

August 18, 2022

Mortality and Death Registration in India

The many holes in the data

By: [Murad Banaji](#), [Aashish Gupta](#), [Vipul Paikra](#)

Underlying the controversy over the estimates of Covid-19 mortality in India is the uncertain quality of mortality statistics and the death registration data. The pandemic is an opportunity to redress the gaps with continued scientific attention and vigorous public discussion.

The World Health Organization (WHO) has estimated that India saw a shocking 4.7 million excess deaths during 2020-2021. These 4.7 million deaths could make up around a third of the global pandemic toll. The WHO estimate is broadly in line with several other estimates of pandemic excess deaths appearing in published research studies or preprints, most of which place excess mortality at between 3 million and 5 million. The estimates are based on various data, but the most important source of data is the India's Civil Registration System (CRS).

The government of India has [strongly objected](#) to these estimates.

To make sense of the numbers – and the objections – we need to answer the question, how many deaths occur in India each year in 'normal' times? If we know this, then we can also calculate 'registration completion', the fraction of deaths registered in the CRS. Having accurate estimates of pre-pandemic mortality and registration completion is crucial to correctly estimating pandemic excess deaths. It is also crucial for our general understanding of mortality in India.

However, estimating annual deaths and registration completion is not entirely straightforward. The data is incomplete and can be confusing. And it is possible to frame different, and even contradictory, narratives around the same data.

To highlight this point, let us start with a [claim by the union health ministry](#) in their rebuttal of the WHO estimates. According to this statement, in 2020:

- An estimated 8.1 million deaths occurred nationally
- 99.9% of these deaths were registered in the CRS.

Although the claim has two parts, one about mortality and one about registration completion, it is actually a single claim. If we believe that only 8.1 million deaths occurred in 2020, then data in the [2020 CRS report](#) confirm that essentially all of these deaths were registered.

So, the question is: where does the estimate of 8.1 million annual deaths come from? And is it correct?

How to estimate mortality and registration completion

Let us step back for a moment and note that there are two ways to estimate total mortality and registration completion.

1. *Directly estimate mortality and indirectly estimate registration completion:* A mortality survey can be used to estimate the crude death rate: how many out of every 1000 people die each year? From this we can estimate total deaths in a year. We then compare estimated total deaths with registered deaths to find what fraction of deaths were registered. For example, if from a survey we estimate a crude death rate of 6.2, we can then multiply by estimates of the national population (say 1.33 billion) to estimate 8.25 million yearly deaths. If there were 7.64 million death registrations that year, then we obtain completion of $7.64/8.25 = 93\%$. This approach is the basis of official estimates of registration completion.

2. *Directly estimate registration completion and indirectly estimate mortality:* A survey can ask people whether deaths that occurred in their families were registered. The resulting estimate of completion can be used alongside civil registration data to estimate total deaths. For example, if a survey estimates that 80% of deaths were registered nationally, and CRS reports tell us that 7.64 million deaths were registered that year, we would calculate that in fact there had been $(7.64 \text{ million})/0.8 = 9.55 \text{ million}$ deaths in total.

These two approaches are both valid and logically distinct. It is, of course, possible for a single survey to estimate both mortality and completion directly.

The government [Sample Registration System \(SRS\)](#), a continuous annual survey in randomly sampled villages and urban blocks, provides direct estimates of mortality. During 2018, 2019, and 2020, the SRS estimated the crude death rate in India to be 6.2, 6.0, and 6.0 per 1,000, respectively. Taking India's population to be around 1.35 billion, this gives 8.1–8.4 million deaths annually. On registration, 6.9 million deaths were registered in 2018, 7.6 million in 2019, and 8.1 million in 2020. Using this data, without acknowledging the likely errors in the crude death rate, we would estimate around 85% completion in 2018 and we would find over 90% completion in 2019. In fact, if we use the 2019 estimate of the death rate, we estimate over 95% completion. In 2020 we would obtain 100% completion, as official estimates of registration completion from the government claim.

|| [W]e appear to have two different sources of data telling very different stories about pre-pandemic mortality in India. One gives us a little over 8 million annual deaths; the other around 9.5 million...

On the other hand, the [National Family Health Survey \(NFHS\)](#) directly asks the question of whether deaths in the family were registered. We can use recently released NFHS-5 data to estimate, for example, that registration completion was around 73% in 2018, considerably lower than the 85% estimate from SRS data. From NFHS completion data we find there must have been around 9.5 million annual deaths before the pandemic. If we assume that there was little change in total mortality between 2018 and 2019, then around 80% of deaths were registered in 2019, immediately before the pandemic.

So, we appear to have two different sources of data telling very different stories about pre-pandemic mortality in India. One gives us a little over 8 million annual deaths; the other around 9.5 million during the years immediately before the pandemic. One says that over 95% of deaths were registered in 2019, the other that only around 80% of deaths were registered in 2019. Which is closer to the truth?

Evidence that the SRS underestimates mortality

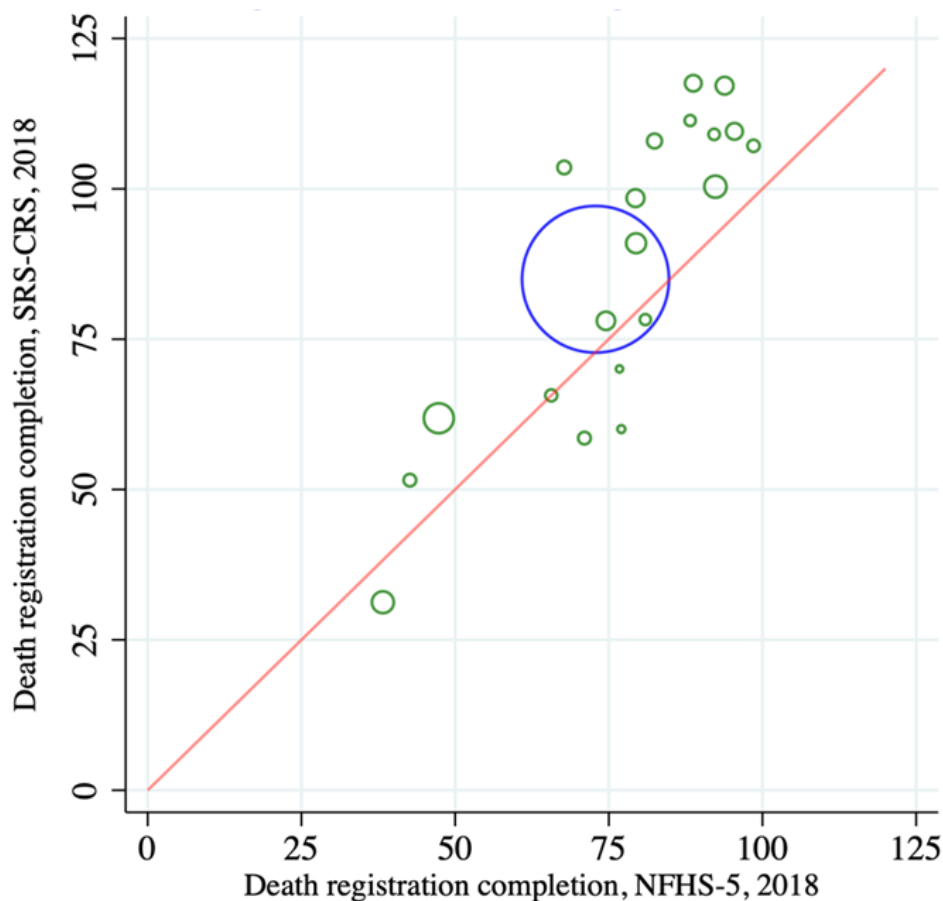
If we look a little closer, we see striking evidence *within the SRS data itself*, that the SRS-based estimate of a little over 8 million annual deaths is too low. And consequently, that estimates of registration completion based on SRS data are too high.

First, there are just too many death registrations in some states. In 2019, Andhra Pradesh, Gujarat, Karnataka, Maharashtra, Odisha, Tamil Nadu, and West Bengal together saw over 500,000 more death registrations than the death toll expected from SRS death rates. The most likely explanation for this is that the SRS underestimated the crude death rate in these states.

When state civil registration systems register more deaths than the total number of estimated deaths (on the basis of the SRS), India's official estimates for registration completion assume that death registration was 100%. This is wrong, because the NFHS shows some deaths went unregistered even in states where official estimates of registration completion are 100%.

And the problem is not limited to a few states. In fact, in 2019 many states saw more deaths registered than expected based on SRS estimates. If we compare completion based on the SRS (indirect) and the NFHS-5 estimates (direct), we find that across the larger states and union territories SRS-based estimates are mostly higher than NFHS ones (Figure 1).

Figure 1: Comparison of death registration completion according to the SRS and the NFHS



Registration estimates from NFHS-5 for 2018 are calculated from NFHS micro-data. Green circles show major states, as classified by the SRS, and the blue circle shows India. Circles sizes represent 2018 population, as estimated by the National Committee on Population.

Another clue that the SRS-based estimates of completion nationally are too high comes from examining two large states where civil registration is weak: Bihar and Uttar Pradesh. These states, together, comprise a quarter of the national population. According to the SRS estimates of death rates, we should expect around 2.1 million deaths every year in these two states. On the other hand, according to the CRS report, Bihar and UP together saw around 1.3 million death registrations in 2019. In other words, *according to the SRS itself*, around 800,000 deaths went unregistered in these two states. Going through each state in turn, we find at least 1.2 million missing death registrations in 2019. There is no way that registration can be above 95% nationally when the data indicates that such a large number of deaths go unregistered.

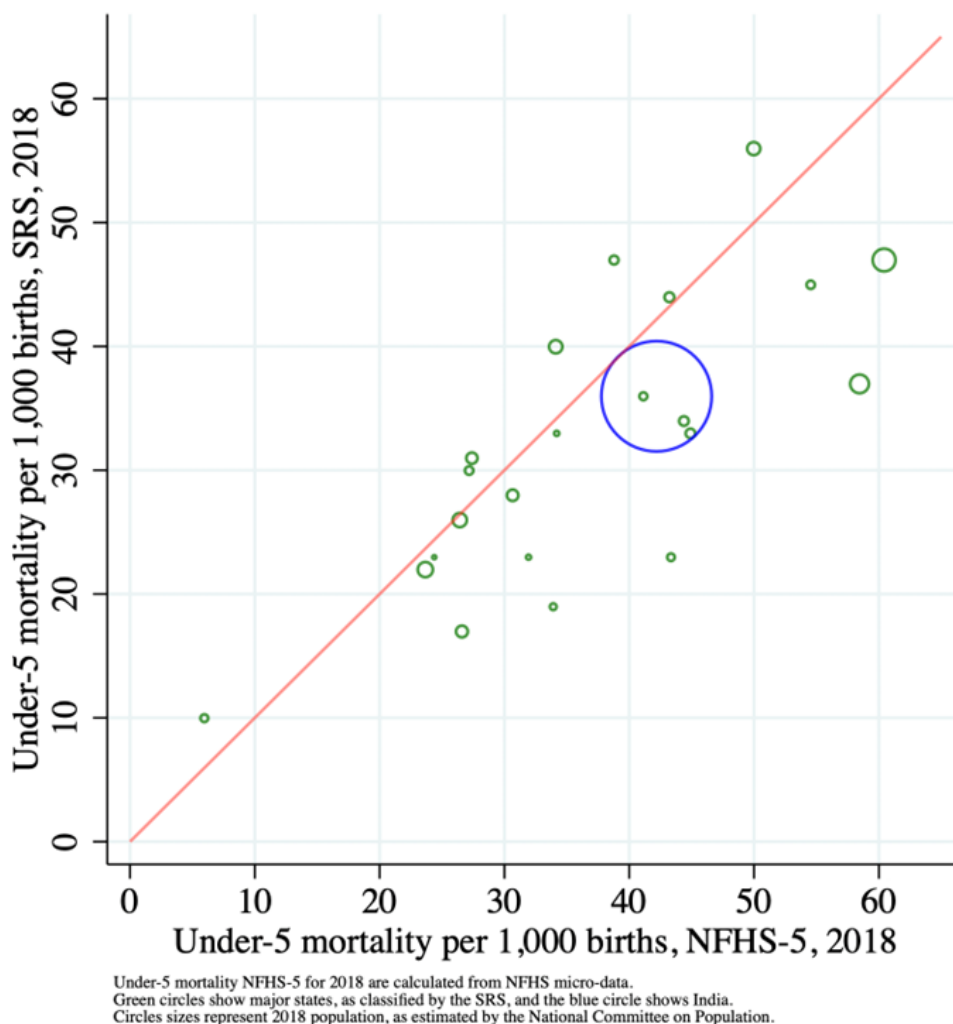
Given these clear gaps, how does the government arrive at an estimate of near complete death registration for 2020 based on SRS and CRS data? By an effective sleight of hand: the large number of unregistered deaths in states such as Bihar and Uttar Pradesh are offset against the “extra” death registrations over and above SRS expectations in states such as Andhra Pradesh and Karnataka. This manoeuvre effectively hides from view two problems clearly visible in regional SRS data. First, the strong evidence that the SRS under-estimates the crude death rate in several large states where the CRS registers more deaths than expected from SRS death rates. Second, the low registration coverage in states where many fewer deaths are registered than expected from SRS death rates.

There is no way that registration can be above 95% nationally when the data indicates that ... a large number of deaths go unregistered.

It is also instructive, to look at under-5 mortality, a key component of overall mortality. Comparing under-5 mortality rates from the NFHS with the SRS shows that the SRS very likely under-estimates under-5 mortality. According to SRS, there were 36 under-5 deaths per 1,000 births in 2018. However, when calculating under-5 mortality rates from the [NFHS-5 using standard approaches](#), we observe 42.2 under-5 deaths per 1,000 births in 2018. The discrepancy was the highest in Uttar Pradesh and Bihar. The SRS reported an under-

5 mortality rate of 37 deaths per 1,000 births in Bihar, whereas NFHS-5 observed 58.4 deaths per 1,000. In Uttar Pradesh, the SRS reported an under-5 mortality rate of 47 deaths per 1,000 births. According to NFHS-5, it was 60.4 under-5 deaths per 1,000 births. This suggests that the problem of under-estimation of mortality is not limited to states where registration coverage is high. It is likely to be widespread across India.

Figure 2: Comparison of under-5 mortality in NFHS-5 and SRS, 2018



Why does the SRS underestimate mortality?

We have seen some fairly compelling evidence, much of it within SRS data itself, that the SRS often underestimates mortality. But why? This is hard to answer. We should bear in mind that survey-based estimates always carry risks of possible bias. For example, if the surveyed population doesn't include enough older people, we risk underestimating mortality. Because death rates are highest amongst the elderly, getting their number wrong can cause serious errors in estimating mortality. Could this be occurring?

The SRS reports provide data about estimated population and death rates in each age group. According to the 2018 and 2019 SRS reports, only 5.2% of the population was over 65. This could be too low: already the fraction of over-65s in the 2011 census was higher than this, and government [population projections](#) suggest it should have reached 6.2% in 2016 and 6.9% in 2021. We would expect perhaps 6.4%-6.7% of the population to have been over 65, by 2019.

Indeed, if we use SRS data on death rates in each age-bracket, along with population projections for numbers in each age-bracket, we expect around 9.5 million deaths annually, in line with estimates derived from NFHS data.

Therefore, part of the explanation for the low death rates in the SRS could be that the its sample simply doesn't include a sufficient number of older people. If this is the case, it remains to be explained why this is so. Moreover, this may not be the whole story; there may be other factors at play. As we have seen, there is evidence that the SRS under-reports mortality in children, an effect which is unlikely to be explained by a sample with an incorrect age-structure. Understanding fully why the SRS underestimates mortality is an important puzzle that remains to be solved.

How many people died in 2020?

Let us return to the question of mortality during the first phase of the COVID-19 pandemic. The government assertion that 8.1 million deaths occurred in 2020, all of them registered, would imply that the first year of the pandemic saw no excess mortality at all. And that death registration became complete for the first time during that year.

Do these remarkable claims hold water? Clearly not. First, they are very much at odds with what the NFHS tells us about total annual deaths, even without a pandemic. Second, even though they are based on SRS data, they are contradicted by regional data from the SRS, and age-wise data from the SRS.

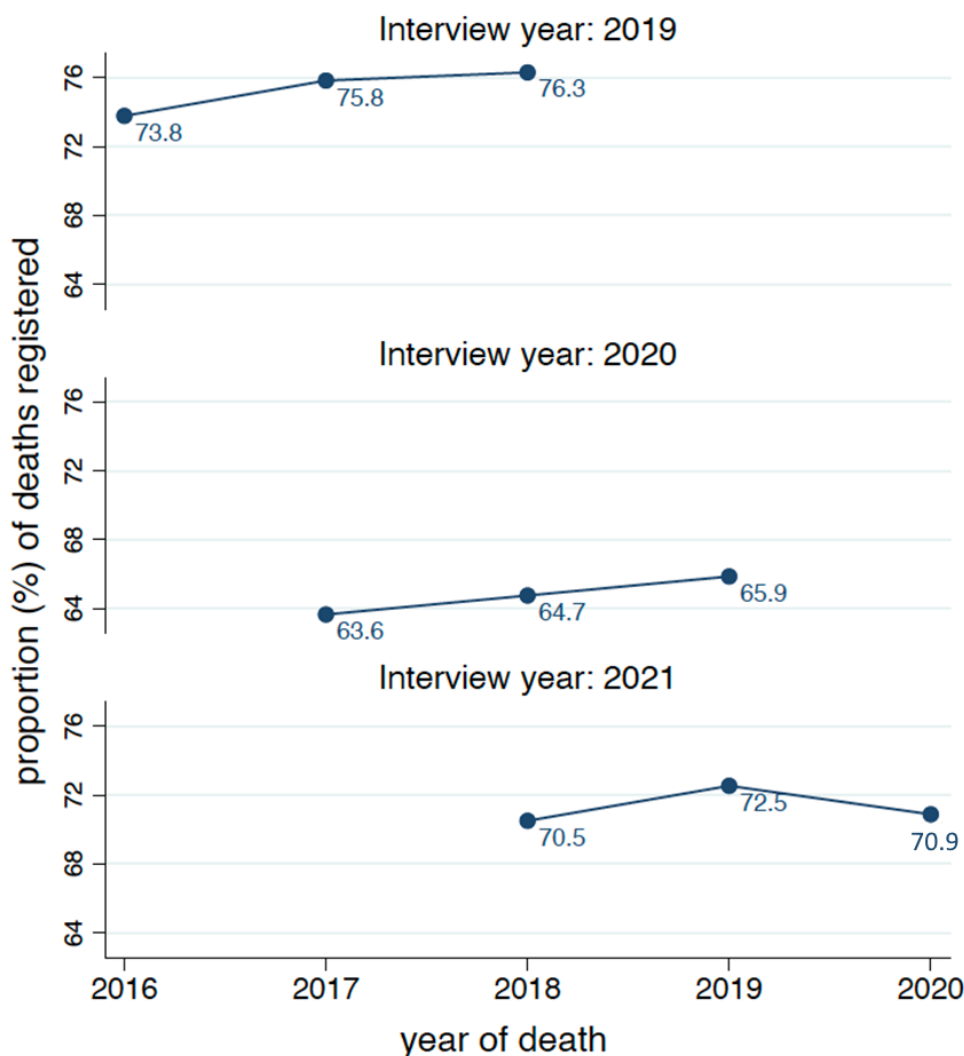
The absurdity of the claim of complete registration in 2020 is highlighted by data from Bihar and Uttar Pradesh. According to the SRS and CRS data itself, just as in 2019, around 800,000 deaths went unregistered in these two states alone in 2020. It is, of course, a logical impossibility for registration nationally have been complete when so many deaths went unregistered in these two populous states.

If the government claim is false, then what was the true death toll nationally during 2020? And how many excess deaths occurred during that year because of the pandemic? To answer these questions, we need to know what happened to registration completion during 2020. The story is likely to have been quite variable across the country. There was certainly some disruption to death registration. For example, according to the 2020 CRS report, Uttar Pradesh saw 7.5% fewer registered deaths in 2020 compared to 2019, strongly suggesting that completion fell.

While there is strong evidence during 2020 of disruption to registration of deaths (and, indeed, births) in some states, it is possible that in some states, registration held steady, or even rose. If we want to infer a national figure, we need to know how the trends added up nationally. Although it is hard to arrive at a national estimate of completion in 2020, there is some evidence within NFHS data that nationally registration completion could have reversed its previous rising trend.

NFHS-5, which spanned the years 2019-2021, visited states in a phased manner. NFHS-5's first phase in 2019 allows us to observe registration completion for 2016, 2017, and 2018 in Phase 1 states. Phase 2 of the survey was disrupted due to the pandemic's first wave in 2020. Interviews in Phase 2 states were conducted in the early part of 2020, and then in the early part of 2021. From phase 2 interviews in 2020, we can observe registration completion for 2017, 2018, and 2019; and from 2021 interviews, we can observe death registration completion for 2018, 2019, and 2020. Looking at this data shows that registration completion fell by a few percentage points during 2020 at least in the states visited in 2021 (see Figure 3).

Figure 3: Trends in registration in India, NFHS-5



NFHS-5 (2019-21) was disrupted by the COVID-19 pandemic. Interviews were conducted in phases, over 3 years: 2019 (June-Dec 2019), 2020 (Jan-Mar 2020 & Nov-Dec 2020), and 2021 (Jan-May 2021). NFHS-5 asked questions on deaths of any members in the household in the last couple of years, and if the death was registered. Because NFHS-5 visited states in a phased manner, levels of registration across interview years are not comparable. However, for states interviewed in 2021, the NFHS design allows a comparison of death registration completion in 2020 compared to previous years. Because some deaths are registered months after the death, we drop those deaths that occur in the same calendar year as the interview, as well as those deaths that occurred three months prior to the interview. Deaths upto three years prior to the month of interview are included. Places visited by NFHS-5 in 2021 include parts of Punjab, Uttarakhand, Haryana, Delhi, Rajasthan, Uttar Pradesh, Arunachal Pradesh, Jharkhand, Odisha, Chhattisgarh, Madhya Pradesh, Tamil Nadu and Puducherry.

Birth registration data from the CRS also point towards disruption. Birth registration has greater demand than death registration and was increasing before the pandemic. However, 2020 saw 2.4% fewer birth registrations than 2019, a drop which strongly suggests that more births went unregistered in 2020 than in 2019.

Thus, both NFHS data and birth registration data suggest that nationally registration may have suffered during 2020. We can examine how different scenarios for changes in death registration lead to different estimates of excess deaths in 2020. If we take completion during 2019 to be around 80% (as inferred from NFHS data), we then find, from the 2020 CRS report:

- If there was no change in completion between 2019 and 2020, then around 600,000 excess deaths occurred in 2020.
- If completion dropped by around 2 percentage points to 78% in 2020, then around 850,000 excess deaths occurred in 2020.
- If completion dropped by around 4 percentage points to 76% in 2020, then around 1.1 million excess deaths occurred in 2020; and so forth.

It is not yet entirely clear which, if any, of these estimates is the most accurate. In any case, excess deaths in 2020 of 0.6-1.1 million (as compared to around 150,000 recorded Covid-19 deaths in 2020) are in line with several of the published estimates of pandemic

excess mortality during the first period of the pandemic. The bulk of the India's pandemic excess mortality of course occurred during the huge second surge in April-June 2021.

Summary

First, all sources of data, including the SRS, agree that up to 2019 annual mortality was changing only slowly. Meanwhile, all agree that the level of completion was steadily increasing. There is agreement about the trends, but disagreement about the actual levels.

Most crucially, there is considerable evidence that the SRS estimates of death rates are too low, both nationally and often locally. We can infer this from NFHS registration data, from NFHS child mortality data, but also from the SRS and CRS reports themselves. This implies that official estimates of registration completion were exaggerated even before the pandemic period.

In fact, a variety of calculations suggest that prior to the pandemic India saw around 9.5 million deaths annually, considerably more than the 8.1 million government estimate. It is worth mentioning that the higher number also agrees broadly with [UN estimates of India's death rate](#) and other [independent estimates](#). By these estimates, in 2019, registration completion was 80% or slightly lower.

Against this backdrop, estimating total deaths during 2020 depends on finding out what happened to death registration. Reasonable scenarios (a 0-4 percentage point drop in death registration) give us 0.6-1.1 million excess deaths in 2020. This is in line with several studies of pandemic mortality, and entirely at odds with government claims about low mortality during the first phase of the pandemic.

|| [I]t is troubling that our understanding of mortality rates and death registration levels remains so poor.

Most studies agree that the second pandemic wave was considerably more deadly than the first. But there is still work to be done to evaluate the scale of the tragedy. "Official" civil registration data for 2021 has not yet been published, and many of the current estimates are based on partial and local data, often obtained by journalists. When official data is released, if it is to be used to update estimates of pandemic mortality, it will be crucial to look carefully at each region, consider pre-pandemic registration completion, and consider what happened to completion during the pandemic.

If we take yearly mortality to be around 9.5 million deaths, then the WHO estimate of 4.7 million pandemic excess deaths during 2020 and 2021 equates to around 6 months' worth of deaths in normal times. This amounts to a staggering mortality crisis. But it lies within the bounds obtained by a variety of other studies.

Finally, it is troubling that our understanding of mortality rates and levels of death registration remains so poor. The pandemic has focused attention on mortality statistics and registration completion, and this could provide an opportunity to redress this. Continued scientific attention and vigorous public discussion are key to moving forward our understanding of mortality in India.

Murad Banaji is a London-based mathematician. Aashish Gupta is a demographer at Harvard University. Vipul Paikra is a research fellow at the research institute for compassionate economics (r.i.c.e). A technical document containing more detail, data sources, and calculations used in this piece is available [here](#).

References:

Banaji, M., & Gupta, A. (2021). Estimates of pandemic excess mortality in India based on civil registration data. *medRxiv*.

Government of India (2022). Sample Registration System Statistical Report 2020. *Office of the Registrar General, New Delhi*.

Government of India. (2022). Vital statistics of India based on the civil registration system, 2020. *Vital Statistics Division Office of the Registrar General of India*

International Institute for Population Sciences (IIPS). (2017). National family health survey (NFHS-4), 2015–16: India. *Ministry of Health and Family Welfare, Government of India*

Knutson, V., Aleshin-Guendel, S., Karlinsky, A., Msemburi, W., & Wakefield, J. (2022). Estimating Global and Country-Specific Excess Mortality During the COVID-19 Pandemic. *arXiv preprint arXiv:2205.09081*.

Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K., Timæus, I. M., & Zaba, B. (2013). *Tools for demographic estimation*. International Union for the Scientific Study of Population.

Rao, C., John, A. J., Yadav, A. K., & Siraj, M. (2021). Subnational mortality estimates for India in 2019: a baseline for evaluating excess deaths due to the COVID-19 pandemic. *BMJ global health*, 6(11), e007399.

United Nations Department of Economic and Social Affairs, Population Division. (2022). *World Population Prospects 2022*. UN DESA/POP/2022/TR/NO. 3.