



State of Aadhaar 2019: Detailed Methodology Note

Contents

- 1 Overview 3**
- 2 Pre-Analysis..... 5**
 - 2.1 In-Depth Survey 5
 - Questionnaire design 5
 - Sampling design..... 6
 - Pilots 8
 - Data collection and quality checks 8
 - 2.2 Pulse Survey 11
 - Questionnaire design 11
 - Sampling design..... 11
 - Pilots 12
 - Data collection and quality checks 12
 - 2.3 Human centred design research..... 13
 - Key research methods 13
 - Research locations 14
 - People we met..... 14
- 3 Analysis 16**
 - Quality Checks 16
 - Statistical Analyses 16
 - Business rules and assumptions 17
- 4 Post-Analysis 21**
 - Reporting..... 21
 - Internal and external validation..... 21
 - Dataset..... 24
- 5 Limitations..... 25**
- 6 Glossary of Terms 25**

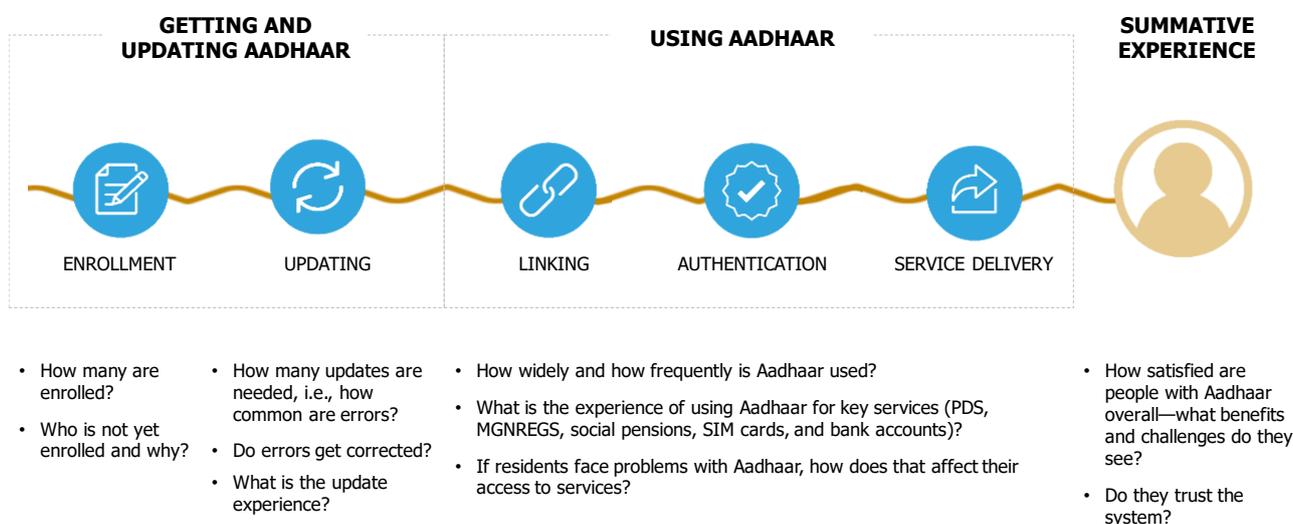
1 Overview

This document outlines the research design, methodology, and analysis for the State of Aadhaar 2019 study. Most elements of the research design were developed in late 2018 and early 2019. Data collection occurred between May and September 2019. Data cleaning and analysis occurred between September and November 2019.

The *State of Aadhaar 2019* report is the third entry in the *State of Aadhaar* series that launched in 2016. The objective of the 2019 edition is to generate datasets on people’s experiences with Aadhaar across India and promote a data-driven discourse on Aadhaar.

To that end, this edition followed a user journey framework to guide our enquiry on the following key research questions outlined in Figure 1.

Figure 1: User journey framework and key research questions



Through our research, we aimed to present a nation-wide view on the use of Aadhaar by the residents of India and how it impacts their lives. For that purpose, the study also followed those who dropped off at various points along this journey for reasons including but not limited to lack of Aadhaar, failure to link Aadhaar to service IDs and authentication failure.

This report relies on two nationwide surveys conceptualised, designed, managed, and analysed by the Dalberg team with support for data collection from specialist organisations. These are:

1. A pulse survey of a panel of 147,868 households in 28 states and union territories
State of Aadhaar added a 10-minute questionnaire to the triannual CHPS household panel survey conducted by the Centre for Monitoring Indian Economy (CMIE). These questions captured critical metrics on the adoption of Aadhaar and people’s experience around it. For part of this questionnaire (Aadhaar enrolment status, errors and some usage questions), we captured data for all household members, a total of 575,127 people. This survey was conducted between May and August 2019.
2. An in-depth 45-minute survey of 19,209 households across 16 states and 1 union territory
State of Aadhaar conducted a separate deep-dive survey to uncover the nuances of people’s experience with Aadhaar. It included a wide range of questions on enrolment, updates and use of Aadhaar for services—particularly three welfare services (PDS, MGNREGS and social pensions) and two non-welfare services (telecom and banking). Considering the Supreme Court ruling of 2018, the survey also included some questions on schools to test on-ground implementation of the court’s orders. This survey employed stratified random sampling across the country and oversampled populations who are typically not covered by household surveys.

Dalberg created the research design and questionnaire. Kantar Public conducted the survey between July and September 2019, with close day-to-day oversight from Dalberg. The raw data were exclusively accessed and analysed by Dalberg.

After completing data collection and preliminary analyses for the surveys, we followed up with Human Centred Design (HCD) research.

This helped us probe deeper into the survey findings and uncover the *why* and *how* behind some of the most interesting trends displayed by the data. We interviewed over 100 people in their homes, communities and workplaces using a range of HCD research methods, such as 1:1 in-depth interviews, intercepts, small-group discussions, activity-led sessions, and observational walk-throughs. This research took place in the states of Maharashtra, Karnataka, Assam and Bihar. These efforts enabled us to better understand the human stories behind the data and uncover further areas for research.

The surveys and report leaned on guidance and inputs from subject matter experts as well as the general public.

We were guided by two panels of independent experts as the study evolved:

1. The Technical Panel provided content and technical guidance through extensive reviews of our research design, questionnaire, analyses and report drafts. This Panel consisted of:
 - Dr Sonalde Desai
 - Dr Rinku Murgai
 - Dr Shamika Ravi
 - Dr Pronab Sen

2. The Advisory Panel helped us frame the report, identify the research questions and review report drafts. This Panel consisted of:
 - Yamini Aiyar
 - Dr Subhasis Banerjee
 - Rahul Matthan
 - Dr Anit N Mukherjee
 - Ananth Padmanabhan

Additionally, we incorporated inputs from more than 30 experts including government officials and policymakers, legal advisors, technologists, researchers, and others.

Dozens of residents shared their experiences and perspectives during our exploratory research and survey pilots, contributing greatly to our understanding of Aadhaar and its role in people's lives.

We thank all the panel members, experts and residents who engaged with us throughout the study, for their insights and perspectives. Any errors remain our own. We particularly thank the respondents of our surveys and HCD research.

This methodology note covers the pulse survey, the in-depth survey, and the HCD research, but focuses most on the in-depth survey.

2 Pre-Analysis

2.1 IN-DEPTH SURVEY

Questionnaire design

In the in-depth survey, we focussed on the following themes, with selected indicators highlighted below

- Getting Aadhaar (enrolment and updates)
 - Aadhaar enrolment
 - Reasons why residents have not enrolled in Aadhaar
 - Number of trips to enrolment centre
 - Number of trips to update centre
 - Aadhaar error rate (name, date of birth, gender, address, photograph, mobile number, biometric)
 - Whether it is easy or difficult to enrol in Aadhaar, update Aadhaar, link Aadhaar to social services (e.g., MGNREGS, social pensions), link Aadhaar to private services (e.g., DBT and bank account)
- Using Aadhaar
 - Aadhaar usage (e.g., MGNREGS, PDS, Social Security and Pensions, loans, insurance, applying for another ID, age proof)
 - Uses of different forms and features of Aadhaar (e.g., card, photocopy, fingerprint, QR code, OTP)
 - Number of authentication attempts for various services (e.g., MGNREGS, PDS, SIM card)
 - Reasons why residents are not receiving services (e.g., MGNREGS, PDS, school enrolment), categorized into Aadhaar and non-Aadhaar related reasons
 - Whether it is easier or more difficult to access welfare services (e.g., MGNREGS, PDS) and private services (e.g., DBT and bank account, SIM card) with Aadhaar
- Perceptions, satisfaction, and trust
 - Benefits and challenges of Aadhaar that are perceived as most important
 - Whether Aadhaar was used to gain access to services (e.g., bank account, SIM card, PDS) for the first time
 - Aadhaar trust rate
 - Aadhaar satisfaction rate
 - Whether residents prefer Aadhaar if given a choice of which ID to provide
- Awareness
 - Whether residents felt they understood processes around enrolment and update
 - Whether residents understood when Aadhaar is legally mandated (e.g., bank account, SIM card, PDS)
 - Reasons why residents provide Aadhaar for private services (e.g., the government made Aadhaar mandatory, the service provider only accepted Aadhaar, Aadhaar is my only ID or address proof)
 - Whether residents who encountered problems were aware of Aadhaar helplines and if they used them
- Children's access to education and welfare
 - Whether children experienced a delay in school enrolment due to Aadhaar-related reasons
 - Whether children were out of school due to Aadhaar-related reasons
 - Whether children missed out on mid-day meals due to difficulties with Aadhaar

Furthermore, we captured the following key descriptive variables of residents

- State and district
- Age
- Rural or urban
- Gender, including third-gender
- Religion

- Caste
- Homelessness
- Education level
- Occupation

The questionnaire was translated into 10 languages: Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Odiya, Telugu, and Tamil.

Sampling design

Stratified random sample

We conducted the survey across 16 states and one union territory, namely Andhra Pradesh, Assam, Bihar, Chhattisgarh, Delhi, Gujarat, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal. We selected these states to achieve regional representativeness, whilst maintaining a high degree of operational feasibility. Data collection was conducted between July and September 2019.

We first split the country into states, then NSSO regions, then districts, and further into urban wards and rural villages, based on the 2011 Census.

We selected the states for this study to cover all regions of India, a high share of the population of the country, and to ensure that we covered specific states of interest for our research question (e.g. Jharkhand where a substantial amount of research on Aadhaar has already been conducted).

Figure 1: Sample Size formula considered for state level

$$N = Deff \times \frac{Z^2 * (p) * (1-p)}{C^2}$$

N = the sample size

P = assumed to be 0.5 for yielding maximum sample size

Z = Standard normal value for one tailed test at 98% level of confidence, this value is 2.33

C = Margin of error 5% (standard value of 0.05)

D = Design effect, this value has been taken as 1.5 for multistage sampling

A provision of 20% additional sample has been kept for non-response at state level.

Keeping above statistical factors in consideration, the sample for each state is derived at 977 which is rounded off to 1000 for operational feasibility and rounding off of sample distribution at PSU level.

The sample for Delhi was 500. This was due to operational reasons and because socio-cultural variance was considered to be lower than in other states.

We divided each state into socio-cultural regions (SCR) according to National Sample Survey Office's (NSSO) regions and ensured that we covered each region in each state.

The number of districts selected from each SCR depended on the population proportion of each SCR within the state. We selected 5 districts per state using probability proportional to size. If a state had more than 5 SCRs then the number of districts chosen was equal to the number of SCRs.

In all but two states, we selected five districts per state. 20 wards or villages were selected in each districts. Each district was also split into four strata, namely 2 urban and 2 rural strata, based on the classification of wards and

villages in each district in the 2011 Census. Ten households were then randomly sampled in each ward or village. This was done by first segmenting each ward or village into equal segments, then (with help of the CAPI system) randomly selecting two segments and using a right-hand rule and skip to randomly select 5 households in each segment.

In the other two states, namely Madhya Pradesh and Maharashtra, we selected six districts per state, instead of five, because these states had six SCRs than that of other states. One district per NSSO zone was selected. Furthermore, in each district, 16-17 wards or villages were selected, instead of 20. This ensured that 100 wards and villages, representing 1,000 households, were sampled across each state.

Figure 2: Urban and rural strata

Urban strata	Urban/Town Class	Population	Rural strata	Rural Class	Population
Urban A	Class I (Tier 1)	> 1,00,000	Rural A	Class A	> 5,000
	Class II (Tier 2)	50,000 – 99,999	Rural B	Class B	1,000-4,999
	Class III (Tier 3)	20,000 – 49,999		Class C	< 1,000
Urban B	Class IV	10,000 – 19,999			
	Class V	5,000 – 9,999			
	Class VI	< 5,000			

One eligible respondent (over 18 years of age) from each household was randomly selected for the interview. In case there was more-than-one eligible respondent available in the household, the CAPI system randomly selected one individual based on a Kish grid algorithm.

Over-sampling

Beyond the above random stratified sample, we oversampled special groups, with the goal of interviewing least 400 elderly people (>70 years of age), 400 homeless people, 400 third-gender people, and 500 labourers. Oversampling targets were equally distributed across the 17 states and union territories and across the selected districts within them.

We set these targets in order to gather enough information on population groups that are traditionally harder to reach through a household survey. Specifically, Kantar Public undertook the following process.

On arriving in each ward and village, the field team checked with key informants whether there were any relevant settlements of special groups (e.g. third-gender communities, homeless settlements, construction sites with on-site temporary settlements). These settlements were chosen for over-sampling, if they were not already selected as part of the random stratified sample. Over-sampling of elderly respondents did not require visits to separate settlements.

If there were no settlements of particular special groups in that ward or village, the field team checked with key informants for nearby settlements and visited those for over-sampling.

This process was continued in each ward and village within each selected district of every state, until state-level targets for special groups were met.

Across the entire in-depth survey, a total of 845 elderly individuals, 459 people who identify as third-gender, 3,271 labourers, and 478 people who were homeless were included in the sample.

Given the method used to achieve over-sampling, our survey findings for these special groups are less representative and we did not create or use any survey weights (except for normalization by state population) to draw conclusions about these population groups.

Pilots

We conducted two pilots – the first with pen-and-paper questionnaires and the second on tablets using computer-assisted personal interviewing (CAPI) questionnaires.

The pen-and-paper pilot was conducted to test the flow of the questionnaire and whether the questions were easy to understand for the target group of respondents. They were performed in and around Mumbai and Delhi (in Marathi and Hindi respectively), totalling 20 respondents, in May 2019.

The CAPI pilots were conducted to test the translated and scripted questionnaires. Therefore, the CAPI pilots were conducted in ten states, in regions that predominantly spoke each language. Hindi was tested in Uttar Pradesh, Marathi in Maharashtra, Gujarati in Gujarat, Tamil in Tamil Nadu, Kannada in Karnataka, Telugu in Andhra Pradesh, Malayalam in Kerala, Bengali in West Bengal, Assamese in Assam, and Odiya in Odisha. These pilots were conducted in June, in a mix of rural and urban areas and included individuals from special groups. The sample for the pilot in each state was ten, totalling to 100 households overall. The feedback from these pilots was used to make the translations colloquial and adapt the CAPI script.

We also conducted two separate translation tests with native speakers of all 10 languages, run independently by Kantar Public and Dalberg. And Dalberg and Kantar Public both conducted extensive in-house script testing.

Data collection and quality checks

Kantar Public performed the data collection. During the data collection, quality checks were performed based on four key data processing indicators, namely the response rate, accessibility rate, replacement rate, and supervisor matching. They are as follows

Response rate (complete interviews)

The number of consents to participate would inform the response rate from the interview attempts. Over and beyond the number of interviews done, we further assessed the proportion of interviews conducted for the stratified random sample and over-sampling. Unmet quotas would be flagged.

Response rate (Stratified random sample)	Percentage	Count
Yes	97%	16,891
Interviewer consent on behalf of the respondent in case the respondent just wants to give verbal consent	2%	419
No	0%	28
Total	100%	17,338

Response rate (Over-sampling)	Percentage	Count
Yes	96%	1,724
Interviewer consent on behalf of the respondent in case the respondent just wants to give verbal consent	4%	70
No	0%	5

Total	100%	1,799
-------	------	-------

Accessibility rate

The number of successful respondents, accessible on the first attempt.

Accessibility rate (Total)	Percentage	Count
Yes	98%	19,171
No	2%	425
Total	100%	19,596

Besides the general outlook, this was further broken down into interviews from the stratified random sample and over-sampling.

Accessibility rate (Stratified random sample)	Percentage	Count
Yes	98%	17,398
No	2%	399
Total	100%	17,797

Accessibility rate (Over-sampling)	Percentage	Count
Yes	99%	1,773
No	1%	26
Total	100%	1,799

Replacement rate

Respondent replacement was tracked to assess the number of drop-offs at the consent question. These interviews were then checked across the enumerators with the aim of flagging alarming patterns such as numerous replacements by certain enumerators. Incorrect replacements by gender and age group were also flagged.

Replacement rate	Percentage	Count
Yes	2%	344
No	98%	17,453
Total	100%	17,797

Supervisor matching and backchecking

Over 25% of the enumerator interviews were validated through a tool to check for matches for a set of factual questions.

Rate of supervision	Percentage	Count
No	69%	13,546
Yes	31%	6,050
Total	100%	19,596

The data collected were then uploaded from the field immediately after completion of interviews (or if connectivity was weak, at a minimum daily) to Dalberg's servers, who had exclusive access to the raw responses. The data were cross-checked centrally to ensure plausibility of the entries (e.g. checking skip logics of questions).

2.2 PULSE SURVEY

Questionnaire design

In the pulse survey, we focussed on a subset of key themes covered in the in-depth survey. This not only allowed us to use the pulse survey to corroborate our findings in the in-depth survey, but also to generate more reliable statistics given the pulse survey larger sample size. The key themes are

- Aadhaar enrolment
- Aadhaar error rate (name, date of birth, gender, address, photograph, mobile number)
- Aadhaar usage (e.g., MGNREGS, PDS, Social Security and Pensions, SIM card, age proof)
- Whether it is easier or more difficult to access welfare services (MGNREGS, PDS, Social Security and Pensions) due to Aadhaar
- Reasons why residents are not receiving welfare services (MGNREGS, PDS, Social Security and Pensions), categorized into Aadhaar and non-Aadhaar related reasons
- Satisfaction with Aadhaar

Furthermore, in addition to demographic variables obtained in the in-depth survey (e.g., age, gender, education level), we also obtained important financial and economic inclusion variables through the pulse survey. These include

- Financial inclusion (e.g. whether resident has credit card, bank account, provident fund account, life insurance, savings in any source, and mobile phone)
- Economic inclusion (e.g. total household income, household income from government transfers, total household expenditure, total household health expenditure)

Sampling design

CMIE conducted the survey across the country, except for some relatively small north eastern states, namely Arunachal Pradesh, Nagaland, Manipur, Mizoram, Andaman & Nicobar Islands, Lakshadweep, Dadra & Nagar Haveli, Daman & Diu. It was conducted between May and August 2019.

CMIE deployed a stratified multi-stage survey design draw its sample of households, with weights based on the projected population as of May 2019.

CMIE stratified the country into Homogeneous Regions (HRs), which is a set of neighbouring districts that have similar agro-climatic conditions, urbanisation levels and female literacy, approximately of the same size.

For urban areas, HRs were split further into four strata of towns, namely small towns (less than 20,000 households), medium-size towns (20-60,000 households), large towns (60-200,000 households), and very large towns (more than 200,000 households). CMIE then randomly selected towns from each of these strata, for a total of 329 towns. Within each town, CMIE randomly selected an average of 21 Census Enumeration Blocks (CEBs), which represented clusters of about 100-125 neighbouring households, for a total of 7,920 CEBs. 110,975 urban households were then randomly sampled from these CEBs.

For rural areas, HRs were split into rural villages, based on the 2011 Census, for a total of 3,965 villages. 63,430 rural households were then randomly sampled from these villages.

Overall, 110,975 urban households and 63,430 rural households were sampled, of which a total of 147,868 households representing 575,127 people responded.

For more information on sampling design, visit CMIE's website [here](#).

Pilots

Once the questionnaire was finalised, CMIE translated the questions into ten regional languages and developed the software necessary to conduct the survey in CAPI. The software was tested in-house and the survey team comprising interviewers and supervisors were trained to administer the Aadhaar-related questions. CMIE then conducted a pilot using its field team to test the software, efficacy of the questions, and complications faced by respondents in understanding these questions.

The pilot was conducted over a period of over two weeks in April, conducted in CAPI form. Responses were captured in the GPS-enabled electronic devices using a specially developed software for the purpose.

It was conducted in all states, except the few states not surveyed by CMIE. This yielded data from 10,253 households, comprising 3,073 rural households and 7,180 urban households. This pilot sample size was large enough and diverse enough geographically to provide useful insights on the efficacy of the questions.

The pilot yielded useful insights regarding the performance of several questions. These insights led to modifications in the questionnaire. In some cases, questions that were addressed to individuals became questions for the household (e.g., use of Aadhaar for PDS). In other cases, the nomenclature was changed (e.g., banking and payments was changed to two options – opening a bank account and making payments). Furthermore, some open-ended questions became specific questions (e.g., other reasons became “age proof” and “land/vehicle/house/marriage registration”).

Data collection and quality checks

CMIE performed the data collection and quality checks. Specifically, data collection was performed on hand-held GPS enabled phone devices. During the data collection, quality checks were performed using multiple tools, including GPS location tracking, supervisor data-checks, call-backs, and ground supervision. GPS location tracking and supervisor data checks were conducted on 100 per cent of the respondents. Both were conducted in near real-time by supervisors close to the field execution using specially developed software. Call-backs were made regularly to random sets of households.

The survey was organised such that all data checks of households surveyed during a day are completed by the end of the day of the survey. This ensured that all clarifications are sought and obtained while the team is still in the field.

This enabled CMIE to provide a website to Dalberg that was updated every day with the latest updates on survey execution and outcomes in terms of distribution of the responses for individual questions. This transparency of operations was also an indirect tool to ensure quality.

The response rate is as follows

Response rate (complete interviews)

Response rate	Percentage	Count
Yes	85%	147,868
No	15%	26,537
Total	100%	174,405

2.3 HUMAN CENTRED DESIGN RESEARCH

Human centred design (HCD) research uses a combination of design and qualitative research methods with an aim to deeply understand users: their underlying needs, motivations, behaviours and aspirations. We used HCD research to strengthen and deepen the survey insights by providing the “why” and “how” behind it at a smaller scale. It was conducted after the first phase of survey analysis was completed

We further used HCD research to identify and fill in gaps in interpreting some of the survey data, such as highlighting the nuances around people’s sentiment towards Aadhaar. We were also able to leverage the research themes to dive deeper into some new areas such as the service providers’ perspective, as well as uncover a few new learnings that could drive further areas of research.

We interviewed 103 individuals across the 4 states of Maharashtra, Karnataka, Bihar and Assam. While the sample size in our study is small by design, its strength lies in tracing common themes and patterns across a diverse range of Aadhaar users. For example, we were able to explore what is the enrolment experience for an ‘average’ user of Aadhaar: men and women across rural and urban divides, with differing education levels and socio-economic backgrounds; vs those with special needs such as an elderly or a disabled person; and beyond to those that live at the margins of society such as a third-gender or homeless person. We were able to listen deeply to the everyday experiences of these diverse people with Aadhaar, from their sources of satisfaction towards Aadhaar to their experience of denial.

We interviewed people in their homes, communities, and workplaces using a range of HCD research methods. This approach allowed us to not just understand what people ‘think, feel and say’ about Aadhaar, but also observe what they ‘do’ through real life observations of Aadhaar enrolment and update centres, ration shops, SIM card shops, etc. This study by no means claims to be exhaustive in capturing all of the people’s sentiments and experiences with Aadhaar. However, it does aim to strengthen and deepen our survey-based understanding of the ‘State of Aadhaar’, as experienced by its everyday users.

Key research methods

We employed a mix of four distinct research methods: in-depth interviews, intercept interviews, small group discussions and observation sessions.

- **28 in-depth interviews** with pre-recruited participants, to observe and document attitudes and behaviours around Aadhaar. Each in-depth interview was ~2 hours long.
- **Small group discussion sessions** with 6 to 7 participants each, to understand diverse opinions, preferences and needs of a representative set of participants. Each group discussion was ~3 hours long.
- **50 intercept interviews** for quick, additional insights, conducted in a variety of settings. These were spontaneous conversations with non-recruited participants, and allowed us to connect with people while they were immersed in specific contexts. Each intercept interview was 15 to 30 minutes long.
- **Observational walk-throughs** which involved shadowing select stakeholders in their natural environments (e.g. e-seva kendras, ration shops, schools, SIM card shops, in both rural and urban areas), to get a nuanced understanding of their activities and interactions. Each observational walk-through was 0.5 to 1 hour long.

The table below shows the break-up of participants across interview types:

Method	Number of participants
In-depth interviews	28
Small group discussions	25
Intercept interviews	50
TOTAL	103

Research locations

We selected locations that would allow us to learn from diverse contexts. We considered two key factors:

Geographic diversity across India, with representation from:

- The North (the Hindi-belt)
- The South
- The Northeast
- The West

Insights from survey data, with a focus on states that had interesting or contrasting findings across metrics like:

- Aadhaar penetration
- Awareness around Aadhaar
- Access to and exclusion from services due to Aadhaar
- Trust in Aadhaar, among others

Figure 3: Data used to shortlist HCD locations (Source: Dalberg State of Aadhaar survey)

Location	Geographic region	Penetration	Awareness around Aadhaar	Tried to enrol but failed	Denial of midday meals	Exclusion from PDS (ration)	Satisfaction with Aadhaar	Trust in Aadhaar
1. Bihar	North	95.5%	Low	Very high	Low	High	High	Medium
2. Karnataka	South	98.2%	Very high	High	Very high	Medium	Medium	Medium
3. Assam (only those who have Aadhaar)	Northeast	18%	Very low	Low	Very Low	Very low	Low	Low
4. Maharashtra	West	95.5%	Low	Medium	Medium	Medium	High	High

Note: Other states on our shortlist included Rajasthan (high performance with Aadhaar and related services), Orissa (high performance with Aadhaar and related services, but poor grievance redressal), Jharkhand (poor performance with Aadhaar and related services but already heavily researched), Kerala (moderate performance with Aadhaar and related services), Meghalaya (poor penetration of Aadhaar but not part of Dalberg's survey). These were eliminated to ensure geographic diversity and cover a range of interesting insights from the survey data.

People we met

Participants were recruited for both in-depth interviews and small group discussions. We focused on participants with a range of experiences, diverse backgrounds, varying literacy levels and age groups. We laid down precise baseline criteria for participant recruitment:

Gender: Equal representation of male and female, and a few third gender participants

Age: Mix of age groups (18+ years)

Location: Equal representation of rural and urban participants

Income levels: Mix of income levels with a focus on low income consumers. We did not have a strict guideline for SEC classification, but made sure there was a preference for SEC C, D and E, with no more than 2 people from SEC A and SEC B each per region

Professions: Mix of professions, such as farmers, daily wage labourers, entry level service professionals, taxi/auto drivers, tailors, women from SHGs (self help groups), women involved with MFIs (micro finance institutions). We also included interviews with migrants and homeless people.

Family structure: Mix of joint family, nuclear family, families with and without children, people living alone, etc.

Religion: Mix of religious beliefs- Hindus, Muslims, Sikhs, Christians, etc.

Caste: Mix of people from general category, ST, SC and OBC

Literacy: Mix of literacy levels– graduate, diploma, high/secondary/primary school, no formal education

Mobile and internet usage: Mix of no mobile, feature phones and smartphone owning participants. Mix of participants who use the internet and those who don't.

Recruiters filled a screener document for each participant prior to the interview, to ensure a sufficiently diverse set of people. The screener includes a high degree of specificity (e.g. access to mobile phone, including type of mobile phone), so that we could cross-check and verify the screener responses with the interview participants themselves.

In addition to the demographic characteristics listed above, we detailed illustrative participant profiles, detailing specific behaviours and experiences, for the recruiters to keep in mind while selecting people. This was to ensure that we included both the average and the more 'extreme' experiences in enrolling in, updating and using Aadhaar. Some of the illustrative participant profiles were:

- Someone who is homeless and has Aadhaar
- A migrant worker/daily wage labourer who has been unable to get an Aadhaar/unable to access services due to Aadhaar related issues
- A third gender individual who doesn't have Aadhaar and has never tried to enrol
- Someone who was able to apply for agricultural/fertilizer subsidy due to Aadhaar
- Someone who started receiving pension due to Aadhaar
- Someone who has not received their pension due to Aadhaar related issues
- Someone who was denied a service due to Aadhaar
- Someone whose biometric/fingerprint authentication failed while using Aadhaar
- Someone who was denied ration due to Aadhaar related problem
- Someone whose child has benefited from the midday meal scheme for the first time post-Aadhaar
- Someone whose child has been denied enrolment in school due to lack of Aadhaar/Aadhaar related problem

3 Analysis

We followed a hypothesis-led, statistically rigorous, and reproducible approach towards analysing the data. Initial hypotheses - based on expert interviews and secondary research - were iterated upon, refined, and tested. While doing so, we strived to achieve rigor and transparency, performing comprehensive quality checks, using commonly accepted statistical tools, and publishing our assumptions and business rules in an open-source manner.

Quality Checks

We conducted comprehensive quality checks.

We checked for inconsistencies within responses. For example, we asked residents for their first, second, and third-ever ID, among a list of IDs, such as an Aadhaar card, PDS ration card, and MGNREGS job card. We also asked residents for the exact date that they received these IDs. Therefore, we performed quality checks to ensure that what respondents reported as their first, second, and third-ever ID matched to the dates when they reported receiving these IDs. No inconsistencies were identified.

In addition, we checked and removed some outliers in the in-depth survey, following the rules outlined in Table 1. We also checked for outliers in the pulse survey, analysing statistics such as household income, household expenditure, and government transfers, among others. No outliers were identified in the pulse survey.

Table 2: List of outlier thresholds

#	Metric	Maximum threshold	Units
1	Total area of the PSU	1000	sq. kms
2	Number of Aadhaar enrolment centres that are present in the village/ward	1000	centres
3	No of trips required for linking or accessing a service (such as PDS, MGNREGS, SIM cards) with Aadhaar	50	trips

Statistical Analyses

We performed rigorous statistical analyses to identify the most salient insights from the data.

We first developed a set of hypotheses, based on expert interviews and secondary research, which formed the foundation of our analyses. This limited the scope for data dredging and reduced the risk of identifying spurious correlations in the data. We then undertook analyses that allowed us to identify statistically significant findings, which accounted for the effects of stratification, clustering, sample weights, and confounding variables.

In the pulse survey, we accounted for clustering at a town and village level, along with stratification at the level of homogeneous regions, to prevent spurious correlation. We also used sampling weights to ensure that our findings were nationally representative. Furthermore, we performed linear regressions to control for variables, such as gender, religion, caste, and years of education, to identify spurious correlations from confounding variables.

In the in-depth survey, we accounted for stratification, clustering, and sample weights. We calculated state-level normalised weights for the over-sampled dataset for increased precision of the results. Moreover, given the in-depth survey consists more granular data relative to the pulse survey, we also used non-linear tools such as classification and regression trees (CART) to identify homogenous segments of key indicators, to capture non-linear relationships in the data.

Statistics behind our analyses

In general, we follow commonly-accepted statistical methods, outlined in the Lumley (2019) through the “survey” package in R. For details, see <https://cran.r-project.org/web/packages/survey/survey.pdf>.

We first calculated standard errors using the delta method or from re-sampling, to account for the clusters, strata, and weights of the survey design. We then used Welch’s unequal variances t-test, to test if differences between averages of two groups are statistically significant. We also used a generalized linear model to incorporate these design-based standard errors and inverse-probability weighting, to test for statistical significance while controlling for confounding variables. Finally, we used classification and regression trees to create decision trees that outline sequences of criteria (e.g. whether a resident is literate or illiterate) to estimate different statistics (e.g. whether a resident has Aadhaar or not), to identify homogenous segments of key indicators.

Business rules and assumptions

We followed certain assumptions in the defining variables and segments of interest in our analysis, listed in Table 2.

Table 3: Description of business rules

#	Metric	Description of assumption
SoA In-depth survey		
1	First time access to a bank account, after getting Aadhaar	YES , if the respondent has Aadhaar, they have a bank account currently, they got Aadhaar before the first bank account, and used Aadhaar for their most recent access to the account
2	First time access to a SIM card, after getting Aadhaar	YES , if the respondent has Aadhaar, they have a SIM currently, they got Aadhaar before the first SIM card, and used Aadhaar for most recent access to a SIM.
3	First time access to MGNREGS, after getting Aadhaar	YES , if the respondent has Aadhaar, they have MGNREGS currently, they got Aadhaar before MGNREGS, and used Aadhaar for most recent access to the service.
4	First time access to PDS after getting Aadhaar	YES , if the respondent has Aadhaar, they currently receive PDS rations, they got Aadhaar before getting PDS, and used Aadhaar for most recent access to the service.
5	First time access to social pensions, after getting Aadhaar	YES , if the respondent has Aadhaar, they get social pensions currently, they got Aadhaar before getting the pensions, and used Aadhaar for most recent access to the service.
6	First time access to any service, after getting Aadhaar	YES , if any of the above five assumptions holds
7	Trust that Aadhaar prevents unauthorised access to my money	NO , for people who think it is easier for others to access money in their name due to Aadhaar, while YES for people who think it is more difficult

#	Metric	Description of assumption
8	Wanted to enrol into Aadhaar	YES , for everyone who has Aadhaar or everyone who has tried to get Aadhaar. Amongst those who have never tried, everyone except those who explicitly said, " <i>I do not wish to enrol in Aadhaar</i> " or " <i>I don't need Aadhaar</i> ". In addition, people are assumed to want to enrol if, they said " <i>I do not wish to enrol in Aadhaar</i> ", AND " <i>I want to enrol, but I haven't gotten around to it yet</i> " OR " <i>It's too difficult to enrol (e.g. takes too much time, is too expensive)</i> "
9	User tried to enrol into Aadhaar	YES , for everyone who has Aadhaar or everyone who has tried to get Aadhaar.
10	Error in Aadhaar	YES , if individuals have Aadhaar and there is a mistake in any field collected for Aadhaar (i.e. Name, Date of Birth, Address, Photo, Biometrics, Gender, Mobile number, or Biometrics). The card error rate does not include errors in mobile number or biometrics.
11	User wanted to update Aadhaar	YES , if individuals ever wanted to update any errors in data collected for Aadhaar (i.e., Date of Birth, Address, Photo, Biometrics, Gender, or Mobile number) OR tried to update these fields
12	User tried to update Aadhaar	YES , if an individual tried to update any data collected for Aadhaar (i.e. Name, Date of Birth, Address, Photo, Biometrics, Gender, or Mobile number)
13	Frequency of usage of Aadhaar	The <i>minimum</i> frequency usage of Aadhaar is estimated as the maximum of number of times people reported using any single Aadhaar feature (Aadhaar card, Aadhaar number, Fingerprint, Iris scan, Photocopy, mAadhaar, Masked Aadhaar, Virtual Aadhaar number, Aadhaar OTP, and Aadhaar QR code) in the last three months The <i>maximum</i> frequency usage of Aadhaar is estimated as the <i>sum</i> of number of times people reported using any Aadhaar feature in the last three months
14	Used newer digital features of Aadhaar	YES , if an individual has used either mAadhaar, Masked Aadhaar, Virtual Aadhaar number, or Aadhaar QR code in the last three months
15	Exclusion from services: PDS	YES , if an individual is eligible to receive the service (eligibility is self-reported), did not have a PDS ration card due to an Aadhaar-related reason, or their ration card was cancelled due to an Aadhaar-related reason. We also calculated the corollary for non-Aadhaar related reasons
16	Exclusion from services: MGNREGS	YES , if an individual is eligible to receive the service (eligibility is self-reported), did not have a MGNREGS job card due to an Aadhaar-related reason, or their job card was cancelled due to an Aadhaar-related reason. We also calculated the corollary for non-Aadhaar related reasons
17	Exclusion from services: Social pensions	YES , if an individual is eligible to receive the service (eligibility is self-reported), did not have social pension due to an Aadhaar-related reason, or their pension was invalidated due to an Aadhaar-related reason. We also calculated the corollary for non-Aadhaar related reasons

#	Metric	Description of assumption
18	Exclusion from services: SIM	YES , if an individual did not have a phone due to an Aadhaar-related reason. We also calculated the corollary for non-Aadhaar related reasons
19	Exclusion from services: Bank account	YES , if an individual did not have a bank account due to an Aadhaar-related reason. We also calculated the corollary for non-Aadhaar related reasons
20	Exclusion from any service	YES , if the individual was excluded from any services mentioned above
21	Exclusion from any welfare service	YES , if the individual was excluded from any welfare service mentioned above (PDS, MGNREGS, Social pensions)
22	Denial from services: MGNREGS	YES , if the individual has a valid MGNREGS job card, tried to get a MGNREGS job in the last year, but did not receive their wages due to an Aadhaar-related reason (such as issues with linking Aadhaar with the job card, linking Aadhaar to the bank account etc.). We also calculated the corollary for non-Aadhaar related reasons
23	Denial from services: Social pensions	YES , if the respondent has a did not get their pension the last time they expected to receive it due to an Aadhaar-related reason (such as issues with linking Aadhaar with the bank account etc.). We also calculated the corollary for non-Aadhaar related reasons
24	Denial from services: PDS	YES , if the respondent has a valid PDS card, tried to access the service, faced biometric authentication failure the last time they tried to access the service, and did not get their rations. The corollary is if an individual faced a failure in biometrics the last time they tried, but received their allocated rations
25	Denial from any service	YES , if they were denied any service mentioned above
26	SC judgement date (2018)	We take the date of 30th September 2018 as the cut off for the impact of the Supreme Court verdict
27	SECC Classification	<p>Households are classified as automatically excluded from certain welfare benefits if they match one of the following conditions; 1) owns a motorized vehicle or boat, 2) has a mechanised three or four-wheeler agricultural equipment, 3) has a kisan credit-card with a limit of INR 50,000 and above, 4) respondent is government employee, 5) owns a non-agricultural enterprise registered with the government, 6) any member earns INR 10,000 or above per month, 7) any member pays income or professional taxes, 8) has a refrigerator, 9) owns a landline, 10) lives in a house with more than three pucca rooms, 11) has more than 5 acres of irrigated land, or 12) has more than 7.5 acres of total land.</p> <p>Among those who are left, households are classified as automatically included (eligible for certain welfare benefits) if they 1) are homeless / do not have a shelter, 2) live on alms / begging / manual scavenging, 3) contains a legally released bonded labourer, or 4) vulnerable tribal groups.</p> <p>Households that fall into neither are classified as general.</p>

#	Metric	Description of assumption
SoA pulse survey		
28	Net user satisfaction	Difference between (1) total proportion of users who say, " <i>somewhat satisfied</i> " or " <i>very satisfied</i> " and (2) total proportion of users who say " <i>somewhat dissatisfied</i> " or " <i>very dissatisfied</i> "
29	Error in Aadhaar card	YES , if individuals have Aadhaar and there is a mistake in any field printed on an Aadhaar card (i.e. Name, Date of Birth, Address, Photo, or Gender)
30	Denial of service due to Aadhaar-related reasons	YES , if respondent said that they could not access Aadhaar due to Aadhaar-related reasons, namely (1) "We are eligible but could not link Aadhaar to the service" or (2) "We are eligible but biometric authentication didn't work (fingerprint or iris scan)" for any one of these three welfare services: PDS, MGNREGS, and SS and Pensions.
31	Population size estimates	All non-respondents (people who refused to participate) were excluded from the data, reducing sample size by ~15%. Survey weights were rescaled by a factor of 1.2 to 2019 population estimate of 1.34 billion (assuming non-respondents are randomly distributed).

4 Post-Analysis

Reporting

In general, if a statistic can be estimated from both the pulse and in-depth surveys, we report numbers from the larger pulse survey whenever possible. Most findings, however, are from the in-depth survey since the survey is longer and covers more topics. Statistics from the pulse survey are explicitly referenced in the footnotes of the report.

For national-level projections from the pulse survey, such as the Aadhaar enrolment rate in India, we assume that the estimates from states where we conduct our survey can be projected to a national level. Thus, for the in-depth survey, we project statistics based on 16 states and one union territory to those states and the union territory, while for the pulse survey, we project statistics based on 28 states and union territories to the national level.

As far as possible and where relevant, we report only significant relationships and insights in our report, with a minimum of a 5% level of significance. In many cases, given the sample sizes for the surveys, the level of significance was higher – and in some cases less than 1%. We do not estimate family error rates.

Internal and external validation

Internal validation

We compared our results on satisfaction rates, enrolment rates, and error rates between the pulse and in-depth surveys, listed in Tables 3, 4, 5, and 6. The correlations between these results are usually high (greater than 75%). Some differences inevitably exist, which can arise due to changes in question phrasing, sample size and weights.

Table 3: Comparison of results on enrolment rates by demographic groups between SoA pulse and in-depth surveys

Group	SoA Pulse	SoA In-depth
By age		
Children 0 - 5	38.9%	48.2%
Children 6 - 17	80.9%	88.3%
Adult	95.5%	96.1%
By gender		
Male	90.5%	90.7%
Female	90.4%	90.5%
By region		
Urban	94.0%	92.9%
Rural	88.7%	89.1%

Table 4 Comparison of results on card error rates between SoA pulse and in-depth surveys

Error type	SoA Pulse	SoA In-depth
Aadhaar card	4.04%	10.40%
Name	0.50%	3.61%
Date of Birth	2.56%	4.39%
Gender	0.35%	0.89%
Address	0.58%	2.09%
Photo	0.54%	1.28%

Table 5: Comparison of results on satisfaction rates by state between SoA pulse and in-depth surveys

State	SoA Pulse	SoA In-depth
India	91.6%	79.7%
Andhra Pradesh	98.2%	86.0%
Assam	72.3%	42.9%
Bihar	93.5%	85.4%
Chandigarh	95.8%	NA
Chhattisgarh	96.8%	82.7%
Delhi	99.5%	74.5%
Goa	64.7%	NA
Gujarat	99.9%	71.5%
Haryana	98.9%	NA
Himachal Pradesh	99.7%	NA
Jammu and Kashmir	86.3%	NA
Jharkhand	98.3%	93.7%
Karnataka	85.0%	67.6%
Kerala	65.3%	47.0%
Madhya Pradesh	87.7%	77.3%
Maharashtra	92.5%	92.2%
Meghalaya	0.0%	NA
Odisha	99.2%	97.5%
Puducherry	78.3%	NA
Punjab	94.6%	77.2%
Rajasthan	90.0%	51.6%
Sikkim	98.0%	NA
Tamil Nadu	83.0%	76.9%
Telangana	96.7%	NA
Tripura	65.3%	NA
Uttar Pradesh	93.7%	86.1%
Uttarakhand	98.4%	NA
West Bengal	90.7%	84.1%

Note: The pulse survey used a 5-point scale. The in-depth survey used a 3-point scale.

External validation

We also compared our results on enrolment rates and error rates among the pulse survey, in-depth survey, UIDAI, and State of Aadhaar 2017-2018 survey, listed in Tables 6 and 7. The correlations among these results are high.

Table 6 Comparison of results on enrolment rates among SoA pulse and in-depth surveys, and UIDAI

State	SoA Pulse	SoA In-depth	UIDAI
India	90.5%	90.5%	89.2%
Andhra Pradesh	99.8%	96.0%	91.3%
Assam	10.1%	14.5%	15.1%
Bihar	89.5%	83.8%	83.3%
Chandigarh	99.1%	NA	99.0%
Chhattisgarh	99.2%	96.3%	93.9%
Delhi	96.7%	92.9%	116.7%
Goa	95.7%	NA	100.9%
Gujarat	96.9%	93.0%	95.3%
Haryana	94.2%	NA	103.2%
Himachal Pradesh	99.6%	NA	101.9%
Jammu and Kashmir	83.4%	NA	75.6%
Jharkhand	87.5%	93.1%	91.0%
Karnataka	94.4%	96.3%	92.7%
Kerala	97.0%	97.0%	102.2%
Madhya Pradesh	78.0%	95.4%	88.9%
Maharashtra	97.4%	94.0%	93.0%
Meghalaya	39.5%	NA	28.8%
Odisha	96.1%	95.4%	92.7%
Puducherry	99.4%	NA	92.3%
Punjab	96.6%	97.8%	101.2%
Rajasthan	83.1%	91.0%	86.2%
Sikkim	91.6%	NA	85.1%
Tamil Nadu	97.4%	95.9%	93.3%
Telangana	99.8%	NA	100.5%
Tripura	97.8%	NA	88.8%
Uttar Pradesh	91.1%	89.5%	86.7%
Uttarakhand	85.4%	NA	98.9%
West Bengal	92.8%	93.4%	92.9%

Table 7 Comparison of results on common indicators between State of Aadhaar 2017-18 survey and State of Aadhaar 2019 survey

Metric	SoA 2017-18			SoA 2018-19			
	Andhra Pradesh	Rajasthan	West Bengal	National	Andhra Pradesh	Rajasthan	West Bengal
Aadhaar enrolment rate	97.6%	87.5%	93.2%	90.4%	99.8%	83.1%	92.8%
Ease of enrolment for Aadhaar							
% who found the process easy	84.7%	84.4%	77.3%	78.4%	82.7%	59.7%	93.1%
% who found the process neutral	6.3%	5.7%	11.1%	15.1%	11.3%	37.1%	5.1%
% who found the process difficult	9.0%	9.8%	11.6%	5.9%	5.1%	2.8%	1.4%
Aadhaar usage							
Photocopy	93.4%	96.8%	93.9%	72.4%	66.7%	70.2%	70.3%
Card	84.7%	39.9%	24.5%	83.7%	89.6%	80.4%	87.3%
Fingerprint	95.4%	76.7%	48.8%	42.6%	44.2%	44.2%	22.6%
Iris	26.4%	1.0%	2.2%	10.6%	17.4%	10.5%	1.8%
One-Time-Password	2.0%	2.0%	9.3%	9.8%	13.0%	12.3%	8.6%
Used Aadhaar for Bank Account Openings	94.7%	82.2%	75.5%	86.0%	84.4%	94.2%	88.8%
Used Aadhaar for SIM Card	94.4%	83.4%	78.3%	54.6%	44.4%	72.3%	45.0%
Errors and denials							
Card Error Rate	8.0%	4.8%	12.2%	5.0%	0.8%	5.2%	2.7%
Denial of PDS rations due to Aadhaar-related problems	0.8%	2.9%	0.9%	4.7%	0.2%	7.1%	0.0%
Ease of PDS Delivery due to Aadhaar							
% who found it became easier after Aadhaar	55.3%	67.4%	NA	67.9%	72.5%	68.8%	40.3%
% who found it became more difficult after Aadhaar	36.5%	22.2%	NA	2.4%	0.7%	1.1%	0.3%

Dataset

To promote transparency and reproducibility, we have published the raw datasets from both surveys, which are freely accessible at www.stateofaadhaar.in. We have also provided guides, such as a variable codebook, and a "read me" for the dataset.

5 Limitations

We recognize several limitations in using surveys. In particular, surveys do not establish causality and are prone to numerous biases, such as acquiescence bias, recall bias, social desirability bias, and non-response bias.

As far as possible, we have attempted to mitigate these limitations. For example, we conducted extensive enumerator training on survey biases. Furthermore, we focused only on questions that residents were able to answer credibly through a survey format. Moreover, we supplemented our survey findings with insights from human centred design research with over 100 people in Maharashtra, Karnataka, Assam and Bihar, which included 1:1 in-depth interviews, intercepts, small-group discussions, activity-led sessions, and observational walk-throughs. This helped us to develop a more nuanced understanding of our key findings.

Despite our efforts, we recognize that some biases will inevitably remain. Our survey findings should therefore be considered with these inherent limitations in mind.

6 Glossary of Terms

Term	Definition
Denial of service	We define denial of service as not receiving a service (as reported by the respondents) at the time of expected service delivery (such as collection of rations).
Error rate	The incidence of error in people's Aadhaar information. We report error rates for data printed on the card (name, date of birth, sex, address, photograph) and for some data that is captured but not printed on the card (mobile phone, biometrics). Error rates are self-reported.
Exclusion	We define exclusion from a service as lack of access to the service because of inability to enrol or cancellation of existing enrolment
Net satisfaction	Share of people who are satisfied minus share of people who are dissatisfied
Resident	Any person living in India. Our study cannot establish <i>legal</i> residence status