NCAER | SINCE 1956
QUALITY. RELEVANCE. IMPACT

NCAER NATIONAL DATA INNOVATION CENTER

MEASUREMENT BRIEF 2024

NDIC FELLOWS PROGRAMME

From Heaps to Insights: Aadhaar's Role in Enhancing Age Data Quality

SUBARNA BANERJEE

Overview and Measurement Challenge

Many sections of society, particularly those residing in areas with weak state capacity, are often unaware of their exact age. As a consequence, large-scale survey data in emerging market economies such as India suffer from age-misreporting concerns. Acquiring correct age distributions is required for the construction of age-specific mortality and fertility rates, and also population estimates and projections. Further, the dependence of governments and the social sector on accurate demographic parameters to formulate and implement policies makes the necessity of good quality age data salient.

Low literacy, lack of official birth registrations, narrow scope of interacting with formal institutions, limited cultural relevance of birthday celebrations, and proxy reporting are some reasons behind providing incorrect age data by respondents during surveys. This often leads to a phenomenon known as age heaping, where individuals tend to inaccurately report their ages by rounding them to numbers ending in 0 or 5 (e.g., 30, 35, 40), rather than providing their precise age. Age heaping distorts the actual age distribution, which can undermine the reliability of demographic estimates and policy-making based on age-specific data. On the other hand, miscommunication, distortion of correct ages during recording or processing the survey data, and the systematic displacement of respondents from the border of an eligible age group to the neighbouring group to avoid interviews can also compromise age data from the surveyor side.

I study the evolution of age data quality in India over the past 30 years using 5 rounds of nationally-representative DHS data.[1] I also explore if the introduction of a nation-wide biometric identification system – Aadhaar – that serves as a proof of identity and streamlines government processes could also be instrumental in reducing age-heaping caused by unawareness of one's age. Additionally, I use information from annual reports of state-level Aadhaar coverage to see its influence on local age heaping patterns.[2]

KEY RESULTS

- Age misreporting has declined in India over the years. The tendency to report ages ending in certain preferred digits such as multiples of five or even numbers has gone down over time.
- Age heaping at the local level is negatively associated with local self-reported Aadhaar holdings. The latter is highly correlated with state-level variation in Aadhaar coverage. In other words, in areas with more Aadhaar card holders, fewer people misreport their age by rounding it. These associations hold even after accounting for state- and year-wise unobserved heterogeneity and time-varying demographic and economic factors.
- Age heaping has reduced after the introduction of Aadhaar. States with higher Aadhaar saturation (as reported by state governments to UIDAI) witnessed a greater decline in age heaping at the time of surveys.
- The fall in age heaping after the introduction of Aadhaar has been stronger for older individuals. This rules out the possibility of other contemporaneous interventions driving the results such as the push for formal birth registrations that have differentially affected younger individuals over recent years.

Method

I use 5 rounds of India's National Family Health Survey (NFHS) conducted between 1991 and 2021. Administered under the Government of India's Ministry of Health and Family Welfare (MoHFW), NFHS

1 The data can be accessed from https://dhsprogram.com/data/available-datasets.cfm

2 PDFs of these reports are available at https://uidai.gov.in/en/media-resources/uidai-documents/annual-reports.html

is a part of the global Demographic Health Survey (DHS) programme. I use the population recode file that includes individual-level information on age of all members of the surveyed household along with other household characteristics. I also use publicly available annual state-level Aadhaar saturation reports published by Unique Identification Authority of India (UIDAI).

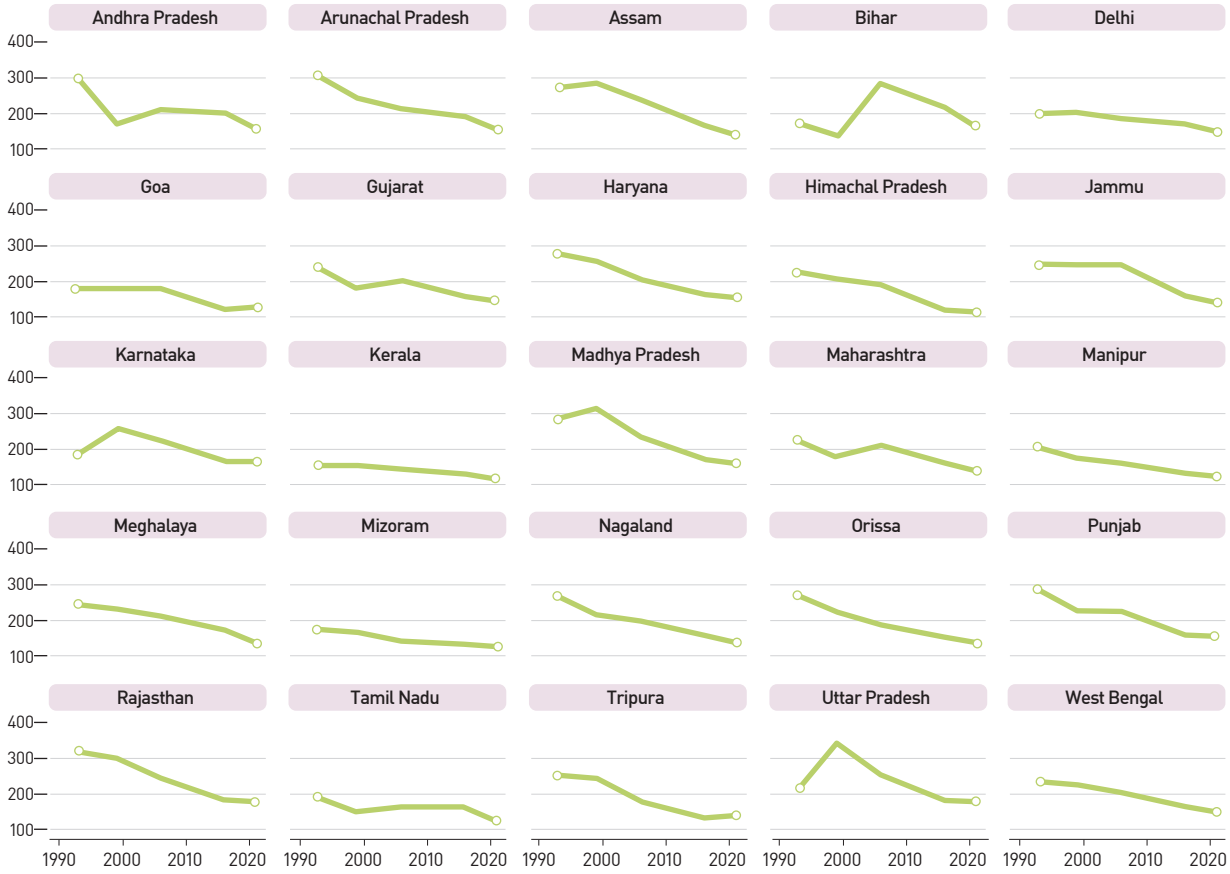
I focus on the tendency to round off ages to attractive digits (such as 0 and 5) while avoiding other digits (such as 1, 3, 7, and 9) in the absence of correct information on age – a phenomenon known as age heaping. The use of indices to measure age heaping has been standard practice in the literature that examines the quality of age statistics. Following established methods in the literature, I construct a Whipple's Index and the Total Modified Whipple's Index at the state and district levels using NFHS data. While the original Whipple's Index measures preferences for ages ending in either 0 or 5, the Total Modified Whipple's Index is an overall summary index that incorporates the degree of preference or avoidance for all 10 digits (0-9). Higher values of both indices correspond to greater age heaping and are my outcomes of interest.

Using a fixed-effects regression framework, I look at the associations between: i) the proportion of surveyed respondents in a district with self-reported Aadhaar holdings and state-level Aadhaar saturation in the previous year, and ii) district-level age heaping indices and district proportions of self-reported Aadhaar holdings.

Next, I use a difference-in-differences specification for a state-level panel to look at the impact of introduction of Aadhaar on age-heaping indices. By using variation in Aadhaar saturation levels across states, I look at the effect of greater Aadhaar coverage in reducing age misreporting after the launch of Aadhaar.

Finally, I rule out concerns that the effects I observe are not attributable to Aadhaar but are due to other central and state-government efforts to formalise birth

Figure 1: Evolution of Whipple's Index for different states over the Years



Source: Author's calculations using 5 rounds of NFHS data.

registrations that would in turn improve age-related misinformation. To disentangle these complementary effects, I use institutional knowledge that while Aadhaar introduction impacted almost all age groups, the push for birth registration cards can only be binding for younger individuals. I create separate age-heaping indices for younger and older individuals to explore heterogeneous effects of the intervention.

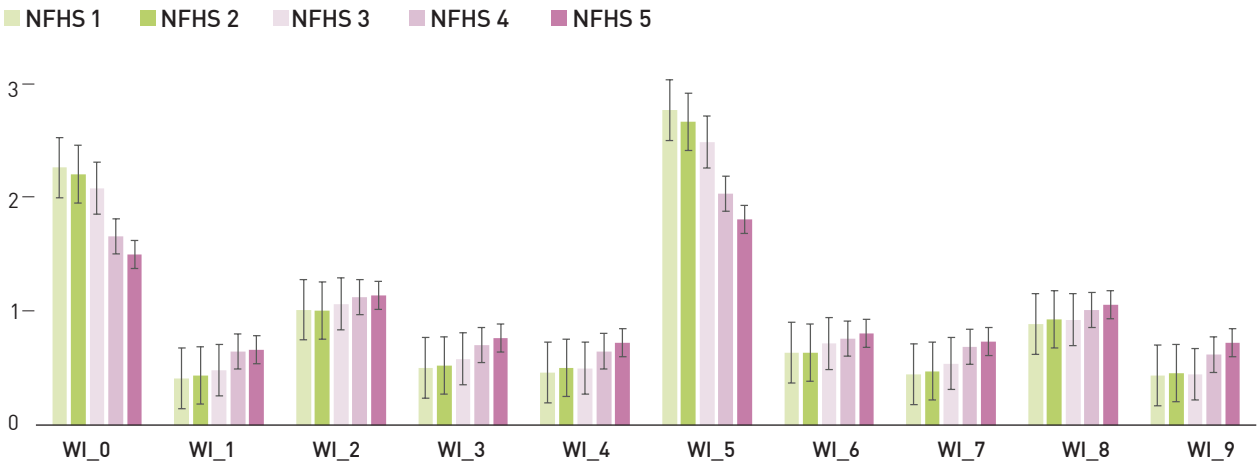
Results

Age misreporting has declined in India over the years. Whipple's Index, which measures preferences to heap ages in numbers ending with 0 or 5, has declined over the years for all the major states in India (Figure 1 on Page 3). The decline of preference for ages ending in 0 or 5 is complemented by a symmetric fall in avoidance of other digits like 1, 3, 4, 6, 7, and 8 over multiple rounds of the NFHS data (Figure 2).

The proportion of surveyed respondents in a district with self-reported Aadhaar holdings is positively correlated with state-level Aadhaar saturation in the previous year (Table 1 & Figure 3). At the same time, age heaping at the local level is negatively associated with local self-reported Aadhaar holdings (Table 2 on Page 5).

A 10 percentage point increase in the previous period's Aadhaar saturation for a state is associated with close to a 5 percentage point increase in the district proportion of self-reported Aadhaar holdings (Table 1). Further, moving from a district with no Aadhaar holdings to a district with 100 percent self-reported Aadhaar hold-

Figure 2: Digit Preferences Over the Years

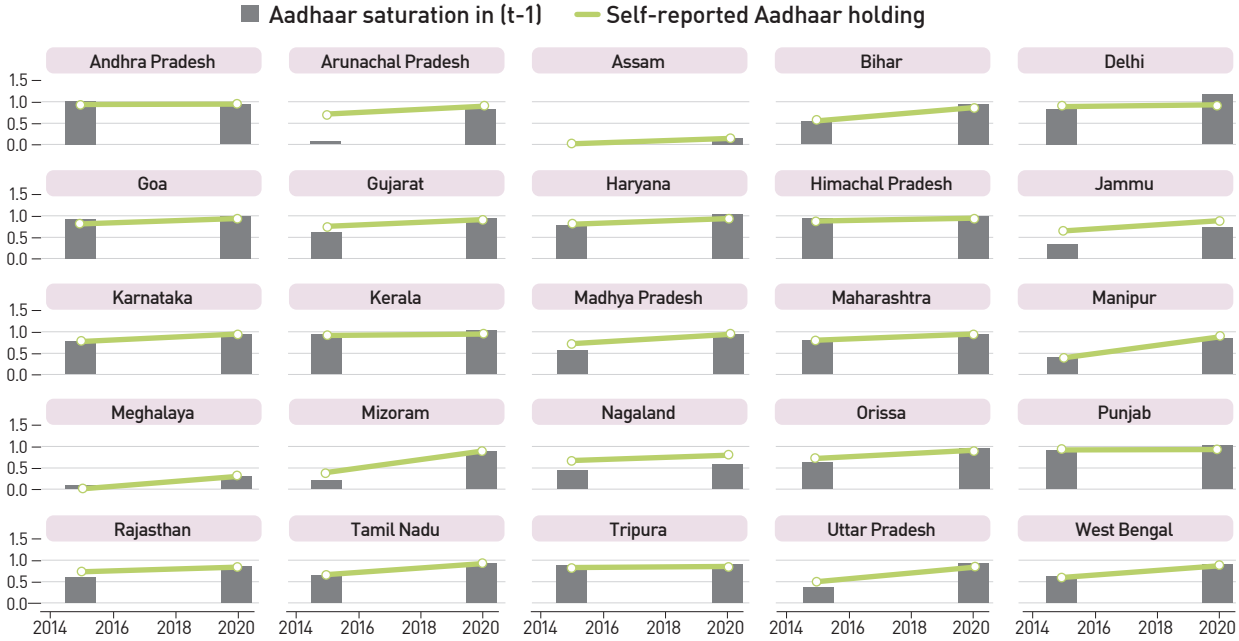


Source: Author's own calculations using 5 rounds of NFHS data.
Notes: 1. The labels indicate digit preference for a given terminal digit. For example, WI_1 indicates heaping of ages around numbers ending with 1. 2. The states and UTs in my sample are Assam, Haryana, Himachal Pradesh, Kerala, Goa, Punjab, Delhi, Andhra Pradesh, Uttar Pradesh, Gujarat, Madhya Pradesh, West Bengal, Tamil Nadu, Orissa, Karnataka, Maharashtra, Bihar, Tripura, Mizoram, Rajasthan, Manipur, Arunachal Pradesh, Jammu & Kashmir, Nagaland, and Meghalaya.
Interpretation: This figure plots the evolution of Whipple's Index for 25 states and UTs in India over the years. The decline of preference for ages ending in 0 or 5 is complemented with a symmetric fall in avoidance of other digits like 1, 3, 4, 6, 7, and 8 over multiple rounds of the NFHS data.

Table 1: State-level Aadhaar Saturation and Local Self-reported Aadhaar Holdings

Dependent Variable: **District-level Self-reported Aadhar Holdings** *(Sample mean = 0.78)*	(1)	(2)	(3)	(4)
State-level Aadhar Saturation at t-1	0.756*** (0.01)	0.623*** (0.01)	0.499*** (0.03)	0.468*** (0.03)
Observations	1,280	1,280	1,280	1,280
State Fixed Effects	No	Yes	Yes	Yes
Year Fixed Effects	No	No	Yes	Yes
State-level demographic controls	No	No	No	Yes
State-level economic controls	No	No	No	Yes

Figure 3: State-level Aadhaar Saturation and State-level Self-reported Aadhaar Holdings



Source: Author's own calculations using NFHS 4 and 5 data.
Interpretation: The state's Aadhar saturation rates in one year strongly predict how many people will report owning an Aadhar card in the next year.

ings is associated with a 21.5 unit decline in Whipple's index and a 1 unit decline in the Total Modified Whipple's Index (Table 2).

Age misreporting has improved after the introduction of Aadhaar.

In Table 3, the coefficient associated with the post-Aadhaar dummy is negative and statistically significant. Thus, both Whipple's Index and the Total Modified Whipple's Index have decreased in the post-Aadhaar period, indicating reduced age heaping overall. The coefficient associated with the interaction of post-Aadhaar and state-level saturation is also negative and statistically significant, suggesting that states with higher Aadhaar coverage experienced a more significant improvement in age heaping after the introduction of Aadhaar. The effect of Aadhaar saturation after the introduction of Aadhaar improved age heaping by 26 percent over the baseline mean (as measured by Whipple's Index) and by 47 percent over the baseline mean (as measured by the Total Modified Whipple's Index).

Table 2: Local Self-reported Aadhaar Holdings Age Heaping at the Local-level

	Whipple's Index	Total Modified Whipple's Index
District-level Self-reported Aadhar Holdings	-21.52***	-1.015***
	(5.32)	(0.23)
Saturation at t-1	1,280	1,280
Observations	158.05	3.39

Table 3: Age heaping after the Introduction of Aadhaar

	Whipple's Index	Total Modified Whipple's Index
Post × Saturation	-19.03*	-1.04**
	(10.51)	(0.48)
Post	-34.13***	-1.43***
	(9.58)	(0.4)
Observations	90	90
Baseline Mean	202.58	5.25
State Fixed Effects	Yes	Yes
State-level demographic controls	Yes	Yes
State-level economic controls	Yes	Yes

Notes:
1. Data comes from NFHS rounds 3, 4, and 5. Standard errors, clustered at the state level are in parentheses for all the specifications.
2. p ← 0.10, ** p ← 0.05, *** p ← 0.01.
3. State-level controls include proportion of Hindus, Muslims, SCs, STs, and OBCs, average family size, average age, population density, proportion covered under a health scheme, proportion that falls under the poorest wealth quintile, night-time luminosity, average travel time to the nearest urban hub.

The fall in misreporting after the introduction of Aadhaar has been stronger for older individuals.

Age heaping indices have fallen more for older individuals in the post-Aadhaar period compared to their younger counterparts. Further, the effect of Aadhaar saturation on age heaping is statistically significant for only the Whipple's Index constructed for an older cohort above 42 years of age. The effect is smaller in magnitude and statistically insignificant for a younger cohort (Table 4).

Table 4: Disentangling the effect for Age heaping of Older vs. Younger individuals

	Whipple's Index (Old)	Whipple's Index (Young)
Post × Saturation	-38.11**	-7.5
	(15.16)	(6.00)
Post	-40.22***	-17.84***
	(12.61)	(4.99)
Observations	90	90
Baseline Mean	250.91	157.64
State Fixed Effects	Yes	Yes
State-level demographic controls	Yes	Yes
State-level economic controls	Yes	Yes

Notes:
1. Data comes from NFHS rounds 3, 4, and 5. Standard errors, clustered at the state level are in parentheses for all the specifications.
2. p ← 0.10, ** p ← 0.05, *** p ← 0.01.
3. State-level controls include proportion of Hindus, Muslims, SCs, STs, and OBCs, average family size, average age, population density, proportion covered under a health scheme, proportion that falls under the poorest wealth quintile, night-time luminosity, average travel time to the nearest urban hub.

Policy Lessons

The introduction and widespread adoption of Aadhaar have played a pivotal role in reducing age heaping in large-scale surveys in India. An important caveat here is that Aadhaar cards may have inaccurate birthdates so even if this reduces the tendency to age heap, the reliability of age data might still remain a concern for effective policy planning and implementation.

While Aadhaar has significantly reduced older individuals' age heaping, ensuring accurate age reporting for younger populations requires a complementary approach. Formal birth registration systems are crucial in this regard, because they provide accurate age data right from the outset. Strengthening these registration mechanisms and incentivizing timely registrations are imperative steps to bolster demographic accuracy, supplementing the efforts of Aadhaar.

Moreover, addressing the underlying factors contributing to age misreporting, such as low literacy levels, is essential. Implementing targeted literacy programmes can empower individuals to better understand and accurately report their personal information, further enhancing the reliability of demographic data.

It's important to recognize that the impact of Aadhaar varies across states and districts due to disparities in Aadhaar saturation rates. Therefore, localized policy adjustments are necessary to ensure equitable improvements in age reporting accuracy nationwide. States with lower Aadhaar coverage should receive focused efforts to boost enrolment rates, thereby fostering uniformity in the quality of demographic data across the country. Integrating robust birth registration systems and literacy enhancement programmes are crucial for further improvements. By expanding and strengthening these initiatives, India can enhance the quality of demographic data, facilitate more effective governance, and ensure equitable access to government services for all sections of society.

Further Reading

Agrawal, G., & Khanduja, P. (2015). Influence of literacy on India's tendency for age misreporting: Evidence from Census 2011. *Journal of Population and Social Studies [JPSS],* 23(1), 47-56.

Borkotoky, K., & Unisa, S. (2014). Indicators to examine quality of large scale survey data: an example through District Level Household and Facility Survey. *PLoS One, 9(3)*, e90113.

Larsen, A. F., Headey, D., & Masters, W. A. (2019). Misreporting month of birth: Diagnosis and implications for research on nutrition and early childhood in developing countries. *Demography,* 56(2), 707-728.

Rosenzweig, S. C. (2021). Age is measured with sys-

tematic measurement error in developing country surveys: A diagnosis and analysis of consequences. *Research & Politics,* 8(3), 20531680211044068.

Spoorenberg, T., & Dutreuilh, C. (2007). Quality of age reporting: extension and application of the modified Whipple's index. *Population,* 62(4), 729-741.

Suggested citation

Banerjee, Subarna (2024). "From Heaps to Insights: Aadhaar's Role in Enhancing Age Data Quality", Measurement Brief - NDIC Fellows Programme (2023-24). NCAER National Data Innovation Centre, New Delhi

Author Bio

Subarna Banerjee is a fourth-year PhD student at Shiv Nadar Institution of Eminence and a fellow at NCAER's National Data Innovation Cell. Her research interests lie in applied work encompassing the fields of education, gender, and political economy. She holds a Master of Science in Economics from Shiv Nadar University and a BSc. Economics (Hons) from Presidency University, Kolkata. Her academic pursuits reflect a deep commitment to advancing evidence-based policymaking through novel data approaches in the Social Sciences.

Funding Support and Disclaimer

This research is supported by the National Council of Applied Economic Research through the NCAER National Data Innovation Centre. The views presented in this paper are those of the authors and not those of NCAER or its Governing Body. Funding for the NCAER National Data Innovation Centre is provided by Bill & Melinda Gates Foundation.

About NCAER NDIC

The NCAER National Data Innovation Centre was set up in December 2017 to promote innovation and excellence in data collection and build research capacity to strengthen the data ecosystem in India. The NDIC is envisaged as a hub for providing expertise to policymakers, government statistical agencies and private data collection agencies. NDIC is involved in pursuing three primary activities:

1. To pilot innovative data collection methods and mainstream successful pilots into larger data collection efforts;
2. To impart formal and informal training to a new generation of data scientists; and
3. To serve as a resource for data stakeholders, including Government data agencies and ministries.

About NCAER

Established in 1956, NCAER is India's oldest and largest independent, non-profit, economic policy research institute. NCAER's work cuts across many sectors, including growth, macro, trade, infrastructure, logistics, labour, urban, agriculture and rural development, human development, poverty, and consumers. The focus of NCAER's work is on generating and analysing empirical evidence to support and inform policy choices. It is also one of a handful of think tanks globally that combine rigorous analysis and policy outreach with deep data collection capabilities, especially for household surveys. More on NCAER is available on www.ncaer.org.

Designed by: How India Lives (www.howindialives.com)

NCAER SINCE 1956 QUALITY. RELEVANCE. IMPACT

NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH

NCAER India Centre, 11 Indraprastha Estate, New Delhi 110 002 (INDIA)

Tel: +91-11-2345 2657, 6120 2698, Email: info@ncaer.org

www.ncaer.org