

Infirmities in NSSO Data for Nagaland

ANKUSH AGRAWAL, VIKAS KUMAR

The samples of the National Sample Survey Office are unlikely to be representative of Nagaland. This is so not because of the size of the sample, but because of the arbitrary restriction to villages within 5 kms of bus routes, whereas the bulk of the rural and tribal population is located farther from the roads. Further, it is argued that the Census of India data, which is used as the sampling frame, is found to be unreliable for Nagaland.

Most national-level household surveys in India do not cover the smaller states of north-eastern India. For instance, most states in the region were not covered in various waves of Rural Economic and Demographic Surveys. Some national surveys cover these states irregularly. The District Level Household and Facility Survey did not cover Nagaland in 2007-08, for instance. In fact, even the National Sample Surveys – one of the most comprehensive, reliable and widely used sources of information on Indian households – that cover the north-eastern states of India regularly do not have sufficiently representative samples to generate reliable estimates for all the states of the region. The *Report of the Committee on Optimum Sample Sizes for North Eastern States* discussed the problem of inadequate sample sizes for the region and related administrative issues (GoI 2011). The report, however, does not deal with other factors that could affect the representativeness of samples. This note draws attention to systematic biases, which are independent of the problem of sample size in the National Sample Survey Office (NSSO) samples for Nagaland.

Biased Sample

Even six decades after it was founded, the surveys conducted by the NSSO exclude “interior villages of Nagaland situated beyond five km of any bus route” (GoI 2013a: D1, also see GoI 1996: 1). Since Nagaland is neither the only insurgency-affected state nor the only state with a difficult terrain,¹ it is not clear why NSSO has imposed this distance-based restriction only in case of Nagaland.² In any case, as shown in Table 1, NSSO over-samples the villages within

5 kms of bus routes in Nagaland to ensure that it meets its state-level requirement for sample size. Note that compared to north-east India and the country as a whole, Nagaland has a much smaller proportion of villages in the sampling frame even though a much higher proportion of the villages in the sampling frame are sampled.

The arbitrary restriction applied in Nagaland skews NSSO estimates of socio-economic characteristics like the monthly per capita consumption expenditure (MPCE). There are a number of reasons for this. First, in Nagaland a great majority of villages are located more than 10 kms away from roads, while the NSSO surveys are restricted arbitrarily to villages within 5 kms of bus routes. Since bus routes are a subset of roads, NSSO is not even covering all villages within 5 kms of roads.

Furthermore, the bus network is not uniformly distributed across the state, with the network being denser in areas closer to the Assam border compared to the areas closer to the Myanmar border. This results in overestimation of average MPCE as households in villages near roads are likely to have a better standard of living. Second, the average MPCE statistic is biased also because the districts closer to the Myanmar border not only have a higher share of rural population, but also a higher share of remote villages.

In addition, districts closer to the Myanmar border are relatively socio-economically backward compared to

Table 1: Sample Size for Different Rounds of NSSO Surveys (2004-05 to 2010-11)

Description	Nagaland	North-east	India
Total number of villages	1,317	42,258	6,38,596
Villages in NSSO frame	371	41,149	6,36,127
Proportion of villages in the frame	28.17	97.38	99.61
Villages sampled in different rounds			
61st	96	1,276	8,124
62nd	47	649	4,847
63rd	32	696	5,601
64th	128	1,136	7,984
65th	80	1,248	8,188
66th	88	1,156	7,508
67th	104	1,008	8,380
Average proportion of villages sampled to that in the frame	22.14	2.49	1.14
Average proportion of villages sampled to total villages	6.24	2.42	1.13

Source: GoI (2011).

The authors are thankful to their affiliate institutes for support.

Ankush Agrawal (ankush@hss.iitd.ac.in) is with the IIT Delhi and Vikas Kumar (vikasprithvipur@gmail.com) is with the Azim Premji University, Bangalore.

districts closer to the Assam border. Third, the share of tribes in rural population increases with distance from the nearest road. Moreover, there is substantial heterogeneity among tribes in Nagaland and different tribes are concentrated in different parts of the state. NSSO's restriction implies that the backward tribes which inhabit areas relatively closer to the Myanmar border are less likely to be included in the sample compared to the advanced tribes, which are based in areas closer to the Assam border where both road density and level of urbanisation are higher.

Data from the 1981 Census, the most reliable census conducted in the state between 1981 and 2001 (Agrawal and Kumar 2012), support the above claims. First, about 88% villages in the circles bordering Myanmar were located at a distance of more than 10 kms from the nearest bus stop compared to the state average of 54%. Second, average literacy rate in the circles on the Myanmar border was 16% compared to 44% in the circles on Assam border. Third, the share of tribes in the total population was 85% in the circles on the Myanmar border compared to 25% in the circles on the Assam border (GOI 1984). Also, note that the overall MPCOE of the state will be overestimated because rural areas are generally poorer than the urban areas and Nagaland is a largely rural state, whereas NSSO over-samples villages that are relatively better off. In 1981, rural areas accounted for 85% (it was 71% in 2011) of the state's population.

In short, NSSO's sample for Nagaland is not representative because of inadequate coverage of population and presence of selection biases in the sampling process.³ The NSSO sample for Nagaland will not allow reliable estimation of population characteristics. Note that representativeness would not have been affected due to exclusion of villages lying beyond 5 kms of bus routes if the population of the state was homogeneous. However, as argued earlier the population is heterogeneous. Also, the extent of bias is not uniform across NSSO survey years because of changing migration trends and growing density of road network so that estimates of

socio-economic characteristics are not comparable over time.

Sampling Frame

While the arbitrary distance-based restriction biases the sample for Nagaland, another factor compounds the problem. The sampling design of household surveys relies heavily on population censuses. But census results have been shown to be incorrect in case of Nagaland for the past few decades because of manipulation driven by economic and political considerations. More specifically, it has been shown that the discrepancy in headcount varies systematically between hill and plain districts and rural and urban areas (Agrawal and Kumar 2012) and bears systematic association with, among other things, the distance from the nearest town (Kumar and Agrawal forthcoming). In other words, a sampling frame based on the census, especially using 1991 or 2001 Census results, will systematically over-represent certain population subgroups relative to the others. An incorrect sampling frame will affect representativeness of samples insofar as it does not contain the population characteristics of interest in proper proportions.

A few points about the Census of Nagaland, which highlight its unreliability as a sampling frame, are mentioned here. First, while the problem of manipulation of census persisted between 1981 and 2001 (Agrawal and Kumar 2012), the uncritical use of the faulty census estimates for Nagaland continues long after they were rejected by the state government in 2005. The most recent example of uncritical use of the 2001 Census data for Nagaland can be found in the *Report of the Committee for Evolving a Composite Development Index of States* (GOI 2013b). Second, manipulation of headcount resulted in systematic errors in the demographic data for Nagaland. For instance, between 1981 and 2001, the census figures for the overall population of the state and the population of hill districts and tribes almost always exceeded their respective true values. The accumulated error in the overall population reached 25% in 2001 (Agrawal and Kumar 2012). In fact, the census

population of certain hill districts was twice the true population. Third, the degree of manipulation was higher in rural and inaccessible areas, in areas closer to Myanmar border, and in areas with higher tribal population share. In short, in case of Nagaland, the census has been an unreliable sampling frame.

Concluding Remarks

Smaller sample size has often been blamed for unrepresentative samples in the north-eastern states (GOI 2011, for instance). But, there are other factors which are also crucial for sampling design to which this article has drawn attention. We have argued that the NSSO samples are unlikely to be representative for Nagaland because of the arbitrary restriction to villages within 5 kms of bus routes, whereas the bulk of the rural and tribal population is located farther from roads and the share of the population more than 10 kms from road varies both across districts and across censuses. We further argued that census is not a reliable sampling frame in case of Nagaland.

The problem highlighted here has wider implications. First, it raises doubts about the reliability of statistics generated from sample surveys in other conflict-ridden and inaccessible areas. For instance, it is not clear to what extent NSSO is able to reach villages in other conflict zones, where the reach of the government may be limited. Second, while we have only examined NSSO samples, other demographic and socio-economic surveys conducted in Nagaland are also likely to suffer from similar problems. Third, if federal redistribution is linked to the level of poverty measured using

Permission for Reproduction of Articles Published in EPW

No article published in EPW or part thereof should be reproduced in any form without prior permission of the author(s).

A soft/hard copy of the author(s)'s approval should be sent to EPW.

In cases where the email address of the author has not been published along with the articles, EPW can be contacted for help.

NSSO's MPCE data (as in GoI 2013b, for instance), then Nagaland could suffer loss of transfers from the centre because NSSO probably overestimates MPCE for the state.

NOTES

- 1 Difficult terrain affects estimates of the characteristics where seasonality plays an important role. NSSO divides the survey year into four sub-rounds, each of three months' duration, and allots an equal number of first-stage units (viz, villages in the rural sector) to each. "However, because of the arduous field conditions, this restriction was not strictly enforced in Nagaland and some other areas" (GoI 2013a: D1). Also, note that the difficult terrain has affected non-demographic surveys as well, e.g. in Nagaland, "[a]bout 48% of the area could not be covered by systematic geological mapping on 1:50,000 scale owing to inaccessibility" (GSI 2011: 2, 70).

- 2 Certain villages of Andaman and Nicobar are regularly excluded on grounds of inaccessibility. Ladakh region of Jammu and Kashmir is often excluded. In the 1990s, a few districts of Jammu and Kashmir and Punjab "had to be excluded from the survey coverage due to unfavourable field conditions" (GoI 1996: 1).
- 3 See Kruskal and Mosteller (1979) for different notions of representativeness.

REFERENCES

- Agrawal, Ankush and Vikas Kumar (2012): "An Investigation into Changes in Nagaland's Population between 1971 and 2011", IEG Working Paper No 316, Institute of Economic Growth, Delhi.
- GSI (2011): "Geology and Mineral Resources of Manipur, Mizoram, Nagaland and Tripura", *Miscellaneous Publication*, No 30, Part IV, Vol 1 (Part-2), Geological Survey of India, Kolkata.
- GoI (1984): *Census of India 1981, Series 15 Nagaland, District Census Handbook* (for different districts), Directorate of Census Operations

- (DCO) Nagaland, Kohima and Registrar General of India (RGI), New Delhi.
- (1996): *Level and Pattern of Consumer Expenditure, NSS Fiftieth Round, July 1993-June 1994*, National Sample Survey Office, Department of Statistics, New Delhi.
 - (2011): *Report of the Committee on Optimum Sample Sizes for North Eastern States*, NSSO, Ministry of Statistics and Programme Implementation, New Delhi.
 - (2013a): *Key Indicators of Household Consumer Expenditure in India, NSS 68th Round, July 2011-June 2012*, NSSO, MOS&PI, New Delhi.
 - (2013b): *Report of the Committee for Evolving a Composite Development Index of States*, Ministry of Finance, New Delhi.
- Kruskal, William and Frederick Mosteller (1979): "Representative Sampling, III: The Current Statistical Literature", *International Statistical Review*, 47: 3, pp 245-65.
- Kumar, Vikas and Ankush Agrawal (forthcoming): "Nagaland's Demographic Somersault: The Political Economy of India's Government Statistics", IEG Working Paper, Institute of Economic Growth, Delhi.