at **6.2** per 1K. Using population projections (National Commission on Population 2019) and assuming no change in the CDR between 2018 and 2019, this implies to the nearest whole percentage **92%** completion nationally in 2019. There are, however, several reasons to believe that the estimate of CDR at 6.2 could be too low, and hence the estimate of 92% completion too high.

UN estimates. The United Nations estimated CDR in India to be **7.2** per 1K for the period 2015-20 (United Nations, Department of Economic and Social Affairs, Population Division n.d.), based on methodologies described by Gerland 2014. This would imply an estimated registration completion during 2019 of only **80%**.

Combining sub-national estimates. We can combine sub-national estimates of registration completion in the 2019 CRS report to estimate national completion, and hence CDR. These calculations cap registration completion at 100%, i.e., if a region saw more deaths registered than expected from its estimated CDR, then registered deaths are taken as an estimate of total deaths. In each region we use the completion estimate to estimate total deaths, and summing these gives an estimated national death toll in 2019; from this we obtain national estimates of CDR and completion. This process gives an estimated national CDR in 2019 of **6.6** per 1K, and registration completion of around **86%**.

Estimates based on age-stratified mortality rates. The 2018 SRS report gives estimated death rates in different age groups. Projected population pyramids are available for 2016 and 2021, but not intervening years. Using the estimated 2018 age-dependent mortality rates, and the projected 2016 pyramid, we obtain an estimated national CDR of **6.8** per 1K, which would imply registration completion in 2019 of **84%**. Using the 2021 pyramid, we obtain an estimated CDR of **7.5** per 1K, implying registration completion in 2019 of **76%**. (Estimates depend on the fraction of the over-80s who are assumed to be over 85, which is not given in the projected population pyramids; we set this to be 0.375 as estimated in the 2018 SRS.)

Estimates based on NFHS-5. Phase 1 of the National Family Health Surveys were conducted between mid-2019 and early 2020 in 22 states and UTs comprising around 50% of the national population (Government of India 2021a). The survey asked respondents about deaths of any usual family member in the last three years, and whether the death was registered. From this data, we can compute a weighted average registration level in these states and territories of 75.4% (see Banaji 2021). On the other hand, based on sub-national data in the SRS and CRS reports, the registration level in these 22 states and UTs for an equivalent period of three years before the average NFHS interview was 81.7%. The NFHS-5 data thus suggests that the CDR in these regions taken together was about 8% higher than estimates based on sub-national SRS and CRS data. If this holds true nationally, we get an estimated CDR in 2019 of **7.0**, and registration completion of around **80%**.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

We thus find estimates of CDR nationally ranging from 6.2 to 7.5, corresponding to registration completion in 2019 ranging from 76% to 92%, with a median estimate of around 82%. The estimated completion of 86% derived from sub-national data in the SRS and CRS reports may thus somewhat overestimate registration completion in India. We return to this point when discussing our results. More detail and full calculations are available at IndiaCOVIDmapping.org (Banaji and Gupta 2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Appendix 2. Computing excess deaths, and sensitivity of the estimates to parameter changes

In order to estimate excess deaths in a given location over a given period, we need to estimate expected deaths and actual deaths. Given death registration data during some pandemic period and some comparable reference period, we additionally need estimates of:

1. coverage in the data during the reference period
2. coverage in the data during the pandemic periodic
3. expected changes in mortality between the reference period and pandemic period

With quantities in appropriate units, we can then compute:

1. (pandemic deaths) = (pandemic registrations)/(pandemic coverage)
2. (baseline deaths) = (baseline registrations)/(baseline coverage)
3. (expected deaths) = (expected change)*(baseline deaths)
4. (excess deaths) = (pandemic deaths) - (expected deaths)
5. p-score = 100*(excess deaths)/(expected deaths)

Finally, if we wish to extrapolate from some regions to others, we also need estimates of how the mortality impact may have varied between the regions for which we have data, and those for which we don't.

We can examine how small changes in the key parameters affect our central estimates of excess deaths in STAR12 during April 2020-May 2021, and nationally:

- *Changes in coverage*. A 1% (relative) decrease in pandemic period coverage in the data in each state relative to 2019 causes a 4.6% increase in excess deaths estimates in STAR12. Based on pre-pandemic trends and the disruption we see early in the pandemic, we might optimistically hope to see a relative increase in coverage in the data of 5% in states where this is possible; pessimistically, we might expect a relative decrease in coverage in the data of 5%. The reality is most likely that the competing effects of disruption and recovery summed in different ways in different states.
- *Changes in expected deaths*. A 1% increase in expected deaths during the pandemic period relative to 2019 (as a consequence of population growth and changes in CDR) would cause a 3.6% decrease in excess mortality estimates in STAR12. Based on pre-pandemic population growth and trends in CDR, we might expect year-on-year deaths to remain fixed or rise by up to 2%.
- *Errors in estimation of baseline mortality/coverage*. A 1% (relative) decrease in both baseline and pandemic period coverage in the data in each state causes a 1% increase in excess mortality estimates in STAR12. Optimistically, we might hope that the sub-national estimates of registration completion in the 2019 CRS report are accurate. A more pessimistic view would be that they could have been overestimated by 5%.
- *Difference in mortality impact outside STAR12*. If regions not in STAR12 collectively saw a mortality impact 1% higher than in STAR12, this causes a 0.4% increase in the national mortality estimates. We consider it plausible that the mortality impact in STAR12 could differ by up to 20% from that in the remainder of the country, causing an 8% shift in our excess mortality estimates.

When we consider the effects and possible scale of the uncertainties in these parameters, it is clear that shifts in coverage in the data during the pandemic cause the greatest uncertainty in the estimates of excess mortality. By comparison, the likely effects of population growth, changes in CDR, and errors in estimates of pre-pandemic registration coverage are relatively small. There is

medRxiv preprint doi: https://doi.org/10.1101/2021.09.30.21264376; this version posted October 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

also no compelling reason to believe that the mortality impact of the pandemic outside of STAR12 should have been considerably different from than in STAR12.