

Social Diversity and Spread of Pandemic: Evidence from India

Authors:

Upasak Das (corresponding author)
Presidential Fellow in the Economics of Poverty Reduction
Global Development Institute
University of Manchester
Manchester, United Kingdom
and
Affiliate
Centre for Social Norms and Behavioral Dynamics
University of Pennsylvania, USA
(upasak.das@manchester.ac.uk)
ORCID: [0000-0002-4371-0139](https://orcid.org/0000-0002-4371-0139)

Udayan Rathore
Oxford Policy Management
New Delhi, India
(rathore.udayan@gmail.com)
ORCID: [0000-0003-1941-1005](https://orcid.org/0000-0003-1941-1005)

Prasenjit Sarkhel
Associate Professor of Economics
Department of Economics
University of Kalyani
West Bengal, India
(prasenjitsarkhel@klyuniv.ac.in)
ORCID: [0000-0001-5486-6959](https://orcid.org/0000-0001-5486-6959)

Abstract

Existing evidence on the implications of social diversity during pandemics is limited. Compliance with the public health guidelines in such times requires coordinated community actions which might be undermined in diverse areas. In this paper, we assess the relationship between caste-group diversity and the spread of COVID-19 infection during the nationwide lockdown and unlocking period in India. Using data from multiple representative sources, we find that caste-homogeneous districts systematically took more days to cross the concentration thresholds of 50 to 500 cases. Estimates using daily cases further show that caste-homogeneous districts experienced slower growth in infection for about 2.5 months since the beginning of the lockdown. The results hold even after accounting for the emergence of initial hotspots before lockdown, broader diffusion patterns through daily fixed effects, region fixed effects, and dynamic administrative response through time-variant lagged COVID-19 fatalities at the district level. These effects are not confounded by differential levels of testing and underreporting of cases in some states. Consistent estimates from bias-adjusted treatment effects also show that our findings remain robust to potential unobservables. In terms of mechanisms, we find suggestive evidence of higher engagement of community health workers in caste-homogenous localities, which further increased after the outbreak. We also provide evidence of lower movement in these localities during the lockdown period, which would have slowed infection spread. Our findings reveal how caste-group diversity can be used to identify potential hotspots during pandemics and emphasize the importance of community health workers, social cohesion and decentralized policy response.

Keywords: Social homogeneity; India; Caste; COVID-19; Lockdown.

1. Introduction

Ethnic and social group affiliations often shape individual and group identities. These in turn influence cooperation and coordination within and across communities, which is particularly paramount during times of crises. For example, such affiliations serve as an important channel for information transmission (Larson & Lewis, 2017) and are likely to ensure compliance with social norms (Goette, Huffman & Meir, 2006). In this paper, we ask if community-level diversity (or homogeneity) can influence the overall public health outcomes during pandemics? Specifically, we study how caste-group diversity affected the spread of the SARS-CoV-2 viral infection (henceforth COVID-19) during the early phase of the pandemic in India. Here, we first assess if the time taken to cross specific thresholds of COVID-19 cases increases in districts with higher levels of caste-group diversity. We also examine the implications of caste-group diversity on the progression of daily cases during the nationwide lockdown and an equivalent period of unlocking.

During the first wave of infection in early 2020, compliance with the appropriate COVID-19 norms was widely recommended as the only protective measure against infection.¹ Ensuring compliance with these norms requires coordinated community action, more so because private protective effort involves a positive externality. Unless other community members follow compliance protocols, the effectiveness of an individual's effort is negligible. A large body of evidence from developing countries suggests that diversity can undermine voluntary contribution to public goods (Banerjee et al., 2005; Habyarimana et al., 2007). If preference is biased towards in-group members than outsiders (Alesina & Ferrara, 2005) or if more diverse societies lack social capital due to lower social interactions and interpersonal trust, a community's ability to effectively monitor and sanction violations of norms might be restricted, thereby weakening collective action (Banerjee et al., 2005; Miguel & Gugerty, 2005)². These channels suggest that diverse communities may experience a faster spread of COVID-19 infection.

¹ [India commenced its vaccination drive only on 16th January 2021.](#)

² Greater diversity is also associated with divergence in preferences and varied risk perception that may lead to a sub-optimal level of compliance (Alesina et al., 1999).

However, some recent studies have challenged this conventional wisdom and show that diversity, under certain conditions, can also support welfare gains (Gisselquist et al., 2016; Akay et al. 2017). In the context of the pandemic in Russia, Egorov et al. (2021) argue that infected individuals in diverse communities are less altruistic and hence have a lower incentive to self-isolate. This knowledge forces the uninfected population to reduce mobility, thereby arresting the spread of the virus. In addition, infection growth in diverse communities may also be inhibited by lower community interactions (Bosancianu, et al. 2020). These countervailing possibilities turn the potential relationship between social diversity and infection spread into an ambiguous one that necessitates empirical investigation.

The question assumes importance in the context of India as it is also one of the most socially diverse countries in the world that is stratified on the basis of caste and caste-groups (Deshpande 2000; Munshi, 2019). Not only is caste a dominant force in India's social structure, but it is also an essential basis of social network formation (Thorat, 2010; Desai & Dubey, 2012; Munshi, 2019). In this paper, we consider all the four prominent caste groups in India. These include the Upper Caste (Others), the socially deprived Scheduled Caste (SC, formerly untouchable caste), the ethnic and tribal groups belonging to Scheduled Tribes (ST) and the Other Backward Castes (OBC). Literature on residential patterns in rural India, where much of the population resides, shows that households follow distinct spatial segregation whereby members of a common caste live in a particular area (Munshi, 2019). These clusters are associated with high interpersonal trust, cooperation and compliance, which may impede infection transmission at the community level.

During the initial phase of the pandemic, and like other countries, India also relied on mobility restrictions through nationwide lockdowns. These were imposed from March 25 to May 31, 2020, in four continuous phases to minimize mobility and control the spread of the virus, with minor relaxations introduced over time.³ These mobility restrictions at the time were seen as one of the strictest ones imposed anywhere in the world (Gisselquist & Kundu 2020). Thereafter, from June 1, 2020, comprehensive unlocking measures were initiated, and respective state governments were given

³ <https://www.bbc.com/news/world-asia-india-52290761> (accessed on December 12, 2021).

more flexibility to decide on restrictions outside a countrywide negative list that included international travel and the opening of educational institutions.⁴

In this paper, we first check if caste-group diversity exerted any influence on the average number of days a district took to cross certain thresholds of COVID-19 cases.⁵ To track the infection trajectory on a daily basis, we use district level, daily COVID-19 data to examine if there was a significant differential in the spread of COVID-19 cases as a share of its population in caste-group homogeneous districts. In particular, after controlling for a wide range of potential confounders, we ask if caste-diverse districts in India experienced faster infection spread during the lockdown and whether these effects, if any, persisted even after these restrictions were withdrawn. Here, we consider the entire lockdown period that lasted for 68 days (March 25 to May 31, 2020) and an equivalent number of days during the unlocking period (June 1 to August 7, 2020). It is important to note that India recorded more than 2 million cases and close to 44,000 COVID-19 related deaths in this period.

Our identification strategy is based on controlling for a wide range of confounding possibilities to isolate the relationship of interest (see Section 2). In addition to a range of socio-economic, demographic, and movement & migration-related variables which are an important source of infection transmission, we also control for any COVID-19 cases in the district before the nationwide lockdown to identify potential hotspots. Also, to account for region specific characteristics like climatic conditions, that may aid or impede infection spread, we account for time-invariant region-specific fixed effects. For the day-wise analysis, as the administrative machinery and public health authorities are likely to be more sensitive to recent deaths when allocating scarce resources, we control for the rolling number of COVID-19 deaths as a share of the population in the past seven days as a measure of time-variant supply-side response. Moreover, daily fixed effects are taken to account for broader diffusion patterns of infection spread. We also undertake a range of internal validity tests to verify our findings and account for alternate channels of infection transmission (Section 3.3).

⁴ <https://timesofindia.indiatimes.com/india/lockdown-4-0-after-may-31-states-to-have-own-curbs-lists/articleshow/76024524.cms> (accessed on December 12, 2021)

⁵ We use days taken to record 50, 100, 200, 300, 400 and 500 cases as the respective thresholds. The analysis only considers districts which crossed the specified threshold in question. To ensure that districts which did not cross these thresholds did not systematically drop out of the analysis, we use a longer period for analysis, ranging March 25, 2020 to March 21, 2021.

Importantly, we account for the influence of unobserved heterogeneity and use a recent method that provides consistent estimates of bias-adjusted treatment effects to further support our findings (Altonji et al. 2005; Oster, 2019).

We find that caste-homogeneous districts, on average, took significantly more days to breach the thresholds of 50 to 500 COVID-19 cases. Lower absolute number of cases in these regions may have provided more time to public health authorities to mount an appropriate policy response. Our day-wise results indicate significantly slower growth in COVID-19 cases in areas characterized by higher caste-group homogeneity during lockdown. These gains dampen over time but remain significant throughout the lockdown period, lasting for about a week after unlocking. The results are additionally robust to potential biases emerging from factors such as differences in testing rates across states, alternate measures of caste-group diversity and disproportionately lower reporting of cases in some states.

We further offer suggestive mechanisms that can explain our findings. First, we make use of the fact that community health workers (CHW), who are the main point of contact with the primary health care system, were mobilized as frontline workers during COVID-19. Their task included engaging with the community to generate awareness about COVID-19 along with case identification, contact tracing and reporting potential infections to district health teams (Salve et al., 2023). These CHWs are generally recruited from the same community they serve. Hence, their effectiveness in performing the aforementioned tasks can potentially be higher in socially homogenous communities, which are characterised by higher reciprocity and trust that can foster higher cooperation (Gadsden et.al. 2022). Our observations using data from the National Family Health Survey dataset indicate disproportionately higher involvement of CWHs in caste-group homogeneous areas after the first COVID-19 lockdown. We argue that this higher involvement was potentially influential in reducing the spread of the infection.

Second, using the Time-Use Survey conducted in 2019, we find higher social interaction and involvement in unpaid voluntary social activities among the residents in caste-group homogeneous areas. The higher levels of socialization and participation in social activities may facilitate higher social capital formation and strengthen collective action, resulting in slower infection spread in caste

homogeneous regions. Finally, we test the implication of this potential collective action in compliance efforts during the period of study using Facebook Mobility data. The findings show lower movement of people in homogenous areas during the lockdown. This is also found to have a partial mediating effect on lowering the spread of the infection.

Our main contributions to the literature are as follows. One, we contribute to the evidence on social diversity and health outcomes in a pandemic setting and find that spread of COVID-19 was faster in caste-group diverse districts. In doing so, we also contribute to the growing literature examining the implications of social diversity and provisioning of public goods tied to public health emergencies. In the Indian context, our research builds on the work of Chakraborty and Mukherjee (2023) who looked at the relationship between regional economic inequality and spread of COVID-19. Given the vital role of caste in social stratification in India, our analysis explores the influence of caste diversity on spread of COVID-19, after controlling for economic inequality. Two, from a policy perspective, our findings reveal how caste-group diversity can be used to identify potential hotspots during pandemics. We argue that prioritizing the allocation of scarce medical and administrative resources to these vulnerable areas can slow overall infection growth. These policy prescriptions remain pertinent in the wake of the new strains of concern and inform future public health crises. Third, our research indicates that health interventions through CHWs might be relatively more effective in caste-group homogeneous during health emergencies. To tackle future pandemics, a mix of centralized and decentralized health interventions are needed for formulating effective community level health policies that facilitates deeper engagement with CHWs in socially diverse areas. Lower levels of movement in caste homogenous areas during the lockdown is also found to have some mediating effect on arresting the growth of infection. In terms of broader implications, this finding underlines the potentially positive role of social cohesion, which can increase the effectiveness of public health policies and ensure better compliance with health protocols.

The paper is structured as follows. Section 2 describes the data used and discusses the variables used and the identification strategy. Section 3 provides the regression results with robustness

checks and explores the possible mechanism that explains our findings. Section 4 discusses the policy implications, and section 5 concludes.

2. Data, variables and identification strategy

2.1 Data

District-day-level data on COVID-19 cases and deaths come from the Development Data Lab's (DDL) COVID India database, available for 684 of 718 (95%) districts in India, which is the unit of analysis.⁶ First, we estimate the number of days taken to cross thresholds from 50 to 500 at the district level, counted from the beginning of the lockdown.⁷ For example, if a district crossed 50 cases on March 28, 2020, the number of days in our case would be 4 (March 25 to March 28, 2020).⁸ The duration of daily analysis ranges from March 25 to August 7, 2020. This covers two distinct periods: (a) the first nationwide lockdown, which ended on the May 31, 2020 and, (b) the unlocking phase of an equivalent duration (68 days) starting from June 1 to August 7, 2020.

We restrict our analysis to these 68 days of the lockdown and the equivalent period of unlocking for multiple reasons. First, even though the second wave of COVID-19 in 2021 was more severe, India recorded more than 2 million cases and close to 44,000 COVID-19 related deaths in the first wave during the period of consideration. Thus, the first wave of infection is important to analyze as it set the wheels in motion for subsequent spread and mutations of the virus that devastated India in 2021. Second, the Union (Central) Government of India imposed, and later eased, mobility restrictions during the lockdown based on a common set of criteria across all Indian states. These mobility restrictions at the time were seen as one of the strictest ones imposed anywhere in the world (Gisselquist & Kundu 2020). In contrast, during the second wave of infection in 2021, state

⁶ The Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) is an open source platform to facilitate data sharing between researchers working on India (<http://www.devdatalab.org/shrug>). These numbers are taken from covidindia.com, a crowd sourced project.

⁷ The district level cases are available from beginning of February 2020 but no district recorded a total of 50 cases before lockdown (March 25, 2020).

⁸ Here, we consider 557, 555, 548, 547, 544 and 542 districts for the thresholds of 50, 100, 200, 300, 400 and 500 cases, respectively. To prevent selection bias, we consider a longer duration upto March 21, 2021 for this analysis. This ensures that districts that, which took more time to cross the respective thresholds do not drop out systematically from the analysis.

governments imposed vastly differentiated, state-specific lockdowns and interventions based on localized surges in infections, fatalities, and medical preparedness, which was dynamic in nature. Restricting the analysis to the first nationwide lockdown and corresponding unlocking phase then allows us to minimize various state-specific unobserved influences and tease out the effects of caste-group diversity on infection spread more cleanly.⁹ Third, it can also be argued that knowledge about the spread of the virus during the first wave was limited, and the potential effects of not abiding by the COVID-19 appropriate protocols were largely considered detrimental, something that may not hold for later time periods (Brown et al., 2021). Notably, in the months right before the second wave of infection, a pre-mature victory over the pandemic was declared, which may have induced a varied and a false sense of protection from the virus and thereby confound the relationship of interest.¹⁰

We use multiple district-level datasets to obtain the set of explanatory variables (discussed in section 2.2.2). Data on the caste-group composition comes from the National Family Health Survey (NFHS-4) with appropriate sampling weights, which is mapped to daily district-level COVID-19 data.¹¹ The NFHS-4 survey conducted in 2015-16 collected information on 601,509 households across 640 districts at the time and is one of the most extensive and district-level representative household surveys in India.¹² Accordingly, relevant district representative measures for a range of socio-economic, demographic, amenities, health behavior and movement & migration-related variables are taken from this survey.¹³ Information on supply-side health indicators like the number of doctors and hospitals as a share of district-level population is taken from the latest Census round conducted in 2011. Average distances of the districts from state capitals and number of flights entering these

⁹ For the subsequent waves of infection, ascertaining the relationship of interest is further complicated by factors such as emergence of new strains that were not uniformly active across all Indian districts, events like the local body (Uttar Pradesh) and state elections that witnessed crowded political campaigns (Bihar, West Bengal, Tamil Nadu, Kerala, Puducherry and Assam), and the nationwide protests on the three farm laws whose intensity varied across different states, among others.

¹⁰ Please refer <https://timesofindia.indiatimes.com/education/news/pm-modi-cherishes-indias-dual-victory-over-covid-19-and-australia-praise-young-india/articleshow/80402121.cms> (accessed on March 21, 2023)

¹¹ We are able to map caste-group homogeneity and other controls for 590 of 684 districts for which COVID-19 data is available (86%).

¹² It is the Demographic and Health Survey (DHS) equivalent of India.

¹³ We did not use data from NFHS-5 (2019-2021) to create these district level measures as this survey was completed only for 22 states before the COVID-19 outbreak. For the remaining 14 states, data collection was halted in March 2020, which resumed only in October 2020 and was completed by May 2021. Thus, the estimated population level characteristics for districts from these states is likely to be endogeneous.

districts for the duration of analysis are taken from Google Earth Pro and Airport Authority of India, respectively.

To understand the potential pathways behind the relationship of interest, we first use the household level data from the NFHS-5 (2019-2021) survey. Here, we utilize information on household's involvement with community health workers from those districts where survey was conducted both, before and after the lockdown. In addition, to assess time allocation towards social interaction and unpaid voluntary community work, we use the Time Use Survey (TUS) conducted by the National Sample Survey Organization in 2019. This survey uses the International Classification of Activities for Time-Use Statistics, 2016, to provide information on the time duration of each activity performed by the sampled household members six years and above on the day before the survey. To track the mobility among residents during the lockdown and unocking period, we use the Data for Good project from Facebook. This provides daywise district level data on movement based on the facebook application and GPS location. More information on the data can be accessed from <https://dataforgood.facebook.com/dfg/about> (accessed on July 11, 2023).

2.2 Variables

2.2.1 Dependent Variables

We first check for the relationship between the number of days taken by different districts to cross the threshold of 50 cases and caste-group homogeneity. We also explore the same using the threshold of 100, 200, 300, 400 and 500 cases. Then, using high-frequency daily data, we study the relationship between caste-group homogeneity and the spread of COVID-19 cases that allows us to assess the changes in the effect temporally. Here, we use the number of daily cases as a share of the population for each of the 136 days, starting from March 25 till August 7, 2020, as the primary outcome of interest.¹⁴

2.2.2 Explanatory variables

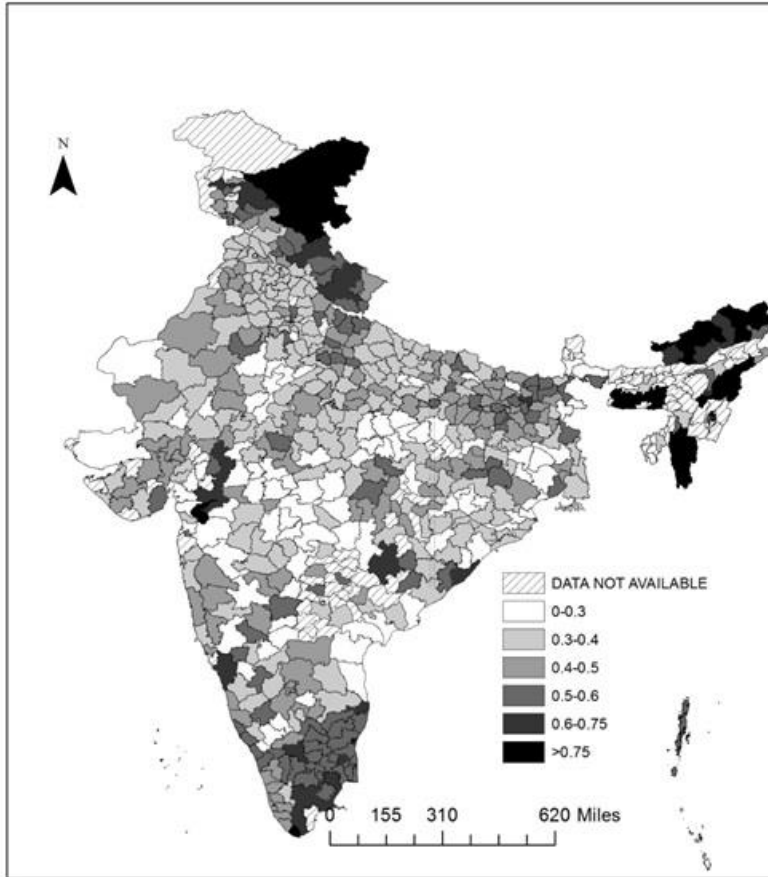
¹⁴ Appendix figure- F1 shows the evolution in the cumulative number of cases across the districts over the four lockdown phases.

Our main variable of interest is caste-group homogeneity, measured by the Herfindahl-Hirschman (HHI)/ fractionalization index at the district level. This index measures the probability that any two individuals chosen randomly from the residential unit belong to two different social groups. The HHI has previously been used extensively to measure social diversity or fragmentation in different contexts (Andreoni et al. 2016; Churchill et al. 2019). As mentioned previously, we use caste-groups that include the historically disadvantaged communities categorized under SCs, STs and OBCs alongside the privileged castes aggregated under “Others” to create this index.¹⁵ Individuals from the ST community in India belong to particular social groups that maintain their traditions, norms and belief system, which are bounded by the religious, social and blood ties (Majumdar, 1958). Individuals from the SC community belong to the traditional Hindu society and have historically been part of the social group disparagingly referred to as *untouchables*. The Constitution of India (Article numbers 341 and 342) has recognized the SCs and STs as socially deprived groups. Those from the OBC community are also socially and economically disadvantaged but relatively better off than the STs and SCs. Individuals from upper castes have historically fared better on social, economic or educational outcomes as against other caste-groups, on average (Maity, 2017). The HHI or fractionalization index is given by $\sum_{n=1}^4 s_n^2$ where s_n is the population share of n^{th} caste-group as described above. Figure-1 gives a spatial overview of the caste-group homogeneity across the Indian districts.

Figure-1.

Caste-group homogeneity across Indian districts

¹⁵ The social networks within the Indian caste system are splintered into various sub-caste or jatis. In this paper, we restrict ourselves to the broad caste groupings that is representative at the district level.



Note: ArcGIS software used to generate the figure. Caste-group homogeneity across Indian districts is measured using Herfindahl Hirschman Index. NFHS-4 dataset is used.

We use multiple control variables in the regression framework to account for confounding possibilities, as discussed in Section 2.3. As the emergence of initial hotspots is potentially endogenous, we also control for any cases reported in the district before the lockdown.¹⁶ This is critical to address the selection on unobservables as it subsumes the influence of superior international connectivity and integration of certain districts into the global supply chain, among other factors. These are likely to be associated with residential patterns that shape diversity measures and infection spread at the district level. To account for the dynamic supply-side administrative responses in the day-wise analysis, we control for the rolling number of COVID-19 deaths as a share of the population in the past seven days in the district. Notably, this also potentially captures time-varying district-level unobservables that might confound our estimates. Public health in India is within the ambit of

¹⁶ The dummy variable takes a value of 1 if any cases were reported in the district on or before 24th March 2020 and 0 otherwise. Out of 590 districts used for analysis, 104 (17.6%) reported a total of 536 cases in this period. In an alternate specification and robustness check, we replace this dummy with the cases as a share of population before the lockdown.

legislative and executive powers of the state and not the federal government. Here, each district is placed under the stewardship of a bureaucrat (district collector) that reports directly to the Chief Minister of a state. In such an administrative structure, the allocation of scarce administrative resources is likely to be sensitive to recent COVID-19 related casualties.

Additionally, we control for a range of socio-economic factors in the dataset at the district level that includes shares of households that are: Hindus- the major religion category in India, have at least one member with an education level above senior secondary level, live in concrete houses and owner-owned dwellings as well as the share of households where members are covered by health insurance. We also control for the average numbers of rooms available for sleeping per household. This indicates the ability of household members to isolate in case of infection. To account for average district-level economic well-being, we use a standardized wealth index and the Gini coefficient that measures wealth inequality. Notably, studies have looked into the effects of economic inequality on COVID-19 spread (Chakraborty and Mukherjee, 2023) and here we account for the inequality to assess the implications of caste-group homogeneity. To control for access to public service announcements, share of households where a woman reads newspapers, listens to radio or watches television regularly is considered. For demographic profile and age-specific vulnerabilities to infection, we control for the share of households in a district with any elderly members (≥ 60 years of age), the density of population as well as the number of households in the district. To proxy for hygiene and health habits, we use share of households where soap was found in the handwashing area and where some member smokes at least on a weekly basis.

Moreover, we account for access to a range of amenities, which is likely to influence the economic resilience to the COVID-19 shock and the ability of the households to follow COVID-appropriate behavior. This includes share of households with access to improved drinking water, piped water inside the dwelling, toilets inside the housing compound, and clean cooking fuel. Also, we account for share of households that have access to social assistance in the form of below poverty line (BPL) cards, unique identification (Aadhar) documents and registered bank accounts, which may be used to identify the vulnerable and provide direct assistance. To control for dependence on public health

facilities, we include share of households that usually go to public health centers when someone is unwell as a covariate. On the supply side, hospitals and doctors, as a share of the district-level population, are controlled for to account for physical and human health infrastructure. Although earlier studies in the pre-COVID-19 times have documented low levels of inter-district and inter-state migration in India (Munshi & Rosenzweig, 2009; Munshi, 2019), mobility and migration are likely to be key transmission channels. We control for short- and long-term migration from these districts to account for this possibility.¹⁷ Moreover, we also account for long-term residential patterns of households in the district and the average distances of these districts from their state capital. The total number of flights entering the districts during the analysis period is also controlled for. Finally, in all the regressions, we take intra-state, administrative region fixed effects to account for the time-invariant, region-level unobserved characteristics. A detailed description of all variables used in the analysis and their summary statistics are provided in Appendix Table-A1 and Appendix Table-A2, respectively.

2.3 Empirical Strategy

First, to study the relationship between days taken to record specific threshold of cases and caste-group homogeneity, we use the following regression specification:

$$NumDays_{dR} = \beta_0 + \beta_1 YDum_{dR}^0 + \beta_2 HISG_{dR} + \beta_3 C_{dR} + \pi_R + \epsilon_{dR} \quad (1)$$

Here, $NumDays_d$ is the number of days taken to record a specific number of cases starting from March 25, 2020, in district, d in region, R . The thresholds considered are 50, 100, 200, 300, 400 and 500 cases, respectively till March 21, 2021. So, if a district does not reach these thresholds by this date, it is not considered in our analysis. $YDum_{dR}^0$ is a dummy variable for any cases recorded before the nationwide lockdown that can account for initial hotspots. $HISG_{dR}$ represents standardized caste-group homogeneity measured through the HHI or fractionalization index. We use the standardized measure to ensure comparability across districts. The vector of time-invariant control variables at the

¹⁷ India witnessed reverse migration from bigger cities during the nationwide lockdown and unlocking period. Controlling for mobility measures before the crisis would control for the potential exposure from reverse migration.

district level, as discussed in Section 2.2.2, is captured by C_{dR} and ϵ_{dR} is the error term. Important to here is that the National Sample Survey Office (NSSO) classifies clusters of districts within a state under common NSSO regions (Government of India, 2001). These regions consist of contiguous districts with common geographical features, rural population density and cropping patterns, which may also influence the disease environment and spread of infection. Therefore, we account for the time-invariant region-level observed and unobserved characteristics through regional fixed effects given by π_R . Here, estimation is done using the standard Ordinary Least Squares (OLS) method with robust standard errors and β_2 represents the coefficient of interest. For the day-wise estimates, we use the specification as in Lee et al. (2021) which examines the evolving relationship of interest on a daily basis. The regression equation is given by:

$$Y_{dRt} = \beta_0 + \beta_1 YDum_{dR}^0 + \beta_2 D_{DRt(lag7day)} + \sum_{t=1}^{136} \sigma_t (HISG_{dR} * Day_t) + \beta_4 C_{dR} + Day_t + \pi_R + \epsilon_{dt} \quad (2)$$

Here, Y_{dRt} represents the total number of cases as a share of population (per 10,000) for district d from region, R on day t . $YDum_{dR}^0$ is as specified in equation 1 and accounts for initial hotspots before the lockdown.¹⁸ $D_{DRt(lag7day)}$ represents total deaths per 10,000 population reported in the seven days before day, t .¹⁹ As discussed previously, $HISG_{dR}$ represents standardized caste-group homogeneity measured through the HHI or fractionalization index. Day_t takes the value of 1 when the day is t and 0 otherwise. Note that t takes the value from 1 (for March 25, 2020, the first day of the lockdown) to 136 (for August 7, the last day of our analysis). σ_t is the vector of coefficients of interest. The vector of time-invariant control variables at the district level and the region fixed effects, as discussed previously, are captured by C_d and π_R respectively. Day_t is the vector of day fixed effects, which

¹⁸ For all our estimates, if we replace this dummy with the number of pre-lockdown Covid-19 cases in the district per 10,000 population the results remain statistically comparable.

¹⁹ In our robustness analysis, we changed this to a three-day lag and find similar results.

accounts for broader diffusion of COVID-19 transmission and trends in national administrative policies. ϵ_{dt} is the error term and standard errors are clustered two-way at day and district level.^{20 21}

Addition of region-level fixed effects, emergence of initial hotspots before lockdown, lagged time-varying death count, and other extensive covariates account for the potential confounders. Moreover, we carry out further checks to establish credibility of the inferences. To ensure that the unobserved heterogeneities do not confound our findings, we examine the potential variation in the estimates after accounting for unobservables through bias-adjusted treatment effect estimates that assumes selection on observables are proportional to selection on unobservables by a hypothetical parameter δ (Altonji et al. 2005; Oster, 2019). The method uses another hypothetical parameter (R_{max}^2), which comes from explained variation in the scenario where all potential unobserved variables are incorporated into the linear regression. While there can be various options to compute the bound on R_{max}^2 , Oster (2019) proposes $R_{max}^2 = 1.3 * R_0^2$ (R_0^2 is the R -squared value of the full model with observed control variables) using data from a comprehensive sample of randomized experiments.²² The literature on this method assumes the effect of observables to be at least as important as unobservables and thus considers a bound on δ that lies between $[-1,1]$ (Altonji et al. 2005, Rathore and Das 2021). The argument is as follows: if $|\delta|$ goes beyond 1, the variation in the outcome explained by unobservable(s) has to be higher than the combined influence of all the observables put together to push the effect size of the coefficient of the variable of interest to zero. With a comprehensive set of controls, this is likely to be improbable. Accordingly, using R_{max}^2 , we estimate the value of δ for which the coefficient of interest turns zero. In addition, we also conduct a range of other robustness checks that are discussed in section 3.3.

3. Results

²⁰ In STATA 16, *reghdfe* command is used with *noabsorb* option for this.

²¹ For robustness check, we add state-level day-wise testing data to the list of controls. See Results section for details. As testing figures are not consistently available for all states across the duration of the analysis, these are used for validation and not used in main estimation equation (2). The results remain valid if standard errors clustered two-way at day and state level.

²² According to Oster (2019), the bounding value of the cut off at 1.3 allows effects from at-least 90% of the random experiments to survive.

3.1 Days to cross specific infection threshold

First, we check if caste-group homogeneous districts, controlling for other confounding possibilities, took more days to cross specific infection thresholds. Here, we first present the regression estimates for the average number of days taken by a district to report 50 cases, as outlined in equation (1). The regression results are presented in Table-1 where we introduce different layers of controls in six phases. Model 1 only controls for any pre-lockdown cases and district-level socio-economic variables. Model 2 adds migration, residential pattern, and mobility controls to the specification. In Model 3, we include covariates for demographic and health behavior controls. Models 4 and 5 also include amenities and physical and human health infrastructure, respectively. Model 6 represents the full specification where administrative region-specific time-invariant characteristics are additionally controlled. The findings from these set of regression estimates (see Table-1) show that a district with one standard deviation increase in standardized caste group homogeneity took about 4.5 to 16 additional days to record a total of 50 COVID-19 cases. Notably, these results remain positive and statistically significant with staggered introduction of control variables. Thus, after accounting for multiple confounding influences, we find that caste-group homogeneous districts took a significantly greater number of days to record a total of 50 cases. In the context of the contagious nature of the virus, this would then suggest a distinct advantage in favor of caste-group homogeneous districts.

Table-1: Estimation for days taken to record 50 cases on caste-group homogeneity

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|------------------------|------------------------|-----------------------|-----------------------|-----------------------|----------------------|
| Standardized HHISG | 14.5719*** (2.2829) | 16.3619*** (2.2984) | 7.7133*** (2.3584) | 9.0350*** (2.4389) | 9.1502*** (2.5263) | 4.4665** (2.0536) |
| Controls for | | | | | | |
| -Dummy for any pre-lockdown cases | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Socio-economic variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Migration, Residential pattern and Mobility | × | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Demography, Health behaviour | × | × | ✓ | ✓ | ✓ | ✓ |
| -Amenities | × | × | × | ✓ | ✓ | ✓ |
| -Human and Physical Health Infrastructure | × | × | × | × | ✓ | ✓ |
| -Administrative Region | × | × | × | × | × | ✓ |
| Fixed Effects | | | | | | |

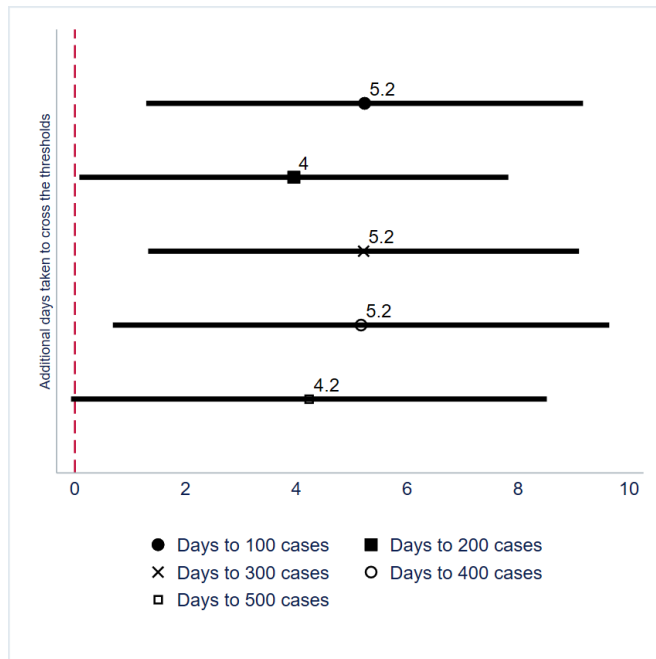
| | | | | | | |
|----------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Constant | 71.4973*** (1.5242) | 128.8278*** (26.4537) | 338.6619*** (32.7956) | 331.0897*** (36.7681) | 335.8885*** (37.5932) | 241.1084*** (39.0651) |
| Observations | 557 | 557 | 557 | 557 | 557 | 557 |
| R ² | 0.357 | 0.379 | 0.484 | 0.503 | 0.503 | 0.820 |

Notes: OLS estimates of the effects of 1 SD increase in caste homogeneity are presented along with robust standard errors in the parenthesis. Standardized HHI SG refers to the Z-score of Herfindahl Hirschman Index for caste-group homogeneity. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10.

We also verify this finding using a range of different thresholds of total COVID-19 cases. Using the same specification as in equation 1, we find a positive and statistically significant relationship for other thresholds ranging from 100 cases to 500 cases (see Figure-2). All these results point to initial infection concentration taking a relatively higher number of days in more caste-homogeneous regions. A natural follow-up question emerges on whether the overall spread of infection was slower in caste-group homogeneous districts, especially during the nationwide lockdown and whether these gained also sustained in the unlocking period.

Figure-2.

Estimation for days taken to record 100-500 cases on caste-group homogeneity



Notes: OLS estimates of the effects of 1 SD increase in caste homogeneity are presented along with 95% confidence intervals calculated from robust standard errors

3.2 Day-wise regressions

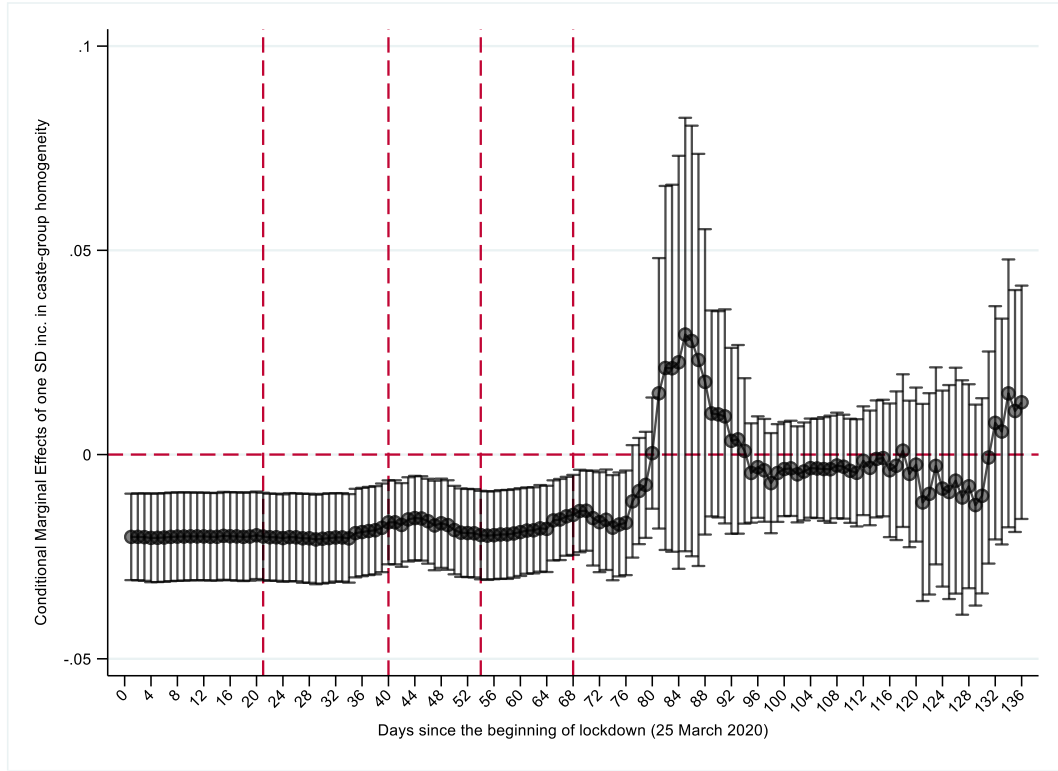
We now look at the relationship of interest on daily basis through daily regression results, as discussed in equation (2). Figure-3 presents the day-wise marginal effects. Here, in addition to controls outlined in the previous specification discussed above, we add lagged total deaths in the past seven days as a share of district-level population and daily fixed effects, which account for time-variant administrative measures and broad COVID-19 diffusion patterns, respectively.²³ The findings reveal that the relationship between caste-group homogeneity and COVID-19 cases per 10,000 population is negative and significant at 1% level for each day of the lockdown over the four phases, *ceteris-paribus*. In the initial days of the lockdown, the daily effect size was found to be stronger in comparison to those towards the end of the lockdown. On average, during the entire period, the effect ranges from -0.015 to -0.021. Back of the envelope calculations suggests that these effects translate to about 1,744 to 2,452 lower cases per day, on average, for one standard deviation increase in caste-group homogeneity.²⁴ The statistically significant relationship continues to hold for about eight days after unlocking, albeit with a weaker average effect size of -0.016, which translates to about 1,879 lower cases per day on average and turns insignificant thereafter. The results are statistically similar when the outcome variable (COVID-19 cases based on 7-day rolling averages) is replaced with 3-day rolling averages (Appendix Figure-F3).

Figure-3.

Relationship between caste-homogeneity and daily COVID-19 cases between 25 March and 7 August 2020

²³ Results with phase-wise introduction of controls are given in Appendix Figure-F2.

²⁴ This is based on 590 districts considered in the analysis assuming an average population of 2 million per district.



Note: Vertical lines denote end of each of the 4 phases of nationwide lockdown in 2020, dependent variable is reported COVID-19 cases per 10,000 population in the district. Marginal effects of caste-group homogeneity*Day are plotted (as in Eq. (2)). 95% confidence intervals calculated after double-clustering of the standard errors by districts (590) and days (136). All controls outlined in appendix table-A1 are used in the regression.

3.3 Robustness checks

3.3.1 State-wise COVID-19 testing

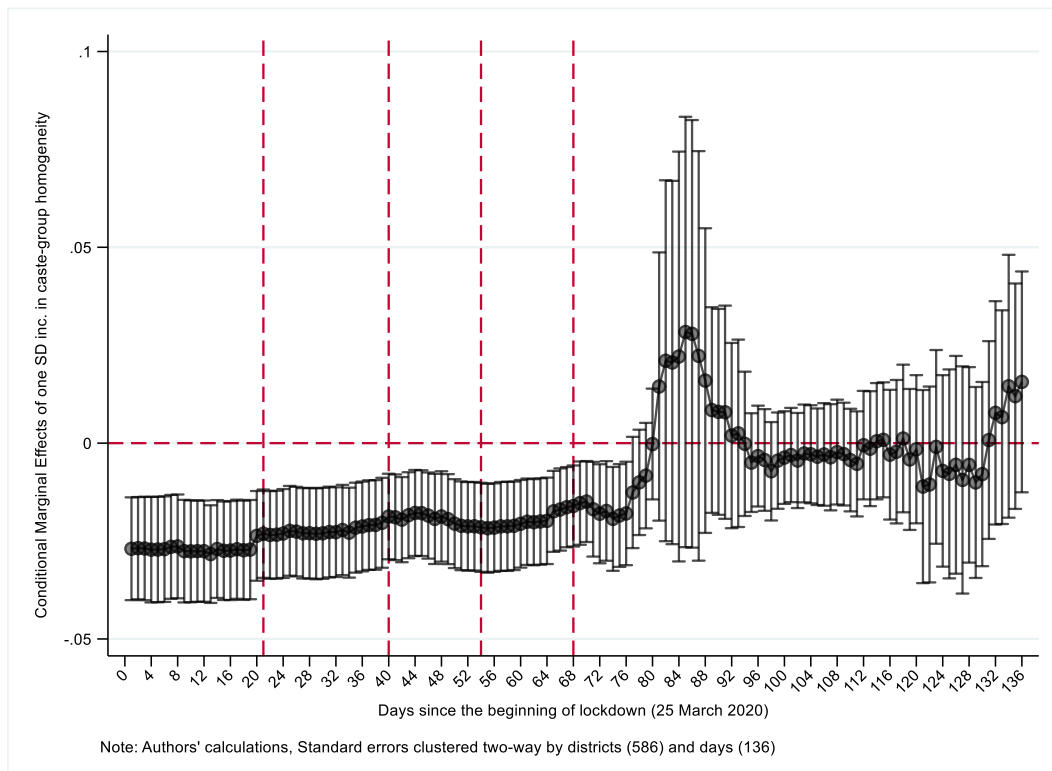
It is possible that regions having higher caste-group homogeneity may also have systematically lower testing, because of which the reported number of COVID-19 cases in these regions is low. In this case, our findings may be driven by fewer tests in these areas. To examine this possibility, we incorporate daily data on the number of COVID-19 tests conducted across the states during the lockdown and the unlocking periods. Despite missing data for some states in the early period of the lockdown, we are able to use testing data for 32 of 36 states and Union Territories (UT).²⁵ For missing data, we use a missing dummy that takes the value of 1 for every district whose testing data is not available for a particular day and 0 otherwise. Figure-4 presents the day-wise regression estimates

²⁵ Testing data is fully available for the duration of the analysis for 26 states and UTs and is partially available for six states and UTs.

after additionally controlling for the data on COVID-19 testing. The findings remain statistically valid and comparable in terms of effect size to what we observe in Figure-3.

Figure-4.

Relationship between caste-homogeneity and COVID-19 cases between after controlling for state-wise testing data between 25 March and 7 August 2020.



Note: Vertical lines denote end of each of the 4 phases of nationwide lockdown in 2020. Dependent variable is reported COVID-19 cases per 10,000 population in the district. Marginal effects of caste-group homogeneity*Day are plotted (as in Eq. (2)). 95% confidence intervals calculated after double-clustering of the standard errors by district and day. All controls outlined in appendix table-A1 are used in the regression.

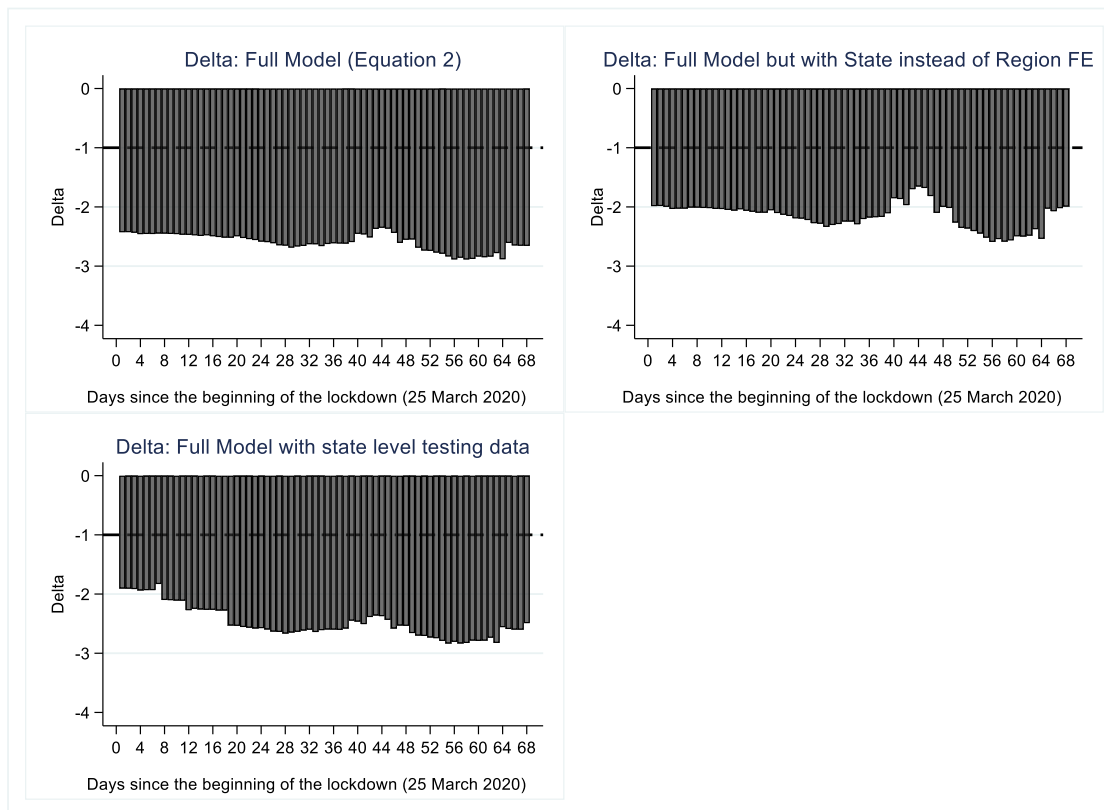
3.3.2 Testing for potential unobserved heterogeneities

To ensure our regression estimates are robust to unobserved heterogeneities emanating through omitted variable bias, we estimate the bias-adjusted treatment effects that use the “selection on observables versus unobservable” as explained earlier. The R-square value obtained from the regression with day interaction (our final specification, as mentioned in equation (2)) is 0.639, which implies that the R^2_{\max} will be about 0.830. With this information, we plot the δ obtained for each of the 68 days of the lockdown period (Figure-5) for specification in equation (2). As one can observe, the absolute values of δ range between 2.3 to 2.9 during the lockdown period, which is well above the

cut-off value of 1. This indicates that the influence of unobservables will have to be 2.3-2.9 times that from the observables to ensure that the null hypothesis is not rejected. Given the wide range of controls used in the estimation, these exceptionally high values of δ offer credibility to the validity of our estimates. Additionally, we incorporate testing data as done in Figure-4 and re-estimate the δ s. The absolute values continue to be well over 1 for each of the 68 days of the lockdown period (1.8 to 2.8).

Figure-5.

Accounting for potential OVB (Plotted values of δ)



Note: “psacalc” command in STATA is used to generate the values of δ . Absolute value of $\delta = 1$ implies that combined effects of unobservables will have to be as large as the effect of all the controls used in the regression put together negate our findings. Even higher absolute values of δ lend further credibility to our estimates as combined influence of unobservables would have to be even larger to invalidate our results.

3.3.3 Shannon’s entropy measure

We additionally consider Shannon’s entropy measure (SEM) of diversity for district, d instead of $HISG_d$. Mathematically, this can be presented as $SEM_d = \sum_{i=1}^k P_i \log_2(1/P_i)$ where P_i is the share of each caste-group in the overall population, which can be divided into k mutually exclusive groups. Unlike the $HISG_d$ that ranges between 0 and 1, with a higher value implying higher homogeneity, SEM_d ranges from 0 to infinity, with higher values representing greater diversity (Frenken et al. 2007). Thus, although $HISG_d$ and SEM_d are inversely related, the coefficients are not directly comparable. The result indicates that the relationship of interest is robust to this measure of diversity as well (Appendix Figure-F4). Note that in some of the districts, a few of the caste-groups may not be present, which would make their share in the population zero. As a result, districts for which SEM_d cannot be computed for the 40 districts, which drop out of the analysis. To validate that this selection is not driving our findings, we re-estimate the model using $HISG_d$ for the sub-set of 550 districts for which the SEM_d can be computed. We find that the relationship of interest remains robust (Appendix Figure-F5).

3.3.4 Regression with non-capital districts

State capitals are usually clusters of higher economic activity. As a result, they may witness larger inflow of migrants from the entire nation, which can also result in higher caste-group diversity in these areas. As a result of higher economic activity and domestic and international movements, these are likely to be the major hotspots of infection spread. Thus, these regions could potentially drive our results. Though we adequately controlled for these characteristics in our specifications by accounting for average distance of the district from the state capital, we now remove these districts (39 out of 590) from our analysis and re-estimate the day-wise regression results for sensitivity check, which is adequately satisfied (Appendix Figure-F6).

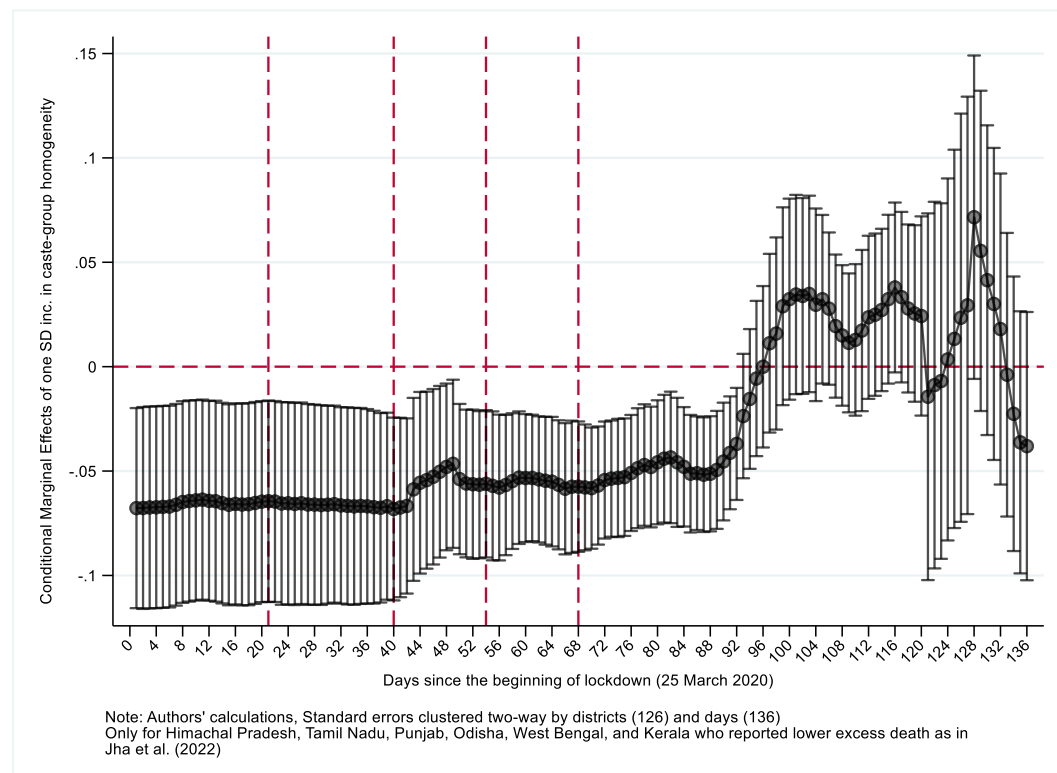
3.3.5. Accounting for underreporting of cases

Emerging evidence from the second wave of infection in India suggests that COVID-19 related deaths in this period (April to July 2021) were systematically underreported (Jha et al., 2022). If such underreporting confounds with caste-group homogeneity, this will introduce bias into our estimates.

Findings of Jha et al. (2022) suggest that on average, predominant share of these underreported deaths (84%) occurred between April and July of 2021, a period outside our study duration. This suggests that underreporting is less of a concern for our estimates. However, for robustness check, we undertake the same analysis using only the states where Jha et al. (2022) estimated excess (unaccounted) death to be comparatively lower.²⁶ Our results are statistically robust even if we restrict our analysis to these select states (Figure-6).

Figure-6.

Estimates for states where underreporting of deaths was comparatively lower



Note: The regression sample includes only the states of Himachal Pradesh, Tamil Nadu, Punjab, Odisha, West Bengal and Kerala, where excess death ratio is $\leq 100\%$. Vertical lines denote end of each of the 4 phases of nationwide lockdown in 2020. Dependent variable is reported COVID-19 cases per 10,000 population in the district. Marginal effects of caste-group homogeneity*Day are plotted (as in Eq. (2)). 95% confidence intervals calculated after double-clustering of the standard errors by district and day. All controls outlined in appendix table A1 are used in the regression.

3.4 Exploring Potential Pathways

²⁶ This includes states where estimates of excess death was less than 100% (Himachal Pradesh, Tamil Nadu, Punjab, Odisha, West Bengal and Kerala).

As indicated earlier, various population characteristics associated with caste-homogeneous districts could have individually or jointly contributed to the lower spread of COVID-19 cases. We now provide suggestive evidence of pathways that may have contributed to slower infection spread in caste-group homogeneous areas.

3.4.1 Engagement with local health workers

The first channel studies the potential involvement of local health workers in curbing the spread of the infection. In India, the role of frontline health workers has been paramount in running successful vaccination campaigns and eradicating life-threatening diseases like Polio (Deutsch et al. 2017). During the pandemic, ground reports documented that community health care workers played a major role in spreading awareness, communicating health protocols and facilitating treatment for COVID-19 patients.²⁷

Simultaneously, there were media reports documenting frontline workers on COVID-19 duty being attacked in some areas, which may have affected their level of motivation and engagement during the pandemic. To check if such instances varied across district-level caste-group homogeneity, we conducted preliminary text analysis and identified multiple news reports from 16 districts and found that 13 of these had median or lower level of caste-group homogeneity (Table-2).²⁸ This would then suggest that after controlling for human health infrastructure, engagement with and trust on local health workers could have been systematically lower in more caste-group diverse areas and may have mediated differential spread of COVID-19 cases across Indian districts.

Table-2. Summary of text analysis on violence against frontline workers across Indian districts

| Sr No | Date | District mentioned | Phrase | Decile of zHHI (Data) |
|-------|------------|------------------------|--|-----------------------|
| 1 | 03-04-2020 | Indore | <i>"Videos of health care workers being pelted with stones as they are chased away from a locality in India went viral on Thursday, taking the country by storm and garnering global attention. The incident occurred in</i> | 2 |

²⁷ Please refer <https://www.gavi.org/vaccineswork/iwd2021/international-womens-day-caring-everyone-asha-workers-covid-19-story> (accessed on March 24, 2022)

²⁸ This was undertaken by using a combination of keyword search using 'India', 'Covid' and 'violence' with 'frontline workers' or 'health workers' on google news.

| | | | | |
|---|-------------|---|--|------------------|
| | | | <i>Indore, a city in India's central Madhya Pradesh state, and the area was Tatpatti Bakhal, a hotspot of coronavirus cases in the city."</i> | |
| 2 | Report 2021 | North East Delhi, South Delhi, Bhopal | <i>"Several ASHA workers were attacked in South and North-east Delhi; Hajira Baji, an Anganwadi worker Bhopal district Madhya Pradesh, noted that media disinformation linking the spread of Coronavirus with Muslim community after the Tablighi Jamaat incident adversely impacted the Anganwadi workers who were Muslims."</i> | 3,3,3 |
| 3 | 12-06-2020 | Bengaluru (urban), Gulbarga (Kalaburagi), Bagalkot, Chikaballapur and Belgaum (Belagavi); Kamrup district | <i>"ASHAs across Bengaluru (urban), Gulbarga, Bagalkot, Chikaballapur and Belgaum districts of Karnataka were attacked by community members while on COVID-19 duty;</i> <i>"We are called 'Corona walas' (Corona carriers) and kept at a distance by our neighbours because we are in touch with community members without proper protective gear", said Minara Begum, an ASHA from Kamrup district in Assam"</i> | 4, 9, 9, 2, 2, 3 |
| 4 | 08-06-2020 | Gorakhpur | <i>"An Asha worker in Gorakhpur was allegedly abused and assaulted when she had gone to a village to collect details of local residents for COVID-19 vaccination."</i> | 5 |
| 5 | 22-06-2021 | Thane | <i>"When Sudarshana Satvi, a 40-year-old Accredited Social Health Activist (ASHA) worker, along with community health officer Prakash Kawade and his team, arrived at Ganje Dhekale, a remote village in Palghar Taluka, in the Thane district of Maharashtra, on June 1, they were met with extreme hostility."</i> | 4 |
| 6 | 22-04-2020 | Nagpur | <i>"Health workers visiting some coronavirus containment zones in Maharashtra's Nagpur have complained that they were spat upon by some people, pelted with stones and abused at times while conducting surveys."</i> | 1 |
| 7 | 11-05-2020 | Faridabad, Gurugram | <i>"Gurugram On April 23, Rekha Sharma, 39, an ASHA (Accredited Social Health Activists) worker in Ballabgarh, had gone to a society in Sector 65, Faridabad, to conduct a door-to-door survey for coronavirus disease (COVID-19) outbreak. As she was questioning the members of a family about their travel history, they started insulting her. Within minutes, a crowd gathered around her and started abusing her."</i> | 4,5 |
| 8 | 31-05-2021 | Mahoba | <i>"Another ASHA worker working in a Mahoba village told Newsclick over the phone that people in her village had been threatening her if she carried out testing or vaccination there"</i> | 7 |

Note: hyperlinks for specific articles provided above.

This can happen because of multiple reasons. First, as these workers are locally appointed, we expect that in a relatively homogeneous neighborhood with greater interpersonal trust, involvement of community workers and confidence in them would also be higher. Second, when these workers operate in homogeneous areas with similar beliefs, preferences and levels of risk aversion, the marginal effort needed for effective engagement is likely to be lower. Finally, because of potentially greater collective action in homogenous communities, collective demand for health intervention might be higher. Given their historical role in promoting health outcomes, it is possible that caste-group homogeneous regions, which observed stronger frontline worker engagement experienced slower growth of infection during the early period of the outbreak.

To assess this aspect, we utilize the NFHS-5 survey, which collected data from 636,699 households across 707 districts in 2019-21. The survey asks every woman between age 15 to 49 years in these sampled households on whether they have visited an *angandwadi*, Accredited Social Health Activists (ASHA) or other CHWs in the last three months.²⁹ We code these women who have visited them at least once as 1 and 0 otherwise and run a probit regression on district-level caste homogeneity. In addition, instead of district-level caste-group homogeneity, we run the same regression with the homogeneity measure defined at the Primary Sampling Unit (PSU). The findings in columns 3 and 4 of table-3 indicate a significantly positive association between engagement between local health workers and individuals in caste-group homogenous areas. We repeat the same exercise using NFHS-4 survey data and found similar results (columns 1 and 2 in Table-3), indicating that the above relationship is statistically robust.

Importantly, NFHS-5 also allows us to measure health worker involvement during the pandemic and explore whether this engagement has disproportionately been higher in areas with high caste-group homogeneity. For this, we utilize a crucial characteristic of NFHS-5 survey, which is as follows: In 57 districts from 14 states, some part of survey was conducted between December to

²⁹ *Anganwadi* (hindi) translates to “courtyard shelter”. It is childcare center that provide health information, pre-school education, supplementary nutrition and health check-up facilities. ASHA workers are local trained female health activist. For more information, refer <https://www.nhm.gov.in/images/pdf/communitisation/task-group-reports/guidelines-on-asha.pdf> (accessed on December 18, 2021)

March 2020. The survey was then halted because of the outbreak of the COVID-19 pandemic. After the first wave, the remaining survey resumed in January 2021 and was completed by May 2021. This allows us to make a direct comparison of engagement with community health workers between the sampled households within these 57 districts surveyed before the COVID-19 outbreak with those surveyed post the first wave. By construction, using data only from these 57 districts would account for the time-invariant district-level confounders.

More formally, we use the following regression to estimate the disproportionate changes in engagement with community health workers or ASHAs among those from PSUs with higher caste group homogeneity:

$$Y_{ip} = \alpha + \beta.(Post\ COVIDwave_{ip} * HISG_p) + \gamma Post\ COVID_{ip} + \delta HISG_p + \varepsilon_{ip} \quad (3)$$

Here, Y_{ip} is the outcome variable, which takes the value of 1 if woman, i between age 15 to 49 years from PSU, p has visited an *angandwadi*, ASHA or other community health worker in the last three months prior to the survey and 0 otherwise. $Post\ COVIDwave_{ip}$ is 1 if the survey data of the household falls in between January 2021 to May 2021 (post first wave) and 0 if it falls in between December 2019 to March 2020 (before outbreak). $HISG_p$ is the caste group homogeneity defined at the PSU level. ε_{ip} is the error term. We are interested in estimating β , which is the coefficient associated with the interaction of $Post\ COVIDwave_{ip}$ and $HISG_p$.

Table-3 presents the estimation results from above regression (column 5 and 6). We find that the engagement with local health workers has been significantly higher after the start of the pandemic among those from PSUs with higher caste-group homogeneity. Importantly, the results remain similar if we use district-level caste-group homogeneity measures. This provides suggestive evidence of slower COVID-19 spread being mediated via greater engagement of local health workers in caste homogenous areas, which also has been disproportionately higher after the COVID-19 outbreak.

3.4.2 Higher coordinated action in homogeneous communities

We also posit that greater coordinated action because of higher cohesion in socially homogenous areas could have led to a slower spread of the virus. If so, socialization and involvement in community-level activities among the residents of caste-group homogeneous regions should also be higher in the pre-pandemic time-period. To examine this, we utilize data from the Time-Use Survey (TUS) conducted by the National Sample Survey Organization of India in 2019.³⁰ The survey collects information on time allocated to different activities by the members of the sampled households on the day prior to the survey. From the survey, we use two relevant measures of involvement in community-level activities: (i) socializing (socializing and communication, community participation and religious practice) and (ii) unpaid volunteering work (unpaid volunteer, trainee and other unpaid work). As in the earlier case, we consider caste-group homogeneity defined at the PSU and district level: For individuals aged 18-50 years, we run similar regressions of time allocated to socializing and unpaid volunteering work on caste-group homogeneity. We find a positive and significant relationship indicating that individuals from caste-group homogeneous regions, at the local or district level, spent more time socializing and being involved in community-level activities.

Table 3: Potential pathways

| | All districts | | | 57 districts pre-post COVID-19 | | All districts | |
|---|------------------|---------|------------------|--------------------------------|------------|---------------|----------|
| | NFHS 4 (2015-16) | | NFHS 5 (2019-21) | | TUS (2019) | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Standardized HHI (PSU level) | 0.005*** | | 0.004*** | | -0.006** | | 0.003*** |
| | (0.001) | | (0.001) | | (0.003) | | (0.001) |
| Standardized HHI (district level) | | 0.001 | | 0.02** | | 0.002 | 0.005*** |
| | | (0.001) | | (0.001) | | (0.003) | (0.001) |
| Post COVID-19 first wave | | | | | -0.008 | -0.014*** | |
| | | | | | (0.005) | (0.009) | |
| Standardized HHI (PSU level)* Post COVID-19 first wave | | | | | -0.009* | | |
| | | | | | (0.005) | | |
| Standardized HHI (district level)* Post COVID-19 first wave | | | | | | 0.017*** | |
| | | | | | | (0.005) | |

³⁰ The survey was administered to 138,805 households having 518,751 individuals. For more information on TUS, refer <https://mospi.gov.in/web/mospi/time-use-survey> (accessed on December 17, 2021)

| | | | | | | | | |
|--------------|---------|---------|---------|---------|--------|--------|---------|---------|
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 667,074 | 667,074 | 343,449 | 343,449 | 38,932 | 38,932 | 278,332 | 278,332 |

Note: All regressions include controls for age, marital status, education, caste dummies, rural/ urban dummy, wealth quantiles/ log of monthly per capita expenditure and state fixed effects. The marginal effects from OLS regressions are presented along with robust standard errors in the parenthesis. For the estimates presented in (5) and (6), state fixed effects are not used. These estimates are given for 57 districts where the survey was conducted before and after the first wave of COVID-19. To generate the estimates in column (1) and column (2), NFHS-4 2015-16 dataset is used, for columns (3), (4), (5) and (6), data from 57 districts from the NFHS-5 2019-21 survey is used and to generate estimates given in column (7) and column (8), TUS 2019 is used. . *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10.

During the lockdown, it is possible that the gains in social capital and community involvement in caste homogeneous areas was instrumental in slowing the spread of the infection. This requires higher compliance with the COVID-19 norms and protocols, for example lesser movement. To assess this, we use data from the Data for Good project by Facebook. This data gathers information of the mobility of people with smartphones having the Facebook application and its GPS turned on. Despite limitations, in absence of representative compliance data, studies have used it to understand movements patterns after the outbreak of COVID (Chadha and Raghu, 2021; Chakraborty and Mukherjee, 2023). The dataset offers district level daywise information on the two key variables concerning mobility: “*positive proportion of users staying put within a single location*” and “*positive or negative change in movement relative to baseline*”. Here the baseline time period considered is that particular date in February 2020, when mobility was unaffected by the pandemic. We use these two variables as the outcomes and regress it on district level caste group homogeneity index along with the other covariates for the time period pertaining to the lockdown. Appendix table- A3, which presents the regression estimates, indicate that the proportion of people to stay put within single location is significantly higher if they are from regions with higher caste homogeneity. We also find that the change in movement relative to the baseline has fallen in these areas. This demonstrates the potential role played by caste-group homogeneity in improving private compliance efforts during the early lockdown period.

Next, to understand if the reduced movement had a mitigating effect on the spread of infection in the homogenous areas, we use the regression outlined in equation 1 and introduce district level mean proportion of users staying put within a single location in the first phase of the lockdown (March 25 to April 15, 2020). If the fall in mobility had a mediating role in arresting the infection

spread, the secular effect of $HISG_d$ on the number of days to reach 50 cases should reduce once the variable on mobility is introduced in the regression specification. The regression estimates presented in table 4 indicates a fall in the effect size with the mobility variable added as an additional covariate when compared with that without introduction of this variable. When the same regressions are repeated for number of days to reach 100 and 200 cases, we can observe a similar decline. This underscores the positive role of caste-group homogeneity, in reducing mobility which potentially facilitated slower infection spread in these areas. Please note that we do not use daywise regression here because daily cases may not be mapped one-on-one with daily movement because of a uncertain incubation period. This period was known to range from 2 to 14 days and in some cases even more than 14 days.³¹

Table 4: Mediating effect of mobility in the first phase of the lockdown

| | Days to reach 50 cases | | Days to reach 100 cases | | Days to reach 200 cases | |
|--|------------------------|------------------------|-------------------------|------------------------|-------------------------|------------------------|
| | Without mobility | With mobility | Without mobility | With mobility | Without mobility | With mobility |
| Standardized HHISG | 4.466** (2.054) | 3.935* (2.032) | 5.228*** (2.005) | 4.789** (2.003) | 3.951** (1.969) | 3.931** (1.959) |
| Controls for | | | | | | |
| -Dummy for any pre-lockdown cases | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Socio-economic variables | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Migration, Residential pattern and Mobility | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| - Demography, Health behaviour | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| - Amenities | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Human and Physical Health Infrastructure | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| -Administrative Region | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Fixed Effects | | | | | | |
| Mobility in the first 15 days | × | ✓ | × | ✓ | × | ✓ |
| Constant | 241.108*** (39.065) | 270.423*** (44.733) | 254.634*** (41.773) | 290.717*** (48.180) | 291.563*** (39.700) | 343.028*** (44.292) |
| Observations | 557 | 548 | 555 | 546 | 548 | 540 |
| R^2 | 0.820 | 0.818 | 0.783 | 0.778 | 0.793 | 0.791 |

Notes: OLS estimates of the effects of 1 SD increase in caste homogeneity are presented along with robust standard errors in the parenthesis. Standardized HHI SG refers to the Z-score of Herfindahl Hirschman Index for caste-group homogeneity. Mobility in the first 15 days indicates district level mean proportion of users staying put within a single location in the first phase of the lockdown (March 25 to April 15, 2020). *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10.

³¹ Refer to <https://www.webmd.com/covid/coronavirus-incubation-period> for discussion on incubation (accessed on July 14, 2023)

4. Discussion and Policy Implications

The main finding of this paper is that caste-group homogeneity predicts the spread of COVID-19 cases across Indian districts during the nationwide lockdown. Moreover, these gains last for more than two months and up to a week after the unlocking procedures were initiated. We conduct multiple sensitivity checks to account for various observed and unobserved confounders and find our estimates to be robust. Although the data used does not allow us to gauge the exact underlying mechanism for this relationship, the results are suggestive of the importance of strong community cohesion or “social capital” that may work through caste-group networks and facilitate socially beneficial compliance behavior. In addition, the disproportionately higher involvement of health workers in caste-group homogenous areas after the first wave of the outbreak may have also played a role in curbing the COVID-19 transmission. Unavailability of disaggregated data, particularly community action restricts our ability to assess these underlying mechanisms precisely.

Our analysis covers a period when vaccines were not yet on the horizon and the nature of the pandemic was new to the majority of the population. Hence, compliance with COVID-19 protocols was critical to curtail the spread of infection. Such compliance was likely to be deeply intertwined with community-level cohesion. We thus postulate that social cohesion and reciprocity were critical channels through which these effects were mediated. Importantly, these gains weakened towards the end of the lockdown and became statistically insignificant a week after unlocking began is something that further supports this argument. A plausible reason could be the economic hardship and fatigue imposed by lockdown measures that have been extensively documented in the Indian context (Kesar et.al.,2021; Ray and Subramanian, 2022), which may have increased the cost of sustaining community coordination. This can then reduce socially desired levels of compliance with time. Moreover, easing of lockdown restrictions increased mobility, especially from outside areas, and would have made enforcement and imposing social sanctions more challenging. However, these are conjectures, and a detailed analysis of the underlying pathways remains a question for future research.

Our results have significant policy prescriptions to offer. First, at the onset of such public health crisis, the government should pay extra attention to more vulnerable and socially diverse areas. As a short-term crisis management response, scarce administrative resources (like law enforcement and surveillance) can be prioritized for diverse blocks to compensate for weaker community cohesion, which makes these areas more vulnerable and poses larger spillover risks for the rest of the community. Strengthening the network of community health workers, prioritizing their well-being and incentivizing higher engagement with residents in these diverse areas are some vital interventions that can serve as key policy instruments.

Regarding the administrative response to a sudden public health emergency, identifying areas with weaker community cohesion and allocating scarce resources to these regions is likely to provide a vital window to mount an appropriate response. In this context, local self-governments have a paramount role to play. These local bodies at the rural and urban areas have long been established in countries like India, and they can support decentralized administration under the tutelage of the state and the union governments. These local bodies can help identify the pockets with higher community-level diversity, earmark potential hotspots and take early precautions to avoid more significant outbreaks.

Second, in relatively homogenous areas, as the effect of existing social ties weakens with longer lockdowns and opening up of the economy, strategies to further strengthen community networks might be undertaken. Research indicates that a positive perception about the community is found to enhance private compliance behavior concerning the pandemic (Bicchieri et al. 2021; Das et al. 2020). One possible way to facilitate this could be through decentralized health interventions through community-level representatives that promote greater participation. Here, targeted intervention and educational messages catering to the relevant community through local-level workers may prove to be an important mitigation measure. If implemented, the effectiveness of these policies would be an important area of future research.

Third, we observe that the engagement of CHWs with the households was disproportionately higher in caste-group homogeneous communities after the outbreak, thereby underscoring their key role in limiting the spread of the infection in these areas. This emphasizes the importance of decentralized health interventions suited to local contexts in formulating effective health policies. To this end, strengthening of institutional capacity primarily through the CWHs assumes significance.

The gains from caste homogeneity through lower mobility during the lockdown period also have important policy implications. First, effective interventions need to recognize the unique characteristics of socially diverse regions, and tailor public health measures accordingly. For example, these could be done by broadcasting community tailored public health campaigns, deploying community health workers who are sympathetic towards multi-caste cultural norms, and working to enhance community's acceptance to CHWs from diverse backgrounds..

5. Conclusion

In a pandemic setting, curtailing the growth of infection needs coordinated community action that can be mediated via social cohesion and strong community ties. However, these are constrained by social diversity that places obstacles in realizing such trust and cooperation. Using daily data on COVID-19 infection at the district level from India, we find that caste-group homogeneity played a significant role in predicting lower growth in cases, especially during the nationwide lockdown imposed from March 25 to May 31, 2020. However, these marginal gains appear to dampen over time and evaporate a week after the initiation of the unlocking process. The results are statistically consistent to a range of robustness and internal validity tests. We find suggestive evidence of engagement with CHWs being more effective in caste-group homogeneous areas. We also observe lower mobility in caste homogenous areas during the lockdown. This is found to partially mediate the spread of the infection through caste-group homogeneity, thereby underscoring the potential implications of social cohesion in facilitating private compliance efforts. These findings offer possible policy levers to the government and public health officials to identify and prioritize allocation of scarce resources to socially diverse regions that are unable to effectively initiate and sustain coordinated community

action. We stress on community-strengthening efforts and argue for decentralized public health interventions during pandemics as these are better suited to address differentiated local requirements instead of a ‘one size fits all’ approach.

Declarations

Funding: Authors declare that did not receive any funding for carrying out this study.

Conflicts of interest/Competing interests: The authors declare that they have no conflict of interest.

Availability of data and material: Development Data Lab’s (DDL) COVID India database used for the study is publicly available and freely downloadable from <http://www.devdatalab.org/shrug>. Data from Good project from Facebook can be accessed freely from <https://dataforgood.facebook.com/dfg/about>. Data from Census is also publicly available and freely downloadable. The Time Use Survey dataset is freely downloadable from Ministry of Statistics and Programme Implementation, Government of India. The fourth wave of National Family and Health Survey used for the study is available on request with permission from the Demographic Health Survey.

References

- Akay, A., Constant, A., Giulletti, C., Guzi, M. (2017). Ethnic diversity and well-being. *Journal of Population Economics*, 30, 265-306.
- Alesina, A., Baqir, R., Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly journal of Economics*, 114, 1243-1284.
- Alesina, A., Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43, 762-800.
- Altonji, J. G., Elder, T. E., Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113, 151-184.
- Andreoni, J., Payne, A. A., Smith, J., Karp, D. (2016). Diversity and donations: The effect of religious and ethnic diversity on charitable giving. *Journal of Economic Behavior & Organization*, 128, 47-58.
- Chadha N, Raghu S (2021) Measuring migration-related movements from daily mobility data (No. 2021/1), India Development Foundation Working Paper
- Chakraborty, T., Mukherjee, A. (2023). Economic geography of contagion: a study of COVID-19 outbreak in India. *Journal of Population Economics*, 1-33.
- Banerjee, A., Iyer, L., Somanathan, R. (2005). History, social divisions, and public goods in rural India. *Journal of the European Economic Association*, 3, 639-647.

- Bicchieri, C., Fatas, E., Aldama, A., Casas, A., Deshpande, I., Lauro, M., ... Wen, R. (2021). In science we (should) trust: Expectations and compliance across nine countries during the COVID-19 pandemic. *PloS one*, 16, e0252892.
- Bosancianu, C. M., Dionne, K. Y., Hilbig, H., Humphreys, M., Sampada, K. C., Lieber, N., Scacco, A. L. (2020). Political and social correlates of covid-19 mortality. 16 June. Accessed November 10, (20. doi:10.31235/osf.io/ub3zd.
- Brown, R., Coventry, L., & Pepper, G. (2021). COVID-19: the relationship between perceptions of risk and behaviours during lockdown. *Journal of Public Health*, 1-11.
- Churchill, S.A., Farrell, L., Smyth, R. (2019). Neighbourhood ethnic diversity and mental health in Australia. *Health Economics*, 28, 1075-1087.
- Das, U., Sarkhel, P., Ashraf, S. (2020). Love Thy Neighbour? Perceived Community Abidance and Private Compliance to COVID-19 Norms in India. *South Asia Economic Journal*, <https://doi.org/10.1177/13915614211053928>
- Desai, S., Dubey, A. (2012). Caste in 21st century India: Competing narratives. *Economic and Political Weekly*, 46, 40.
- Deutsch, N., Singh, P., Singh, V., Curtis, R., Siddique, A. R. (2017). Legacy of polio—use of india’s social mobilization network for strengthening of the universal immunization program in india. *The Journal of Infectious Diseases*, 216, S260-S266.
- Egorov, G., Enikolopov, R., Makarin, A., Petrova, M. (2021). Divided we stay home: Social distancing and ethnic diversity. *Journal of Public Economics*, 194, 104328.
- Frenken, K., Hanusch, H., Pyka, A. (2007). "The Elgar companion to Neo-Schumpeterian economics." In *Entropy and Information Theory*, by H. Horst and A. Pyke, 544–555. Cheltenham: Edward Elgar.
- Gadsden, T., Maharani, A., Sujarwoto, S., Kusumo, B. E., Jan, S., & Palagyi, A. (2022). Does social capital influence community health worker knowledge, attitude and practices towards COVID-19? Findings from a cross-sectional study in Malang district, Indonesia. *SSM-Population Health*, 19, 101141
- Gisselquist, R. M., Leiderer, S., Nino-Zarazua, M. (2016). Ethnic heterogeneity and public goods provision in Zambia: Evidence of a subnational “diversity dividend”. *World Development*, 78, 308-323.
- Gisselquist, R., & Kundu, A. (2020). Horizontal inequality, COVID-19, and lockdown readiness: Evidence from India.
- Goette, L., D. Huffman, and S.Meir (2006): “The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups,” *American Economic Review*, 96, 212–216.
- Government of India. (2001). Concepts and Definitions Used in NSS. Golden Jubilee Publication. URL: http://www.mospi.nic.in/sites/default/files/publication_reports/concepts_golden.pdf

- Habyarimana, J., Humphreys, M., Posner, D. N., Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review*, 101, 709-725.
- Jha, P., Deshmukh, Y., Tumbe, C., Suraweera, W., Bhowmick, A., Sharma, S., Brown, P. (2022). COVID mortality in India: National survey data and health facility deaths. *Science*, eabm5154.
- Kesar, S., Abraham, R., Lahoti, R., Nath, P., & Basole, A. (2021). Pandemic, informality, and vulnerability: Impact of COVID-19 on livelihoods in India. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, 42(1-2), 145-164.
- Larson, J. M., Lewis, J. I. (2017). Ethnic networks. *American Journal of Political Science*, 61(2), 350-364
- Lee, J. N., Mahmud, M., Morduch, J., Ravindran, S., Shonchoy, A. S. (2021). Migration, externalities, and the diffusion of COVID-19 in South Asia. *Journal of Public Economics*, 193, 104312.
- Maity, B. (2017). Comparing health outcomes across Scheduled Tribes and Castes in India. *World Development*, 96, 163-181.
- Majumdar, D. N. (1958). *Caste and communication in an Indian village*. Bombay: Asia Publishing House.
- Miguel, E., Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, 89, 2325-2368.
- Munshi, K., Rosenzweig, M. (2009). Why is mobility in India so low? Social insurance, inequality, and growth (No. w14850). National Bureau of Economic Research.
- Munshi, K. (2019). Caste and the Indian economy. *Journal of Economic Literature*, 57, 781-834.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37, 187-204.
- Rathore, U., Das, U. (2021). Health Consequences of Patriarchal Kinship System for the Elderly: Evidence from India. *The Journal of Development Studies*, 1-19.
- Ray, D., Subramanian, S. (2022). India's Lockdown: An Interim Report. In: Dutta, M., Husain, Z., Sinha, A.K. (eds) *The Impact of COVID-19 on India and the Global Order*. Springer, Singapore.
- Salve S, Raven J, Das P, Srinivasan S, Khaled A, et al. (2023). Community health workers and Covid-19: Cross-country evidence on their roles, experiences, challenges and adaptive strategies. *PLOS Global Public Health* 3(1): e0001447. <https://doi.org/10.1371/journal.pgph.0001447>
- Thorat, A. (2010). Ethnicity, caste and religion: Implications for poverty outcomes. *Economic and Political Weekly*, 47-53.