

LEAD

# Everyone is afraid of data

**Sonalde Desai**

FEBRUARY 12, 2019 00:02 IST

**UPDATED:** FEBRUARY 12, 2019 00:34 IST

**There needs to be robust infrastructure for official statistics so that governments do not suppress inconvenient truths**

**O**ver the past two weeks, headlines have focussed on declining **employment** between 2011-12 and 2016-17; loss of jobs under the National Democratic Alliance government, particularly post-demonetisation; and the government's refusal to release a report using the Periodic Labour Force Survey (PLFS) documenting this decline, leading to resignations of two members of the National Statistical Commission. In a pre-election, politically charged environment, it makes for eye-catching headlines.

**Five trends**

Let us step back from this episode and recall similar controversies over official data in the past. Past experiences tell us five things.

First, suppression of results seems to be a problem common to all political parties. Census 2011 data on religious distribution of the population was not released until 2015. It is widely believed that these data were ready before the 2014 election, but the United Progressive Alliance government was worried about inciting passions around differential population growth between Hindus and Muslims and chose not to release the tables. Similarly, the UNICEF conducted the Rapid Survey on Children (RSOC) 2013-14 on behalf of the Ministry of Women and Child Development but the report was held up by the new government, allegedly due to the fear that it showed Gujarat in poor light. Sometimes these concerns lead to lack of investment in data collection itself, as is the case with the National Sample Survey, or the NSS's Employment-Unemployment surveys (not conducted since 2011-12), forcing public policy to rely on non-comparable statistics from other sources such as the data from the Employees' Provident Fund Organisation (EPFO). These episodes are likely to recur, and hence, we need a more comprehensive strategy for dealing with them.

Second, the fear of having statistical reports misquoted is legitimate. We live in a world where appetite for news is incessant and the news cycle is very short. Statistics that don't always lend themselves to rapid unpacking into sound bites and headlines are easily misinterpreted.

When Census 2001 results on religion were released, in September 2004, a newspaper led with a story that although the Hindu growth rate between 1981-1991 and 1991-2001 had declined from 25.1% to 20.3%, that for Muslims had gone up from 34.5% to 36%. Media reports paid little attention to the actual report that highlighted that the 1991 Census was not conducted in Jammu and Kashmir and that after adjusting for it, growth rates for both Hindus and Muslims had declined. When the mistake was discovered, it was blamed on the then Registrar General and Census Commissioner, J.K. Banthia, a highly competent demographer. He was sent into bureaucratic exile while the news media moved on to a new story.

Third, it is impossible to bottle up the genie once data are collected and reports prepared. In a world dominated by WikiLeaks, suppressing reports seem to create an even bigger problem, since it allows individuals with exclusive access to act as the interpreters for others. In the instance of the RSOC mentioned above, suppression of the report, coupled with leaked data encouraged speculation by *The Economist* (July 2015) that the data were being suppressed because Gujarat must have fared poorly on reducing malnutrition. It stated that Bihar had made much greater progress since the proportion of children who go hungry had been cut from 56% to 37% between 2005-6 and 2012-13, while the decline was much smaller in Gujarat, from 44.6% to 33.5%.

Fourth, sometimes leaked results create speculation that is far worse than full disclosure would warrant. The Economist cherry-picked its comparisons. Nutritional status is measured by weight-for-age (underweight) and height-for-age (stunting). The final report showed that about 41.6% of children in Gujarat were stunted (had low height for their age), This is higher than the nationwide average of 38.7%. However, improvement in stunting in Gujarat between 2005-6 and 2013-14 was of similar and slightly higher magnitude as that for the nation as a whole: 10.1 versus 9.7 percentage points. Moreover, stunting decline in Gujarat was greater than that in Bihar, 10.1 percentage points as opposed to 6 percentage points. Usually statistics on underweight and stunting should provide a similar picture; when they do not, greater care is required in interpretation. This was not possible because only *The Economist* seemed to have access to the report and led the headlines.

## The employment picture

Fifth, statistics often deal with complicated reality and require thoughtful analysis instead of the bare bones reporting contained in typical government reports. The headline in *Business Standard* on February 3 based on the leaked PLFS report claims that more than half the population is out of labour force; however, the statistics they present show that the trend is dominated by women and the rural population. If the full report were available, I think it would be rural women who would drive the employment story. This is very much a continuation of the trend between 2004-5 and 2011-12 documented by the NSS, under a different government.

Between 2004-5 and 2011-12, work participation rates for rural women of working ages (25-64) fell from 57% to 43%. However, much of this decline was in women working on family farms and in family businesses, from 42% to 27%; decline in wage work was much smaller, from 24% to 21%. If lower engagement of women with family-based activities such as farming, rearing livestock or engaging in petty businesses drives the decline in employment, we may need to look at declining farm sizes and increasing mechanisation as the drivers of this decline. One can blame the government for not creating more salaried jobs for women pushed out of farming and related activities, but it would be hard to blame it for eliminating jobs.

If the full report and unit level data for the PLFS were available, it is possible that we will find a continuation of the trend that started in 2004-5. This is not to say that demonetisation may not have had a negative impact, particularly in urban India, where *Business Standard* reports that employment fell from 49.3% to 47.6%, but this is a much smaller decline. It is also important to note that the urban comparison between the NSS and the PLFS requires caution, particularly for unemployment figures. Whereas the NSS contains independent cross-sectional samples for each sub-round, the PLFS includes a panel component in urban areas

where the same households are re-interviewed every quarter. Since it would be easier to find unemployed individuals than employed individuals for interview, attrition adjustment is necessary before drawing any conclusions. Without access to the full report, it is difficult to tell whether attrition adjustment was undertaken.

So how do we get out of this vicious cycle where fear of misinterpretation leads to suppression of data, which in turn fuels speculation and suspicion and ultimately results in our inability to design and evaluate good policies? The only solution is to recognise that we need more openness about data coupled with deeper analysis, allowing us to draw informed and balanced conclusions. The onus for this squarely lies with the government. Simply placing basic reports in the public domain is not sufficient, particularly in a news cycle where many journalists are in a hurry to file their stories and cherry-pick results to create headlines.

### **Spread the net wider**

Understaffed and underfunded statistical services cannot possibly have sufficient domain expertise to undertake substantively informed analyses in all the areas for which statistical data are required. A better way of building a robust data infrastructure may be to ensure that each major data collection activity is augmented by an analytical component led by domain experts, recruited from diverse sources, including academia.

*Sonalde Desai is Professor of Sociology at University of Maryland, and Professor and Centre Director, NCAER-National Data Innovation Centre. The views expressed are personal.*