

# *An avoidable controversy over sample surveys*

**T**here was controversy recently over how sound data collection procedures in India are, especially when it comes to some of the important national level surveys. An article (by Shamika Ravi, Member of the Economic Advisory Council to the Prime Minister), which was published in a leading national daily on July 7, 2023, cast doubts on the soundness of data collection procedures of some of the surveys such as the National Sample Survey (NSS), National Family Health Survey (NFHS) and Periodic Labour Force Survey (PLFS). The writer proceeded to suggest ‘a major sampling overhaul’ so that the survey estimates reflect the ground reality in the country. In support of her argument, the writer said that from 2011-12 till 2019-21, out of 11 surveys listed in her article, ‘every survey (except NFHS-4 of 2015-16) underestimates the proportion of the urban population or overestimates the rural population significantly’. Consequently, the estimates from these surveys ‘systematically underestimate the improvements across the country’.

The publication of this article led to articles countering her point of view.

It is important to re-emphasise the point that all the surveys listed do adopt scientific sample designs. This is widely acknowledged, even at the international level. However, there can be no denying the fact that there is always a scope for improvement of the sample designs. In fact, the sample designs of the NSS have undergone periodic revisions after due deliberations in the meetings of the NSS round-specific working groups, with final approval by the National Statistical Commission (and by the governing council of the National Sample Survey Office earlier). These committees and bodies have been chaired by or have had as their members some of the most eminent economists, statisticians and demographers that India has ever produced.

## **Bias in population estimation**

On the issue of an underestimation of the proportion of the urban population or an overestimation of the proportion of rural population in surveys (as pointed out by the writer in her article on July 7), it is important to remember that the sample designs of the NSS or the PLFS are not aimed at estimating the number of households or population. Instead, they are meant primarily to estimate the major



**G.C. Manna**

is a professor at the Institute for Human Development (IHD), New Delhi. He was earlier the Director General of the Central Statistics Office and the National Sample Survey Office. He was also a Member of the National Statistical Commission

The allegation that many of India’s official nation-wide surveys are based on unsound data collection frameworks is misplaced

socio-economic indicators that relate to the subjects of interest. The estimates of the number of households or population are auxiliary information. Data users appropriately adjust the survey-based estimates separately for the rural and urban areas by using projected population figures based on the Census. Even if this sounds repetitive, it must be mentioned that the rates and ratios relating to the major survey characteristics for rural or urban areas based on the NSS, broadly reflect the ground realities. This fact has been acknowledged by the National Statistical Commission (Rangarajan, 2001).

Nevertheless, the fact of underestimation of population has been a perennial problem in the NSS. What is worrying is that the extent of underestimation, particularly for the urban areas, is quite significant for which remedial/corrective measures are necessary. In this context, it is worth mentioning that as against the estimated population, estimates of the number of households based on the NSS are in close agreement with the Census-based number of households. The corollary is that even if no adjustments are carried out, the average level of performance for rural and urban areas combined (based on these surveys) should be fairly reliable as far as household level indicators are concerned.

The allegation by the writer on samples based on these surveys not being representative in view of their use of outdated sampling frames also loses its relevance significantly because: first, these surveys primarily depend on the population census list of villages and towns/urban blocks (available once in 10 years only) for sampling purpose which, in any case, is complete in coverage. And second, for sampling of urban blocks, the NSS and PLFS use the latest list of Urban Frame Survey (UFS) blocks (i.e., the counterpart of the list of urban census enumeration blocks) covering all the towns of the country – this partially corrects the frame for urbanisation taking place subsequent to the census by way of State government notifications. On the issue of rural-urban classification of geographical areas, all these surveys treat census towns as part of the urban sampling frame.

## **Systematic bias in response rate**

The writer rightly mentions that a refusal to part with the information is ‘never random’ and the

response rate falls with the growth in the level of income of households. This is a problem also faced by similar surveys internationally. However, the survey methodology prescribes for substitution of such households by more or less similar households to the extent feasible. Here, of course, the possibility of relatively lower levels of income of the substituted households cannot be ruled out causing some downward bias in the overall estimates which are closely associated with the level of income. However, given that a majority of the welfare programmes of the government are targeted towards households in the lower income brackets, the said problem of non-response – with a very low non-response rate in these surveys – is not likely to have a serious impact on the overall household level indicators of interest, as estimated through these surveys.

## **Room for improvement**

Sample design and the data quality are two different components of a survey. Both are important. When it comes to sample design, a lot of care is taken generally by adopting a scientific sample design in these surveys. However, on the issue of sampling frame, given the apprehension over inadequate representation of rich households, it may be worth exploring whether a list of such households can be developed by tapping alternative sources and covering a representative sample of them in conjunction with the traditional survey of the residual population.

Further, it may be worth examining the coverage of the UFS frame, given the extent of underestimation of the urban population. Setting up a methodological study unit to undertake other similar studies oriented towards improvement of the survey design may also be a step in the right direction.

The aspect of training of field personnel, field inspection, concurrent data validation and publicity measures may be strengthened to improve the quality of primary data, which is most crucial in any survey.

Finally, while there is always scope for the improvement of the survey results, criticising all large-scale official surveys on the ground/basis that they do not capture improvements adequately is akin to throwing the baby out with the bathwater.