The Great Indian Identity Crisis?

Exclusions & Intersectionality in the Indian Aadhaar System

Aditi Ramakrishnan

May 2021

University of California, Berkeley

Department of Economics

Undergraduate Honors Thesis

Advised by Professor Benjamin Faber

Abstract

As India leads the charge in digitizing its ID system under the biometric Aadhaar umbrella, this paper studies which sections of the population are being left behind in this process. I study the unique experiences of women, 'lower' caste members, and particularly the intersection of the two – lower caste women – in enrolling in Aadhaar as well as availing services using the ID. I find that women are more likely to enroll in Aadhaar by 1.8 percentage points, but less likely to later use the ID to avail financial, educational, political, and nutritional benefits. Conversely, lower caste members show mixed results. Lower caste women show the exact opposite pattern: lower association with Aadhar enrollment, but higher likelihood of use to avail financial, nutritional, and educational services. One caveat, however, is that this analysis does exclude the experiences of people who do not identify with the binary genders outlined above, leaving room for further study. Further, a follow-up survey would facilitate panel data creation to perform a diff-in-diff and assign causality.

I. Introduction

In the 21st century, having access to identification mechanisms is a central tenet of the human societal experience. Alongside the psychological validation of one's identity and their belonging within society, identification provides a pathway to access various services and fundamental rights, including political (the right to vote), financial (banking and loan services), and educational (school and college enrollment). Indeed, so fundamental is identification to development that the United Nations' Sustainable Development Goal (SDG) Target 16.9 reads "to provide legal identity for all, including birth registration" by 2030.

Even as countries roll out identification systems faster than ever, developing countries are leading the charge in this endeavor. As Gelb and Metz 2018 show, over 60 developing countries have established national identification systems since 2000, harnessing digital technology and biometric identifiers to do so. Digitally forward policy has been hailed as a potential equalizer in this landscape for being unbiased, accessible, and accurate. In line with these benefits, India introduced the Aadhaar program in 2009. As the world's largest biometric ID system, it would provide all citizens who voluntarily signed up with a 12-digit unique ID number which could then be used to access a range of services, in contrast to the prior system of requiring separate IDs for each type of service. For instance, in the past, Indian citizens would require an electoral ID card to cast their vote, a ration card to access food support mechanisms, an Indian passport to prove citizenship etc. The Aadhaar number, issued by the UIDAI (Unique Identification Authority of India), is a non-mandatory attempt to unify access to all such services under one umbrella. In other words, enrollment in the Aadhaar system, while useful, is entirely voluntary.

Per the UIDAI website, Aadhaar "does not profile people based on caste, religion, income, health, and geography", and it "facilitates financial inclusion of the underprivileged and weaker sections of the society and is therefore a tool of distributive justice and equality". This paper aims to study these statements in some depth, quantifying the differences in experience with the Aadhaar program along gender and caste lines.

In 2019, Dalberg, a development consulting company, conducted a nationwide audit of the Aadhar program by means of a pulse survey of 147,868 respondents as well as an in-depth survey of 19,209 respondents. Through both of these mediums, the study brought to light several promising findings, including 95% coverage by Aadhaar among adults in India, improved service delivery for 80% of beneficiaries, and trust in the system among 90% of respondents. However, it also found exclusions along income level and gender lines. In this paper, I explore the gender dimension further to study the enrollment rates of men versus women in the Aadhaar system, as well as the next step: access to services upon enrolling. In addition – and particularly in light of recent sociopolitical events in the country – I study the same two things along caste lines, exploring whether belonging to a Scheduled Caste (SC), Scheduled Tribe (ST), or Other Backward Caste (OBC) has any impact on enrollment and access to benefits. Finally, I explore the intersectionality of gender and caste to see if there are exacerbated effects, or a so called "double impact", for minorities along both dimensions.

Overall, using data from Dalberg's nationally representative pulse survey, this paper aims to answer the following two questions:

1. How are gender, caste, and the intersection of the two correlated with possessing an Aadhaar number or card, i.e., with Aadhaar enrollment?
2. Having enrolled in the Aadhaar system, how does each type of individual (based on gender and caste) use the Aadhaar ID to access different services? Here, I look at four broad categories: financial services (through bank accounts and debt/loan access), political rights (through passport/voter ID), education and scholarships, and governmental programs such as the midday meal scheme.

The empirical analysis of these questions reveals that without any controls, gender has no statistically significant impacts on enrollment, but when controlling for literacy rate, being female is associated with a 0.32 percentage point increase in enrollment rate, and when controlling for years of education, this rises to a 1.8 pp increase. Belonging to the abovementioned minority caste categories shows mixed results in Aadhaar enrollment dependent on the use of controls. Finally, if the 'lower' caste individual is also female, the coefficient is negative, indicating an association with lower Aadhaar enrollment. Further, a deep dive into the subpopulation of people who possess an Aadhaar card shows statistically significant declines in the likelihood of female respondents using the ID for all four categories mentioned above. On the other hand, members of marginalized castes who possess the ID show increased likelihood of using it to access midday meals, mixed impacts for financial services, and no significant effects for political rights and educational opportunities. These results are further elucidated below.

This research is contributive for two main reasons. Firstly, it encourages a continued focus on incentivizing enrollment in the Aadhaar program, since the subpopulation of lower caste women are less likely to enroll, even though upon enrollment they are making effective use of the ID to avail various services. In addition, though, I show that the government's responsibility does not end with incentivizing enrollment and creating services – it must bridge the gap between these two steps. I show that the overall population of women, despite enrolling in Aadhaar in higher numbers, are subsequently not using the ID to access the services they are eligible for. In essence, this means that the government must also incentivize and support the use of the ID to avail its benefits, and must empower women to take ownership of their digital identity.

Apart from this, my research also provides a potential intervention for lower caste populations, as it shows that they are less likely to use their Aadhaar to avail financial benefits and school/college enrollment. Upon studying the cause for this relationship, if it is determined that it is not because they are using other IDs to avail these benefits, the government can introduce schemes or informational programs to support the availing of these services.

The rest of this paper proceeds as follows: Section II will provide a brief overview of India's history with ID, societal gender roles, and the archaic yet still highly prevalent caste system. Section III summarizes current literature on digital IDs and minorities and indicates the potential contribution of this paper to this literature. Section IV describes the dataset, relevant variables, and empirical strategy for addressing the two questions detailed above. Section V discusses the results in further depth alongside potential concerns as well areas for further study, and Section VI concludes. Following that are tables, figures, and graphs, as well as references.

II. Institutional Context

2.1 Identification Systems in India

India has a long, close relationship with identification mechanisms. As Sriraman 2018 shows, this relationship started with the introduction of the ration card in the 1940s and has persisted through a laundry list of other forms of identity, including but not limited to passports, PAN cards for income tax, driving licenses, birth certificates. The need for the UIDAI-run digital ID program was, therefore, twofold. First, such a large and unorganized system of IDs could not serve a population of 1.3 billion, since it lacked standardization and was privy to bureaucratic whims. Second, as Zelazny 2012 details, as of 2012 India's subsidy programs amounted to nearly 14% of GDP, but only 51% of the country's low-income population – or those eligible for subsidies – were benefiting from programs such as the Public Distribution System (PDS). Per the 2008 Planning Commission report, 36.7% of grain intended for low-income households was instead sold in the regular market due to leakages and fraud. This problem persisted, so much so that 42% of grain was diverted in 2011-12 (Drèze and Khera, 2015). There was therefore a pressing need to implement a system that would uniquely identify every individual in the country and allow the government to audit distribution systems to gauge whether eligible beneficiaries were, in fact, benefiting from the welfare schemes designed to support them.

2.2 History of Gender-related Development Statistics & Exclusions

The need to study the gender-based disaggregation of Aadhaar enrollment and subsequent welfare benefits stems from India's performance on female development metrics. In the context of social norms working against women even as legal mechanisms attempt to empower them, an identification system that unequivocally supports women as much as men is crucial. Here are example statistics from India: while 48% of the population is female, World Bank statistics show that in 2012, only 27% of women were employed or seeking employment, compared to 79% of men. The World Bank also shows that only 26.5% of females in India held an account at a formal financial institution in 2012, compared to 46.6% in the world. Per a Women's World Banking report in 2019, only 51% of women are literate compared to 77% of men, and there was a 14% gender gap in bank account ownership in 2014, which lessened to 6% in 2017. Overall, a number of demographic, educational, financial, and health-related statistics work against women in society, and the Aadhaar system could prove to be a powerful tool to address these, if it is done right. One problem faced by such digital ID systems, though, is the risk of exclusion. Gelb and Metz 2018 detail how women in some countries face barriers in obtaining official ID, and in some cultures are actively discouraged from obtaining one by family members who fear that it will be a tool for empowerment, independence, and increased decision-making power. Therefore, an audit of the Aadhaar system by gender as well as a further deep dive into what the system yields in terms of development statistics for females is both helpful and necessary. That said, this paper does have the flaw of treating gender as a binary and exploring it only through male-female breakdowns of the Aadhaar system. I fully acknowledge the incompleteness of this analysis as it does not incorporate the experiences of people who do not identify with either of the above.

2.3 History of Caste System, Caste-based Violence, and the Need for Aadhaar

India's hierarchical system of castes is said to have its origin in Hinduism's division of people on the basis of karma (work) and dharma (duty). Historically the caste system dictated members of society's position in the sociopolitical food chain, acting as a system of segregation within the country. As Aneja 2020 says, this historic social segmentation has lasting economic consequences as well due to "centuries of mistreatment", leading lower caste members of society to suffer worse outcomes than upper caste groups on educational, employment, and other such metrics. Further, as per Mosse 2018, today most of India's capital wealth lies in the hands of the upper castes, leaving the lower castes to participate in the economy as wage laborers and leading to them having lower per-capita income as well as lower access to "high-status occupations". This translates to a lack of access to opportunities for upward mobility for the latter group. Even beyond this, a growing body of literature documents crimes against Scheduled Caste/Scheduled Tribes by members higher in the caste hierarchy, including but not limited to sexual violence and abuse of women, physical abuse by police personnel, forced evictions etc. (Sharma, 2015). In the context of such discrimination, marginalization, and violence, Aneja 2020 studies whether political representation reduces the violence that these minorities suffer, with the support of police response and determent of future crimes. This paper finds that an increase in representation by 10 percentage points reduces the murder rate by 3 percentage points. This finding, as well as the ready availability of caste-related data from the pulse survey, inspired me to study the caste-based experience with identity mechanisms. Despite the Constitution instituting affirmative action 'quotas' in schools, colleges, and other institutions to overcome the historic disadvantage faced by these communities, in practice there are mixed opinions as to the real-world impacts of these attempts. This makes it even more important that a digital ID system with the goal of unequivocally identifying every resident of the country to ensure that the services designed for them actually reach them, does indeed serve those disadvantaged by the caste system. This paper will explore whether the Aadhaar system achieves this below.

III. Literature Review

To begin with, this paper uses Dalberg's State of Aadhaar survey report from 2019 as the groundwork upon which it is built. This report, the third of its kind, aimed to collect a nationally representative sample of responses about people's experiences with enrollment in the Aadhaar program, accuracy of data collected, and next steps following enrollment. As stated before, Dalberg conducted a pulse survey of a panel of 147,868 households in 28 states and union territories of India. In addition, they also conducted an in-depth survey of 19,209 houses which allowed them to gain more nuanced insights, but this sample lacked representativeness as it covered only 16 states and 1 union territory. Based on their analysis, multiple key insights emerge. The most relevant to my study is the finding that women's experience with Aadhaar is similar to men’s and that the two groups had comparable levels of enrollment. They also find that Aadhaar appears to support increased independence for women, and that 40% of women felt they had more control over their money with Aadhaar. I wanted to further quantify this claim using the nationally representative sample. In addition, I wished to explore the reason why women felt that they had more control over their money with Aadhaar by exploring what financial services Aadhaar had made accessible to them. Further, the dataset allows an exploration of women’s experience with a number of other services. I chose the four detailed in the Introduction because they cover key development metrics: education, nutrition, financial inclusion, and political rights. The Dalberg report also briefly notes that the quantity of people in a household not enrolled in Aadhaar is higher than the national average for Scheduled Castes and Scheduled Tribes, and about at par for Other Backward Castes. Here, I will quantify this and study the likelihood of enrollment within these caste categories. Further, Dalberg does not explore what having an Aadhaar card means for members of lower castes, nor what services they access using it. This is another contribution my research makes to their study. Perhaps most importantly, I also study the intersectional impacts of gender and caste – something entirely unique from this dataset.

Outside of this, economic literature on Aadhaar-related inclusions and exclusions does not exist in large numbers, although several studies have been conducted on the effectiveness of various distribution systems after the implementation of Aadhaar. Given this context, my literature review draws on a range of writings both directly economics-related and some more policy- or legal-focused and qualitative. First, in a paper exploring the relationship between biometric ID systems and social exclusions using ethnographic field work in the southern states of Karnataka and Andhra Pradesh, the authors explore both how people engage with biometric technologies during the process of welfare access, and what they do when the technology fails. They used qualitative interviews with beneficiaries and observed fair price shop owners (FPSOs) using Aadhaar’s biometric ID system, finding three types of exclusion issues: authentication (wherein fingerprint IDs are required to avail benefits, but the scans sometimes fail and need to be redone at a later time or using different methods), quantity fraud (when beneficiaries receive less rations than they are eligible for), and issues in Aadhaar-ration card linking (the ration card is the original ID required to avail grains from PDS, the public distribution system). This study actually motivates my own, because it highlights the centrality of fingerprint scanning as a biometric identifier during the Aadhaar enrollment process. I hypothesize that this could be one factor preventing members of lower castes from enrolling in the program, since several of them are daily wage workers for whom fingerprint scanning or regular Aadhar center appointments may be infeasible. Secondly, this study details exclusions in the ability to use Aadhaar to access grains from PDS, a large ration-distribution scheme. My paper expands this insight by exploring how Aadhaar helps access other services, such as financial institutions and education.

A highly similar study to the above was conducted in two villages in Western Tamil Nadu, wherein the authors used past research, in-depth interviews, home and work visits, and observation to understand people's experiences with the reformed PDS. The difference, though, is that in Tamil Nadu, beneficiaries use a 'smartcard', the electronic authentication of which does not require biometric ID or fingerprint scanning. Here, the authors find some lack of comfort with the digitized smartcard, particularly through testimonials from women who were highly comfortable with the previous manual book entry when rations were received. The new text-message receipt system creates language barriers and relies on technological connectivity, thus alienating portions of the population unfamiliar with these tenets. This has led to exclusion errors. The authors detail the exact kinds of edge cases that leak through the nationwide implementation of smartcards and that policymakers often describe as "teething issues", but, as the authors claim, are too frequent to be classified as such. For instance, they document examples of families providing each individual's Aadhaar number but still facing exclusion of some members from receiving smartcards; rural residents not knowing that it is possible to enroll in Aadhaar online and so on. Again, my research is indirectly related, but further studies the grievances faced by women in the larger Aadhaar program, as well as exclusions in service access when the service is tied to a digital system. In addition, my research is contributive because of its quantitative nature: I am able to add to the findings of this paper by first studying the entire country at a macroscopic level as opposed to a single state (although I account for inherent state-level differences), as well as by quantifying the change in experience when the beneficiary is female.

The final paper of this kind I draw upon is another mixed methods study looking at Aadhaar-related opinions and experiences in two villages, Konia and Deendayalpur, in the northern Indian Varanasi district. They are relevant because most of the sampled population belongs to the SC/ST/OBC caste categories. The authors collected both quantitative household surveys and qualitative in-depth interviews from a random sample of houses with at least one child, looking at their experience (particularly as households with children) with birth certificates and Aadhaar. Overall, one of the key misconceptions identified was that Aadhaar was mandatory, and that "without [it] nothing could be done". In practice this is legally not true, as the Indian Supreme Court has ruled that Aadhaar should not be considered mandatory to avail any services. Children in particular indicated a belief that Aadhaar was mandatory for school enrollment, access to SIM cards etc. Several people, despite lamenting the costs related to Aadhaar, indicated either a desire to get enrolled in the future, or a mounting social pressure to do so. This is interesting because my study looks at the experience of people from lower caste categories and finds a lower enrollment rate in the program. If this paper is representative, then we can eliminate at least one cause for this lower enrollment: a lack of desire by members of lower caste households. Thus, placing my work in tandem with this paper, it becomes all the more important to identify the exact barriers faced by lower caste members (such as the costs mentioned above), and subsequently introduce policies or subsides to eliminate them.

Outside of these mixed method state-level studies, my research is also inspired by a UC San Diego paper wherein the authors performed a large-scale randomized natural experiment in the Indian state of Jharkhand to study the impacts of biometric ID on leakages, corruption, and reconciliation within the PDS. The authors worked with the government of Jharkhand to randomize the order in which Aadhaar-based biometric authentication (ABBA) was introduced across the state so as to establish treatment and control groups. The paper finds that ABBA did not decrease government spending. However, beneficiaries faced increased costs, including a doubling in the number of unsuccessful trips to fair price shops. The authors also found that the probability that a beneficiary received no commodities at all in a given month increased by 2.4 percentage points, and exclusion was concentrated among households who had not linked their ration cards with Aadhaar in the beginning. In essence, the study advocated not against ABBA itself, but against its rapid implementation without guarding against exclusion errors during program design, and without instituting grievance address mechanisms. This research is further down the pipeline than mine, i.e., the factors I am studying are precursors to the process that these authors are auditing. However, there are two key learnings. The first is the grievances caused to low-income households by virtue of corruption and leakage within the system, even when its intentions are to fix this very problem. Even though I am not studying PDS, but rather other indicators of development, similar motivations could be causing the lack of usage of Aadhar to avail these other services. Secondly, the biggest learning from their paper is to ensure thoughtful implementations of digital policy with reparations mechanisms in place. Even as my paper advocates for the inclusion of women, lower caste individuals, and intersectional minorities in the program, in truth these are often the communities most adversely impacted by leakages and corruption. Thus, my position draws upon this research to advocate for inclusion that does not compromise on access to basic services, and I contribute to these authors' research by quantifying beneficiaries who may face such leakages and corruption in other areas.

IV. Data and Empirical Strategy

4.1 Data

In this paper, I have worked with individual-level cross-sectional data from the year 2019 collected by Dalberg about 147,868 households in India. This translates to data about 575,127 individuals, which is the total size of my dataset. The dataset contains demographic variables such as age, gender, literacy, occupation, education, employment status, and caste, as well as qualitative variables surrounding enrollment in Aadhaar ("Do you have an Aadhaar card?") and accuracy of data presented on the card (in terms of correctness of name, date of birth, gender, address etc.). In addition, the survey presented a number of questions about what the respondent has or has not used the card for (e.g., "Have you personally used Aadhaar for Mid-day Meals?") yielding another set of qualitative binary variables. Finally, the data is organized by state, facilitating state-level clustering of standard errors as well as state-level controls in the regressions described below.

Within this dataset, my units of observation are individuals within the households sampled across 28 states and union territories. A summary statistics table can be found in Section VI detailing the population breakdown from each state within the dataset. For reference, India has 28 states and 8 union territories. The states left out of the sample are the northeastern states of Arunachal Pradesh, Manipur, Mizoram and Nagaland. The union territories not represented are Andaman and Nicobar Islands, Dadra and Nagar Haveli and Daman and Diu, Ladakh, and Lakshadweep. These areas collectively represent about 0.6% of the country's population, leading me to conclude that the national representativeness remains intact since the most highly populated regions of the country are sampled adequately. That said, the non-included regions serve as an urgent area for further research, because other organizations and audits have reported far lower coverage by Aadhaar in the Northeast than elsewhere in the country.

4.2 Outcome Variables

For my first question of how gender, caste, and their intersection are associated with possessing an Aadhaar card or number, the outcome variable of interest is the binary variable hasaadhaar, which answers the question of whether or not the respondent has an Aadhaar card. Within Dalberg's original dataset, this question is answered through the quantitative variable has_aadhaar, which, when tabulated, has 91.2% "yes" responses, with the remaining distributed between "no", "no response", "don't know", and "I lost it". Using these, for the sake of my quantitative analysis I created a quantitative binary variable (hasaadhaar) which takes on the value of 1 for "yes" responses and 0 for all others. The only concern here arises from whether to treat "I lost it" the same way I am treating "yes", since these people have technically enrolled in the Aadhaar program and then lost their credentials. However, I chose not to do so for two reasons: first, the "I lost it" responses are 0.02% of the total (109 in number), which I believed would not statistically significantly affect my analysis. Secondly, since I sought to study whether having an Aadhaar later translated to utilizing it to avail services, losing the card would not facilitate this nor enable me to accurately study this question.

For the second question addressing who uses the Aadhaar card to avail which services upon enrolling in the program, there are six outcome variables of interest depending on the service category. For financial inclusion, I first chose to look at whether or not the person used their Aadhaar to open a bank account, since this is a core financial service. Again, Dalberg's dataset qualitatively answers this question with "yes", "no", "don't know", "no response", or "N/A", and I created a binary quantitative variable that equals 1 for yes and 0 otherwise. The second financial outcome is debtloan, which takes on 1 if the respondent has used their Aadhaar to support the process of seeking debt relief and/or loans. The tabulated results from both of the above can be found in Section VI. Using similar methods, for the governmental nutritional support category I looked at Midday Meal access, using the variable mdm which equals 1 if the respondent used their Aadhaar to access midday meals. For political rights, the outcome variable is otherid, gauging whether the respondent used their Aadhaar to sign up for or create another ID (such as a passport or voter ID). Finally, for education access, there are two outcome variables of interest: enrollment which equals 1 if the Aadhaar was used to enroll in schools and colleges, and scholarships for whether they used Aadhaar to access academic scholarships.

While the dataset does grant access to other metrics such as accessing employment opportunities and other governmental schemes through Aadhaar, I chose to focus on these four broad categories on the basis of necessity and interest.

4.3 Explanatory Variables

The independent variables across both questions and all regressions in this paper are gender1, caste1, and a new variable called gencaste.

As mentioned, the Dalberg survey collected data about the gender of the respondent. In the pulse survey, for uniformity these are classified entirely as male and female, and the survey is 46.99% female and 53.01% male. Based on this, I created a binary quantitative variable called gender1 which will take on the value of 1 if the respondent is female and 0 otherwise.

The caste data is less binary. Using the variable caste_category, the dataset organizes respondents into "Intermediate Caste", "OBC" (Other Backward Castes), "SC" (Scheduled Castes), "ST" (Scheduled Tribes), "Upper Caste", "Not Stated", and "N/A" categories. Since I am particularly interested in studying the experiences of lower caste members with digital ID, my binary variable, caste1, takes on the value of 1 for SC, ST, and OBC, and 0 otherwise. One potential area of concern is how to treat the "not stated" category, but outside of N/A it is the smallest category of respondents, representing 1.29% of the total. I reran the regressions with "not stated" being treated as 1, and it did not change the significance of my results. The tables with this version can be found in section VI.

Finally, I generated a variable called gencaste which is the interaction of the two abovementioned variables, i.e., gencaste = gender1 * caste1. This allows the regression to reflect the change in Aadhaar enrollment likelihood (for question 1) or usage of Aadhaar to avail various benefits (for question 2) for the subpopulation wherein gender1 = 1 and caste1 = 1, i.e., for women from SC/ST/OBC lower caste categories.

4.4 Controls

In order to study and address different potential omitted variable biases, I created new quantitative variables to use as controls using the available qualitative information in the dataset. They are as follows:

1. Lit1 is a binary variable encoding literacy. Lit1 takes on a value of 1 if the respondent is literate and 0 for no or N/A. In the original dataset, literacy is a qualitative variable with responses Y (yes), N (no), and N/A.
2. Emp codifies employment status. It equals 1 if the respondent is employed, and 0 if they are unemployed, not willing/looking for a job, or similar arrangements.
3. schoolage is a binary variable encoding whether the respondent falls within schooling age or not. It is 1 if they are 6-17 years old, and 0 if they are 0-5 or in the adults category.

4.5 Empirical Model & Strategy

Having an Aadhaar card

As follows is the regression equation for the Aadhaar enrollment question:

$$hasaadhaar_i = \alpha_0 + \beta_1(gender1_i) + \beta_2(caste1_i) + \beta_3(gencaste_i) + \varepsilon_i$$

Here, I use a linear probability model to regress hasaadhaar for individual i on gender, caste, and the interaction term. β_1 will reflect the change in Aadhaar enrollment rate between males and females, β_2 will represent the same change between upper/intermediate castes and lower castes, and β_3 will quantify the experience of lower caste females. ε_i is the error term, and α_0 represents the individual constant term for each state, since I am controlling for state fixed effects. The state fixed effects are not reflected in the above regression, but there exists a separate dependent dummy variable for each state taking on the value of 1 if the respondent is from that state. Thus, α_0 represents the overall constant plus the individual coefficient for the state dummy variable that the respondent belongs to. The disaggregation of these individual state coefficients can be found in section VI, but section V will present only β_1, β_2, and β_3, having accounted for state fixed effects. In running this regression, my standard errors are also clustered by state.

Accessing services using the Aadhaar card

First, at this stage I restrict the population to those for whom hasaadhaar = 1, i.e., those who answered “yes” to possessing an Aadhaar card. From there, the equations for the four categories follow the same structure:

$$Metric_j = \alpha_N + \beta_1(gender1_j) + \beta_2(caste1_j) + \beta_3(gencaste_j) + \varepsilon_j$$

In this OLS regression, each individual within the subpopulation of interest is characterized as 'j'. *$metric_j$* is a placeholder for bank account, debt/loan access, midday meal access, other ID creation, school/college enrollment, and scholarships. Similar to the above equation, α_N is the constant term for each state using state fixed effects. Again, all standard errors are clustered by state.

V. Results, Discussion, & Potential Concerns

Enrollment Likelihood

For the first question about the likelihood of having an Aadhaar card based on gender and caste, the regression results with no controls are below:

VARIABLES	(1) hasaadhaar
gender1	0.000953
	(0.00169)
caste1	-0.0135**
	(0.00535)
gencaste	-0.00303**
	(0.00139)
Constant	0.928***
	(0.00374)
Observations	575,127
R-squared	0.143

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This initial regression indicates no statistically significant impact along gender lines, i.e., the difference between female and male enrollment in Aadhaar is not statistically significantly different from 0. However, belonging to a lower caste is associated with a 1.35 percentage point decline in Aadhaar enrollment which is significant at the 95% level, and being a lower caste woman is associated with a 0.303 percentage point decline, significant at the 99% level. This confirms the original hypothesis that Aadhaar enrollment is lower for members of lower castes who are already disadvantaged by the economic, social and legal systems, and could therefore benefit even further from enrolling in a digital ID system such as this one. This would indicate that India's digital policy is exclusive along caste lines. One potential reason for this is enrollment failures: Aadhaar requires biometric data to aid deduplication. Enrollment fails if biometric identifiers fail, such as fingerprint scans. As mentioned before, many lower caste members are daily wage workers employed in industries such as construction, and can therefore not provide accurate fingerprint data. The UIDAI itself, in a report about the use of biometric technology, states that the biometric failure to enroll rate is 0.14%. A second reason could be low awareness about the benefits of Aadhaar among members of lower caste households, but this is unlikely due to the massive scale of the operation as well as the responses seen in the literature review from Varanasi. Finally, the prerequisite documents and bureaucratic necessities may also act as a disincentivizing factor – for instance, residence/address proof is required during the enrollment process in most cases (except in "introducer mode"), and if individuals are unable to attain this, they may choose not to enroll. In addition, some states require documents which themselves require an Aadhaar or similar ID to procure, leading to the creation of a bureaucratic loop.

One concern with this regression, however, is the prevalence of omitted variable bias. For instance, I hypothesize that even after controlling for state fixed effects, factors such as literacy and education may play into Aadhaar awareness and therefore enrollment. To that end, I ran two more regressions controlling for these factors, shown below. Column 1 is identical to the above regression without controls, column 2 controls for literacy (lit1), and column 3 controls for years of education (cal_education_yrs).

VARIABLES	(1) hasaadhaar	(2) hasaadhaar	(3) hasaadhaar
gender1	0.000953	0.00319*	0.0178***
	(0.00169)	(0.00167)	(0.00353)
caste1	-0.0135**	-0.00657	0.0114*
	(0.00535)	(0.00481)	(0.00617)
gencaste	-0.00303**	-0.000476	-0.00355**
	(0.00139)	(0.00168)	(0.00156)
lit1		0.373***	
		(0.0456)	
cal_education_yrs			0.0138***
			(0.00218)

Constant	0.928***	0.568***	0.794***
	(0.00374)	(0.0429)	(0.0215)
Observations	575,127	575,127	575,127
R-squared	0.143	0.234	0.188

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Comparing these 3 columns leaves us open to mixed interpretations. On the one hand, controlling for education does not significantly change the coefficient on gencaste. However, controlling for literacy, gencaste's coefficient loses its significance. Conversely, controlling for both of these things increases the significance of the coefficient on gender1, indicating a 0.3 pp increase in likelihood of enrollment when controlling for literacy and a 1.8 pp increase when controlling for years of education. Finally, the direction and significance of the caste coefficient changes across the board.

The increased significance of the gender coefficient indicates that gender was likely correlated with both literacy and years of education, and was absorbing their impact in column 1. Separating the effects of the two allows the gender coefficient to no longer capture the partial effect of literacy and education. In fact, controlling for years of education seems to account for a significant portion of the explanation, demonstrated through an increase in the R^2. With the caste variable's change in direction however, it would be helpful to write out the expectation equations for the change in hasaadhaar when caste = 1 and when caste = 0. This would involve the difference between the expectation of years of education for lower and upper castes, which is likely to be negative under the assumption that lower caste members on average receive fewer years of schooling due to lower income levels. In essence, belonging to a lower caste alone seems to have a negative impact on Aadhaar enrollment, but including education makes the coefficient positive, perhaps due to targeted governmental programs.

Financial Inclusion: Bank Account & Debt/Loan Access

The regression table for using Aadhaar for bank accounts or debt/loan support are below:

VARIABLES	(1) bankacc	(2) debtloan
gender1	0.000910	-0.0287***
	(0.00219)	(0.00634)
caste1	-0.0147	-0.0211***
	(0.0105)	(0.00557)
gencaste	-0.00749***	0.0105**
	(0.00262)	(0.00389)
Constant	0.880***	0.0873***
	(0.00686)	(0.00532)
Observations	528,329	528,329
R-squared	0.069	0.046

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From this, it is clear that upon possessing an Aadhaar, being female has no statistically significant impact on having a bank account, but is associated with a 2.87 pp decline in using the card to access debt and loan support. The same is true of caste, although being from a lower caste background has a slightly smaller decline of 2.11 pp. Finally, the interaction term demonstrates that being a lower caste female is associated with a 0.75 pp decline in bank account ownership, but a 1.05 pp increase in the likelihood of accessing debt/loan support.

One explanation for the lack of significance for using the Aadhaar to facilitate bank account ownership is if bank accounts are readily and equally accessible to both men and women, or to members of all castes. This is possible given the array of banks across the private and public sector, as well as local-level financial institutions. The double impact is not necessarily surprising, though, since more often than not, particularly in disadvantaged or lower income families, women are not the ones who seek out bank accounts on behalf of their families. Insofar as debt/loan is concerned, the decline in likelihood of using the Aadhaar to obtain this kind of support could either be because women and lower caste members are not seeking out loans, or because the loan request was denied due to low credit scores or other such prerequisites. However, the positive coefficient on gencaste is counterintuitive. This could be either because of the availability of credit support for the most marginalized in some communities, or because of some omitted variable.

One potential concern with the debt/loan variable is that only 5.79% of respondents answered "yes" to this question – 85.9% had not used their Aadhaar to access this benefit. Therefore, for among the people who answered yes, there is likely some omitted variable not reflected in the above regression equation that led them uniquely to seek out loans while most others didn't. One such variable could be existing income, which provides some information on the need for loans. Since this dataset does not provide income data, I looked at employment status and attempted to add this as a control. The employment tabulation can be found in section VI. The results are below, where column 2 controls for employment status:

VARIABLES	(1) debtloan	(2) debtloan
gender1	-0.0287***	-0.00689
	(0.00634)	(0.00477)
caste1	-0.0211***	-0.0231***
	(0.00557)	(0.00580)
gencaste	0.0105**	0.0123***
	(0.00389)	(0.00399)
emp		0.0459***
		(0.00829)
Constant	0.0873***	0.0622***
	(0.00532)	(0.00495)
Observations	528,329	528,329
R-squared	0.046	0.051

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Here, we see that controlling for employment status marginally changes the magnitude of the coefficient for caste and gencaste, but that the coefficient for gender is no longer statistically significant. This indicates that in the first regression, employment was an omitted variable and was highly correlated with gender. In the future, one way to avoid this omitted variable bias is through the use of instrumental variables, given that this is a cross-sectional dataset. One potential instrument for income is enrollment in women's employment guarantee programs, which would satisfy the exclusion restriction in that in most cases, such enrollment does not impact loan access or bank account use through any channel other than income. In addition, interaction terms between gender and employment could be used to avoid this problem. Finally, employment could be further disaggregated as a quantitative variable indicating years of employment or industry to serve as a more accurate proxy for income in this context.

Midday Meals

The regression table for using Aadhaar to access midday meals is below:

VARIABLES	(1) mdm
gender1	-0.00276**
	(0.00105)
caste1	0.00713**
	(0.00267)
gencaste	0.00165**
	(0.000784)
Constant	0.0152***
	(0.00173)
Observations	528,329
R-squared	0.016

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

I find a 0.28 pp decline in the likelihood of females accessing the Midday Meal (MDMS) program using Aadhaar, a 0.7 pp increase for lower caste members, and a 0.16 pp increase for the intersection. There could be a number of reasons for this. First, in 2017, the government of some states (such as Uttar Pradesh) made it mandatory to have an Aadhaar to avail MDMS. This posed a problem because at that time, a vast majority of the people who did not possess an Aadhaar were school-aged children. There may be a different gender-based disaggregation of Aadhaar possession for children which could be statistically significant – perhaps girls are less likely to have Aadhaar cards than boys, leading to the negative coefficient. However, one potential issue with this analysis could be another omitted variable – age. It could merely be that the sample contains a higher average age for females than males, who therefore cannot access MDMS as they are out of school. I reran this regression controlling for age, using a dummy variable that took on a value 1 if the person was of school age. Column 1 represents the controlled results.

VARIABLES	(1) mdm	(2) mdm
gender1	-0.00159*	-0.00276**
	(0.000837)	(0.00105)
caste1	0.00425*	0.00713**
	(0.00216)	(0.00267)
gencaste	0.00140*	0.00165**
	(0.000720)	(0.000784)
schoolage	0.0917***	
	(0.0222)	
Constant	7.64e-05	0.0152***
	(0.00512)	(0.00173)
Observations	528,329	528,329
R-squared	0.082	0.016

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Here, we see that the direction of impact and the significance do not change, but the magnitude shifts – for instance, gender's coefficient changes by 0.117. For all 3 variables, controlling for school age leads to a drop in magnitude of the coefficient. This indicates the need to have removed school age from the error term, as the other regression attributes more explanatory power than is accurate to the other three independent variables. Potential further controlling by income level (thereby to determine eligibility for MDMS) or school type (i.e., government vs. private, since many private schools do not have MDMS) could further refine these results, but this dataset does not provide information on the same. Further, it may be helpful to also study whether the adults have used their Aadhaar for their children's midday meal.

Other ID access

The regression table for using Aadhaar to access passports, voter IDs, and other such documents is below:

VARIABLES	(1) otherid
gender1	-0.0322***
	(0.00798)
caste1	-0.0108
	(0.0140)

gencaste	0.00228
	(0.00558)
Constant	0.402***
	(0.0104)
Observations	528,329
R-squared	0.222

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Here, we see no caste-based impact, but we do see a 3.2 pp decline in the likelihood that women use Aadhaar to access another ID. Once again one potential omitted variable here could be income, since higher income typically corresponds to the need for passports (for travel etc.), or higher education level. In addition though, needing a voter ID may be associated with literacy rates and education years as well, since more literate people are likely to be more politically active. The regression tables with these controls are below:

VARIABLES	(1) otherid	(2) otherid	(3) otherid
gender1	-0.0317***	-0.0222**	-0.0322***
	(0.00794)	(0.00838)	(0.00798)
caste1	-0.00974	0.00332	-0.0108
	(0.0139)	(0.0131)	(0.0140)
gencaste	0.00282	0.00210	0.00228
	(0.00555)	(0.00546)	(0.00558)
lit1	0.0849***		
	(0.0149)		
cal_education_yrs		0.00785***	
		(0.000889)	
Constant	0.318***	0.323***	0.402***
	(0.0175)	(0.0148)	(0.0104)
Observations	528,329	528,329	528,329
R-squared	0.223	0.226	0.222

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Interestingly, not much changes with these controls except for the magnitude of the coefficient on gender, which is reduced with both controls indicating some correlation between gender and literacy or education years as we attribute explanatory power away from the gender coefficient. One further question for exploration is why there is no correlation between caste and these variables in the context of accessing another ID, given the drastic shifts when these controls are added for the Aadhaar enrollment question.

Educational services

VARIABLES	(1) enrollment	(2) scholarships
gender1	-0.0303***	-0.0176***
	(0.00596)	(0.00420)
caste1	0.00294	-0.000265
	(0.00427)	(0.00481)
gencaste	0.00668**	0.00528***
	(0.00313)	(0.00168)
Constant	0.188***	0.0883***
	(0.00357)	(0.00380)
Observations	528,329	528,329
R-squared	0.018	0.032

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

These results indicate a 3 pp decline in likelihood of using Aadhaar to avail enrollment and a 1.8 pp decline for scholarships for females. There is no statistically significant impact by caste, but the interaction of the two changes the direction of the relationship, suggesting a 0.7 pp increase for enrollment and 0.5 pp increase for scholarships for low caste women. This could be because of targeted programs incentivizing women of lower caste categories to enroll in school or avail scholarships, but more likely here age will act as an omitted variable again, since many women or members of lower castes perhaps enrolled in school prior to the introduction of Aadhaar, rendering this question irrelevant to them. I did rerun this regression using the school age control from the MDMS question, and its results are below, where columns 1 and 3 include controls:

VARIABLES	(1) enrollment	(2) enrollment	(3) scholarships	(4) scholarships
gender1	-0.0239***	-0.0303***	-0.0152***	-0.0176***
	(0.00530)	(0.00596)	(0.00386)	(0.00420)

caste1	-0.0129**	0.00294	-0.00537	-0.000265
	(0.00495)	(0.00427)	(0.00397)	(0.00481)
gencaste	0.00530	0.00668**	0.00445***	0.00528***
	(0.00318)	(0.00313)	(0.00160)	(0.00168)
schoolage	0.504***		0.174***	
	(0.0309)		(0.0343)	
Constant	0.104***	0.188***	0.0592***	0.0883***
	(0.00692)	(0.00357)	(0.00722)	(0.00380)
Observations	528,329	528,329	528,329	528,329
R-squared	0.274	0.018	0.091	0.032

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Doing this yields small changes in magnitude for gender indicating some correlation with age, but including controls for school age make the coefficient on caste significant, indicating that members of lower caste are 1.3 pp less likely to use Aadhaar to enroll in school after controlling for age. There is no change in significance along caste lines for scholarship. In addition, for low caste women, controlling for age removes the significance of the coefficient on gencaste for enrollment but retains it with small changes in magnitude for scholarships. From this we can infer that controlling for age was an omitted variable correlating with gender for both educational attainment variables, but with caste only for school enrollment.

One caveat here is that schoolage only controls for students who are enrolled in school, whereas the question asks about whether they used Aadhaar to enroll in school or college, and does not detail whether scholarships stop at a certain education level. However, the data does not allow disaggregation above the age of 17, categorizing everyone as an adult beyond this point and making it impossible to study college enrollment. This would be a useful future area of study though and will likely change the coefficients even further.

Overall

In general, the size of my dataset makes the usage of OLS regression less risky in answering the questions posed above than doing so with a smaller, less representative dataset. However, I am still only able to identify correlations and associations, and cannot assign causation to any of these relationships. Doing so would require further developments and analysis, some of which are detailed below.

Having discussed each individual regression result, the biggest concern overall is omitted variable bias and the need for IVs in some cases and further data collection or disaggregation in others. The identification – and data collection – on appropriate instruments that satisfy relevance, exogeneity, and the exclusion restriction would bring me a step closer to causation e.g., is belonging to a lower caste causing people to enroll in Aadhaar in smaller numbers than the majority population? One potential instrument for caste could be using enrollment in educational institutions through the 'quota' affirmative action system targeted towards lower caste members, but this would not satisfy the exclusion restriction if such enrollment requires the possession of an Aadhaar specifically in the first place. If, however, one can enroll in these quotas using other IDs, such as birth certificates, the validity of this instrument could be preserved.

A second way to approach causality would be to create a panel dataset featuring information about the same households across multiple years to track Aadhaar-related metrics over time. While Dalberg does have reports and datasets from previous years (starting from 2016), the respondents differ each year due to new random sampling, as does the size of the dataset. Therefore, this renders artificially creating a panel dataset impossible, but doing so could facilitate rigorous fixed effects analysis or difference-in-differences at critical change points (for instance, the introduction of incentives for enrollment, or the necessity of Aadhaar to procure a midday meal). At an even larger scale, if the country's census collected information about Aadhaar enrollment, we could test for parallel trends between the two groups (hasaadhaar = 1 and hasaadhaar = 0) before the introduction of Aadhaar in 2009, and see if there are divergences in the outcomes described above after its introduction. This would be possible once we have ensured that nothing else changes around the time of the introduction of Aadhaar, so that the differences in omitted variables in the treatment and control group remain constant over time.

Finally, outside of causality, one other improvement to the linear probability methodology detailed in this paper would be to establish a logit or probit model to properly quantify the probabilities of Aadhaar enrollment as well as the utilization of the ID for different subpopulations and minorities in the country.

In addition to these, one big area of potential improvement is that as mentioned previously, this dataset treats gender as a binary, and does not allow any discussion of nonbinary or gender nonconforming people. In fact, Dalberg's complementary in-depth survey includes a "third gender" category, but as mentioned this data is not as representative. In practice, the Aadhaar program does allow for this gender to be indicated during enrollment, but many transgender people have faced exclusion from the program by virtue of discrimination, not identifying with the gender on their birth certificates, and numerous other reasons. For this reason, a rigorous quantitative study of the experiences of people who do not identify within the above gender binary is crucial. Similarly, the current dataset excludes several Northeastern states, as well as some union territories. Given the means to collect data on these regions, expanding the study to understand the experiences of these populations would be invaluable, since I hypothesize a very low and statistically significant coefficient on the state dummy variable for these regions, given reports of low coverage.

VI. Conclusion

In a 2018 Supreme Court statement, the SC stated that the Aadhaar system was "constitutionally fair" and gave "dignity to the marginalized". This paper aimed to study the validity of this claim by quantifying the experiences of two marginalized or disadvantaged communities in India: women and members of lower caste. In addition, I sought to study the impact of intersectionality in the country: i.e., of being female and belonging to a lower caste. In doing so, I looked at two questions: how these groups' enrollment in India's digital ID Aadhaar program differs from their corresponding 'advantaged' groups (i.e., male and upper caste), as well as how service access using Aadhaar differs across each of these groups.

Overall, in answering my first question, I find that without any controls, gender has no statistically significant impacts on enrollment, but when controlling for literacy rate, being female is associated with a 0.32 pp increase in enrollment rate, and when controlling for years of education, this increases to a 1.8 pp increase. On the other hand, caste has a negative impact without controls, but is associated with a 1.1 pp increase in enrollment controlling for years of education. Finally, the intersection yields negative impacts across the board; most notably being associated with a 0.36 pp decrease in enrollment for lower caste women when controlling for education.

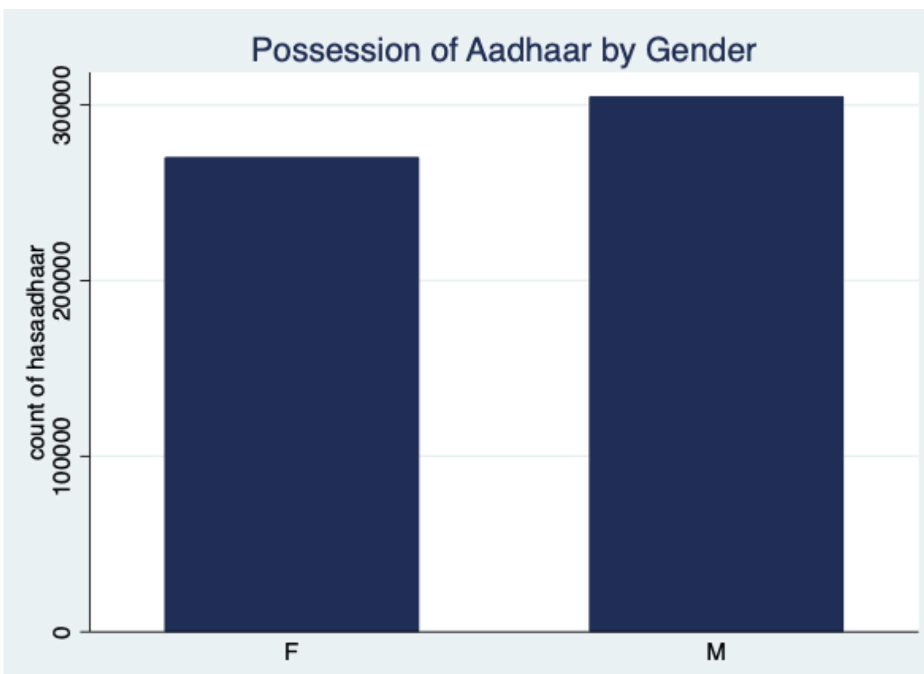
In answering the second question, I looked at four categories. Within financial inclusion, lower caste females are 0.75 pp less likely to have used Aadhaar to open a bank account. Looking at debts and loans with employment status acting as a proxy for income level controls, we see a 2.3 pp decline in the likelihood that lower caste respondents used their Aadhaar to access loans/debts, but a 1.23 pp increase in likelihood for lower caste women, perhaps because of targeted schemes. Within governmental welfare scheme access, using school age as a control, girls are 0.16 pp less likely to use Aadhaar to avail the midday meal program benefits, whereas lower caste members are 0.43 pp more likely to, and lower caste females are 0.14 pp more likely to as well. The third category of services studied was using Aadhaar to access other forms of ID, such as passports and voter IDs. Here as well, after controlling for literacy level and educational years, women are between 2-3 pp less likely to have used their Aadhaar to get other IDs, but there are no significant impacts for any of the other categories. The final category was educational services, and we note that with or without controlling for school age, women are 2-3 pp less likely to have used Aadhaar to enroll in school/college, and about 1.6 pp less likely to have used it for scholarships. Conflicting results emerge for castes, but lower caste women are 0.45 pp more likely to have used their ID to access scholarships, perhaps due to the existence of scholarships for this category of people.

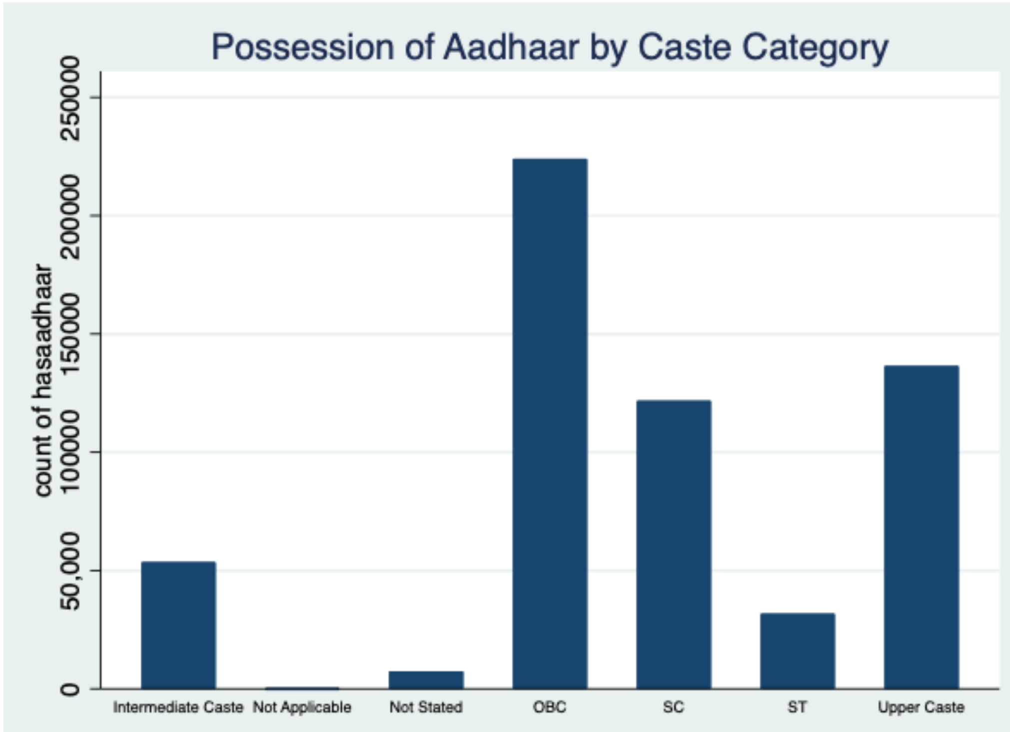
Overall we see that women suffer disadvantages across the board in bank account access, midday meal benefits, accessing political ID, and availing enrollment and scholarships, despite being more likely to enroll in Aadhaar within this sample. This is an interesting conundrum, because it either points to the existence of unidentified omitted variables or indicates that women enroll in the program but are not able or willing to enjoy its benefits as much as men along these four dimensions. Along caste lines, lower caste members are less likely to use their Aadhaar to avail debt/loan support and school/college enrollment, but more likely to access midday meal benefits. These mixed results indicate a potential area for intervention for the government. Finally, lower caste women are less likely to enroll in Aadhaar to begin with, and then less likely to use it to open bank accounts, but more likely to use it for debts/loans, midday meals, and scholarships, in complete contrast to the experience of women overall. Perhaps targeted schemes exist to support these women, but the impact could be maximized even further if incentives were created for them to enroll in Aadhaar to the same degree as the rest of the population.

The policy implications of this paper are, in my view, twofold. The first goal should be to empower women to maximally utilize their digital ID as opposed to merely signing up for it. There are ways to extrinsically motivate this, such as by offering monetary incentives, but the more powerful intervention would be to facilitate intrinsic motivation. The government could create loan programs specifically designed for women, midday meal schemes with the unique nutritional requirements of girls (e.g., iron content) in mind, or political awareness mechanisms to make women interested in seeking a voter ID and making their voice heard in politics. All of these should happen in tandem with household-level empowerment and policies to circumvent archaic gender roles. The second goal should be to make the process of Aadhaar enrollment easier for lower caste women specifically, as well as generally for all members of lower caste communities. It is clear that the desire to sign up and utilize the ID is in place, but the government must remove the barriers – both institutional/bureaucratic as well as financial – that stand in the way of increased signups. For instance, alternatives for address proof – such as popularizing the "introducer mode" – could be put in place, or the creation of village-level Aadhaar representatives who can help members who are not digitally fluent sign up for Aadhaar online. Or, Aadhaar sign ups could be combined with other local duties, such as attending village or town council meetings or procuring agricultural supplies.

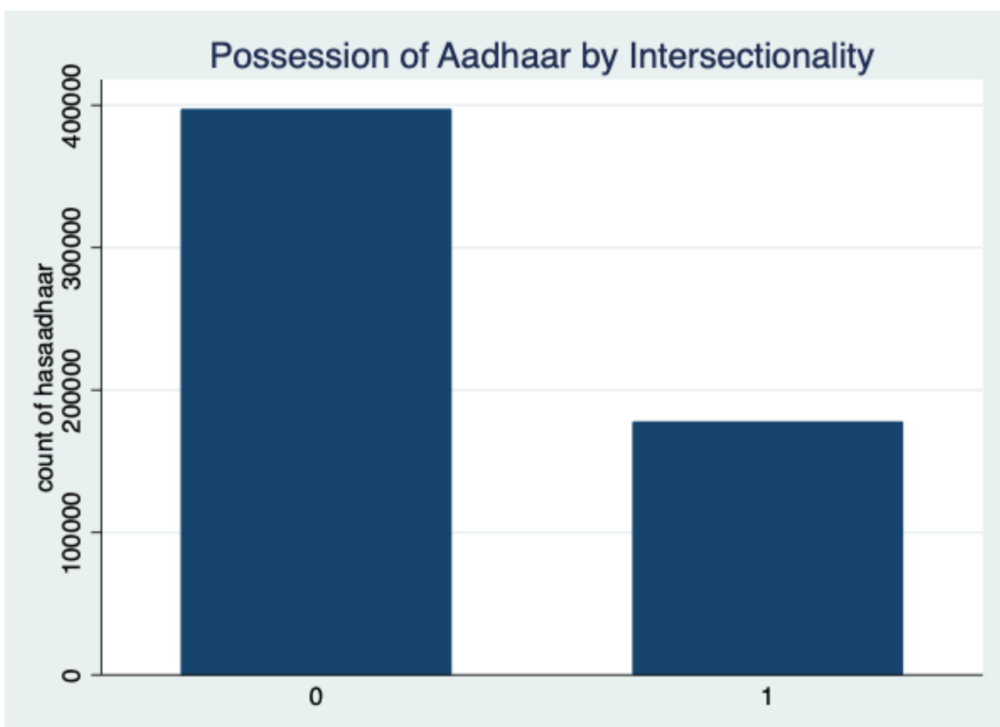
In summary, digital policy can be a transformative and powerful tool for social change, but can also further widen existing gaps in society between the historically privileged and the historically marginalized. The Aadhaar program has the potential to bridge the gender- and caste-based digital divide in India, but only if carried out with appropriate policy interventions that legitimize the needs and nuances of various subpopulations and facilitate their digital inclusion.

VI. Tables & Figures





Possession of Aadhaar by Caste Category
count of hasaadhaar
0
50,000
100000
150000
200000
250000
Intermediate Caste
Not Applicable
Not Stated
OBC
SC
ST
Upper Caste



Possession of Aadhaar by Intersectionality
count of hasaadhaar
0
100000
200000
300000
400000
0
1

Has_aadhaar dataset disaggregation

has_aadhaar	Frequency	Percentage (%)	Cum. (%)
Don't Know	4347	.76	.76
I Lost it	109	.02	.77
No	34875	6.06	6.84
No Response	7467	1.3	8.14
Yes	528329	91.86	100
Total:	**575127**	**100**	

State breakdown in dataset

state	Frequency	Percentage (%)	Cum. (%)
Andhra Pradesh	19154	3.33	3.33
Assam	4870	.85	4.18
Bihar	35532	6.18	10.36
Chandigarh	1529	.27	10.62
Chhattisgarh	16750	2.91	13.53
Delhi	6189	1.08	14.61
Goa	2356	.41	15.02
Gujarat	33815	5.88	20.9
Haryana	19479	3.39	24.29
Himachal Pradesh	4806	.84	25.12
Jammu and Kashmir	6652	1.16	26.28
Jharkhand	16245	2.82	29.1
Karnataka	27780	4.83	33.93
Kerala	14060	2.44	36.38
Madhya Pradesh	30611	5.32	41.7
Maharashtra	64245	11.17	52.87
Meghalaya	5292	.92	53.79
Odisha	21472	3.73	57.52
Puducherry	3372	.59	58.11
Punjab	22018	3.83	61.94
Rajasthan	38471	6.69	68.63
Sikkim	1525	.27	68.89
Tamil Nadu	29034	5.05	73.94
Telangana	15048	2.62	76.56
Tripura	3579	.62	77.18
Uttar Pradesh	91929	15.98	93.16
Uttarakhand	6122	1.06	94.23
West Bengal	33192	5.77	100
Total:	**575127**	**100**	

Caste-related breakdown within dataset

caste_category	Frequency	Percentage (%)	Cum. (%)
Intermediate Caste	53584	9.32	9.32
Not Applicable	11	0	9.32
Not Stated	7425	1.29	10.61
OBC	224026	38.95	49.56
SC	121797	21.18	70.74
ST	31793	5.53	76.27
Upper Caste	136491	23.73	100
Total:	**575127**	**100**	

Employment categories

cal_employment	Frequency	Percentage (%)	Cum. (%)
Daily Wage worker/ Casual labour	44557	7.75	7.75
Data Not Available	492	.09	7.83
Not Applicable	96028	16.7	24.53
Salaried - Permanent	29431	5.12	29.65
Salaried - Temporary	23370	4.06	33.71
Self-employed	85642	14.89	48.6
Unemployed, not willing and not looking for a job	271712	47.24	95.85
Unemployed, willing and looking for a job	17797	3.09	98.94
Unemployed, willing but not looking for a job	6098	1.06	100
Total:	**575127**	**100**	

"Have you used an Aadhaar to obtain a bank account?"

used_open_bank_account	Frequency	Percentage (%)	Cum. (%)
Don't Know	302	.05	.05
No	68820	11.97	12.02
No Response	308	.05	12.07
Not Applicable	46689	8.12	20.19
Yes	459008	79.81	100
Total:	**575127**	**100**	

“Have you used an Aadhaar for debt/loan?”

used_debt_loan	Frequency	Percentage (%)	Cum. (%)
Don't Know	704	.12	.12
No	494117	85.91	86.04
No Response	319	.06	86.09
Not Applicable	46689	8.12	94.21
Yes	33298	5.79	100
Total:	**575127**	**100**	

Regression results including ‘not stated’ as 1 for caste

VARIABLES	(1) hasaadhaar
gender1	0.00107
	(0.00173)
caste1	-0.0138**
	(0.00516)
gencaste	-0.00320**
	(0.00148)
Constant	0.928***
	(0.00370)
Observations	575,127
R-squared	0.143

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Has_aadhaar regression including state fixed effect coefficients

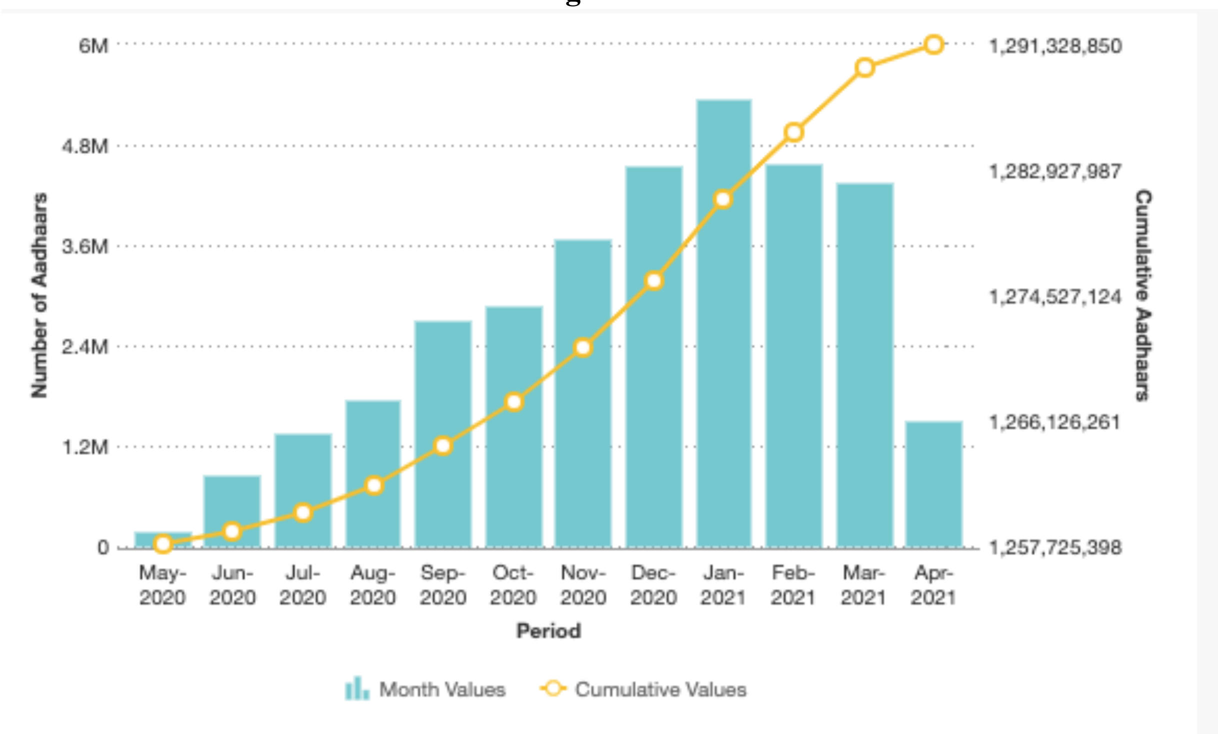
VARIABLES	(1) hasaadhaar
gender1	0.000953
	(0.00169)
caste1	-0.0135**
	(0.00535)
gencaste	-0.00303**
	(0.00139)
Assam	-0.813***
	(0.000364)
Bihar	-0.0914***
	(0.00123)

Chandigarh	-0.00987***
	(0.000915)
Chhattisgarh	0.000734
	(0.00163)
Delhi	-0.0331***
	(0.000651)
Goa	-0.0526***
	(0.00126)
Gujarat	-0.0265***
	(0.000878)
Haryana	-0.0465***
	(0.000754)
Himachal Pradesh	-0.00588***
	(0.000970)
Jammu & Kashmir	-0.165***
	(0.00139)
Jharkhand	-0.111***
	(0.00167)
Karnataka	-0.0542***
	(0.000393)
Kerala	-0.0233***
	(0.00122)
Madhya Pradesh	-0.254***
	(0.000603)
Maharashtra	-0.0229***
	(0.000224)
Meghalaya	-0.555***
	(0.00148)
Odisha	-0.0293***
	(0.000383)
Puducherry	-0.000184
	(0.00200)
Punjab	-0.0258***
	(0.000305)
Rajasthan	-0.164***
	(0.000147)
Sikkim	-0.0764***
	(0.00211)
Tamil Nadu	-0.0187***
	(0.00184)
Telangana	0.00277***
	(0.001000)
Tripura	-0.0160***
	(0.000150)
Uttar Pradesh	-0.0832***
	(0.000398)

Uttarakhand	-0.127***
	(0.000745)
West Bengal	-0.0654***
	(0.000103)
Constant	1.006***
	(0.00328)
Observations	575,127
R-squared	0.143

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Aadhaar generation trend



Source: UIDAI government website

VII. References

1. "Inclusive and Trusted Digital ID Can Unlock Opportunities for the World's Most Vulnerable." *World Bank*, www.worldbank.org/en/news/immersive-story/2019/08/14/inclusive-and-trusted-digital-id-can-unlock-opportunities-for-the-worlds-most-vulnerable.

2. Gelb, Alan H, and Anna Diofasi Metz. *Identification Revolution: Can Digital ID Be Harnessed for Development?* Washington, D.C., Center For Global Development, 2018.
3. Tarangini Sriraman. *In Pursuit of Proof: A History of Identification Documents in India*. New Delhi, India, Oxford University Press, 2018.
4. "An Indian History of Identity Infrastructure." *The Wire*, thewire.in/books/identity-proof-and-paper-infrastructure-an-indian-history. Accessed 6 May 2021.
5. Zelazny, Frances. *The Evolution of India's UID Program Lessons Learned and Implications for Other Developing Countries*, 2012.
6. Dr`eze, Jean and Reetika Khera. *Understanding Leakages in the Public Distribution System*. Economic and Political Weekly, February 2015, 50 (7).
7. "Women in India's Economic Growth." *World Bank*, 2018, www.worldbank.org/en/news/speech/2018/03/17/women-indias-economic-growth.
8. "Indian Women Aren't Using Their Bank Accounts. This Is How and Why Women's World Banking Plans to Change That." *Women's World Banking*, 22 Jan. 2019, www.womensworldbanking.org/insights-and-impact/indian-women-arent-using-their-bank-accounts-this-is-how-and-why-womens-world-banking-plans-to-change-that/. Accessed 6 May 2021.
9. BBC news. "What Is India's Caste System?" *BBC News*, 20 July 2017, www.bbc.com/news/world-asia-india-35650616.
10. Mosse, David. "Caste and Development: Contemporary Perspectives on a Structure of Discrimination and Advantage." *World Development*, vol. 110, Oct. 2018, pp. 422–436, 10.1016/j.worlddev.2018.06.003.
11. Sharma, Smriti. "Caste-Based Crimes and Economic Status: Evidence from India." *Journal of Comparative Economics*, vol. 43, no. 1, Feb. 2015, pp. 204–226, 10.1016/j.jce.2014.10.005. Accessed 9 May 2019.

12. Aneja, Abhay and Ritadhi, S.K., Minority Representation and Protection from Targeted Violence: Evidence from Low-Caste Political Parties in India (June 10, 2019). Available at SSRN: https://ssrn.com/abstract=2816668
13. Hundal, Hartej Singh, and Bidisha Chaudhuri. "Digital Identity and Exclusion in Welfare." *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*, 17 June 2020, 10.1145/3392561.3397583. Accessed 6 May 2021.
14. Carswell, Grace, and Geert De Neve. "Transparency, Exclusion and Mediation: How Digital and Biometric Technologies Are Transforming Social Protection in Tamil Nadu, India." *Oxford Development Studies*, 31 Mar. 2021, pp. 1–16, 10.1080/13600818.2021.1904866. Accessed 7 May 2021.
15. Bhatia, Amiya, et al. "'Without an Aadhaar Card Nothing Could Be Done': A Mixed Methods Study of Biometric Identification and Birth Registration for Children in Varanasi, India." *Information Technology for Development*, vol. 27, no. 1, 23 Nov. 2020, pp. 129–149, 10.1080/02681102.2020.1840325. Accessed 7 May 2021.
16. Muralidharan, Karthik, et al. *Identity Verification Standards in Welfare Programs: Experimental Evidence from India*. Apr. 2020, www.nber.org/papers/w26744.
17. Venkatanarayanan, Anand. "Aadhaar Enrollment Rejections Are Accelerating." *Medium*, 30 Nov. 2017, medium.com/karana/aadhaar-enrollment-rejections-are-accelerating-5aa76191d9a9. Accessed 6 May 2021.
18. Unique Identification Authority of India. *Role of Biometric Technology in Aadhaar Enrollment*.
19. "1.05 Billion Aadhar Cards Issued, Challenge to Enrol Remaining 20 Crore: UIDAI." *NDTV.com*, 29 Sept. 2016, www.ndtv.com/india-news/1-05-billion-aadhar-cards-issued-challenge-to-enrol-remaining-20-crore-uidai-1468140. Accessed 6 May 2021.

20. Aadhaar 2019, State Of. “Key Findings: State of Aadhaar 2019.” *State of Aadhaar*, 2019, stateofaadhaar.in/top-10-insights.php. Accessed 7 May 2021.

21. Dalberg. *State of Aadhaar*, 2019, https://stateofaadhaar.in/assets/download/SoA_2019_Report_web.pdf?utm_source=download_report&utm_medium=button_dr_2019