Meeting Report: Why do we need data science and statistics to count deaths during a pandemic

Shahid Jameel and Vijay Chandru

“Any doctor who is called to a house to treat a severely wounded person or one suffering from unwholesome food or drink shall report the fact to the gopa and the sthanika. If he makes a report, he shall not be accused of any crime; if he does not, he shall be charged of the same offence (which he helped to conceal).”
– Kautilya’s Arthashastra (350-275 BCE)

The Indian Academy of Sciences organized a session on “Why do we need data science and statistics to count deaths during a pandemic” at its 87th Annual Meeting in November 2021 [1]. It was motivated by a lack of data transparency in India, where over 900 scientists signed a petition petition [2] to the Indian Prime Minister in April 2021 demanding access to granular pandemic data available with the government. Despite assurances and several attempts, senior researchers lament about continued data opacity. However, striking volunteer efforts [3] continue to provide access to federated health data for research.

Early in the pandemic, several of the epidemiological models used in India were anchored on the basis of official mortality figures. These were considered more reliable compared to infection rates, which depended on test volume and veracity limited at best to major cities. The first wave of infections also seemed largely focused in urban centres and hence this assumption appeared reasonable.

India suffered greatly during its second wave of COVID-19, yet there were about 400,000 reported deaths by the time it subsided in July 2021 [4]. This amounted to 0.29 per thousand, far lower than several other countries – USA, 1.8; Brazil; Italy and UK, >2; Mexico, >3; and Peru, >5. This despite India’s infection rate at the same time reported to be 67.6% by the government’s 4th National Serosurvey [5].

Pandemics have happened with regularity across human history. Despite great leaps in science and technology, and information flow, the human response continues to include, denial, distorted facts, politics and lack of equity. Countries under colonial occupation suffered greatly in the 1918 Flu. Western Samoa lost 22% of its population and India 6% - the latter being far more populous lost an estimated 18 million people [6]. Yet, the images that remind us of that pandemic come largely from the West. In 2021 science delivered vaccines at amazing speed, but access and equity remain key challenges. On average less than 10% people in low-income countries have received one vaccine dose compared to over 80% in high-income countries [7].

The session included three short talks and a panel discussion.

Arvind Subramanian (Brown University, USA), a former Chief Economic Advisor to the Government of India used data from three sources to calculate all-cause mortality. This included States' Civil Registration Systems (CRS), Indian seroprevalence plus international infection fatality rate (IFR), and a Consumer Pyramid Household Survey (CPHS). His results showed 3.5 to 5 million deaths against the official count of about 400,000, i.e., about 8 to 10-fold possible undercounting of COVID-19 deaths in India [6].

Prabhat Jha (University of Toronto, Canada) quantified all-cause excess mortality in India by comparing deaths during the first (July-December 2020) and second (April-June 2021) wave.

waves of COVID-19 to deaths in 2015-19 from three sources - CRS mortality reports covering 37% of India's population; deaths in 0.2 million health facilities; and a COVID deaths survey of 0.14 million adults. He also concluded that India had 2.7-3.4 million estimated excess deaths, which is about 6-8 times the official figures, and that while 2020 deaths were largely urban, those in 2021 were both rural and urban [7].

S. Rukmini, an independent Indian data journalist from Chennai, India shared undercounting all deaths and low rates of medical certification as key to scepticism of the official mortality figures. The CRS data tallied with data from crematoria and burial grounds showed over 12,000 excess COVID-19 deaths in 2020 over the official 4,000 deaths in Chennai city; the 2021 mismatch was much larger. The mismatch in two Indian states – Madhya Pradesh and Andhra Pradesh were 4.2 and 3.4 times, respectively [8].

In the discussion that followed, Manindra Agrawal (IIT Kanpur, India), an author of the government’s SUTRA model of COVID-19 spoke about the CRS mortality data from India’s most populous state Uttar Pradesh (UP). He suggested waiting for CRS 2021 data before coming to any conclusions and believed that additional mortality estimates would be much lower. Murad Banji (Middlesex University, London) spoke of the undercounting of mortality in rural India due to structural reasons and brought out early indications of mismatch between epidemiological data, serosurveys, official death counts and significant variation between the states' mortality data. Madhuchhanda Bhattacherjee (University of Hyderabad, India) addressed the interpolation of data gaps and how probabilistic modelling can be used for adjusted cumulative fatality rates. She also provided insights from statistical modelling that warn of continuing spikes of infection on the path to endemicity.

Further discussions revolved around several connected issues as follows

i. Role of competitive federalism that incentivised improvement in data collection, analysis and reporting by states on CRS, and to guard against centralization.

ii. In a pandemic even quick and crude estimates are very useful if communicated quickly.

iii. Need for tight protocols for reporting of deaths from around 5,000 medical facilities across the country, better sampling and accounting for rural deaths.

iv. Need for “statistical independence” for surveys in India with autonomy from the executive.

v. The trade-off between privacy and “sunlight” in publishing of data which could improve transparency.

vi. Media reporting and social media needs to be improved.

vii. Timely reporting of data is critical. Provisional CRS data be put out instead of waiting for annual reports.

viii. Harmonisation in statistical methodologies, benchmarking, training of data generation agencies across the nation.

ix. Going forward, India could organize cohorts of citizens profiled for vaccination status, demographics and monitored serological immunity levels to address questions on impact of vaccination on IFRs.

The speakers agreed that the actual death rate in India would be around 1.8-3 per thousand instead of 0.29, making it consistent with the reported high seropositivity, and like countries with poor health infrastructure. This calls for robust preparedness for any future waves, such as the new threat from the Omicron variant. India paid a heavy price for not having good real-time data on deaths, especially during the first wave that led to complacency and a terrible toll in the second wave. They also agreed that there is no substitute for granular data that is reliable, timely and official, and strongly recommended autonomy to agencies for
providing easy access to data from surveys as public goods. The recommendations were to rely on science and the scientific method – to use all available sources, be transparent on assumptions and methodology, understand biases and limitations, and to be careful about conclusions. More eyes looking at data are better than fewer.

References:

1. Why do we need data science and statistics to count deaths during a pandemic. YouTube Link https://www.youtube.com/watch?v=WR5vQ5WzLJ8 (IASc Channel)
2. https://sites.google.com/view/corona-appeal/home

Author details:

Shahid Jameel is a Fellow at OCIS and Green Templeton College, University of Oxford, UK, and a Visiting Professor at Ashoka University, India. He is a former Chair of the Scientific Advisory Committee of the Indian SARS-CoV-2 Consortium on Genomics (INSACOG).

Vijay Chandru is co-Founder, Strand Life Sciences and Adjunct Professor at the Indian Institute of Science, Bengaluru where he teaches computational epidemiology. He is also a Commissioner of The Lancet Citizen’s Commission on Reimagining India's Health System.

Statement of Competing Interest:

Both authors are Fellows of the Indian Academy of Sciences, Bengaluru, India.