IIMB-WP N0. 710/2024

WORKING PAPER NO: 710

Caste inequality in medical crowdfunding in India

Arpit Shah
Indian Institute of Management Bangalore
arpit.shah@iimb.ac.in

Year of Publication – July 2024

Caste inequality in medical crowdfunding in India

Arpit Shah[1]

[1]Centre for Public Policy, Indian Institute of Management Bangalore

(Accepted, to appear in the Journal of Development Studies)

Abstract

Medical crowdfunding has gained significant popularity and importance, yet researchers argue that it might not offer equitable financial assistance to vulnerable groups. Studies in Western countries have highlighted disparities in medical crowdfunding based on racial, ethnic, and socioeconomic dimensions. Despite this, the equity implications of crowdfunding in India have not been thoroughly investigated. We present the initial empirical evidence indicating caste-based inequalities in medical crowdfunding within India. Leveraging a comprehensive dataset comprising 5,527 medical crowdfunding campaigns from one of India's largest platforms, we evaluate the impact of the recipient's caste identity on campaign outcomes. Our methodology utilises administrative data to deduce caste from the recipient's surname. Campaigns by individuals from dominant caste groups tend to garner higher funds compared to those from marginalised castes, largely due to higher average donations. Furthermore, individuals from Scheduled Castes and Scheduled Tribes (marginalised groups in India) initiate disproportionately fewer campaigns than dominant castes. Our findings remain robust across various performance metrics and alternative model specifications, even after adjusting for multiple campaign features and recipient locations. In summary, our study highlights how crowdfunding on for-profit platforms may exacerbate existing caste-related health disparities in India by disproportionately favouring privileged groups.

Keywords: caste, inequality, donation-based crowdfunding, India

1. Introduction

One of the most popular categories in donation-based crowdfunding is fundraising for emergency medical expenses (Young and Scheinberg, 2017). Theoretically, crowdfunding holds vast potential for social good in this area. Healthcare expenditure shocks frequently drive vulnerable families into long-term poverty, making recovery challenging (Keane and Thakur, 2018). In addition to being an immediate source of funds, retail donations through online crowdfunding platforms can substitute for health insurance, which many impoverished families lack (Panjwani and Xiong, 2023; Keane and Thakur, 2018). For instance, GoFundMe, a major global crowdfunding platform, pledges to provide 'immediate help with medical bills' (GoFundMe, 2023). Additionally, Ketto, an Indian crowdfunding platform established in 2011, aims to 'make healthcare more accessible to everyone' (Ketto, 2023).

However, similar to all emerging technologies, crowdfunding carries the potential to exacerbate existing societal inequalities. This effect has been well-documented in the case of AI chatbots, which tend to amplify prevailing social biases against minorities and vulnerable groups (Ray, 2023). Empirical research, primarily centred on Western countries, has illustrated inequalities based on race/ethnicity and income in medical crowdfunding (Igra, 2022; Kenworthy and Igra, 2022; van Duynhoven et al., 2019). In our literature review, we found no studies on inequality within the context of crowdfunding in developing country settings. There is a clear need for empirical investigation in these areas, as healthcare expenditures in developing countries often act as a significant income shock impacting vulnerable households.

This paper studies medical crowdfunding between 2020 and 2022, aiming to understand whether it exacerbates existing societal inequities in the Indian context. Medical crowdfunding has gained substantial popularity in India. Reports indicate that the top three donation-based crowdfunding platforms in India amassed approximately INR 40 billion (USD 500 million[1]) in donations in 2020, with medical crowdfunding being the dominant category across these platforms (ISB, 2021). The COVID-19 pandemic has further catalysed the growth of online crowdfunding in India. Economic shocks, widespread displacement, and lockdowns have made it particularly appealing for recipients seeking to remotely raise funds.

We are specifically interested in the impact of a recipient's caste identity on crowdfunding campaign outcomes. Caste functions as a dynamic force that has historically shaped political and economic trajectories in India across multiple generations (Kothari, 1995). Recent research indicates that caste continues to have pervasive and persistent impacts in India's modern market economy (Mosse, 2020; Chakraborty et al., 2023). Moreover, caste is intricately linked to health outcomes, with documented caste-based disparities in access to healthcare services and utilisation in India (Gupta and Coffey, 2020; Thapa et al., 2021). Caste also plays a pivotal role in determining access to financing for health-related issues (Debnath and Jain, 2020). Similar effects may potentially manifest in accessing funds from online platforms like Ketto. Consequently, the implications of online crowdfunding on caste-driven health inequalities remain uncertain. Crowdfunding could po-

[1]Exchange rate: USD 1 = INR 80

tentially function as a means for communities to rally together and offer mutual support (Solnit, 2010), or conversely, it could exacerbate prevailing inequalities, as observed in Western contexts (Igra, 2022).

To understand the impact of caste on medical crowdfunding, we draw upon a dataset encompassing medical crowdfunding campaigns hosted on the prominent platform Ketto between February 1, 2020, and December 31, 2022. Our selected start date aligns approximately with the onset of the COVID-19 pandemic (India's first confirmed COVID case was reported on January 27, 2020) (Andrews et al., 2020). From a public health standpoint, this period potentially witnessed substantial healthcare funding requirements, as individuals and families encountered health and income-related shocks due to the pandemic. Our analysis leverages comprehensive data from medical crowdfunding campaigns during this period.

We find evidence for caste-based disparities in medical crowdfunding campaign outcomes on Ketto, with recipients from marginalised caste groups raising lower funds as compared to recipients from dominant caste groups. Considering the pre-existing caste-based inequalities in health outcomes in India, our results show that the emergence of crowdfunding could aggravate the problem. In doing so, we contribute to the emerging literature on inequality in medical crowdfunding (Igra et al., 2021; Igra, 2022). In addition, we build on the literature on charitable giving (List, 2011). Previous work on charitable giving from around the world has documented the importance of recipient identity - donors tend to contribute less to recipients from stigmatized or marginalised groups (Linos et al., 2021). Similar effects have been recorded with regard to caste in India (Deshpande and Spears, 2016). To the best of our knowledge, our work is the first examination of caste-based inequality in a newly emerging technology that continues to grow into an important source of funding for medical treatments in India.

2. Background

2.1. Caste in modern India

The caste system in India is a complex and deeply entrenched social stratification system that has shaped the region's history and continues to influence its present (Kothari, 1995; Mosse, 2020). Over centuries, this system has permeated every aspect of Indian life, affecting social mobility, economic opportunities, and access to resources (Kothari, 1995). Despite substantial efforts to eradicate the caste system, its influence remains prevalent in contemporary India. Fieldwork conducted in Southern India by Mosse (2020) highlights the pervasive effects of caste in labour markets and the business economy.

Caste also holds significance in the context of healthcare, with numerous studies documenting caste-related disparities in health outcomes in India. Marginalised caste groups exhibit poorer performance across various health indicators compared to dominant castes, even after accounting for various socio-economic factors (Borooah, 2010). Access to healthcare services is also uneven, as illustrated by Acharya (2010), who highlight caste discrimination in rural India's public healthcare services. Additionally, Lamba and Spears (2013) establish that villages led by chairmen from marginalised caste groups are less likely to receive government awards for eradicating open defecation. Using ethnographic methods, O'Reilly et al. (2017) reveal caste's significant role in the failure of sanitation interventions in Southern India. They note that schemes are often caste-targeted, resulting in inferior policy support for marginalised caste groups in sanitation initiatives.

Relevant to our study is the insight that caste influences access to health financing in India. Utilising administrative data from a publicly financed health programme, Debnath and Jain (2020) demonstrate the importance of caste networks in determining programme utilisation. Individuals are more inclined to claim from the programme if other members of their caste group have previously made similar claims.

2.2. Inequity in online crowdfunding

According to Varun Sheth, CEO, and co-founder of Ketto, 'crowdfunding is one of the most convenient ways to raise money for medical treatments'. In India, with a high burden of hospitalizations and limited penetration of medical insurance (Dhar, 2018), the demand for medical funding is notably high. Public expenditure on healthcare remains relatively low at 2.1% of GDP (Economic Survey, 2020). In this scenario, medical crowdfunding has gained significance, especially since the onset of the COVID-19 pandemic. Although the fundraising figures are lower compared to leading countries like China and the United States, the industry has experienced substantial growth (ISB, 2021).

As medical crowdfunding has expanded, there have been debates among researchers regarding its fairness in distributing financial resources to vulnerable individuals and families (Snyder et al., 2016). Investigations focused on GoFundMe in Western countries reveal that medical crowdfunding is particularly crucial in areas where safety nets such as healthcare systems and insurance coverage are less robust. However, most campaigns fail to achieve their funding targets, and their outcomes mirror prevailing social and economic disparities linked to race, income, and geographical location (Kenworthy and Igra, 2022; Davis et al., 2023). In the United States, Igra (2022) demonstrate that campaigns featuring White and Asian recipients tend to perform better than those with Black and Hispanic recipients. Additionally, Kenworthy and Igra (2022) reveal that campaigns in the US raise fewer funds in regions with higher medical debt, lower insurance coverage, and reduced income levels. Similarly, van Duynhoven et al. (2019) find comparable patterns in Canada, with crowdfunding disproportionately favouring regions with greater socioeconomic advantage.

Inequities in medical crowdfunding can compound existing health-related disparities. Across the world, existing healthcare infrastructures often exhibit inequities in terms of access and service provision, impacting the most vulnerable populations, including Black communities in the United States or marginalised caste groups in India (Abedi et al., 2021; Thapa et al., 2021). These vulnerable groups are more susceptible to bearing the impact of health-related crises and tend to perform poorly on health indicators (Stewart et al., 2008).

As a rapidly expanding for-profit industry striving to offer a widespread platform for accessing funds for medical treatment, crowdfunding possesses the potential to influence health inequities. While studies in Western contexts portray a concerning narrative of crowdfunding exacerbating existing disparities, its impact in India warrants further investigation. This paper aims to explore this issue within the context of India's caste system. Specifically, we aim to determine whether the recipient's caste influences the performance of medical crowdfunding campaigns in India.

3. Data and methods

3.1. Data description

Our research is based on data obtained from Ketto, one of India's largest crowdfunding platforms. Initially focused on medical crowdfunding, Ketto has expanded its categories to include education, animals, creativity, environment, children, and community development. However, medical crowdfunding remains the predominant category on the platform. Over time, Ketto has established itself as a prominent crowdfunding platform in India, raising an estimated total of INR 11 billion (approximately USD 147 million) in 2020 alone (ISB, 2021).

Our study centres on health-related crowdfunding campaigns on Ketto that commenced between February 1, 2020, and December 31, 2022. To identify health-related campaigns, we compiled a list of all campaigns listed under the 'Medical' category on the website. Additionally, we noticed that numerous campaigns in other categories, like 'Children' or 'Women', also addressed health-specific crowdfunding needs. To incorporate such campaigns in our sample, we conducted searches within campaign texts for health-related keywords. These keywords encompassed terms such as 'medical', 'medicine', 'health', 'disease', 'corona', 'COVID', and 'illness'. From a total of 15,603 campaigns initiated on Ketto during the study period, our filtering criteria resulted in a sample of 8,721 campaigns related to health. Subsequently, campaigns initiated by non-governmental organisations (NGOs) were excluded as our primary focus revolves around the caste identity of individual campaign recipients.

We employed custom Python scripts to gather comprehensive data on these campaigns, following a methodology previously utilised by researchers examining medical crowdfunding in the United States (Igra, 2022; Kenworthy and Igra, 2022; Davis et al., 2023). Examples of medical crowdfunding campaigns on Ketto can be found on the platform website: https://www.ketto.org/. Our data collection encompassed the campaign's text (title and description), launch and end dates, recipient's name, location, specified target amount, raised amount at the time of data collection, number of entities contributing to the campaign, indication of 'Urgent' or 'Trending', number of campaign images, and whether the campaign included a video.

3.2. Variables

3.2.1 Location of the recipient

The location details provided by campaign recipients on Ketto varied, with some specifying their state of residence within India, while others mentioned their city or town. To standardise this information, we employed geocoding techniques on the address details provided by the recipients. Custom scripts utilising OpenStreetMap on R were used for this purpose (Cambon et al., 2021). Given the variability in location information across recipients, the geocoding process aimed to determine the state as the primary location for each recipient.

Out of the 8,721 campaigns analysed, successful geocoding was achieved for 7,643 campaigns, allowing us to standardise and extract state-level geographical information.

3.2.2 Caste identity of the beneficiary/recipient

Assessing demographic information directly from campaign pages presents challenges. To operationalise the caste identity of campaign recipients, we utilised the names provided on the campaign pages. We categorised individual recipients into Scheduled Castes (SC), encompassing the formerly marginalised 'untouchable' caste groups at the lower hierarchy, Scheduled Tribes (ST), representing India's indigenous tribes, and a residual category labelled as OTHERS, which includes predominantly the 'upper' castes and the Other Backward Castes (a politically dominant group in India currently) (Borooah, 2005).

Our approach to inferring caste identity from recipients' names is based on the understanding that last names in India exhibit varying prevalence across caste groups. For instance, last names like 'Sharma' and 'Mehta' are highly likely to belong to dominant caste groups (OTHERS). Conversely, names like 'Basumatary' and 'Boro' are associated with the ST group, while 'Jatav' and 'Valmiki' are surnames commonly linked to the SC group. This strategy aligns with prior research that utilised last names for caste inference in the Indian context (Banerjee et al., 2009; Bhagavatula et al., 2023).

We employed the open-source 'outkast' package to deduce caste identity from recipients' last names (Laohaprapanon and Sood, 2020). 'Outkast' utilises parsed data from the Socio-Economic Caste Census (SECC), encompassing 140 million Indian names across 19 states (Sood and Laohaprapanon, 2018). The SECC data provides caste identities (SC, ST, and OTHERS) aligned with individual names and their respective states of residence. It's crucial to note that the mapping of last names to castes isn't a one-to-one relationship, and the distribution of last names among castes varies across geographical regions or states. For example, the mapping of the 'Dubey' surname to caste might differ between Uttar Pradesh and Orissa, where the surname is prevalent. The 'outkast' package utilises SECC data to estimate probabilities of a specified last name belonging to a particular caste group in a given Indian state. These probabilities are based on the actual distribution of caste identity associated with the specified last name in the SECC data. For instance, in the north-eastern states, 98.4% of individuals with the last name 'Basumatary' belong to the SC group in the SECC data, while 99% of 'Dubey' surname holders in Uttar Pradesh belong to the OTHERS group (Laohaprapanon and Sood, 2020).

For our research, we inferred caste identity for Ketto campaign recipients based on the estimated probabilities for their last names obtained from the 'outkast' package. Specifically, we assigned the caste group that the last name had the highest probability of belonging to in the SECC data. We utilised state-level information (derived from the geocoding process) to determine the mapping of last names to caste groups for all campaign recipients. Out of our sample of 7,643 campaigns, the 'outkast' package successfully mapped caste groups for recipients in 5,527 unique campaigns.

3.2.3 Gender of the campaign recipient

We employed the open-source package 'naampy' to probabilistically infer gender from the campaign recipient's first name (Laohaprapanon et al., 2022). 'Naampy' utilises India's electoral rolls data from 2017, where every state in India is mandated to publish a comprehensive list of registered voters for all elections (Sood and Dhingra, 2023). These electoral rolls contain first and last names, along with gender, age, and location information of all registered voters. 'Naampy' utilises this dataset to map first names to genders (male, female, and third gender). This inference is based on the premise that various first names in India have distinct likelihoods of belonging to a specific gender. Earlier research in the United States has utilised Census data to infer gender from first names by considering the distribution of names across genders in the population (Tzioumis, 2018). 'Naampy' is customised specifically for Indian names, providing an advantage due to the unique associations of most Indian first names with particular genders. While some names might be unisex, the majority of Indian first names tend to uniquely correspond to a specific gender. For example, names like 'Rahul' and 'Arjun' exhibit a 99.9% probability of being male, whereas 'Ritu' and 'Parvati' show a 99.9% probability of being female.

Similar to our approach for inferring caste, we mapped the recipient's first name to the gender it had the highest probability of belonging to based on the Indian electoral data. In our dataset, all recipients were classified into the 'male' and 'female' categories, with no recipients assigned to the third gender.

3.2.4 Other control variables

Beyond location, caste, and gender, our analysis encompasses additional control variables identified as relevant in prior crowdfunding literature. Specifically, we consider the number of images in the campaign, whether a video is included, the length of the campaign text, sentiment within the campaign text, campaign duration (adjusted for ongoing campaigns until the data collection point), and the requested amount by the campaign (Mahmood et al., 2019; Wang et al., 2020; Su et al., 2023; Igra, 2022). In addition, recipients are given a choice to mark their requirement as 'Urgent' as Ketto. Campaigns are also identified by whether they achieved 'Trending' status, which would typically refer to a campaign that has generated significant donation interest. We use both indicators as control variables in our models.

For health-related campaigns, Ketto allows recipients to specify the disease or illness they are seeking funds to treat. If the disease aligns with a standard set of illnesses, the campaign is tagged accordingly for search functions. We include a binary control variable denoting whether the campaign includes a recognised disease/illness tag.

While most control variables are directly derived from campaign information, we utilise an open-source algorithm called 'Vader' to infer the sentiment of the campaign text (Hutto and Gilbert, 2014). Vader's applicability spans various textual inputs and has been utilised by several researchers for sentiment analysis (Berger et al., 2020). Notably, the Vader model considers valence, distinguishing sentiments even when negations are present - for example, 'happy' signifies positive sentiment, whereas 'not happy' implies a negative sentiment. This accounts for valence, an improvement over traditional dictionary approaches that solely count word occurrences to gauge sentiment. The Vader model categorises the campaign text (title and description) into three sentiment categories: positive, negative, and neutral (Hutto and Gilbert, 2014).

3.2.5 Outcome variables

Our primary outcome variable centres on the funding raised by the campaign. Additionally, we explore models utilising alternative outcome variables. These models include the proportion of the target amount raised and a binary variable indicating whether the campaign achieved its target. Furthermore, we test models employing the number of donors to a campaign and the average amount per donation as outcome variables. Our objective is to develop an understanding of the factors that lead to variations in campaign performance across recipient caste categories.

3.3. Regression model

We estimate the impact of the recipient's caste identity on campaign performance using log-linear models specified in Equation 1.

$$Y_i = \beta_0 + \beta_1 \cdot Caste_i + Z + \phi + \epsilon_i \tag{1}$$

Where Y_i represents the outcome variable measuring campaign i performance (such as amount raised, achieving campaign target, proportion of target amount raised, number of donors, and amount per donor). $Caste_i$ denotes the campaign recipient's caste identity inferred through the recipient's last name. Z incorporates control variables, detailed in Sections 3.2.3 and 3.2.4. Notably, the campaign category control is omitted as 98.5% of the sample campaigns belong to Ketto's 'Medical' category. However, the results are robust to controlling for category, and also if campaigns not in the 'Medical' category are excluded from the analysis.

The term ϕ includes fixed effects for the model. In particular, healthcare policy-making powers in India are divided between the federal and State governments. Public health and sanitation is one of the subjects under which the State may make laws as per the Constitution of India (Chokshi et al., 2016). As a result, medical requirements and funding may vary at the State level, potentially impacting the demand and supply of medical crowdfunding. To account for this, we include State-level fixed effects in our models. In addition, we include fixed effects for the month and year in which the campaign was launched on Ketto.

Several continuous variables in our analysis exhibit right-skewed distributions. Following recent studies on modelling curvilinear relationships, we adopt log-transformed values for these variables in our models (Haans et al., 2016). Similar methodologies have been employed in examining medical crowdfunding in the United States (Igra, 2022). Figure 1 illustrates histograms for selected variables that approximate normal distribution when log-transformed.

Our hypothesis is that the coefficient β_1 will be negative for SC and ST when compared with the OTHERS group, indicating inequality in medical crowdfunding in India. We expect medical crowdfunding to reflect broader patterns of caste inequality observed in the country. It is important to make explicit an unstated assumption in our analysis. Campaigns on Ketto do not provide any information about the recipient's caste identity. However, we are assuming that the recipient's name serves as a marker of caste. While the data from the SECC provides evidence that last names in India systematically map to different caste groups (Laohaprapanon and Sood, 2020; Sood and Laohaprapanon, 2018), it is worth noting that donors *also* need to be able to infer caste from last names for crowdfunding inequality to exist. Prior research suggests that this is indeed the case, with respondents in India being able to associate a high proportion of last names with caste categories, atleast for the regions that they belong to (Banerjee et al., 2009). Prior work on charitable giving in the Indian context has also used last names as a proxy for caste groups to understand caste-based inequality in contributions (Deshpande and Spears, 2016).

4. Results

4.1. Initial results

Health-related campaigns raised INR 13.8 Bn (∼USD 172 Mn[2]) from 595,200 donations on Ketto between February 2020 and December 2022. 250 out of 5,528 campaigns achieved their funding targets (a success rate of 4.5%). The low success rates are similar to medical crowdfunding outcomes in the United States (Igra et al., 2021). Highly successful campaigns dominate fundraising on Ketto - the top 1%, 5% and 25% of campaigns accounted for 21%, 44% and 80% of all funds raised, respectively. In contrast, the bottom 16% of campaigns raised less than INR 1,000 each and accounted for only 0.02% of total fundraising.

Table 1 provides an initial indication of caste inequality in Ketto. Fundraising is dominated by the OTHERS group. OTHERS accounted for 89.9% of all campaigns started and 91.6% of all fundraising on Ketto during the study period. As compared to their share of India's population, SCs and STs are underrepresented on Ketto (Census of India, 2011). Figure 2 shows the average outcome for a Ketto campaign by the caste identity of the recipient. A campaign started by an individual from the OTHERS group raised 30.6% more than the average SC campaign and 10.7% more than the average ST campaign. The variation in success rates across caste groups also suggests caste inequality in crowdfunding outcomes. No ST campaign achieved its target, while 3.2% and 4.8% of campaigns by SC's and OTHERS achieved their funding targets, respectively.

[2] USD 1 = INR 80

Table A1 provides descriptive statistics for the variables used in the regression models (Table S1 in the Supplementary Information provides campaign counts by state; Table S2 provides correlation statistics for the regression variables). It is worth noting that the average SC campaign had a higher fundraising target than OTHERS, while the average ST campaign had a smaller fundraising goal. Male recipients accounted for more than 60% of all recipients across caste groups. A majority of campaigns (>95%) marked themselves as 'Urgent'. In contrast, most campaigns did not achieve 'Trending' status. A unique feature in Ketto is that campaigns specify the start and end dates for fundraising (unlike peer platforms such as GoFundMe where recipients are not mandated to specify an end date). Interestingly, ST campaigns had a lower average duration of 54 days, compared to 70 and 74 days for OTHERS and SC recipients, respectively. ST campaigns were likely to have fewer images, were less likely to specify the disease, and were more likely to have a negative sentiment in the campaign text. 99% of campaigns across caste groups did not have a video.

4.2. Regression model results

Table 2 provides the results of the regression models for all dependent variables. The models in Table 2 include all control variables in the analysis, in addition to accounting for State, year of launch, and month of launch fixed effects.

Model 1 in Table 2 provides the results for an OLS model with the amount raised by the campaign as the dependent variable. The coefficients for SC and ST-led campaigns are negative and significant, implying lower amounts raised compared to those led by individuals from the reference group of OTHERS. An increase in the amount targeted by a campaign leads to an increase in the funds raised. As expected, campaigns that achieve 'Trending' status raise more funds. Marking campaigns as 'Urgent' is also beneficial in terms of fundraising outcomes. We find no evidence of gender discrimination, with the coefficients for 'Male' led campaigns not significant. However, >60% of campaigns across caste groups are led by males (as shown in Table A1), implying that while gender inequality may not operate at an individual campaign level, male-led campaigns still dominate overall fundraising. Specifying a disease in the campaign and having a greater number of images aid campaign outcomes. In contrast, the coefficients for the length of the campaign text, having a video in the campaign, and the sentiment of the campaign text are not significant.

It is important to note that the coefficients for SC and ST groups in Model 1 operate on logarithmic scales. For instance, consider an urgent, not trending, male-led campaign from the State of Maharashtra with two campaign images, targeting a fundraising amount of INR 1 million. The predicted values for the funds raised are INR 0.21 million (OTHERS-led campaign), INR 0.1 million (SC-led campaign), and 0.08 million (ST-led campaign).

Model 2 provides the results for a logistic model with success as a binary dependent variable. We find that ST-led campaigns are significantly less likely to succeed in raising their targeted funds as compared to OTHERS. While an increase in the target amount reduces the likelihood of success, trending campaigns are more likely to be successful. Once again, we find no evidence for gender inequality at a per-campaign level, and our results show that an increase in duration reduces the likelihood of success. Model 3 changes the dependent variable to the proportion of target funds raised by the campaign. The results are similar to Model 1.

There are two possible explanations for lower fundraising by SC and ST individuals. Firstly, it could be that individuals contributing to SC and ST campaigns belong to lower socio-economic groups than contributors to OTHERS-led campaigns. This could be especially true in the case of caste-related homophily in donations, as administrative data from India clearly indicates that SC and ST communities are economically worse off as compared to the dominant caste groups (Thorat and Neuman, 2012). Prior research has clearly demonstrated the role of network effects and homophily in shaping economic outcomes (Jackson, 2021), including in crowdfunding (Igra, 2022). In such a case, the average donation for SC and ST campaigns should be smaller. Alternately, SC and ST campaigns could attract smaller networks of potential donors, which could result in fewer donors backing such campaigns. Model 4 tests the number of donors as a dependent variable. We find that the coefficients for both SC and ST groups are negative but not significant. We find stronger evidence for the impacts of caste in Model 5, which shows that the average donation for both SC and ST-led campaigns is significantly lower than that for the OTHERS group.

4.3. Exploring explanatory mechanisms

We conduct a mediation analysis to examine whether campaign characteristics explain the caste-based disparities in campaign outcomes. Since our independent variable (caste) is categorical with multiple levels (SC, ST and OTHERS), we conduct a mediation analysis for each level separately (Hayes and Preacher, 2014; Imai et al., 2011). The dependent variable is the amount raised by the campaign. Results are provided in Table 3. Whether a disease is recognized has a negative mediating effect on campaign outcomes for SC individuals. No other mediator (for SC and ST) is significant at the 95% level. Total effects and direct effects for SC and ST are negative and significant, indicating that factors other than observed campaign characteristics also contribute to the disparities between caste groups. As a supplementary analysis, we also report and discuss results of regression models that include interaction terms between caste and campaign characteristics in Table S3 in the Supplementary Information.

4.4. Robustness checks

4.4.1 Varying the confidence of last name–caste mapping

As we discuss in Section 3.3, caste inequality in an online crowdfunding setting like Ketto is contingent on donors being able to map last names to specific caste groups. In our main analysis, we assign caste identity based on the maximum probability of the recipient's last name occurring in a particular caste category in the SECC data. While prior research does support the argument that the average Indian citizen can map last names to caste (Banerjee et al., 2009), this might become difficult for particular last names where the last name–caste mapping is not very clear. For instance, the last name 'Ahire' from Maharashtra is distributed across caste groups as follows: 42% (OTHERS), 34% (SC), and 24% (ST). While our analysis would map 'Ahire' to the OTHERS category, it would be difficult to argue that inequality in the case of such last names on Ketto is being driven by caste, especially in the absence of other caste-related information.

As a robustness check, we conduct a subsample analysis that considers campaigns with higher confidence in assigning the recipient to a particular caste category. Specifically, we use a subset of campaigns where the last name has a >90% probability of being assigned to a unique caste group. Out of the total sample, 2,541 campaigns fulfill this requirement. The results of the analysis are presented in Table 4. The results provide stronger evidence for caste-driven inequality than the full sample analysis. The coefficients for SC and ST groups are negative and significant across all dependent variables, and the magnitude of the coefficient is larger across all models. As one might expect, caste-driven inequality is stronger when attributing the recipient's last name to a caste category is easier.

To provide further evidence, we conduct a second subsample analysis with campaigns where it is more difficult to map the recipient's last name to a unique caste group. Specifically, we consider all campaigns where no caste group has a >90% probability of mapping onto the recipient's last name. 2,986 campaigns fulfill this requirement. The results, provided in Table 5, are revealing. The coefficients for caste are no longer significant in some models (for instance, ST in Model 5). In cases where the coefficients are significant, the magnitude is smaller as compared to the analysis in Tables 2 and 4. When donors cannot map the recipient's last name to a particular caste group with a high degree of confidence, we do not observe the same level of caste inequality that was present in the full sample.

4.4.2 Alternate model specifications

For all dependent variables, we also check for alternate model specifications that include different combinations of the control variables and fixed effects. The results are provided in Tables S4, S5, S6, S7, and S8 in the Supplementary Information file. In each table, Model 1 tests the impacts of caste identity on campaign outcomes with no control variables. Model 2 adds the target amount specified by the campaign as a control. Model 3 adds all other controls specified in the analysis. Finally, Model 4 further incorporates fixed effects for the State in which the recipient is located, the year of launch, and month of launch of the campaign. Model 4 in Tables S4 to S8 corresponds to the final model for the dependent variables presented in Table 2. Across all model specifications, our results are qualitatively unchanged.

5. Discussion

During the period between February 1, 2020, and December 31, 2022, medical crowdfunding on Ketto proved highly competitive, with the top 1% of campaigns gathering 24% of all funds raised. Conversely, the lowest 16% accounted for only 0.02% of the funding on the platform. While several explanations could account for the poor performance of the majority of campaigns, our findings underscore the influence of caste in determining fundraising outcomes. At its simplest level, we discern inequality in campaign creation - while SC and ST groups constitute nearly a quarter of India's population, they account for only 10.2% of all campaigns initiated on Ketto. Additionally, inequality at a per-campaign level implies that SC and ST-led campaigns accumulate just 8.4% of the funds raised. Specifically, campaigns initiated by individuals from the OTHERS group typically raise 30% more than average SC campaigns and 10% more than average ST campaigns. This inequality persists despite considering various textual and visual factors related to the campaign and controlling for the time of the campaign launch and the recipient's State. Our findings also highlight that lower average donations contribute to decreased fundraising for SC and ST groups. Overall, our findings contribute to the literature on inequality in crowdfunding, primarily concentrated on Western countries (Kenworthy and Igra, 2022; van Duynhoven et al., 2019).

The subsample analysis we conduct in Section 4.4.1 provides supporting evidence that caste, inferred through last name, drives the observed inequality. When testing the model for a set of campaigns where the confidence of last name–caste mapping for the recipient is high, the differences in outcomes between SC/ST and OTHERS increase. Conversely, the observed inequality diminishes when the confidence of last name–caste mapping is low. In contrast, our preliminary mediation analysis suggests that other campaign characteristics (such as textual and visual features) do not provide a strong explanation for the caste inequality that we observe.

While our results depict a discouraging scenario of crowdfunding's capability to address systemic inequalities in India's healthcare system, our study presents a preliminary examination that has several limitations. Firstly, the data does not encompass other major players in the Indian crowdfunding market, such as Milaap and ImpactGuru (ISB, 2021). The analysis is also confined to factors observable from the campaign pages of specific campaigns, possibly neglecting other unobserved factors that influence campaign outcomes. Secondly, we lack self-reported caste information and assign caste identity to campaign recipients using probabilistic estimates. It is noteworthy that the same fact is also valid for individuals donating on Ketto - if caste influences their donation decisions, the recipient's last name is the most likely indicator of their caste identity. Thirdly, our results reflect a segment of society that has the abilility to start online crowdfunding campaigns. This potentially understates caste-based health financing inequality in India as many marginalised groups individuals and families may not have the means to access online crowdfunding. Finally, the open-source algorithm we use to assign caste to recipients cannot identify names for around 20% of campaigns. Although the algorithm identifies every name present in India's SECC data, which encompasses 140 million individuals, there might be selectivity in the kinds of names that remain unidentified.

6. Conclusion

Previous studies have highlighted the detrimental impact of caste on health in India. Individuals from marginalised castes often exhibit poorer health indicators, are more susceptible to health crises (such as the pandemic), and have limited access to healthcare services (Borooah, 2010). Our findings suggest that online retail crowdfunding exacerbates these existing inequalities. Interestingly, our results hint at a disparity between the medical crowdfunding needs and the actual outcomes. While individuals from marginalised castes may have a greater requirement for financial aid to meet their medical needs, funds raised on platforms like Ketto disproportionately favour dominant caste groups.

There are multiple channels through which the observed inequality could operate. Marginalised caste groups in India typically face disadvantages in accessing education and other resources (Thorat and Neuman, 2012). Initiating a campaign on Ketto requires some level of English proficiency, access to the internet, possession of a bank account, documentary evidence of identity (which Ketto mandates for starting a campaign), and supporting data to justify their need (TOI, 2022). These requirements present barriers to entry, possibly explaining the disproportionately low share of campaigns started by SC and ST individuals.

However, the fact that fundraising for SC and ST recipients is lower even after they successfully launch a campaign is possibly a matter of greater concern. Because of lower access to education and resources, it is possible that campaigns by SC and ST individuals are poorly designed and do not communicate their needs well. While our mediation analysis provides indicative results, further research is required to understand the exact mechanisms driving inequality. Our findings (especially the subsample analysis in Section 4.4.1) suggest that a primary driver of caste-based inequality on Ketto is through the recipient's last name, which in the Indian context is an indicator of their caste identity.

The implications of mapping last names to caste could be manifold. On one level, caste-based inequality on Ketto might be a consequence of active (or implicit) discrimination by donors. However, an alternative explanation arising from last name–caste mapping is plausible. SC and ST-led campaigns might attract donors with lower financial capacity compared to dominant caste groups. If there is donor homophily in this context, it could explain the smaller average donation sizes for SC and ST campaigns. For instance, Igra (2022) demonstrate systematic differences in donor networks based on the race/ethnicity of the campaign recipient in the United States. While campaign images and names might hint at race and ethnicity in the American context, we believe the recipient's last name serves as the primary source of caste identity information in India.

Nevertheless, it's essential to acknowledge that crowdfunding occurs on a for-profit platform that operates based on its economic imperatives. Platforms like Ketto may favour highlighting campaigns with higher chances of getting funded since their revenue is tied to donations. Our results indicate that the 'trending' status improves campaign outcomes, and previous studies have shown that higher positions on a webpage receive more clicks/purchases (Abdollahpouri, 2019). While biases like "presentation bias" and "ranking bias" are established in internet studies (Caliskan et al., 2017), further research is imperative to explore whether similar biases exist in crowdfunding and if they are caste-driven in the Indian context.

Overall, our findings resonate with broader insights from crowdfunding and the impacts of technological advances from across the world. Researchers have noted that technological advances such as crowdfunding are not neutral and tend to amplify existing societal biases. For instance, consider the racial biases demonstrated by artificial intelligence algorithms, including in the field of healthcare (Obermeyer et al., 2019). Our results underscore the need for steps to make technological advancements such as crowdfunding more equitable.

We offer a few preliminary policy recommendations in this regard. First, researchers and policymakers should ask for data from crowdfunding platforms to become transparent and publicly available to enable improved research and policymakers. Given the rapid growth of the sector, there is a clear need to understand and address inequities before they become a mjor concern. Second, policymakers should call for transparency in the algorithms used by for-profit crowdfunding firms to rank and show campaigns on their webpages (as placement of campaigns can influence results). Finally, there is a need for health systems to become more robust and equitable to address the needs of vulnerable groups more effectively and alleviate dependence on crowdfunding.

Acknowledgments

Data and code for the article will be made available to bonafide researchers upon request.

References

Abdollahpouri, H. (2019). Popularity bias in ranking and recommendation. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pages 529–530.

Abedi, V., Olulana, O., Avula, V., Chaudhary, D., Khan, A., Shahjouei, S., Li, J., and Zand, R. (2021). Racial, economic, and health inequality and covid-19 infection in the united states. Journal of racial and ethnic health disparities, 8:732–742.

Acharya, S. (2010). Caste and patterns of discrimination in rural public health care services. Blocked by Caste: Economic Discrimination in Modern India (New Delhi: Oxford University Press, 2010), pages 208–229.

Andrews, M., Areekal, B., Rajesh, K., Krishnan, J., Suryakala, R., Krishnan, B., Muraly, C., and Santhosh, P. (2020). First confirmed case of covid-19 infection in india: A case report. The Indian journal of medical research, 151(5):490.

Banerjee, A., Bertrand, M., Datta, S., and Mullainathan, S. (2009). Labor market discrimination in delhi: Evidence from a field experiment. Journal of comparative Economics, 37(1):14–27.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., and Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. Journal of marketing, 84(1):1–25.

Bhagavatula, S., Bhalla, M., Goel, M., and Vissa, B. (2023). Social diversity in corporate boards and firm outcomes. Journal of Corporate Finance, 83:102499.

Borooah, V. (2010). Inequality in health outcomes in india: the role of caste and religion.

Borooah, V. K. (2005). Caste, inequality, and poverty in india. Review of development economics, 9(3):399–414.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334):183–186.

Cambon, J., Hernangómez, D., Belanger, C., and Possenriede, D. (2021). tidygeocoder: An r package for geocoding. Journal of Open Source Software, 6(65):3544.

Chakraborty, T., Mukherjee, A., Saha, S., and Shukla, D. (2023). Caste, courts and business. Journal of Economic Behavior & Organization, 212:333–365.

Chandramouli, C. and General, R. (2011). Census of india 2011. Provisional Population Totals. New Delhi: Government of India, pages 409–413.

Chokshi, M., Patil, B., Khanna, R., Neogi, S. B., Sharma, J., Paul, V., and Zodpey, S. (2016). Health systems in india. Journal of Perinatology, 36(3):S9–S12.

Davis, A. R., Elbers, S. K., and Kenworthy, N. (2023). Racial and gender disparities among highly successful medical crowdfunding campaigns. Social Science & Medicine, 324:115852.

Debnath, S. and Jain, T. (2020). Social connections and tertiary health-care utilization. Health Economics, 29(4):464–474.

Deshpande, A. and Spears, D. (2016). Who is the identifiable victim? caste and charitable giving in modern india. Economic Development and Cultural Change, 64(2):299–321.

Dhar, S. (2018). Why indians are crowdfunding medical treatments. Times of India, https://tinyurl.com/2p8f5y4n, accessed June 6, 2023.

Economic Survey (2020). Outkast. https://github.com/appeler/outkast. accessed 18 May 2023.

GoFundMe (2023). Get help with medical fundraising. https://www.gofundme.com/start/medical-fundraising. accessed 10 June 2023.

Gupta, A. and Coffey, D. (2020). Caste, religion, and mental health in india. Population Research and Policy Review, 39:1119–1141.

Haans, R. F., Pieters, C., and He, Z.-L. (2016). Thinking about u: Theorizing and testing u-and inverted u-shaped relationships in strategy research. Strategic management journal, 37(7):1177–1195.

Hayes, A. F. and Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. British journal of mathematical and statistical psychology, 67(3):451–470.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media, volume 8, pages 216–225.

Igra, M. (2022). Donor financial capacity drives racial inequality in medical crowd-sourced funding. Social Forces, 100(4):1856–1883.

Igra, M., Kenworthy, N., Luchsinger, C., and Jung, J.-K. (2021). Crowdfunding as a response to covid-19: Increasing inequities at a time of crisis. Social Science & Medicine, 282:114105.

Imai, K., Keele, L., Tingley, D., and Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. American Political Science Review, 105(4):765–789.

ISB (2021). Donation-based Digital Crowdfunding—The Changing Face of Giving in India. https://tinyurl.com/mu87sxju. accessed 24 May 2023.

Jackson, M. O. (2021). Inequality's economic and social roots: The role of social networks and homophily. Available at SSRN 3795626.

Keane, M. and Thakur, R. (2018). Health care spending and hidden poverty in india. Research in Economics, 72(4):435–451.

Kenworthy, N. and Igra, M. (2022). Medical crowdfunding and disparities in health care access in the united states, 2016–2020. American Journal of Public Health, 112(3):491–498.

Ketto (2023). What ketto is made of. https://www.ketto.org/new/about/about-us. accessed 10 June 2023.

Kothari, R. (1995). Caste in Indian politics. Orient Blackswan.

Lamba, S. and Spears, D. (2013). Caste,'cleanliness' and cash: effects of caste-based political reservations in rajasthan on a sanitation prize. Journal of Development Studies, 49(11):1592–1606.

Laohaprapanon, S. and Sood, G. (2020). Outkast. https://github.com/appeler/outkast. accessed 18 May 2023.

Laohaprapanon, S., Sood, G., and Chintalapati, R. (2022). naampy: Infer Sociodemographic Characteristics from Indian Names. https://github.com/appeler/naampy. accessed 14 May 2023.

Linos, K., Jakli, L., and Carlson, M. (2021). Fundraising for stigmatized groups: A text message donation experiment. American Political Science Review, 115(1):14–30.

List, J. A. (2011). The market for charitable giving. Journal of Economic Perspectives, 25(2):157–180.

Mahmood, A., Luffarelli, J., and Mukesh, M. (2019). What's in a logo? the impact of complex visual cues in equity crowdfunding. Journal of Business Venturing, 34(1):41–62.

Mosse, D. (2020). The modernity of caste and the market economy. Modern Asian Studies, 54(4):1225–1271.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464):447–453.

O'Reilly, K., Dhanju, R., and Louis, E. (2017). Subjected to sanitation: caste relations and sanitation adoption in rural tamil nadu. The Journal of Development Studies, 53(11):1915–1928.

Panjwani, A. and Xiong, H. (2023). The causes and consequences of medical crowdfunding. Journal of Economic Behavior & Organization, 205:648–667.

Ray, P. P. (2023). Chatgpt: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems.

Snyder, J., Mathers, A., and Crooks, V. A. (2016). Fund my treatment!: A call for ethics-focused social science research into the use of crowdfunding for medical care. Social Science & Medicine, 169:27–30.

Solnit, R. (2010). A paradise built in hell: The extraordinary communities that arise in disaster. Penguin.

Sood, G. and Dhingra, A. (2023). Indian Electoral Roll PDF Corpus.

Sood, G. and Laohaprapanon, S. (2018). Socio-Economic and Caste Census 2011.

Stewart, M., Makwarimba, E., Barnfather, A., Letourneau, N., and Neufeld, A. (2008). Researching reducing health disparities: Mixed-methods approaches. Social Science & Medicine, 66(6):1406–1417.

Su, L., Sengupta, J., Li, Y., and Chen, F. (2023). "want" versus "need": How linguistic framing influences responses to crowdfunding appeals. Journal of Consumer Research, page ucad033.

Thapa, R., van Teijlingen, E., Regmi, P. R., and Heaslip, V. (2021). Caste exclusion and health discrimination in south asia: A systematic review. Asia Pacific Journal of Public Health, 33(8):828–838.

Thorat, S. and Neuman, K. S. (2012). Blocked by caste: Economic discrimination in modern India. Oxford University Press.

TOI (2022). Ketto and milaap statements on crowdfunding fraud. Times of India, https://tinyurl.com/yx58yrrx, accessed June 6, 2023.

Tzioumis, K. (2018). Demographic aspects of first names. scientific data, 5, 180025.

van Duynhoven, A., Lee, A., Michel, R., Snyder, J., Crooks, V., Chow-White, P., and Schuurman, N. (2019). Spatially exploring the intersection of socioeconomic status and canadian cancer-related medical crowdfunding campaigns. BMJ open, 9(6):e026365.

Wang, W., Chen, W., Zhu, K., and Wang, H. (2020). Emphasizing the entrepreneur or the idea? the impact of text content emphasis on investment decisions in crowdfunding. Decision Support Systems, 136:113341.

Young, M. J. and Scheinberg, E. (2017). The rise of crowdfunding for medical care: promises and perils. Jama, 317(16):1623–1624.

[]

Table 1: Caste inequality in Ketto

	Share	Population	No of campaigns	Fundraising
1	OTHERS	74.8%	89.9%	91.6%
2	SC	16.6%	6.8%	5.3%
3	ST	8.6%	3.4%	3.1%

Note: (1) Population refers to share in India's population
(2) No of campaigns and fundraising refer to data from Ketto
(3) Source: Author's analysis

Table 2: Regression results - All DVs (full sample)

	Amount raised OLS (1)	Success Logistic (2)	Prop OLS (3)	No donors OLS (4)	Av donation OLS (5)
SC	−0.796***	−0.191	−0.796***	−0.047	−0.749***
	(0.197)	(0.329)	(0.197)	(0.075)	(0.163)
ST	−0.927**	−17.019***	−0.927**	−0.008	−0.919***
	(0.295)	(0.626)	(0.295)	(0.095)	(0.252)
log(Target amount)	0.451***	−1.139***	−0.549***	0.300***	0.150***
	(0.044)	(0.084)	(0.044)	(0.023)	(0.029)
Trending	4.348***	3.222*	4.348***	3.412***	0.936**
	(0.483)	(1.545)	(0.483)	(0.252)	(0.338)
Urgent	1.512***	0.500	1.512***	1.036***	0.476**
	(0.330)	(0.613)	(0.330)	(0.201)	(0.178)
Gender: Male	−0.114	−0.046	−0.114	−0.050	−0.064
	(0.071)	(0.164)	(0.071)	(0.040)	(0.051)
Duration	0.000	−0.006*	0.000	0.000	0.000
	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)
log(Text Length)	0.108	−0.742*	0.108	0.126+	−0.019
	(0.115)	(0.294)	(0.115)	(0.067)	(0.077)
Specified disease	2.042***	1.618***	2.042***	1.271***	0.771***
	(0.081)	(0.343)	(0.081)	(0.053)	(0.053)
No of images	0.058***	0.003	0.058***	0.048***	0.011*
	(0.009)	(0.016)	(0.009)	(0.006)	(0.004)
Has video	−0.226	−0.062	−0.226	−0.167	−0.059
	(0.207)	(0.675)	(0.207)	(0.150)	(0.120)
Sentiment: Neutral	−0.240	−0.446	−0.240	0.107	−0.347
	(0.362)	(0.708)	(0.362)	(0.191)	(0.229)
Sentiment: Positive	0.066	0.111	0.066	0.028	0.038
	(0.077)	(0.195)	(0.077)	(0.047)	(0.051)
Num.Obs.	5527	5527	5527	5527	5527
R2	0.466		0.402	0.546	0.205
R2 Adj.	0.460		0.395	0.541	0.196
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] SC - Scheduled caste, ST - Scheduled tribe

[b] Reference category for caste - OTHERS

[c] Reference category for gender - Female

[d] Reference category for sentiment - Negative

Table 3: Results for mediation analysis

Dependent Variable: log(Amount raised)

Mediation for SC

Parameter	Estimate	SE	P.value
IDE: Log (Amount requested)	-0.006	0.029	0.846
IDE: Trending	0.012	0.013	0.361
IDE: Urgent	0.011	0.014	0.423
IDE: Gender	0.005	0.004	0.248
IDE: Log (Title length)	-0.0002	0.002	0.918
IDE: Sentiment (Neutral)	0.002	0.003	0.479
IDE: Sentiment (Positive)	0.005	0.006	0.352
IDE: Disease recognized	-0.224***	0.055	0.00005
IDE: No of images	0.011	0.013	0.379
IDE: Has video	0.002	0.003	0.539
Direct effect	-0.79***	0.13	0
Total effect	-0.981***	0.149	0

Mediation for ST

Parameter	Estimate	SE	P.value
IDE: Log (Amount requested)	-0.068+	0.041	0.097
IDE: Trending	0.035+	0.019	0.058
IDE: Urgent	-0.022	0.019	0.252
IDE: Gender	0.005	0.005	0.327
IDE: Log (Title length)	-0.003	0.004	0.401
IDE: Sentiment (Neutral)	0.001	0.002	0.704
IDE: Sentiment (Positive)	-0.005	0.005	0.386
IDE: Disease recognized	0.010	0.076	0.892
IDE: No of images	-0.024	0.018	0.184
IDE: Has video	0.0004	0.002	0.823
Direct effect	-0.93***	0.19	0
Total effect	-0.988***	0.209	0.00000

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] The table provides standardized results for the coefficients

[b] IDE: Indirect effect

[c] SC - Scheduled caste, ST - Scheduled tribe

[d] Reference category for caste - OTHERS

[e] Reference category for gender - Female

[f] Reference category for sentiment - Negative

Table 4: Regression results - All DVs (high confidence in name–caste mapping)

	Amount raised OLS (1)	Success Logistic (2)	Prop OLS (3)	No donors OLS (4)	Av donation OLS (5)
SC	−2.153**	−16.833***	−2.153**	−0.398	−1.754**
	(0.787)	(0.960)	(0.787)	(0.242)	(0.668)
ST	−1.618**	−19.661***	−1.618**	0.046	−1.663***
	(0.573)	(1.433)	(0.573)	(0.164)	(0.484)
log(Target amount)	0.427***	−1.490***	−0.573***	0.265***	0.162***
	(0.069)	(0.186)	(0.069)	(0.033)	(0.047)
Trending	4.446***	−11.726***	4.446***	3.500***	0.946+
	(0.724)	(1.994)	(0.724)	(0.387)	(0.490)
Urgent	1.864***	0.594	1.864***	1.270***	0.593*
	(0.520)	(1.164)	(0.520)	(0.297)	(0.280)
Gender: Male	−0.083	0.199	−0.083	−0.015	−0.068
	(0.103)	(0.286)	(0.103)	(0.060)	(0.072)
Duration	0.000	−0.014	0.000	0.000	0.000
	(0.001)	(0.010)	(0.001)	(0.000)	(0.000)
log(Text Length)	0.303+	−0.599	0.303+	0.204*	0.099
	(0.177)	(0.496)	(0.177)	(0.103)	(0.118)
Specified disease	1.982***	1.960**	1.982***	1.230***	0.751***
	(0.123)	(0.607)	(0.123)	(0.081)	(0.082)
No of images	0.058***	0.012	0.058***	0.052***	0.006
	(0.015)	(0.035)	(0.015)	(0.011)	(0.007)
Has video	0.029	0.600	0.029	−0.141	0.170
	(0.295)	(1.084)	(0.295)	(0.191)	(0.180)
Sentiment: Neutral	−0.186	−1.042	−0.186	0.147	−0.333
	(0.642)	(1.527)	(0.642)	(0.345)	(0.379)
Sentiment: Positive	0.087	0.248	0.087	0.032	0.054
	(0.114)	(0.318)	(0.114)	(0.069)	(0.076)
Num.Obs.	2541	2541	2541	2541	2541
R2	0.479		0.411	0.551	0.236
R2 Adj.	0.466		0.397	0.540	0.217
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] SC - Scheduled caste, ST - Scheduled tribe

[b] Reference category for caste - OTHERS

[c] Reference category for gender - Female

[d] Reference category for sentiment - Negative

Table 5: Regression results - All DVs (low confidence in name–caste mapping)

	Amount raised OLS (1)	Success Logistic (2)	Prop OLS (3)	No donors OLS (4)	Av donation OLS (5)
SC	−0.583**	−0.140	−0.583**	−0.025	−0.559***
	(0.200)	(0.350)	(0.200)	(0.082)	(0.163)
ST	−0.528+	−15.324***	−0.528+	−0.102	−0.425
	(0.311)	(0.301)	(0.311)	(0.122)	(0.264)
log(Target amount)	0.464***	−1.031***	−0.536***	0.330***	0.134***
	(0.059)	(0.098)	(0.059)	(0.033)	(0.035)
Trending	4.231***	2.950	4.231***	3.262***	0.969*
	(0.653)	(1.847)	(0.653)	(0.374)	(0.451)
Urgent	1.274**	0.111	1.274**	0.855**	0.419+
	(0.415)	(0.710)	(0.415)	(0.275)	(0.221)
Gender: Male	−0.157	−0.179	−0.157	−0.079	−0.078
	(0.096)	(0.218)	(0.096)	(0.055)	(0.070)
Duration	0.000	−0.004	0.000	0.000	0.000
	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
log(Text Length)	−0.037	−0.739+	−0.037	0.072	−0.109
	(0.152)	(0.399)	(0.152)	(0.088)	(0.104)
Specified disease	2.073***	1.527***	2.073***	1.299***	0.774***
	(0.109)	(0.430)	(0.109)	(0.071)	(0.070)
No of images	0.058***	−0.010	0.058***	0.044***	0.014*
	(0.012)	(0.021)	(0.012)	(0.008)	(0.006)
Has video	−0.359	−0.572	−0.359	−0.120	−0.240
	(0.304)	(0.934)	(0.304)	(0.235)	(0.170)
Sentiment: Neutral	−0.240	−0.343	−0.240	0.092	−0.332
	(0.401)	(0.969)	(0.401)	(0.199)	(0.287)
Sentiment: Positive	0.059	0.156	0.059	0.032	0.027
	(0.105)	(0.265)	(0.105)	(0.063)	(0.071)
Num.Obs.	2986	2986	2986	2986	2986
R2	0.472		0.414	0.553	0.217
R2 Adj.	0.461		0.401	0.544	0.200
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] SC - Scheduled caste, ST - Scheduled tribe

[b] Reference category for caste - OTHERS

[c] Reference category for gender - Female

[d] Reference category for sentiment - Negative

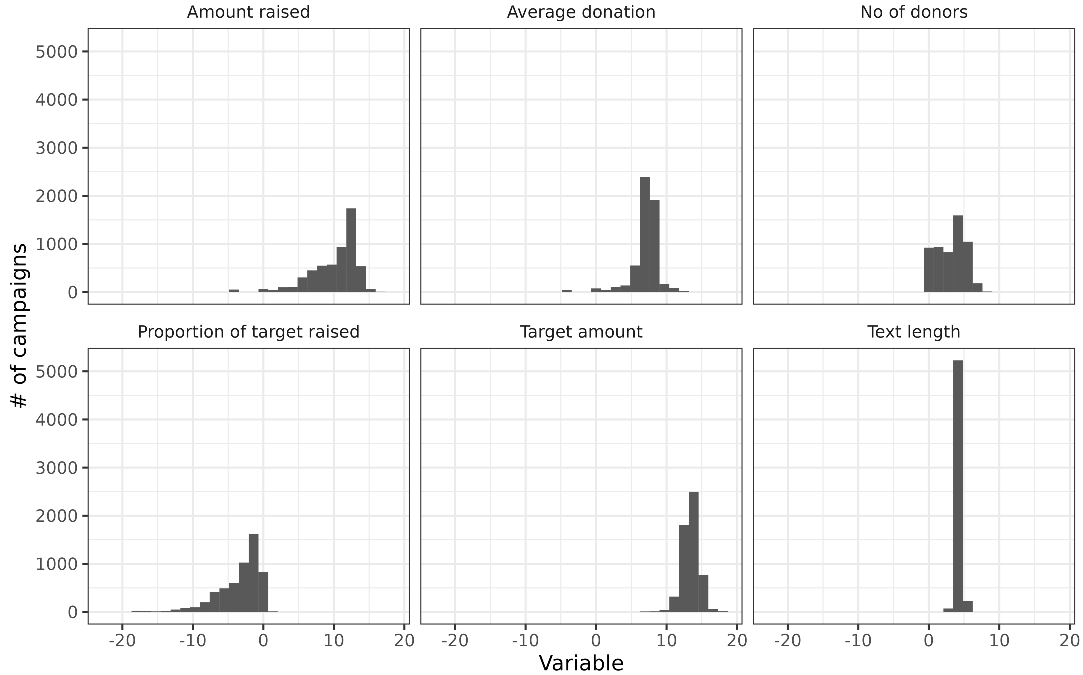

Figure 1: Distribution of selected continuous variables. Log-scales

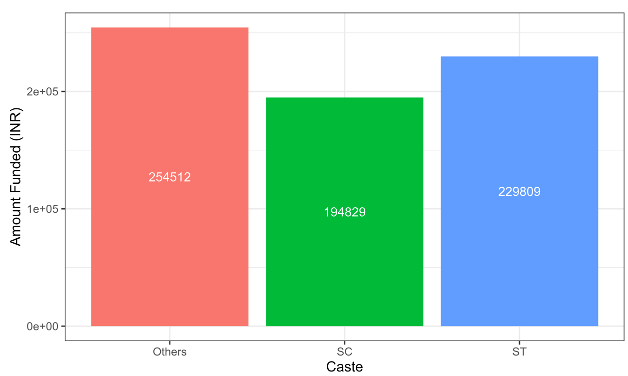

Figure 2: Variations in amount raised by caste group

Appendices

Descriptive statistics

Table A1 presents summary statistics for the variables used in the regression model.

Table A1: Summary Statistics

Variable	Caste	N	Mean	SD	Min	Median	Max
Amount raised (INR)	OTHERS	4967	254512	843841	0.01	106497	3.8e+07
	SC	374	194829	407963	0.01	49326	2817184
	ST	186	229809	464947	0.01	80338	2723477
	All	5527	249642	811552	0.01	1e+05	3.8e+07
Target amount (INR)	OTHERS	4967	1407504	3e+06	0.01	8e+05	5e+07
	SC	374	1594744	4162705	25000	6e+05	5e+07
	ST	186	1166656	1388540	2000	7e+05	1e+07
	All	5527	1412069	3e+06	0.01	8e+05	5e+07
Success (%)	OTHERS	4967	4.8	21	0	0	100
	SC	374	3.2	18	0	0	100
	ST	186	0	0	0	0	0
	All	5527	4.5	21	0	0	100
Trending (%)	OTHERS	4967	0.26	5.1	0	0	100
	SC	374	0.53	7.3	0	0	100
	ST	186	1.1	10	0	0	100
	All	5527	0.31	5.5	0	0	100
Urgent (%)	OTHERS	4967	97	17	0	100	100
	SC	374	98	14	0	100	100
	ST	186	96	20	0	100	100
	All	5527	97	17	0	100	100
Duration (days)	OTHERS	4967	70	101	0	45	1093
	SC	374	74	114	1	45	907
	ST	186	54	73	1	45	557
	All	5527	70	101	0	45	1093
Text Length	OTHERS	4967	78	29	9	71	255
	SC	374	78	31	16	71	255
	ST	186	77	29	7	71	250
	All	5527	78	29	7	71	255
Gender: Male (%)	OTHERS	4967	67	47	0	100	100
	SC	374	63	48	0	100	100
	ST	186	62	49	0	100	100
	All	5527	66	47	0	100	100
Disease specified (%)	OTHERS	4967	47	50	0	0	100
	SC	374	36	48	0	0	100
	ST	186	47	50	0	0	100
	All	5527	46	50	0	0	100
Has Video (%)	OTHERS	4967	1.3	11	0	0	100
	SC	374	0.53	7.3	0	0	100
	ST	186	1.1	10	0	0	100
	All	5527	1.2	11	0	0	100
No of images	OTHERS	4967	2.3	4.1	0	0	64
	SC	374	2.4	3.8	0	0	18
	ST	186	1.8	3	0	0	16
	All	5527	2.3	4	0	0	64
Sentiment: Positive (%)	OTHERS	4967	65	48	0	100	100
	SC	374	73	45	0	100	100
	ST	186	58	49	0	100	100
	All	5527	65	48	0	100	100
Sentiment: Negative (%)	OTHERS	4967	34	47	0	0	100
	SC	374	27	44	0	0	100
	ST	186	41	49	0	0	100
	All	5527	34	47	0	0	100