THE INDIA FORUM
A JOURNAL-MAGAZINE ON CONTEMPORARY ISSUES

April 25, 2023

Making the Bihar Caste Census Better

By: M.R. Sharan

Bihar's caste census can offer new insights into the endurance of caste's links with occupation and economic status. Conducted properly, it has the potential to empower residents and deepen democracy.

Bihar is embarking on a historic caste census. This will be the first such enumeration of castes since the Socioeconomic Caste Census (SECC), which was conducted over a decade ago.

Bihar's new census assumes more importance because much of the SECC's data on caste remains unreleased. The new enumeration promises to go beyond broad categories like Scheduled Castes (SC), Scheduled Tribes (ST), Other Backward Classes, and Extremely Backward Classes. It will collect and release much finer data on jatis, the thousands of sub-castes that form the building blocks of caste society.

The heterogeneity among jatis is crucial to understanding India's complex caste structure and the unique set of upward mobility pathways available to each one. Hence, gathering granular jati data is essential in any comprehensive caste enumeration across the country. At first glance, the enthusiasm for the caste census among those who represent disadvantaged groups suggests that a careful jati-wise enumeration of economic well-being could bring rich political dividends to them.

> The caste census has the potential to assist in the development and targeting of more effective government policies.

However, the upcoming caste census presents a significant opportunity to achieve more.

The data collected can offer new insights into the endurance of caste and its interconnectivity with occupation and economic status. Furthermore, the caste census has the potential to assist in the development and targeting of more effective government policies. Providing individuals access to this data can empower them to advocate for their rights and participate more actively in governance.

Past lessons

The first phase of enumeration in Bihar is complete. But we can still ask: how can the caste census data collection and dissemination exercise be made better so that it can foster a more inclusive and representative society?

To answer that question, we should cast our eyes backwards at the SECC to see what we can learn from its successes and failures.

The SECC was a nationwide survey by the Government of India that conducted a meticulous examination of households based on their economic and social status. In Bihar, the SECC was conducted in 2012. Surveyors interviewed almost two crore households in Bihar, assessing their asset ownership, main source of income, land ownership, jati, and occupation. The primary objective of the survey was to determine households living below the poverty line, living in deprived conditions, and those belonging to socially disadvantaged groups. Some government departments continue to use the SECC as the bedrock of their databases, updating it at regular intervals.

It is important to bear in mind some key features of the SECC, which provide pointers on how the new census can improve over its antecedents.

First, unlike the decennial censuses, the SECC collected detailed data on jatis. But it is known in policy circles that the collection of this data was poor. The entry on jati spellings was not standardised. This made cleaning the jati data a Herculean task, one that most governments were loath to undertake.

Second, to understand a household's economic status, the SECC focused on a range of assets that could easily be observed by enumerators, like quality of housing (kuccha vs pucca) or ownership of items like fridges and vehicles. This means that the SECC (like the current caste census) could not tell apart fine income differences between households.

> The enduring link between caste and occupation means any census of caste must collect clean data on occupation too.

Third, the SECC collected data on occupation of individual members, but again, just as in the case of the jati data, the occupation data in the SECC was not standardised. This meant that it has become hard to analyse the data on occupation and therefore use it in any meaningful way to inform policy.

The link between caste and occupation has been an enduring feature of Indian society, with certain professions being associated with specific castes. This association has had a significant impact on labour market outcomes and occupational choices across the country. In Bihar, for instance, the upper castes, such as Brahmins. Rajputs and Bhumihars, have traditionally held positions of power and privilege, while the lower castes, such as Dalits and OBCs, have been relegated to low-paying manual labour jobs. This dynamic has perpetuated a cycle of economic and social inequality that has been difficult to break. The enduring link between caste and occupation means any census of caste must collect clean data on occupation too.

Lastly, the SECC was conducted at a time when digital data collection was still in its infancy. As a result, although the data was being collected electronically by enumerators throughout the country, the potential to access and run real-time checks on the data was limited. As I will argue later, much can be done today to ensure data collected at scale is clean and usable.

Caste data

What do we know about caste in Bihar?

Despite its shortcomings, the SECC remains the most recent comprehensive source of caste in Bihar. A qualified analysis of the SECC data will help shed light on what the current census may show us.

While we do not have access to the jati variable in the SECC, we can proxy for jati using surnames. The link between surnames and jati is not always straightforward. Some surnames, like Kumar, mask jati entirely, while others belong to multiple jatis. Despite these limitations, using surnames as a proxy for jati can give us a broad understanding of how caste and class overlap in Bihar.

> An aspect of society that needs better enumeration is the status of Muslims in Bihar [...] there is no clear state-wide data on where Muslims stand in the socioeconomic hierarchy.

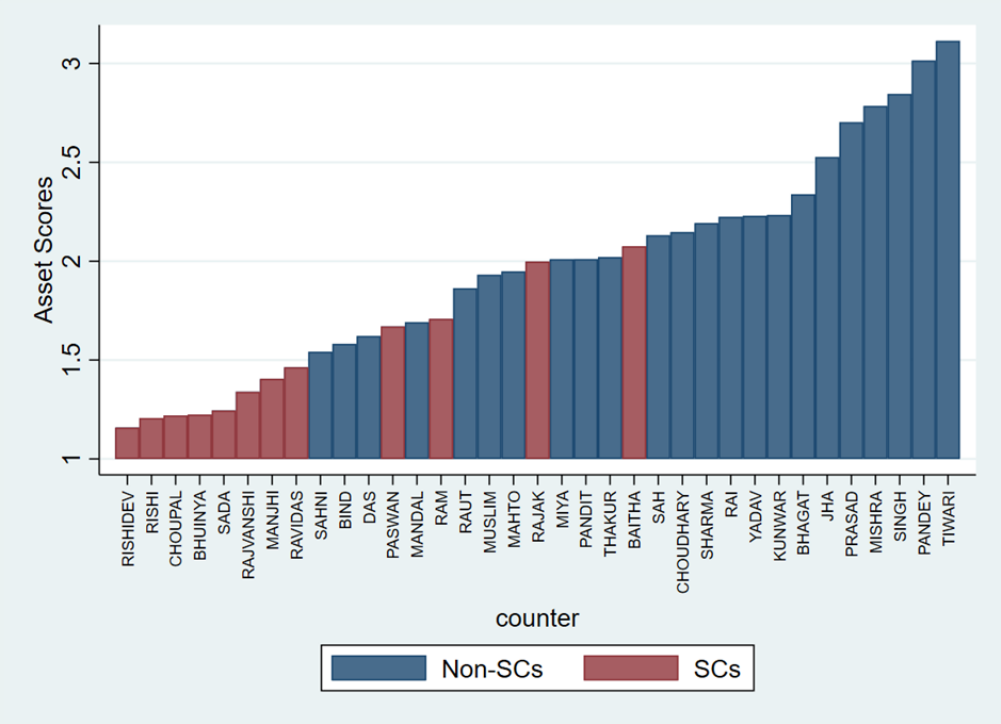
There are 27 surnames that repeat at least 75,000 times and make up nearly three quarters of all caste names recorded in the SECC. Of these, four - Paswan, Sada, Ram and Manjhi - are predominantly used by SCs.

To rank surnames by asset wealth, we turn to six of the most commonly found asset indicators in the SECC data. These were:

- Whether the household owns land
- Type of roof (concrete or not)
- Whether the wall is made of burnt brick or concrete in the main dwelling room of the house structure
- Whether the house has 4 or more rooms
- Whether the household has a phone
- Whether the household owns a vehicle

Giving a value of 1 each if the answer to these questions was "yes", every house can have a score between 0 and 6. Since these asset indicators comprise items that take time to accumulate - like land or vehicles - this asset score is a good proxy for long-term economic status.

We plot the average asset score across large sub-castes, as proxied by 51 surnames that repeat across at least 30,0000 households. (Figure 1).



The first thing that strikes one is that a full six decades after Independence and the enshrining of Dalit upliftment as a key societal goal, Dalit surnames continue to be at the bottom of the distribution. Indeed, the bottom eight poorest surnames all belong to Dalits.

The second thing to note is that the next rung of surnames - Sahni, Bind - are all surnames belonging to what is now termed as Extremely Backward Classes (EBCs). This is followed by a string of middle-castes, then powerful Other Backward Classess (like Yadav) and at the very top at are the Brahmins, Rajputs and Bhumihars.

There are some jatis who have performed economically better than those above them in the social hierarchy: for instance, the more upwardly mobile Dalit castes, like Paswans, Chamars and the Dhobhi jatis are not just outperforming their Dalit brethren, but also some EBC and Baniya castes. However, for the most part, the striking - and unfortunate - picture that emerges is that the economic status of sub-castes closely mimics their place in the caste hierarchy.

Building on the SECC

Have mobility patterns changed in the decade since the SECC? This is a question that the new caste census should strive to answer.

Since 2006, Bihar has implemented a slew of policies targeted at marginalised caste groups.

A non-exhaustive list includes formation of a new Mahadalit Vikas Mission with the objective of promoting the socioeconomic development of the Mahadalit community, which comprises over 20 scheduled caste jatis considered to be the "Dalit among the Dalits", having a dedicated local bureaucrat in every village - called the Vikas Mitra - drawn from the Mahadalit community to help Mahadalits liaise with the state; implementing reservation for Dalits in local government, a process that was stalled for years because of challenges in court; identifying extremely backward classes (EBCs) jatis and creating policies targeted explicitly towards them; the creation and expansion of women's self-help groups (SHGs), first for Dalits, then more broadly, across villages and towns in the state: over a million self-help groups have been formed since 2006 and a staggering 60% of households now have a member in an SHG.

The new caste census could shed light on whether these policies have made a dent on the existing socioeconomic structure.

> Data cannot simply be released on a public portal for it to be empowering. It needs to be made relevant and accessible.

For the censuses to be comparable, the new census should strive to collect at least some of the asset indicators collected by the SECC. New additions will have to be made (type of smartphone could be a good way to distinguish across households) and some redundant indicators dropped (for example: presence of landline phones). However, many of the old indicators continue to be informative regarding a household's status in the wealth distribution and must not be discarded. If anything, care should be taken to make the variables exactly comparable.

A second aspect of society that needs better enumeration is the status of Muslims in Bihar. The SECC did collect data on religion, but that data too was not released. Hence, there is no clear state-wide data on where Muslims stand in the socioeconomic hierarchy.

Using the SECC and name categorisation algorithms, we can establish, broadly, the following facts. First, the average Muslim household occupies a place considerably below the median household in the asset distribution. Second, that Muslims are under-represented in local politics. While nearly 16% of Bihar's population are Muslims, under 10% of Panchayat heads are Muslims. Moreover, recent work by economists Sam Asher and Paul Novosad, drawing from the draft version of the SECC, argues that educational mobility of Muslims has been lower than even Dalits across the country.

These are, at best, estimates. If the new census can document these facts rigorously and also provide a geographical breakdown of trends, that would go a long way in helping design policies towards Muslims.

A third area is the enumeration of EBC castes. While we know what jatis are classified as EBCs, we do not know what their population shares are. Moreover, insofar as the categorisation of jatis as EBC or OBC is also a function of economic well-being of castes, understanding which jatis are economically deprived could help better categorise jatis as EBCs or OBCs.

Identifying population shares could also help refine reservation rules. For instance, currently, reservation for Scheduled Castes and Scheduled Tribes in elected panchayat positions is proportionate to their share in the local population. On the other hand, EBC reservation in panchayat positions is up to 20% irrespective of what their share in the local population is. This creates situations where too few seats are reserved where EBCs are plentiful and too many in areas where they are relatively fewer.

> Data collected from citizens should be used to empower them, deepening democracy and giving voice. If data is more broadly accessible, the new census has the potential to be a living, empowering exercise.

A fourth area the census could shed deeper light on is how jati interacts with migration, both seasonal and permanent. Bihar boasts high rates of migrations. Remittances from Bihari workers working across the country - and, to a lesser extent, the world - account for a substantial share of the state's GSDP. Migration varies by both caste and region. Indeed, as research has shown, migration is deeply linked to jati networks (Munshi and Rosenzweig (2016). The caste census provides a key opportunity to understand what jatis contribute to the state-wide pattern of migration we see and this could then inform policymaking.

Finally, the new census must record data of not just a household's panchayat or revenue village, but also electoral ward in rural areas.

Wards are the last unit of local democracy in Bihar. With over 1 lakh elected representatives (one for roughly 1,000 members), elected ward representatives form the most diverse, locally recognisable and representative group of elected leaders in the state (Sharan 2021). Linking households to wards and building ward and ward-jati level poverty indices could help develop a finely grained view of development in the state.

Moreover, at least on paper, wards are the last mile implementers of policy - building a rich understanding of ward-level metrics could help design and target policies better.

Quicker and better

For the first century and more of census in India, enumerators carried pen and paper from house to house. Data was collected by hand, then sent to an office where it was examined by a team of scrutinisers, before being analysed, either by hand or digitally. For most of the 21st century, most analysed data were first digitally entered by an operator. The whole process would take months - sometimes running into years.

Technology has made things much easier. Data collection and entry are no longer separate entities. Once data is collected on a digital device, it is automatically sent to a centralised server that stores all data.

Bihar's caste census can harness these advances to avoid some of the pitfalls of earlier enumeration methods. Lists of households and members, castes, and religion, even options to answers can now be pre-loaded into the device. This saves enumerators precious amounts of time and resources, as well as reduces entry errors.

Seasoned quantitative researchers are aware of these techniques and these are practised across the world. Yet, owing to institutional inertia, governments can be slow to catch up.

> Marginalised castes should be allowed free access to the data in order to make claims on the state's resources.

In addition to having simple entry errors corrected, Bihar can have a series of skilled analysts looking at the data at the back-end on a real-time basis. These analysts can focus on logical inconsistencies across multiple fields and raise red flags with the enumeration team. For instance, it is unlikely that a landless Dalit household living in a thatched house owns a fridge. If the response to the fridge question has been coded "yes", then this could be an enumerator error, which an analyst can take up with the surveyor.

Constant engagement with seasoned practitioners as the data is being collected will allow bureaucrats to assess errors and resurvey areas before enumerators moved on to another jurisdiction.

Fourth, responses from a random sample of households must be back-checked by a separate back-check team. This back-check team, a mainstay of field surveys across the world, keeps the enumeration team honest.

To conduct real-time data collection and analytics, the government should look at the stellar work the State Election Commission of Bihar did in conducting the Panchayat and Municipal elections. Their commendable use of app-based training techniques of electoral officers, the presence of a well-staffed round-the-clock control room to deal with field issues and innovative use of technology to monitor the electoral process, could provide pointers.

Whose data?

Data collected by the state is too often stowed away behind logins and passwords, put to use only so that the state can "see" better. Data flows from citizens to the state (and researchers), but rarely flows the other way.

Instead, data collected from citizens should be used to empower them, deepening democracy and giving voice. If data is more broadly accessible, the new census has the potential to be a living, empowering exercise.

The first way in which the caste census data can be used is to allow citizens to view themselves in comparison with others, both across geographies and castes. Marginalised castes should be allowed free access to the data in order to make claims on the state's resources. This has echoes in history. Recent work by Pritam Singh (2022) has argued that caste data from the decennial census in the pre-1941 era allow low caste groups to highlight their marginalisation and make claims for power.

Citizens should be able to compare their living standards with others in their villages (or towns) and with respect to others elsewhere. Bihar's panchayats are among the largest in the country. Residents might have a good sense of the land ownership patterns in their wards or villages but are unlikely to know of their relative status across their entire panchayat. It is even less likely that households would know how land ownership varies within blocks or districts or where one's own panchayat stands with respect to others in terms of land inequality. Such data can particularly strengthen the case of marginalised caste beneficiaries to make claims for land to be transferred under government programmes.

> A poorly conducted census is a double blow. It is a waste of the state's precious resources and deprives the people and their government of a crucial opportunity to take a deep, hard look at themselves.

Data cannot simply be released on a public portal for it to be empowering. It needs to be made relevant and accessible. Experts could work with citizens to develop visualisations of the data that could make it easily digestible to citizens in villages, as done with the Pudhu Vaazhvu project in Tamil Nadu. This could be used to spark discussions and spur claim-making.

The other way in which data can be used to deepen democracy is to allow elected representatives - ward members, Mukhiyas, Block Panchayat, Zilla Panchayat members, MLAs and MPs - better access to the data. Within government, data is usually controlled by the

bureaucracy, with elected representatives - especially at lower tiers - kept away from it. Elected representatives should not, of course, be given data on individual-level outcomes, but could learn from aggregated data. When the government releases the state-wide aggregated data and a report, it should also create similar reports for representatives at various tiers of government. These smaller reports should focus on representatives' constituencies, documenting raw averages from their jurisdictions and putting these in a comparative context. The advantage of a census is that, by virtue of covering the entire population, it can be aggregated to relatively small jurisdictions like wards and tolas, which most sample surveys cannot.

Conclusion

As I have argued elsewhere, detailed counts of jatis within towns and villages are already known to many in Bihar, especially those in government and politics. What the new caste census does, therefore, is to make public a facet of Bihari society that is somewhat of an open secret. There is less agreement on the economic well-being of jatis, a gap the new census attempts to fill.

A poorly conducted census is a double blow. It is a waste of the state's precious resources and deprives the people and their government of a crucial opportunity to take a deep, hard look at themselves. For Bihar's caste census to fulfil its promise, the state must conduct the enumeration exercise carefully. Often the hardest parts of such mammoth policy exercises revolve around getting the smallest details right.

M.R. Sharan is an assistant professor in the Department of Agricultural and Resource Economics at the University of Maryland, College Park.

References:

Alkire, S. and S. Seth (2012). "Identifying BPL Households: A Comparison of Methods." *Economic and Political Weekly* 48 (2): 49-57. (12 January).

Ambedkar, B.R. (2014). *Annihilation of Caste*. Verso Books.

Asher, S., P. Novosad and C. Rafkin, C. (2018). 'Intergenerational mobility in India: Estimates from new methods and administrative data'. World Bank Working Paper (September).

Dreze, J. and R. Khera (2010). 'The BPL census and a possible alternative. *Economic and Political Weekly* 45 (9): 54-63. (27 February).

Munshi, K. and M. Rosenzweig (2016). "Networks and misallocation: Insurance, migration, and the rural-urban wage gap." *American Economic Review* 106 (01): 46-98.

Sharan, M.R. (2021). *Last Among Equals: Power, Caste and Politics in Bihar's Villages*. Context.

Singh, Pritam. (2022). "Numbers as a Means to Power: Politics of Caste as a Census Category in Colonial India c 1871-1941." *Economic and Political Weekly* 57 (18) (30 April).