

# Caste and Census: A Forward Looking Strategy

SONALDE DESAI

In modern India, vast quantities of research have documented caste-based inequalities in many dimensions of well-being. If these inequalities are not simply imagined but reflect social processes that deserve public policy attention, incorporating questions about caste in the census is imperative. However, there is a need to devise an accounting framework that has clarity of purpose since there are many complexities involved in collecting caste data.

Opponents of inclusion of caste in the census argue that for a society which seeks to abolish caste-based inequalities, a census that inquires about caste identities is a retrogressive step since it is more likely to solidify caste-based divisions than to obliterate it. Following Benedict Anderson's argument that censuses played an important role in creation of imagined communities that transcend face-to-face associations (Anderson 1983), many scholars have argued persuasively that colonial censuses created caste as enumerated communities (Das 2003) and solidified hitherto fluid identities (Dirks 2001). Resistance to this reification of social difference often emerges in the form of reluctance to collect caste data. In many ways, this reluctance is similar to that observed in collection of racial statistics in other cultures (Zuberi 2001).

While there is some justification to this argument, we are now living with the aftermath of these political processes. In modern India, vast quantities of research have documented caste-based inequalities in many dimensions of well-being, including income, education, health and access to employment (Govinda 2002; Thorat and Newman 2009; Desai et al 2010; Deshpande 2000). If these inequalities are not simply imagined but reflect social processes that deserve public policy attention, then incorporating questions about caste in census is imperative. However, it is easier to suggest that caste be counted (e.g., the EPW editorial of 22 May 2010), than to devise an accounting framework. Much of the difficulty emerges from lack of clarification regarding the purpose of this accounting.

## Why Collect Caste Data?

The most recent demand for a count of the Other Backward Classes (OBCs) has come from a powerful OBC lobby that hopes for an increase in OBC reservations if the

count turns out to be higher than expected. The 27% reservation for OBCs is based on the estimate by the Mandal Commission that OBCs form about 52% of the population and since *all* OBC families are not poor or "backward", a quota limit set at about half the estimated population makes sense. However, the Mandal Commission's claim of 52% of the population being OBC was based on somewhat flimsy empirical evidence and if the Census 2011 identifies more than 52% of the population as being OBC, this would bolster the claims for higher representation. Successive rounds of National Sample Survey (NSS) have documented the number of individuals identifying themselves as OBC at 36% in 1999-2000 and at 41% in 2004-05. Hence it is unlikely that the actual count will exceed 52%. However there exist other, less partisan, arguments for a caste census. Social background continues to define privilege and lack thereof in Indian society resulting in demands for positive discrimination from marginalised groups. Unfortunately, these demands are continually stymied by lack of data. What data exist from sample surveys continue to document disparities in education, income and standards of living between different socio-religious communities. Table 1 (p 11), based on the India Human Development Survey (IHDS), documents substantial disparities in different markers of human development across different social groups. This nationally representative survey of 41,554 households was organised by researchers from National Council of Applied Economic Research (NCAER) and University of Maryland and was carried out in 2004-05.

The IHDS is not unique in documenting these disparities. A vast number of studies based on NSS, National Family Health Survey (NFHS) and other sample surveys have documented inequalities between scheduled castes (SCs), scheduled tribes (STs), OBCs and forward castes. However, utilisation of survey statistics is by no means adequate for evidence-based policy design. Survey data suffer from two shortcomings. First, even in large surveys such as the NSS, sample sizes get extremely small once

Sonalde Desai ([sonalde.desai@gmail.com](mailto:sonalde.desai@gmail.com)) is with the National Council of Applied Economic Research, New Delhi, and the Department of Sociology, University of Maryland, College Park, MD, USA.

we begin to compare groups on several dimensions of disadvantage. For example, if we wanted to address the question of whether Christian dalits are disadvantaged like other dalits or whether they are closer to the general Christian population, this question cannot be easily answered due to the small number of Christian dalits in various surveys. Any advocacy for affirmative action for Muslim OBCs also suffers from this constraint. The only sensible way of collecting data for these smaller communities is to rely on a nationwide census.

Second, advocacy for refining reservations is even more affected by lack of recent data. Castes have been notified as belonging to SC, ST or OBC categories largely (although not solely) based on outdated data from the 1931 Census. This is particularly true for castes identified as SC, ST, or OBC in the early years of reservations. Since a caste, once notified rarely seems to lose this status, their subsequent economic prosperity, if any, remains unnoticed. Where disadvantage is associated with social distance and discrimination, such as with dalits or adivasis, this is not particularly relevant. Even middle class dalits and adivasis continue to suffer from disadvantages and social exclusion (Desai and Kulkarni 2008; Navsarjan Trust and RFK Center 2010). However, where disadvantages are associated with historical exclusion from access to education or productive resources rather than active discrimination, as with some of the OBC communities like the jats, the situation is more fluid. Some OBC communities may continue to be economically disadvantaged; others may be on par with the forward castes. Table 1 also shows that a substantial proportion of OBC families have a college graduate in the household, have access to salaried jobs and live in homes with flush toilets. Under these situations, obtaining accurate data for better targeting of benefits is crucial.

If inappropriate inclusion of some castes in reserved category is a problem, exclusion of others from benefits is also problematic. As Table 1 documents, a substantial proportion of forward caste families also suffer from socio-economic disadvantages. Whether these disadvantages reflect social class or are due to some castes being particularly disadvantaged cannot be

ascertained without better data. Only accurate data reflecting modern Indian reality would allow us to answer the question of whether there is greater variation within or between castes on various dimensions of material disadvantage. If there were greater variation across castes, this would bolster claims for caste-based affirmative action and reclassification.

### Complexities of Collecting Data on Caste

These two imperatives – identifying smaller marginalised communities and better identification of caste groups for policy targeting – place tremendous demands on the nature of data collection. First, in order to have sufficient information on small populations groups, we need to collect the data in census rather than through household surveys. Second, we need to enumerate caste details for everyone rather than just the SC, ST or OBC categories and even there we need to get information at the jati level rather than at the level of broad population groupings.

Collecting data on jatis is hard enough, given thousands of jatis inhabiting the length and breadth of India; making sense of these data after collection is even harder. For example, it is easy for an interviewer to ask a respondent about their caste and write down this response verbatim.

**Table 1: Social Group Differences in Indicators of Human Development**

	Mean Household Income	Males 15-49 with College Degree or Diploma (%)	Children 8-11 Who Can Read (%)	Males 15-59 with Salaried Work (%)	Daily Income from Wage and Salaries Males Aged 15-59		Homes with Flush Toilets (%)
					Rural	Urban	
Forward caste Hindu	72,717	17	71	32	112	243	37
OBC	42,331	8	56	21	77	154	20
Scheduled caste	34,128	5	44	19	69	142	14
Scheduled tribe	32,345	4	46	13	62	180	7
Muslim	44,158	6	45	20	86	114	24
Other religion	1,01,536	16	79	35	147	228	59
All India	47,804	9	54	22	79	173	23

Source: Desai et al (2010) based on India Human Development Survey 2004-05.

However, caste identification can be at a very general level or at an extremely fine-grained level. For example, depending upon her mood, a respondent may say “I am Baniya”, “I am Modh” or “I am Dasha Shrimali Modh Banik”. All of these responses would be correct. If she calls herself Modh and her cousin calls herself Baniya, our computerised analysis would assume they are from different castes. In the IHDS, a sample of 41,554 households

yielded a list of 7,372 castes. Some of these are different spellings of the same group, Jadav vs Yadav; but others refer to totally different but similar sounding groups, e.g, Jat vs Jatav.

Given these complexities, it is not surprising that since independence, the Office of the Registrar General has steadfastly avoided collecting detailed caste data in decennial censuses.

### Collection of Data on SCs/STs in 2001 Census

While the census does not collect detailed data on caste, the 2001 Census did collect information on SCs and STs. The census enumerators were armed with official state-specific lists of SCs/STs and were asked to check whether the caste identification provided by the household matched the list. They were marked as being SC or ST if their caste or tribe was included in the list. Moreover, only Hindu, Sikh or Buddhists were marked as being SC, regardless of their individual claims to SC status. This process follows the official policy in identifying SC/ST households.

However, even here, considerable complexities were observed. In some cases, the castes noted by the enumerators on the census forms did not match the caste list and the data entry personnel were required to check with anthropologists in

the Social Studies Division of the Office of the Registrar General of India (RGI) before deciding on the SC or ST status. The complexities of this process and concern that mistakes in identification could lead to under-enumeration of some tribes and lead to premature conclusions about their impending extinction undergirds the reluctance of the RGI for collecting OBC data.

Moreover, the OBC list is considerably longer than the SC and ST lists. Hence the

burden on the enumerator for flipping through caste codes could be enormous, resulting in many enumerators using a short-cut and simply counting many households as belonging to forward castes in order to avoid having to look through the whole list. Much of the objection to collection of caste data in census has centred on measurement error in self-identification of caste, while measurement error due to deliberate enumerator misrepresentation has received little attention. However, most experienced survey designers would tell us that when work burden can be reduced by misclassification on the screener question, interviewers are apt to do so. For example, demographic and health surveys find that when interviewers have to ask a whole series of questions for children under five, miraculously the number of six-year olds increases! In order to avoid this, the information must be collected for the whole population and not just for the target group. This also serves the broader policy objectives since the case for refining affirmative action targeting requires that caste information be collected for the entire population and not just for individuals who fall in the reserved category.

### What Would It Take?

If we were to design a full caste census, what would it look like? Hitherto, the census approach has been to take the official list of castes and tribes falling under the SC and ST categories and confirm with the respondents whether they belong to one of these. An alternative to this highly structured approach is a fully flexible approach in which the households are asked to provide their caste identification, enumerators would write it down fully and then classify it at the data entry phase. Both of these are extremely difficult in practice. For a structured approach we must start with a list of castes and as of now no such list exists for the whole population. A flexible approach has a potential for turning into a classification nightmare along the lines discussed earlier in this article.

Some of these issues were debated at length in working groups set up in preparation for the 2001 Census but their deliberations were ignored subsequently. One possible strategy for a caste census may be

to adapt the strategy used for the creation of industrial and occupational classification systems. The occupational classification system provides an interesting illustration. In 2004, a new system of National Classification of Occupations (NCO-04) was developed which contains the following hierarchical structure: Division (10), Subdivision (30), Group (116), Family (439), Occupations (2,945).

Under this schema, a bidi furnace operator would be classified with a code 7,416.45, where the major division is seven (craft and related workers), subdivision is 74 (craft and related workers excluding metal workers, building workers and textile and printing workers), group is 741 (food processing and related workers), family is 7,416 (tobacco preparers), and occupation is 7,416.45 (bidi furnace operator). However, if the respondent were to say that he engages in some highly specialised task within bidi preparation that is not included in the list (for example, counting and making bidi bundles), he could be classified as 7,416.90, tobacco worker not elsewhere classified. This classification scheme both provides ease of classification when data are being collected and ease of analysis so that we may aggregate or disaggregate data depending upon the level of detail desired.

What would it take to come up with such fine-grained classification system for caste enumeration? Once again, the processes through which occupational and industrial systems are developed offer some insights. A possible analogous process might be the following:

- Obtain a list of *jatis* residing through the length and breadth of India. This could be done in surveys such as the NSS, NFHS as well as other unrelated

surveys sponsored by the government. This should be the last question in any survey so as not to distort other responses and would simply provide one line for writing down the full caste name.

- This list should be augmented using various other sources where caste data are available such as marriage advertisements, lists of caste associations and so on. The list can be combined with the caste list from the 1931 Census to create a superset.
- This broad list can be taken by sociologists and anthropologists to come up with a classification schema following the occupational classification approach outlined above with sub-castes grouped within broader castes and allowing space for the possibility that the enumerator may encounter a caste for which no predefined category is specified.
- In creation of the caste classification system two issues should be kept in mind. First, caste system should be treated independently from religion allowing for the possibility that we will collect sufficient information to identify Muslim OBCs (or Christian Brahmins). Second, place of residence should be recorded to allow for a creation of district specific caste list.
- This caste classification list should be sorted by district so that the enumerators can be given two lists. One list containing about 100 castes most frequently found in each district, and a second list containing all castes found throughout India. Given the geographic clustering of castes and tribes in various regions of India, it seems likely that in about 80% of the cases, respondents would only need to refer to the district

### G.B. Pant Social Science Institute

Jhusi, Allahabad – 211019

#### ADMISSION IN DOCTORAL PROGRAMME

The Institute invites applications for its doctoral programme from candidates awarded UGC Junior Research Fellowships (JRF) in Economics, Anthropology, Sociology and History. Candidates may send their applications by email at [director@gbpsi.org.in](mailto:director@gbpsi.org.in) or by post by **August 14, 2010**.

No TA shall be paid but accommodation shall be provided for the day of the interview.

Director

specific list of castes and tribes. For example, Garasias will only rarely be found in West Bengal but frequently in Gujarat and Rajasthan.

- Once this classification list is developed it should be tested by incorporation in large nationwide sample surveys.
- This can then be incorporated in the 2021 Census.

As we look at the complexity of creating a full list of jatis and subjatis and including it in an enumeration system, it is clear that it is unlikely that at this late date any data collection system can be designed for its incorporation in the 2011 Census that would do justice to the complexity of Indian social organisation and would meet the policy demands for data. However, unless we start thinking about this now, when the

national attention is centred on the importance of caste data, it is unlikely that we will be ready even by 2021. If we start now, it is not unrealistic that we may be able to obtain caste information in 2021, 90 years after the 1931 Census when caste data were last collected. Fortunately, by 2021 technological advances will allow each enumerator to carry a handheld device or Personal Digital Assistant with her to avoid her having to flip through pages and pages of caste list to find a classification code for the Dasha Shramali Modh Banik respondent she is likely to encounter in Surat but not in Shillong.

#### REFERENCES

Anderson, Benedict (1983): *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (New York: Verso).

- Das, Veena, ed. (2003): *Handbook of Indian Sociology* (New Delhi: Oxford University Press).
- Desai, Sonalde, Amaresh Dubey, B L Joshi, Mitali Sen, Abusaleh Shariff and Reeve Vanneman (2010): *Human Development in India: Challenges for a Society in Transition* (New Delhi: Oxford University Press).
- Desai, Sonalde and Veena Kulkarni (2008): "Changing Educational Inequalities in India: In the Context of Affirmative Action", *Demography*, 45 (2): 245-70.
- Deshpande, Ashwini (2000): "Does Caste Still Define Disparity? A Look at Inequality in Kerala, India", *The American Economic Review*, 90 (2):322-25.
- Dirks, Nicholas (2001): *Castes of Mind: Colonialism and the Making of Modern India* (Princeton, NJ: Princeton University Press).
- Govinda, R (2002): "India Education Report" (New Delhi: Oxford University Press).
- Navsarjan Trust and RFK Center (2010): *Understanding Untouchability: A Comprehensive Study of Practices and Conditions in 1589 Villages, 2010-2010*, available from <http://www.rfkcenter.org/untouchability>.
- Thorat, Sukhdeo and Katherine Newman, ed. (2009): *Blocked by Caste: Economic Discrimination in Modern India* (New Delhi: Oxford University Press).
- Zuberi, Tokufu (2001): *Thicker Than Blood: How Racial Statistics Lie* (Minneapolis: University of Minnesota Press).