


The COVID-19 pandemic death toll in India: can we know better?

M Gupta,¹ Chalapati Rao ,² Arun Kumar Yadav,³ Munita Jat,¹ Rajinder K Dhamija,⁴ Nandita Saikia⁵

To cite: Gupta M, Rao C, Yadav AK, *et al*. The COVID-19 pandemic death toll in India: can we know better? *BMJ Glob Health* 2023;**8**:e012818. doi:10.1136/bmjgh-2023-012818

Handling editor Seye Abimbola

Received 11 May 2023

Accepted 11 August 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Alchemist Research and Data Analysis, Chandigarh, India

²Australian National University, Canberra, Australian Capital Territory, Australia

³Army Medical Corps, 2118 Fd Hospital, c/o 56 APO, India

⁴Institute of Human Behaviour and Allied Sciences, New Delhi, Delhi, India

⁵International Institute for Population Sciences, Mumbai, Maharashtra, India

Correspondence to

Dr Chalapati Rao;
chalapati.rao@anu.edu.au

BACKGROUND

Timely and accurate mortality statistics are essential for health policy, monitoring and evaluation of health programmes, and epidemiological research.¹ The COVID-19 pandemic brought into sharp focus the critical need for reliable mortality data to monitor its impact across time, place and person. Across the world, initial data compilations were conducted according to pandemic mortality surveillance guidelines issued by the WHO.² However, it was quickly realised that there were challenges in health sector reporting of COVID-19 deaths due to gaps in availability of confirmatory testing; variations in practices for ascribing deaths to COVID-19; and the diagnostic challenges for deaths that occur at home, outside the purview of the health sector. Hence, the surveillance community turned its attention to evaluate pandemic impact through the assessment of 'excess' mortality, defined as the increase in observed deaths in a defined time period, as compared with mortality expectations based on pre-pandemic trends. Such assessments require reliable population wide mortality data from Civil Registration and Vital Statistics systems, but such systems are not efficient in many countries. Hence, mathematical models were formed to derive both pre-pandemic mortality patterns as well as excess mortality attributable to the COVID-19 pandemic for such populations.³ The inherent limitation of these mathematical models are the assumptions on which calculations are based, as well as potential inaccuracies in primary data used as model inputs, which together create considerable uncertainty in estimated mortality patterns.⁴

In India too, the national government implemented mortality reporting according to the WHO pandemic surveillance guidelines, but these data were inadequate for the same reasons.⁵ In response to public interest in tracking the pandemic, various

SUMMARY BOX

- ⇒ The COVID-19 pandemic extracted a heavy mortality toll across the world during 2020 and 2021.
- ⇒ The pandemic toll in India was estimated by several statistical analyses that used a range of locally available mortality records for 2020–2021 from different parts of the country.
- ⇒ The validity of estimation methodology as well as results have been a matter of intense debate among scientists, policy analysts and civil society.
- ⇒ This article scrutinises the characteristics and quality of local data that were used for estimation, and assesses the potential for input data biases that could affect the validity of modelled estimates.
- ⇒ The official Civil Registration System (CRS) report for 2021 is scheduled for release during the second half of 2023, and would serve as an invaluable resource for pandemic mortality estimation.
- ⇒ We propose specific standards for demographic and event related variables in the 2023 CRS data release, as well as alternate methods to evaluate and correct for data biases.
- ⇒ The proposed analyses will improve the utility of death registration data for better pandemic mortality estimates for India, and provide recommendations to strengthen mortality data systems at national and subnational levels.

compilations of pandemic period mortality data were reported by independent investigators in national newspapers.⁶ This article evaluates the characteristics of pandemic period mortality data available from different sources for India, and potential biases in them that could impact accuracy of mortality models and excess mortality estimates. The article also suggests alternate statistical methods to estimate pandemic mortality in India, and makes recommendations to strengthen death registration.

PANDEMIC MORTALITY DATA COMPILATIONS

Table 1 describes various data sources that reported mortality compilations for India. The Civil Registration System (CRS) is the

Table 1 Input data sources and their characteristics which recorded deaths during COVID-19 pandemic in India

Input data source	Description	Populations covered	Data period	Potential data biases
Civil Registration System	Data obtained from state CRS department websites/provided through public information requests	Andhra Pradesh, Bihar, Gujarat Chhattisgarh, Haryana, Himachal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Rajasthan, Tamil Nadu, Uttar Pradesh, West Bengal	Monthly data from January 2018 till May 2021 (except) Gujarat—March–May 2021	<ul style="list-style-type: none"> ▶ Website data lower than deaths published in annual CRS reports for 2018–2020 in all states ▶ For 2021, not specified if monthly data are based on date of death or date of registration
News articles/blogs	Data from multiple sources local hospitals, cremation grounds, municipal corporations, local CRS offices	Ahmedabad, Bengaluru, Chennai, Hyderabad, Kolkata, Mumbai, Nagpur	Weekly, monthly and quarterly data from January 2018 till May 2021 (varying periods across cities)	<ul style="list-style-type: none"> ▶ City death data based on place of occurrence hence includes non-residents
Ministry of Health and Family Welfare Annual HMIS reports	HMIS 2020–2021 and 2021–2022—analytical report reports	All reporting hospitals	2018–2019, 2020–2021 and 2021–2022	<ul style="list-style-type: none"> ▶ Increase in reported deaths could potentially be due to improved institutional reporting compliance with government pandemic surveillance orders
C Voter Tracker Survey	Telephonic household survey on COVID-19 deaths from 1 July 2021 to 15 September 2021	Pan India Covered 15 700 households and 57 000 individuals No of deaths reported=2107	Jan 2019–June 2021	<ul style="list-style-type: none"> ▶ Very small sample size—not representative of Indian population ▶ Information bias—no evidence how much valid data was collected telephonically
Consumer Pyramid Household Survey (CPHS)	Household survey	Pan India covered 0.87 million individuals in 0.18 million households No. of deaths reported=8694	2015–2018, 2019–June 2021	<ul style="list-style-type: none"> ▶ Insufficient sample size ▶ Type of questions asked for capturing mortality is uncertain
Railway employees	Railways data on number of COVID-19 deaths	Pan India except mountainous areas 1.3million employees No of deaths reported=1952	March 2020–7 May 2021	<ul style="list-style-type: none"> ▶ Small sample of 1952 deaths ▶ Population characteristics and availability of medical facilities not representative for general population

CRS, Civil Registration System; HMIS, Health Management Information System.

most comprehensive mortality data source, and was established in 1969 under the Office of the Registrar General of India (ORGI), with registrars at state and local level.⁶ State registrars compile and submit data to the national ORGI after 1 year of the reference period, and aggregated deaths by age and sex for all 37 states and union territories are subsequently published in an official CRS report. The most recent report is for 2020, released in June 2022.⁷ Although the CRS has evolved over the past five decades with significantly improved data completeness at national level, the system performance has remained patchy across the country.⁸

The severity of the second wave of the COVID-19 pandemic in India during March to June 2021 prompted

an urgent need to estimate the mortality impact across India. Several independent investigators compiled local mortality data in different locations.⁹ For 14 states, data was procured either from state CRS office websites or through ‘Right to Information’ requests to the state registrar. The data comprised monthly numbers of reported deaths for varying periods between 2018 and 2021 across the states. In addition, investigators also compiled mortality data for several cities from local municipal corporations, local hospitals and crematoriums. These local data for states and cities were efficiently processed and disseminated by the investigators through a series of media reports, which served as timely initial evidence to gauge the mortality impact of the pandemic.⁹ These data were also used in

several mortality estimation studies.^{10–15} Although these state/city data compilations are also sourced from the same CRS death reporting processes, we refer to them as ‘local mortality records’, since they could potentially differ from the final reported data in the official annual CRS report, which include delayed registrations up to 12 months following the reference period. The characteristics and potential limitations of these ‘local mortality records’ are discussed in the following section.

The Health Management Information System (HMIS) compiles reports of deaths that have occurred in health facilities across the country. Annual HMIS reports for all states are published, and data is also available from the Ministry of Health and Family Welfare website.¹⁶ HMIS data for the years 2018–2021 were used by one research team in their excess mortality estimation methodology.¹¹ Prior to the pandemic, HMIS reporting was largely confined to public sector institutions, which limits use of the data to make predictions of population based expected mortality in 2020 and 2021. Also, not all the observed increase in HMIS reported deaths in 2020–2021 should be solely considered to be due to the pandemic, since increased attention to data reporting for pandemic surveillance could have improved private health facility reporting compliance.

The C Voter survey is a nationally representative telephonic sample household survey that regularly polls electoral preferences in the sample population. The survey conducted during July to September 2021 including a substudy of 57 000 people in 13 500 households who were asked to report all COVID-19 and non-COVID-19 deaths that had occurred in immediate family members during January 2019 to June 2021. A total of 2107 deaths were recorded to have occurred in the sample population during the reference period, which were used for analysis of excess mortality.¹¹ Essentially, the C Voter survey mortality data is based on self-reported events without any official verification of records. In addition, the small sample size also limits generalisability of these data in modelling pandemic mortality estimates for India.

The Consumer Pyramid Household Survey is a routine quarterly socioeconomic survey covering 0.87 million individuals from 170 000 households. During 2020–2021, each quarterly survey collected reports of household deaths that had occurred within the previous 3 months. A total of 8694 deaths were recorded in the sample population, and these data were used for mortality estimation by Anand *et al.*¹² Since this was an economic survey, its sample size was not sufficiently powered to derive measures for rare events such as deaths. Also, the survey report lacks information on the specific variables collected for each death, which limits the reliability of these data for modelling excess mortality.

Another primary data compilation used to estimate the pandemic mortality toll was the Indian Railways database which recorded 1952 deaths between March 2020 and May 2021 among a cohort of 1.3 million railway employees.¹⁷ Although a largely pan-Indian sample

(with the exception of mountainous areas), the relatively small sample size, the socioeconomic characteristics of railway employees, and their accessibility to health services and medical facilities limits the generalisability of the mortality experience of this sample to the national population.

LIMITATIONS OF LOCAL MORTALITY RECORDS

From the above description, it is evident that the CRS primary data is most appropriate of all to evaluate population level mortality patterns, given the legally mandated total population coverage which ensures adequate sample size for subnational mortality estimation. A recent review of analyses that reported COVID-19 mortality estimates for India identified that most of the studies used these ‘local mortality records’ as inputs for their statistical models.¹⁸ We compared these ‘local mortality records’ to the official CRS data contained for these 14 states in the annual CRS reports for 2018, 2019 and 2020 (see [table 2](#)).

As can be seen, the ‘local mortality records’ are relatively lower than the annual CRS reports for all states, indicating that these data were possible derived from ‘provisional’ databases maintained at local registration departments, rather than final tallies of verified events at the end of the official reporting period of 12 months from the date of death, submitted to the national ORGI for dissemination in annual CRS reports. Hence, forward projections of these under-reported data from ‘local mortality records’ for the prepandemic period would under-estimate expected deaths during the pandemic, and these lower baselines would tend to exaggerate the magnitude of estimated excess mortality.

From another perspective, even the data from annual CRS reports are biased due to incompleteness of death registration, which is estimated by the ORGI using crude death rates from the Sample Registration System (SRS), combined across both sexes as the evaluation standard for comparison.¹⁹ [Table 2](#) shows the ORGI estimate of completeness for each state for 2019, along with an alternate set of sex-specific completeness estimates derived from an age-stratified weighted analysis using age-specific death rates from the SRS and National Family Health Survey (NFHS) for comparison.²⁰ As can be seen, the age stratified weighted analysis shows that CRS completeness is lower than that inferred from the ORGI aggregated analysis. Another analysis by Saikia *et al* using self-reported responses regarding registration of reported deaths in the NFHS household sample found that only 70% of the reported deaths had been registered.²¹

On the other hand, CRS mortality data trends have shown a steady increase in data completeness over the past decade.¹⁹ Hence, any increase in reported deaths during 2020–2021 could not entirely be attributed to the pandemic, since there could be an element of increased reporting compliance, in keeping with improving CRS performance trends. The likelihood of improved reporting compliance is supported by the fact that the

Table 2 Comparisons of data from local mortality records* with data from annual CRS reports for 2018–2020 for 14 states of India

State	2018			2019			2020			CRS 2019 data completeness (%)	
	Local	CRS	% diff	Local	CRS	% diff	Local	CRS	% diff	ORGI est* (both sexes)	Alternate est† (M/F)
Andhra Pradesh	333 275	375 777	-11	363 414	401 472	-9	428 907	455 000	-6	100	94/85
Assam State	142 605	142 605	0	163 057	163 057	0	187 085	187 085	0	74	78/49
Bihar	182 926	213 989	-15	351 248	359 349	-2	387 521	425 047	-9	52	48/35
Gujarat	NA	433 256	NA	86 410‡	462 284	-	NA	523 892	NA	100	92/89
Haryana	178 634	185 842	-4	183 795	188 910	-3	198 903	212 238	-6	100	90/79
Himachal Pradesh	40 133	41 833	-4	40 970	43 633	-6	42 416	44 449	-5	86	91/79
Karnataka	368 000	483 511-	-24	NA	508 584	NA	544 000	551 808	-1	100	96/88
Kerala	255 594	258 530	-1	264 150	270 567	-2	252 421	250 983	1	100	98/97
Madhya Pradesh	407 172	424 257	-4	449 819	493 328	-9	461 057	524 454	-12	89	80/67
Maharashtra	426 489	667 900	-36	462 028	693 800	-33	565 185	808 783	-30	100	90/82
Rajasthan	216 370	443 173	-51	219 814	451 315	-51	229 564	477 151	-52	99	87/71
Tamil Nadu	536 192	574 006	-7	588 221	633 897	-7	644 291	687 212	-6	100	98/95
Uttar Pradesh	NA	906 653	NA	771 383	944 596	-18	791 485	873 419	-9	63	60/54
West Bengal	361 330	490 530	-26	431 531	551 695	-22	509 325	606 714	-16	100	89/81
Total	3 087 390	4 244 679	-27	3 172 926	4 670 196	-32	3 941 350	5 148 102	-23	92	81/72

*Source: Ministry of Home Affairs, Government of India.¹⁹
 †Source: Rao *et al.*²⁰
 ‡Data for Gujarat state was available only for only 3 months during 2019.
 §Sources: The Hindu, Jha *et al.*, Banaji and Gupta.^{9 11 13}
 CRS, Civil Registration System; ORGI, Office of the Registrar General of India.

national government had announced financial compensation for pandemic victims, for which proof in terms of death registration documents was necessary, and this resulted in a general enforcement of death registration protocols in all states.²²

On comparing mortality estimates from different studies, the highest was 4.7 million excess deaths in India during 2020–2021, from the WHO study.¹⁵ However, the review of all pandemic mortality studies for India noted that the methods did not satisfactorily account for various sources of bias¹⁸ These include the use of prepandemic completeness estimates to adjust ‘local mortality records’ from the pandemic period, without considering likelihood of improvements in reporting compliance during the pandemic. Other areas of uncertainty noted in the review included the limited generalisability of extrapolation of data from the selected reporting states to the whole country, as well as the limited plausibility in projection of data from the first few months of 2021 across the entire calendar year.¹⁸ All these sources of bias raise the likelihood of potential over estimation of the magnitude of pandemic mortality in India, and the need for a more nuanced approach to understand the actual impact.

IN SEARCH OF ROBUST PANDEMIC MORTALITY MEASURES

The characteristics of data sources displayed in tables 1 and 2 emphasise the importance of CRS data for

pandemic mortality measurement in India, despite challenges in quantifying and adjusting for varying levels of data completeness by sex, age and location. As per standard practice, the official CRS report for 2021 is expected to be released sometime during the second half of 2023, which will include all delayed registrations that may have resulted from pandemic related disruptions. As mentioned earlier, the government monetary support scheme for COVID-19 related deaths could have incentivised delayed death registration for both 2020 and 2021, thus improving the prospects of the availability of more reliable data to measure pandemic mortality.²² It will be necessary to conduct a detailed analysis of completeness of the 2020/2021 CRS data and make relevant adjustments as required, to improve the potential accuracy of excess mortality estimates.

Despite these prospects, however, there are some background aspects of CRS data from official reports, that serve as constraints for detailed mortality analysis, as follows:

1. CRS forms record both ‘place of usual residence’ of the deceased and ‘place of occurrence’ of the death, but the CRS report provides data only by place of occurrence.¹⁹ Data should be made available from both aspects, since statistics by place of usual residence eliminate distortions arising from temporary migration and help understand disease exposure patterns,

while analysis by place of occurrence could guide the planning of critical care services.

- CRS provides adult mortality data according to 10-year age groups (15–24, 25–34, ... 55–64, 65–69) with a terminal category of 70 years and above. In view of the known associations between older ages and COVID-19 mortality as well as non-communicable diseases it is desirable that CRS data are available in standard demographic age categories (0–1, 1–4, 5–9, 10–14, ... 80–84, 85+) to enable in-depth understanding of mortality patterns in the elderly.
- Annual CRS reports include a summary table of delayed registrations aggregated for all previous years, and for both sexes and all ages for each state. However, given the time sensitivity of pandemic mortality metrics, it is desirable that data on delayed registrations are made available for each state by sex, age and year of death occurrence for 2019 onwards. This will facilitate their appropriate integration with previously released annual data, and enable improved time-specific analysis.
- Owing to irregular data submission from some states, there are variations in inclusion of data from states in different years, which hinders subnational mortality trend analysis.
- Finally, although there have been improvements in CRS performance, there still remains the potential for under-reporting during the pandemic period. Such gaps in completeness would preferably require detailed data from the SRS for 2021, as a comparator for evaluation.¹⁹ In addition to completeness estimation methods discussed earlier, record linkage between individual records from SRS and CRS for the SRS sample sites could be used with ‘capture-recapture’ analysis to derive the Chandrasekhar-Deming completeness factor, as direct evidence of the magnitude of such bias.²³ Such record linkage analysis could also be attempted with individual records from other local mortality sources including surveys, HMIS and other administrative databases, ensuring that geographic and time references are of the same scope as CRS records.

CONCLUSIONS AND THE WAY FORWARD

Reliable measurement of the mortality impact of the COVID-19 pandemic in India is essential to understand the epidemiological profile of the disease, the health status of the population and the priorities for health system strengthening into the future.²⁴ The forthcoming annual CRS report for 2021 holds much promise for providing the best possible primary evidence for such mortality measurement, which will reduce uncertainty from current estimates. We propose that the national public health leadership should convene a Task Force of national experts in this field who could be provided access to all the microdata from the CRS, SRS and other relevant data sources as necessary. It is anticipated that the Task Force would then be sufficiently equipped with

all the available and required information to conduct a thorough analysis using standard statistical methods that use empirical data, to provide the required mortality measures at national, state and district level. The activities and outputs of such a detailed analysis would have both immediate and long-term influence on the performance of the death registration system, towards strengthening evidence-based health monitoring and policy action for India.

Twitter M Gupta @DrMamtaGupta7, Chalapati Rao @ChalapatiRao13, Munita Jat @Munitajat05 and Nandita Saikia @nanditajnu

Contributors MG and CR conceptualised the manuscript, and MG prepared the first draft. All authors collaborated in sourcing relevant information, interpretation of findings and revisions to prepare the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Chalapati Rao <http://orcid.org/0000-0002-9554-0581>

REFERENCES

- Oxford University Press. *World development report 1993: investing in health*. New York, 1993.
- World Health Organization. *Public health surveillance for COVID-19: Interim guidance*. Geneva, Switzerland, 2022. Available: <https://www.who.int/publications/i/item/WHO-2019-nCoV-SurveillanceGuidance-2022.2>
- World Health Organization. Methods for estimating the excess mortality associated with the COVID-19 pandemic [Internet]. n.d. Available: <https://www.who.int/publications/m/item/methods-for-estimating-the-excess-mortality-associated-with-the-covid-19-pandemic>
- Ioannidis JPA, Zonta F, Levitt M. Flaws and uncertainties in pandemic global excess death calculations. *Eur J Clin Invest* 2023;53:e14008.
- Government of India. India fights Corona COVID-19 New Delhi. 2021. Available: <https://www.mygov.in/covid-19/>
- Gupta M, Rao C, Lakshmi P, et al. Estimating mortality using data from civil registration: a cross-sectional study in India. *Bull World Health Organ* 2016;94:10–21.
- Vital Statistics Division. Vital statistics of India based on the civil registration system 2020. office of the registrar general and census Commissioner, Ministry of home affairs, government of India. New Delhi; 2022. Available: https://crsorgi.gov.in/web/uploads/download/CRS_report_2020.pdf
- Rao C, Gupta M. The civil registration system is a potentially viable data source for reliable Subnational mortality measurement in India. *BMJ Glob Health* 2020;5:e002586.
- The Hindu. Coronavirus: spike in excess deaths during pandemic. 2021. Available: <https://www.thehindu.com/specials/coronavirus/spike-in-excess-deaths-during-pandemic/article34876318.ece>
- Leffler CT, Lykins V JD, Das S, et al. Preliminary analysis of excess mortality in India during the COVID-19 pandemic. *Am J Trop Med Hyg* 2022;106:1507–10.

- 11 Jha P, Deshmukh Y, Tumbe C, *et al*. COVID mortality in India: national survey data and health facility deaths. *Science* 2022;375:667–71.
- 12 Anand A, Justin S and Subramanian A. Three new estimates of India's all-cause excess mortality during the COVID-19 pandemic [CGD Working Paper 589]. Washington, DC: Center for Global Development; 2021. Available: <https://cgdev.org/publication/three-new-estimates-indias-all-cause-excess-mortality-during-covid-19-pandemic>
- 13 Banaji M, Gupta A. Estimates of pandemic excess mortality in India based on civil registration data. *PLOS Glob Public Health* 2022;2:e0000803.
- 14 Wang H, Paulson KR, Pease SA, *et al*. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *Lancet* 2020.
- 15 Msemburi W, Karlinsky A, Knutson V, *et al*. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* 2023;613:130–7.
- 16 Ministry of Health and Family Welfare. Health management information system. government of India, New Delhi. 2021. Available: <https://hmis.nhp.gov.in/#/standardReports>
- 17 Guilмото CZ. An alternative estimation of the death toll of the COVID-19 pandemic in India. *PLoS One* 2022;17:e0263187.
- 18 Kumar Yadav A, Ahmed T, Dumka N, *et al*. A systematic review of excess all-cause mortality estimation studies in India during COVID-19 pandemic. *Med J Armed Forces India* 15, 2023.
- 19 Ministry of Home Affairs, Government of India. *Vital Statistics Division Office of the Registrar General of India. Vital Statistics of India based on the Civil Registration System, 2019*. New Delhi, 2021. Available: https://censusindia.gov.in/2011-Common/Annual_Report.html
- 20 Rao C, John AJ, Yadav AK, *et al*. Subnational mortality estimates for India in 2019: a baseline for evaluating excess deaths due to the COVID-19 pandemic. *BMJ Glob Health* 2021;6:e007399.
- 21 Saikia N, Kumar K, Das B. Death registration coverage 2019–2021, India. *Bull World Health Organ* 2023;101:102–10.
- 22 National Disaster Management Authority. *Guidelines for Ex Gratia Assistance to next of kin of the Deceased by Covid-19*. Government of India. New Delhi, 2021. Available: <https://ndma.gov.in/sites/default/files/PDF/Final-Guidelines-on-Ex-Gratia-Assistance-to-Covid-19-Victims.pdf>
- 23 Sekar CC, Deming WE. On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association* 1949;44:101–15.
- 24 Rao C. Medical certification of cause of death for COVID-19. *Bull World Health Organ* 2020;98:298–298A.