COVID-19, fake news and religious tensions: experimental evidence from India

Alex Armand[1], Britta Augsburg[2] and Antonella Bancalari[3]

Pre-Analysis Plan

Abstract

While fake news spreading misinformation about COVID-19 is a global concern, it is a particular pervasive problem in India. In this paper, we study how to debunk fake news and combat misinformation in slums of India. While studies found that communication technologies and social media can be effective communication tools, we know little about what role the identity of the messenger plays. Our first research question (RQ) adds to this knowledge gap, by providing evidence on: How effective are doctors' messages to counter misinformation about ways to prevent COVID-19? Accompanying the pandemic, we also experienced globally riots and protests linked to discrimination events. Given religious tensions in India, we next address the question: How does religion identity moderate the processing of new information? We will do so making use of technologies of information through mobile phones. Yet, uptake of messages via phone technologies can be extremely low and hence, the effectiveness of these tools can be limited. Hence, we also study: Can higher financial rewards lead to higher uptake of messages? We will conduct a field experiment to study these research questions in the context of slums in Lucknow and Kanpur, Uttar Pradesh, making use of mobile phone technology. We rely on a recently collected census data of more than 30,000 households and we will collect baseline and follow-up surveys through mobile phones for almost 4,000 randomly sampled households.

[1] NOVA SBR, NOVAFRICA and IFS
[2] IFS
[3] LSE and IFS

I. Introduction

One of the most, if not *the* most, at-risk groups of COVID-19 is the urban poor, living in overcrowded conditions with very limited access to public (health) infrastructure. One billion people live in such settlements, more than half of these in Asia and almost a fifth in India (World Bank 2020). Their ability to follow governments' and scientists' advise on mitigation strategies – such as handwashing, social distancing, and the shielding of elderly and vulnerable groups – has been significantly hampered by the hardships they face on a daily basis, which include lack of access to water and sanitation systems (at home) and overcrowded living (Brown, Ravallion, and van de Walle 2020; Afridi, Dhillon, and Roy 2020).

Along these hardships, misinformation about ways to prevent COVID-19 is widespread. While fake news spreading misinformation about COVID-19 is a global concern[4], it is a particular pervasive problem in India[5]. Misinformation about ways to prevent coronavirus is circulating through social media, ranging from eating vegetarian food to killing the virus with heat[6].

Fake news may generate utility for some slum dwellers (e.g. eating vegetarian may be easier than keeping social distance), but it also imposes private and social costs by making it more difficult to infer the true state of the pandemic. In the presence of widespread misinformation, slum residents are at risk of falling into a false sense of protection and conduct risky behaviours. Furthermore, the pandemic has forced decision-making to take place under great uncertainty, and evidence suggests that individuals are systematically less risk averse under uncertainty compared to certainty (Callen et al. 2014). In the US, for instance, misinformation transmitted through TV shows generated harmful effects by delaying the adoption of preventive behaviour. Because of the large externalities inherent in this pandemic, few viewers affected the disease transmission trajectories for the whole population (Bursztyn et al. 2020).

In this paper, we study how to debunk fake news and combat misinformation in slums of India. The government of India, as many in the World, are currently testing different means through which to communicate the population about the right ways to prevent the spread of this lethal virus. Studies exploring means to release information constraints that affect public health have found that communication technologies and social media can be effective tools (Banerjee et al. 2020; Alatas et al. 2019).

What we know little is what role the identity of the messenger plays. Banerjee et al. (2020) find that a message from Nobel-prize winner Abhijit Banerjee reminding participants to comply with COVID-19 policies is effective at improving behaviour, but the external validity of this study is limited. To what extent can we translate this effect to that of receiving information from other types of messengers (e.g. health experts, religious and political leaders, celebrities)? Our first research question (RQ) adds to this knowledge gap, by providing evidence on:

RQ1: How effective are doctors' messages to counter misinformation about ways to prevent COVID-19?

---

[4] The issue of fake news surrounding the COVID-19 crisis has been highlighted by UN's Secretary General Antonio Guterres (https://www.unbonn.org/news/covid-19-we-are-war-virus-un-secretary-general).

[5] For India, there is evidence of widespread circulation of fake news (https://qz.com/india/1813845/coronavirus-fake-news-rife-on-indian-facebook-whatsapp-twitter/).

[6] The Government of India created a Fact-sheet in a website debunking the most common fake news: https://transformingindia.mygov.in/covid-19/?sector=myth-busters&type=en#scrolltothis

Accompanying the pandemic, we also experienced globally riots and protests linked to discrimination events. In India, the COVID-19 pandemic started in the peak of a conflict between Hindus and Muslims. Fake stories about Muslims creating the virus in a lab and spreading it purposely to kill Hindus circulated in social media[7]. Exposure to different religions, ethnicity, caste and political ideologies affects decision-making (Allcott and Gentzkow 2017; Bazzi et al. 2019; Lowe 2020). Given these religious tensions, it is important to understand if the extent to which the spread of misinformation about COVID-19 can be countered depending on the religion of sender and receivers. We therefore address the question:

RQ2: How does religion identity moderate the processing of new information?

We will do so making use of technologies of information through mobile phones. Yet, uptake of messages via phone technologies can be extremely low and hence, the effectiveness of these tools can be limited. For instance, Banerjee et al. (2020) achieved a viewing rate of only 1.14%, consistent with low rates of other click-through studies (Richardson, Dominowska, and Ragno 2007; Kanich et al. 2009). Hence, we also study:

RQ3: Can higher financial rewards lead to higher uptake of messages??

We will conduct a field experiment to study these research questions in the context of slums in Lucknow and Kanpur, Uttar Pradesh, making use of mobile phone technology.

We rely on a recently collected census data of more than 30,000 households (including geo-codes and mobile phones) located in 142 slums. 1,500 of these households living in the catchment areas of 110 community toilets were interviewed a few more times as part of a completed research study. In this study, we will collect baseline and follow-up surveys through mobile phones for almost 4,000 randomly sampled households.

II. Interventions

A main intervention with variants will be evaluated. We will send messages from different doctors working in renowned health centres in the study area, debunking fake news and reminding the audience about the proven ways to protect against COVID-19. This is called the "Doctor" intervention.

We will send these messages in the form of short videos (approximately 2 minutes) through a Whatsapp chatbot[8]. Given that 50% of the slum households do not have a smartphone with Whastapp, we will send the audio of the videos to study households without Whastapp[9] via voice messages. Each video (and audio) starts with a short clip of a citizen from Uttar Pradesh introducing the doctors' messages.

The variants that we introduce are the following. First, instead of doctors' messages debunking fake news about COVID-19, some households will receive messages debunking fake news about Bollywood stars. This is what we call the "Control" intervention, which will allow us to disentangle the effects of our intervention from receiving a message through mobile phone technologies.

---

[7] This problem attracted the attention of international media: https://www.bbc.co.uk/news/world-asia-india-53165436

[8] Whatsapp chatbot is a software program that runs on encrypted WhatsApp platform. WhatsApp users can communicate with a chatbot through the chat interface as they would talk to a real person.

[9] We use a program that allows us to identify which phone numbers are active on Whastapp.

Second, we will vary the identity of the messenger by changing the religion of the citizen in the introductory clip. Half of the households will receive a “Hindu citizen” intervention and the other half a “Muslim citizen” intervention. We vary the identity by changing the looks, name and surname, body language and phrases used by the citizen. We hold constant any other factors affecting the identity of the messenger by using the same citizen.

Third, in order to incentivise households to watch the video and listen to the audio, we will give participants the chance to enter a lottery. Half of the households will enter a lottery for Rs. 5,000 and others for half. We call the former the “High incentive” intervention, and the latter the “Low incentive”. This will allow us to analyse the effect of a higher expected payoff on the uptake of messages.

Start date: 26/09/2020
End date: Expected 26/10/2020

III. Outcomes

Main outcomes:

1. RQ1: Knowledge about COVID-19 *prevention*: measured as the extent to which the participant agrees with statements on different (confirmed and not confirmed) ways to prevent COVID-19.
2. RQ2: Religion bias: elicited by randomly varying the names of citizens that agree with different statements (not confirmed ways) about how to prevent COVID-19 and asking participants the extent to which they also agree with the statements. We use names that clearly convey the gender and religion of the citizen. For this, we use the most common names in our household Census for each identity.
3. RQ3: Exposure to intervention: number of participants that watched the video and listened to the audio; extent to which participants recall receiving a video through Whatsapp or voice message related to COVID-19.

Secondary:

1. Acquiring and spreading information: the extent of discussion about COVID-19 with other people; time spent in acquiring COVID-19 related information; knowledge about COVID-19 symptoms.
2. Risk perception: the extent to which respondents believe a member of the household can get COVID-19; how anxious they feel about the pandemic.
3. Trust: the extent to which participants trust doctors and people from other religion, in comparison to people in their State in general.
4. Attitudes towards vaccination: the extent to which the participant is willing to vaccinate (having to pay or not) when the vaccine for COVID-19 becomes available; attitude towards people that don’t want to vaccinate.
5. Complying with policy guidelines: behaviour related to better hygiene and physical distance

IV. Experimental design

We address the research questions using a field experiment through mobiles phones in 142 slums in the cities of Lucknow and Kanpur, Uttar Pradesh.

We randomly allocate households (one or more mobile phones in each household) to receive one of the variations of the following three treatments:

T1: Doctor messages vs. Control messages

T2: Hindu citizen vs. Muslim citizen

T3: High incentive vs. Low incentive

We end up having a 2 x 2 x 2 cross-randomized design, resulting in 8 treatment arms.

To allocate households to the treatment arms, we stratify the sample by religion (Hindu or other) and by city of study (Lucknow or Kanpur).

Randomisation method: The statistical software Stata, and specifically the random number generator, will be used to apply this procedure.

Randomisation unit: Randomisation into the experimental arms is conducted at the household level given that the intervention is directed one-to-one through mobile phones. Randomising at the household level allows us to take advantage of greater variation in response to the intervention within slums.

The distribution of households across treatment arms is as follows:
- 498 households allocated to the "Doctor-High Incentive-Hindu" treatment arms
- 505 households allocated to the "Doctor-Low Incentive-Hindu" treatment arms
- 507 households allocated to the "Doctor-High Incentive-Muslim" treatment arms
- 492 households allocated to the "Doctor-Low Incentive-Muslim" treatment arms
- 505 households allocated to the "Control-High Incentive-Hindu" treatment arms
- 479 households allocated to the "Control-Low Incentive-Hindu" treatment arms
- 503 households allocated to the "Control-High Incentive-Muslim" treatment arms
- 502 households allocated to the "Control-Low Incentive-Muslim" treatment arm

Was the treatment clustered? No.

V. Experiment characteristics

Within each slum, we sample up to 60 households, aiming for an average of 30 households per slum. Our sampling procedure is informed by the power calculation used in the initial study registered in the AEA Registry Number AEARCTR-0003087.

Sampling procedure:

We conduct a two-step sampling procedure. First, we sample only from the households that were part of our initial study (approximately 1,500 study households). This allows us to take advantage of the wealth of data available for these households, as well as to focus specifically on more vulnerable households that are forced to leave daily their dwelling to defecate in community toilets. Second, we sample from all the remaining households in the slums that were not part of the study. To deal with high non-response (given the fact that mobile numbers are from two years ago), we randomly sample replacements. We have 1,234 households sampled in the first step and 2,757 households sampled in the second step (a total of 3,991 households).

In order to collect detailed high-quality data about behaviour, while at the same time balancing the need for a pragmatic, short and concise surveys, we randomly allocate households to one of two modules: (i) Hygiene and health; and (ii)Social distance. Primary outcomes and other secondary outcomes are collected for the whole sample.

To deal with misreporting, we collect additional information. First, we measure social desirability bias of each respondent based on a short version of the Marlowe-Crowne Social Desirability Scale (Fischer and Fick 1993). Second, we ask respondents to report the behavior of an intimate neighbour similar to them along socio-demographics, which has proven effective for measuring sensitive public health behaviour (Yeatman and Trinitapoli 2011).

The timing of the survey rounds are:

- Baseline: June 17– July 18 2020
- Follow-up 1: September – October 2020
- Follow-up 2: March 2021

Sample size clusters: 142 slums
Sample size units: 3,991 households, with a mean of 28 households per slum.
Sample size: ~500 units by treatment arm

VI. Analysis Plan

The evaluation design for the comparison of different interventions examines differences in outcomes across households assigned to different treatment groups. Since these households were allocated at random to different treatment groups, they are expected to be identical on average on all their other characteristics, observed or unobserved. A simple comparison of households across groups will give us the impact on household-level outcomes of implementing one versus another intervention.

Identifying the effect of interventions

We start by focusing on the general effect of receiving the “doctor” vs. “control” message, testing differences in mean across the main treatment and control groups.

We next evaluate if there are differential effects by varying the identity of the messenger and the incentives to uptake messages. For individual and household-level outcomes, let $T1_{im}$ be indicator variables that takes value 1 if household is allocated to “doctor” intervention and control otherwise; and $T2_{im}$ will be either:

= 1 if allocated to the “Hindu citizen” intervention; =0 if allocated to “Muslim citizen” intervention.

= 1 if allocated to the “High incentive” intervention; =0 if allocated to “Low incentive” intervention.

In order to estimate the effect of the interventions on the outcome Y_{imt} at time t, we estimate the following model:

$$Y_{imt} = \alpha + \beta_1 T1_{im} + \beta_2 T2_{im} + \beta_3 T1_{im} * T2_{im} + \theta_m + \varepsilon_{im} \quad (1)$$

where θ_m are strata dummy variables capturing the dimensions along which the randomization was stratified. ε_{im} is a residual idiosyncratic error term picking up unobserved determinants of the outcome of interest.

The impact on outcome Y_{imt} of receiving a doctor message, conditional on receiving a citizen video starred by a Hindu or a higher financial incentive to watch the video is given by β_3.

Control variables will be selected by a double-post LASSO procedure. This method prevents over selecting potentially spurious covariates, reduces error, increases statistical power and tests for effectiveness in the randomization. As a robustness check, we will also run regressions without the inclusion of control variables.

Following work by (Mckenzie 2011) for outcomes with high autocorrelation, we will run an ANCOVA specification as a robustness check, where we account for the baseline value of the outcomes considered, namely $Y_{im,t0}$.

Heterogeneous Effects

We also plan to study heterogeneous treatment effects. To this purpose, for each sub-group k in the variable for which we want to study heterogeneity in the effect, we define an indicator d_{ik} that takes value 1 if household/individual *i* belong to sub-group k and 0 otherwise. For binary indicators, the sub-group definition is straightforward. For non-binary dimensions of interest we will split into sub-groups based on the median of the distribution.

The key heterogeneity dimensions that we will look at are:

- Religion of the respondent
- % of Muslims leaving in the slum of the respondent

Procedure for multiple hypotheses testing

We will follow two procedures to address issues related to multiple inference. First, whenever possible, we will build an index capturing different outcomes that are measuring a specific dimension. Second, whenever we have a large set of outcomes, we will adjust p-values for multiple hypothesis testing using the bootstrap-based procedure proposed by (List, Shaikh, and Xu 2016). This has been proven to asymptotically control the family-wise error rate (i.e., the probability of one or more false rejections), and be asymptotically balanced (i.e. the marginal probability of rejecting any true null hypothesis is approximately equal in large samples).

References

Afridi, Farzana, Amrita Dhillon, and Sanchari Roy. 2020. 'How Has Covid-19 Crisis Affected the Urban Poor? Findings from a Phone Survey'. *Ideas for India*, 2020. https://www.ideasforindia.in/topics/poverty-inequality/how-has-covid-19-crisis-affected-the-urban-poor-findings-from-a-phone-survey.html.

Alatas, Vivi, Arun G Chandrasekhar, Markus Mobius, Benjamin A Olken, Cindy Paladines, Marcella Alsan, Nancy Baym, et al. 2019. 'When Celebrities Speak: A Nationwide Twitter Experiment Promoting Vaccination In Indonesia'. 25589. NBER Working Paper. http://www.nber.org/papers/w25589.ack.

Allcott, Hunt, and Matthew Gentzkow. 2017. 'Social Media and Fake News in the 2016 Election'. *Jorunal of Economic Perspectives* 31. https://doi.org/10.1257/jep.31.2.211.

Banerjee, Abhijit, Marcella Alsan, Emily Breza, Arun Chandrasekhar, Abhijit Chowdhury, Esther Duflo, Paul Goldsmith-Pinkham, and Benjamin Olken. 2020. 'Messages on COVID-19 Prevention in India Increased Symptoms Reporting and Adherence to Preventive Behaviors Among 25 Million Recipients with Similar Effects on Non-Recipient Members of Their Communities'. Cambridge, MA. https://doi.org/10.3386/w27496.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, Maisy Wong, Alberto Alesina, Oriana Bandiera, Toman Barsbai, et al. 2019. 'Unity in Diversity? How Intergroup Contact Can Foster Nation Building †'. *American Economic Review* 109 (11): 3978–4025. https://doi.org/10.1257/aer.20180174.

Brown, Caitlin, Martin Ravallion, and Dominique van de Walle. 2020. 'Can the World's Poor Protect Themselves from the New Coronavirus?' *National Bureau of Economic Research*. https://doi.org/10.3386/w27200.

Bursztyn, Leonardo, Aakaash Rao, Christopher P Roth, David H Yanagizawa-Drott, Christopher Roth, David Yanagizawa-Drott, Alberto Alesina, et al. 2020. 'Misinformation During a Pandemic'. 27417. NBER Working Paper.

Callen, Michael, Mohammad Isaqzadeh, James D Long, and Charles Sprenger. 2014. 'Violence and Risk Preference: Experimental Evidence from Afghanistan'. *The American Economic Review* 104 (1): 123–48. https://doi.org/10.1257/aer.104.L123.

Fischer, Donald G., and Carol Fick. 1993. 'Measuring Social Desirability: Short Forms of the Marlowe-Crowne Social Desirability Scale'. *Educational and Psychological Measurement* 53 (2): 417–24. https://doi.org/10.1177/0013164493053002011.

Kanich, Chris, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. 2009. 'Spamalytics'. *Communications of the ACM* 52 (9): 99–107. https://doi.org/10.1145/1562164.1562190.

List, John A, Azeem M Shaikh, and Yang Xu. 2016. 'MULTIPLE HYPOTHESIS TESTING IN EXPERIMENTAL ECONOMICS'. 21875. NBER WORKING PAPER SERIES MULTIPLE. Massachussets, MA. https://github.com/seidelj/mht.

Lowe, Matt. 2020. 'Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration'. 8089. CESifo Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3536734.

Mckenzie, David. 2011. 'Beyond Baseline and Follow-up: The Case for More T in Experiments *'. WPS 5639. Policy Research Working Paper. Washington D.C. https://blogs.worldbank.org/impactevaluations/files/impactevaluations/beyond_baseline_and_followupjde_final.pdf.

Richardson, Matthew, Ewa Dominowska, and Robert Ragno. 2007. *Predicting Clicks: Estimating the Click-Through Rate for New Ads*. https://www.microsoft.com/en-us/research/publication/predicting-clicks-estimating-the-click-through-rate-for-new-ads/.

World Bank. 2020. 'World Development Indicators'. 2020. http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators.

Yeatman, Sara, and Jenny Trinitapoli. 2011. 'Best-Friend Reports: A Tool for Measuring the Prevalence of Sensitive Behaviors'. *American Journal of Public Health* 101 (9): 1666–67. https://doi.org/10.2105/AJPH.2011.300194.