

Alex Armand
Britta Augsburg
Antonella Bancalari
Kalyan Kumar Kameshwara

23/39

Working paper

Social proximity and misinformation: Experimental evidence from a mobile phone-based campaign in India

Social Proximity and Misinformation: Experimental Evidence from a Mobile Phone-Based Campaign in India*

Alex Armand
Antonella Bancalari

Britta Augsburg
Kalyan Kumar Kameshwara

Abstract

We study how social proximity between the sender and the receiver of information shapes the effectiveness of preventive health campaigns. Focusing on shared religious affiliation as a signal of proximity, we implemented a field experiment during the initial phase of the COVID-19 pandemic in two major Indian cities characterized by Hindu-Muslim tensions. We randomly allocated a representative sample of slum residents to receive either “doctor messages” consisting of voice messages promoting recommended practices to prevent virus transmission or uninformative control messages on their mobile phones. The messages, introduced by a local citizen (the sender), were cross-randomized to commence with a greeting signaling either a Hindu or a Muslim identity, thereby manipulating the religion concordance between the sender and receiver. As compared to control messages, doctor messages are effective at improving compliance with recommended practices. However, the impact of these messages is contingent on the presence of religion concordance. In instances of concordance, the take-up of doctor messages is larger, and the intervention demonstrates efficacy in establishing a protective barrier against misinformation related to preventive practices. (*JEL codes: C93; D91; I12; I15; O12*)

Keywords: Health; Campaign; Information; Religion; India; COVID-19; Field Experiment.

***Armand:** Nova School of Business and Economics – Universidade Nova de Lisboa, CEPR, NOVAFRICA, and Institute for Fiscal Studies (corresponding author, e-mail: alex.armand@novasbe.pt); **Augsburg:** Institute for Fiscal Studies (e-mail: britta_a@ifs.org.uk); **Bancalari:** Institute for Fiscal Studies (e-mail: antonella.bancalari@ifs.org.uk); **Kameshwara:** Westminster Business School, University of Westminster (e-mail: K.Kameshwara@westminster.ac.uk). We would like to thank Oriana Bandiera, Maithreesh Ghatak and Michael Callen, as well as seminar participants at the LSE-STICERD Development Economics seminar series and the IFS for helpful comments. This project would have not been possible without the insightful support of our partners in the field, Bhartendu Trivedi, Yashashvi Singh, Tatheer Fatima and the field team of Morsel Research and Development. We are grateful to Julia Loh for outstanding research assistance. We are also indebted to study participants for their willingness to contribute to the study. We gratefully acknowledge financial support from the IGC (grant ref. IND-20077), the LSE Special Grant for Coronavirus Research (grant ref. 106404), and the ESRC’s Centre for the Microeconomic Analysis of Public Policy at the IFS (grant ref. ES/T014334/1). Kameshwara acknowledges the support from the ESRC’s SWDTP Placement Scheme. Armand acknowledges funding from the Fundação para a Ciência e a Tecnologia (UIDB/00124/2020, UIDP/00124/2020 and Social Sciences DataLab - PINFRA/22209/2016), POR Lisboa and POR Norte (Social Sciences DataLab, PINFRA/22209/2016). Ethics approval was secured from the London School of Economics and Political Science (REC ref. 1132). The pre-analysis plan was registered on the AEA RCT registry ([Armand et al., 2020](#)).

1 Introduction

In countries with greater social diversity, the nature and extent of interactions among citizens are significantly influenced by their relative *social proximity*. This tendency to adapt behavior based on shared traits and identities, with serious ramifications, is notably observed in health-related interactions, especially between patients and doctors (Greenwood et al., 2018; Alsan et al., 2019; Greenwood et al., 2020; Hill et al., 2020).¹ Despite its significant influence on aggregate economic outcomes (see, for instance, Alesina and Ferrara, 2005), social proximity is often overlooked in communication campaigns promoting health-related behavior. Consequently, evidence on the effect of social proximity in the spread of information, especially in low-income settings, is scarce. More generally, research on the mechanisms of (mis)information remains limited and predominantly focused on higher-income countries (DellaVigna and Kaplan, 2007; Allcott and Gentzkow, 2017; Lazer et al., 2018; Bursztyn et al., 2023).

This paper explores the effectiveness of an information campaign promoting health-related preventive practices and how introducing social proximity in the campaign enhances its effects. We examine the impact of social proximity based on whether the sender and the receiver of information share the same religion. We document that promoting preventive behavior can increase compliance with recommended practices and beliefs about their efficacy. We find that these effects are driven primarily by a campaign in which the sender and the receiver share the same religion. In this case, the receiver listens to a larger share of the message and is more compliant with recommended practices. We further find that religion concordance inoculates citizens from misinformation shared by individuals of their same religion.

We implement a field experiment in the Indian State of Uttar Pradesh (UP) in the context of a global outbreak of an infectious disease – the COVID-19 pandemic. At the onset of the pandemic, we designed a mobile phone-based information campaign to raise citizens’ awareness about evidence-based practices to mitigate the spread of the virus, and to counteract the sudden rise in misinformation surrounding the pandemic (World Health Organisation, 2020).² To this purpose, between October 2020 and January 2021, we sent two pre-recorded voice messages to a representative sample of slum residents, a largely understudied population (Lilford et al., 2017), in the two major cities of the state. The campaign held particular importance in this context, not only due to the overcrowded living conditions that made physical distancing challenging, but also due to the low-income and marginalized nature of the setting, which limited access to healthcare and adequate hygienic conditions.

Each voice message consists of two components: an introduction by a local citizen, the *sender*, followed by the *content* of the message. Using cross-randomization, we vary both components. To obtain exogenous variation in sender-receiver social proximity, we randomly vary the greeting used by the sender at

¹Alternative settings in which social proximity has been shown to influence behavior are nation building processes (Bazzi et al., 2019; Mousa, 2020; Lowe, 2021), financial decisions (Fisman et al., 2017, 2020), or experimental games (Habyarimana et al., 2007; Bicchieri et al., 2022).

²In India, the spread of misinformation about COVID-19 was so severe that it compelled the PM Narendra Modi to address the nation urging everyone to rely only on credible medical advice and demanding social media companies to curb misinformation on their platforms (Al Jazeera, 2020). Internet penetration rates went from 4% in 2007 to 50% in 2020, raising social media platforms as a primary source of news and as a key mean of communication for all political party actors (Statista, 2021).

the beginning of the message to signal either a Muslim or Hindu identity. This feature creates exogenous variation in the *religion concordance* between the sender and the receiver, while leaving the sender, the remaining part of the introduction, and the content of the message unchanged. Religion is a highly salient dimension of social proximity in our setting, particularly at the time of the experiment. In India, Hindu–Muslim tensions have been present since pre-partition era, and are particularly relevant for UP, home to the largest Muslim population in India (Jha, 2013; Mitra and Ray, 2014). In line with religion being salient in the presence of unpredictable events (Sinding Bentzen, 2019; Atkin et al., 2021), the onset of the pandemic saw a sudden increase in these inter-religious tensions: misleading claims about the role of Muslim citizens in the spread of the virus were the primary driver of fake news on social media, and spurred further violence (see, e.g., Yasir, 2020).

To obtain exogenous variation in the content of the message, we randomize whether the receiver is sent messages about preventive practices or uninformative content. In the former, which we label as *doctor* messages, the content is provided by doctors of locally-renowned hospitals reminding about evidence-based policy recommendations and debunking common misconceptions about the virus. The religion identity of doctors is not revealed. In the latter, which we label as *control* messages, the content consists of Bollywood gossips unrelated to the pandemic. Thanks to cross randomization, both the doctor or the control messages are either religious-concordant or -discordant.³

We gathered information about participants’ behavior related to preventive practices, beliefs over the efficacy of both recommended and non-evidence-based practices, and about participants’ response to misinformation about the pandemic during a baseline and two follow-up surveys. We base our main analysis on intention to treat (ITT) effects, which capture the effect of sending the messages. Using administrative data on the take-up of the interventions, we complement ITT estimates with treatment on the treated (ToT) estimates of the effect among compliers.

The design of the experiment allows us to study first the overall effect of promoting preventive practices, and then to estimate the effect of combining social proximity with it, a novel setup in the literature. Providing informative content via mobile phones is effective at promoting welfare-improving behavior. As compared to control messages, doctor messages increase significantly the compliance with recommended practices and update positively recipients’ beliefs about the efficacy of these practices. However, despite being debunked in the message, doctor messages have no significant effect on the degree to which respondents believe that non-evidence-based practices like relying on vegetarianism or on a stronger immune system can prevent from infection, indicating the persistence of these beliefs to new information.

To assess the added benefit of social proximity, we focus on the sample that was sent the doctor message, and exploit the cross randomization in the religion concordance between the sender and the receiver of the information. First, we find that religion concordance leads participants to listen to a larger share of

³The experimental design also cross-randomized whether the receiver was incentivized with lower or higher monetary incentives to listen to the message. Refer to Section 3.

the doctor message, an increase of 13.3% as compared to religion-discordant messages. Second, the effect on compliance with recommended practices of doctor messages is primarily driven by religion-concordant messages. Third, religion concordance in the doctor messages effectively reduced beliefs over the efficacy of non-evidence-based practices, particularly those with a religious connotation.

The latter results are specific to the combination of informative content provided by the doctor in combination with religion concordance. Studying the differential effects of religion concordance in the control messages, which serves as a placebo test, indicates no effect in any of the outcomes studied. In addition, effects are specific to misinformation. In fact, none of the interventions influence agreement with non-factual opinions about the spreading of COVID-19, by definition more persistent and harder to be influenced by information campaigns as compared to pure misinformation (see, for instance, [Walter and Salovich, 2021](#)). Finally, we provide evidence that spillover effects were not present in the interventions, suggesting that mobile-phone campaigns are effective at targeting individuals rather than communities. To understand drivers behind these impacts, we first analyse fact-checking behavior of respondents, an important determinant of factual knowledge ([Barrera et al., 2020](#)). The findings reveal that doctor messages significantly reduce the likelihood of verifying the truthfulness of information. This reduction is likely because individuals, having heard the messages from doctors, feel more confident in dismissing misinformation. We further use a novel survey instrument to measure whether respondents agree with misinformation shared by other citizens and show that doctor messages reduce the agreement with misinformation shared by citizens *outside* the religious group of the respondent (*out-group* citizens), while keeping unchanged their level of agreement with citizens of the same religion (*in-group* citizens). Religion concordance in the doctor messages is effective at detaching in-group norm compliance in the response to misinformation. When the sender and the receiver share the same religion, doctor messages reduce agreement with misinformation shared by in-group citizens by 4.6%, as compared to religion-discordant messages. This finding aligns with existing research and placed in high-income countries, which emphasize that the perceived credibility of information is influenced by the social distance between the communicator and the recipient ([Tabellini, 2008](#); [Alsan et al., 2019](#)).

These results suggest that the information campaign induces some degree of crowding out of the effort exerted to verify the truthfulness of information, but at the same time it creates a layer of protection against misinformation. However, this layer is crucially affected by group salience, suggesting a high level of in-group norm compliance in our setting (see, e.g., [Akerlof and Kranton, 2000](#)). This compliance can, however, be detached through a carefully designed information campaign that takes into account social proximity with the objective of leveraging social norms, challenging the assumption that in- and out-groups agree with prevailing norms.

Our findings offer novel insights into the design of information campaigns, an instrument that has been extensively used to communicate risk and best practices for health behavior ([Dupas, 2011](#)). We complement available evidence on the effectiveness of communication technology to raise health awareness in the US ([Alsan et al., 2020](#); [Breza et al., 2021](#); [Torres et al., 2021](#)), in the Indian state of West Bengal

(Banerjee et al., 2020), and in rural India and Bangladesh (Siddique et al., 2022). We further the understanding of these interventions by providing novel evidence on how the effectiveness of information campaigns on preventive behavior is crucially influenced by social proximity. Our design is unique in the literature because it allows identifying the effect of the initial signal of social proximity (i.e., the first word of the campaign), while keeping the content of the message indistinguishable in terms of identity. Previous literature focuses instead on *micro-targeting* (i.e., the shaping of both the sender and the content of information to the individual characteristics of the receiver). This approach has been used to influence interactions with patients (Yom-Tov et al., 2018; Alsan and Eichmeyer, 2021).

By linking compliant behavior with beliefs and response to misinformation, we not only provide novel evidence on the drivers of information, but also on the mechanisms of misinformation, whose persistence remain a puzzling result in the literature (Van der Linden et al., 2017; Zhuravskaya et al., 2020). In particular, despite the recognition that understanding how beliefs are affected by information is crucial, few studies explicitly elicit the effect of information on beliefs over practices and on how misinformation is perceived (Kremer et al., 2019).

Finally, highlighting the role of social proximity also complements available evidence on the role of identity in decision-making. The literature shows how identity affects cooperation, political mobilization and violence (Philpott, 2007; Bhalotra et al., 2014; Lowe, 2021), but limited evidence covers information sharing. We reinforce the role of religious identity among interacting citizens, a growing field of study in both economics and political science (Iyer, 2016). Further, the specific focus on the use of religion for spreading information through mobile phones further our understanding on how these technologies stimulate social mobilization (see, e.g., Enikolopov et al., 2020; Manacorda and Tesei, 2020).

2 Conceptual framework

Following the frameworks of Pauly and Blavin (2008) and Baicker and Schwartzstein (2015), we assume that agents have inaccurate beliefs about or salience of the value of preventive health practices in combating COVID-19. Wrong beliefs about the returns of preventive practices can lead to under-adoption, i.e., a lower take-up than the socially-optimal level.⁴ If these are binding constraints to preventive care, an information campaign could promote adoption by correcting beliefs about the returns of these practices or by raising their salience (Haaland et al., 2023).

We study two hypotheses related to this mechanism. The first hypothesis is that messages from doctors are effective at promoting the adoption of preventive practices. This hypothesis depends primarily on three factors. First, whether doctors are considered a credible and trusted source of information (O’Keefe, 2016; Khan et al., 2021). In our information campaign, this is crucial as 95% of the targeted population report doctors as the most trusted source of COVID-19 information. Second, the degree of

⁴Under-adoption in slum settings can also be driven by limited access to clean water, safe sanitation and overcrowding (Patel, 2020; Wasdani and Prasad, 2020; Armand et al., 2023).

malleability of the beliefs that are causing under-adoption. Information campaigns are more effective at influencing beliefs based on misconceptions or incomplete understanding as compared to views that are less grounded on facts or knowledge (Walter and Salovich, 2021). This factor demands distinguishing between these two dimensions in the analysis of the campaign's impacts. Third, whether messages influence an individual's attitude towards checking the truthfulness of new information, which demands studying how the targeted population reacts when facing misinformation. For instance, a campaign may increase fact-checking if individuals become more aware of the degree of misinformation flowing in their social network, or decrease it if the ability to recognize false or inaccurate information is improved. The second hypothesis is that messages from doctors are more effective when the sender and the receiver of the message are socially close. If beliefs or salience are binding constraints to preventive care, then social proximity could enhance the effectiveness of the campaign by increasing the degree of credibility of information, particularly when the target group is more marginalized and less-educated, thus more socially-distant from doctors (Lazer et al., 2018; Bavel et al., 2020). The enhancing effect of social proximity can also operate by raising the salience of group identity, with important consequences on norm compliance (see, for instance, Akerlof and Kranton, 2000; Chen and Li, 2000), but also on beliefs' updating. For instance, social proximity could correct beliefs that have a close connection to the in-group or the out-group identities. In our setting, beliefs over the effectiveness of vegetarianism in protecting against COVID-19 have a strong salience in Hindu communities, but not in Muslim communities.

3 Intervention and experimental design

The intervention is designed to test the hypotheses discussed in Section 2. It targets the population of slum residents in the two largest urban agglomerations in UP, Lucknow and Kanpur. Appendix A shows their geographic location and the distribution of different religions and castes in these populations. Similar to many expanding cities in low and middle-income countries, Lucknow and Kanpur are characterized by a relatively large prevalence of informal settlements, and a prospect of rapid population growth, which makes the setting highly relevant for contagious diseases.⁵

We draw the study population from a census of slum residents conducted in the second half of 2018 in both cities as part of a distinct study. Refer to Solís Arce et al. (2021) and Armand et al. (2023) for further details about this population and the census procedures. The sampling frame carries unique information for more than 30,000 households living in the slums of the study area before the beginning of the pandemic. From the sampling frame, we targeted a random sub-sample of 4,000 households. The resulting study population is comparable to the average slum resident in the state and in the rest of India. The intervention took place during the initial phase of the COVID-19 pandemic. Similar to other states of India, UP was hit hard by the pandemic during the period of the study, with a rapid spread in the number

⁵In 2015, Lucknow and Kanpur were the 129th and 141st cities worldwide in terms of population (United Nations, 2019), with an expected growth in the period 2015–2035 of 59% and 37%, respectively. Across agglomerations of similar size, this growth prospect is comparable to cities such as Accra (Ghana), or Amman (Jordan).

of COVID-19 cases and a steep increase in the number deaths (Appendix Figure A2). Guidelines of social distancing and wearing of face masks remained in place throughout the study period. Saliency of the guidelines were particularly high in UP: out of 29 states, it is the largest (home to 200 million people), the 4th most-densely populated, and the 6th in terms of share of population living in slums, totalling more than 6 million people (Government of India, 2011).⁶ Appendix Figure A3 summarizes the study timeline and compares it with COVID-19 regulations in UP in the corresponding period.

The intervention consists of sharing voice messages via calls targeted at individual citizens using mobile phone technology.⁷ Each message has two components: the *introduction* delivered by a local citizen, the *sender*, and the *content* of the message. The full scripts of the messages are reported in Appendix A.2.

To introduce variation in social proximity associated with the message, we exploit religious diversity in UP. In the slum setting, the representation of religious groups is comparable to that of the whole state, with 79% of the sample being represented by Hindu citizens, and 21% by Muslim citizens. Members of these religious groups tend to use distinct greetings. We exploit this characteristic by introducing two variations in the introduction of the message. The sender either signals a Hindu identity by using the greeting “namaste” at the start of the message, or a Muslim identity using the greeting “salam alaykum”. The remaining part of the introduction is kept constant, including the language spoken. We refer to *religion concordance* of the message when the initial greeting of the sender is signalling the same religion of the receiver of the message, and *religion discordance* when it is signalling a different religion.

To separately introduce variation in the content of the message, we varied the content following the introduction to be either informative (with the objective of raising preventive health awareness) or uninformative. In the informative version, labelled as the *doctor* messages, the content is presented by doctors from locally renowned medical institutions debunking common misconceptions about ways to prevent COVID-19 and reminding about the confirmed ways to protect against infection. Qualified medical practitioners were chosen for the informative content to guarantee that information was shared by trusted sources (see Section 2). We sent two rounds of messages. Each message reminded about the World Health Organization (WHO) recommended practices to avoid contagion, and, in addition, the first message highlighted that eating a vegetarian diet does not protect against COVID-19 (sent in October–November 2020), and the second message debunked the fake-news that the immune system of Indians is resilient to COVID-19 (sent in December 2020–January 2021).⁸ At baseline, relying on vegetarianism

⁶While UP presents a higher poverty rate as compared to the average for India (29.43% versus 21.92%, Reserve Bank of India, 2019), its slum population is highly comparable to the average slum population in the country. The share of adult males (0.53 in UP versus 0.52 in India), of adult females (0.47 versus 0.48), and of children (0.14 versus 0.12), as well as the sex ratio (1.12 versus 1.08) and the share belonging to Scheduled Castes (0.22 versus 0.20) are indicative of close similarities between these two populations. In terms of literacy rates, the average slum in UP outperforms the one of the whole India (0.69 versus 0.78).

⁷Alternative remote approaches include live phone calls (Sadish et al., 2021), communication via instant messaging platforms (Bowles et al., 2020), or pedagogical interventions (Badrinathan, 2021).

⁸The content for these messages was built by first asking several doctors from renowned local institutions to reply unscripted to the questions: “is it true that eating a vegetarian diet protects against COVID-19?” and “is it true that the immune system of Indians is resilient to COVID-19?”. Responses were collated ensuring that every message was composed by a first part debunking the misconception and a second part on policy recommendations.

and on the Indian immune system were the two most prevalent non-evidence-based preventive practices to avoid contagion from COVID-19 (Appendix Figure A4). All participants allocated to the *doctor* messages received messages from the same set of three doctors. We did not randomize the religious identity of doctors in order to disentangle the effects of identity from other doctor-specific characteristics (for example, doctor from religion A also being more charismatic than doctor from religion B). Instead, we share messages from religious-neutral doctors (i.e., doctors do not reveal their religion identity, neither through salutation nor their name).⁹

In the uninformative version, labelled as the *control* messages, the recording begins with the same introduction of the local citizen as in the doctor message, but the message content is an unsubstantiated gossip concerning Bollywood stars. Sending a control message, rather than no message, allows us to disentangle the effects of the intervention from the effect of receiving a message.

The length of the recordings was 1.58 minutes (or 95 seconds) for the first round of the doctor message, and 1.55 minutes (or 123 seconds) for the second round of the doctor message. The length of the control message was shorter, at 1.28 minutes (or 77 seconds) in both rounds. In addition to sharing voice messages, the original intervention consisted also of sending the video underlying the voice messages through a WhatsApp chatbot, i.e., a software purposely programmed for the intervention that runs on the encrypted WhatsApp platform and in which users can communicate with the software through the chat interface. In addition to the variation induced by the initial greeting, videos also varied the name (as printed in the video) and the clothes of the sender to signal either a Muslim or Hindu identity. Yet, videos were only visualized by a very small share of participants due to the WhatsApp policy that requires to start each chat with a generic greeting “Hi”, and share the rest of the chatbot message and the video message only if the respondent replied to the initial greeting. Previous studies using WhatsApp make use of subscribers to by-pass this pre-condition (see, for instance, Bowles et al., 2020). We sent the video message to all phone numbers in the sample, 38.5% received the chatbot message saying “Hi” (namely, this share had a smartphone, WhatsApp installed on their phone, and a data package activated or an internet connection), and just 2.5% replied to the initial greeting and received the rest of the chat message and the video. We cannot verify the share that downloaded and watched the video, but, in line with the literature, we can reasonably assume to be much smaller than 2.5%.¹⁰ Including controls for the receipt of the video message on WhatsApp or excluding these participants from the sample does not alter any of the results. We did not face these limitations when sharing the audio messages via phone calls, allowing us to measure which participants answered the call, and the duration of the audio message that was played (see Section 6.1).

To reduce the risk of low uptake of the information campaign, all messages were incentivized to increase

⁹It is possible that respondents infer that the doctors providing the answers belong to the same religious group as the sender. This dimension is not observed in our data. We take a conservative approach and interpret the results as the religion concordance between the participant and the sender only, as intended in the intervention design.

¹⁰The low uptake is a common risk not only in information experiments (see, for instance, Azrieli et al., 2018), but also in mass information campaigns. In the context of unincentivized video messages sent to Indian citizens by SMS and urging them to comply with COVID-19 policies, Banerjee et al. (2020) achieved a viewing rate of just 1.1%.

attention paid to the message by giving participants the chance to enter a lottery if they replied correctly to a follow-up question about the message. The research design is therefore a $2 \times 2 \times 2$ randomized controlled trial using household-level randomization after stratifying by religion of the household head and city of residence. We adopted the following procedure: first, we randomly allocate targeted households to receive either doctor or control messages; second, we cross randomized households in both the doctor and control message groups to receive a message introduced by a Hindu or a Muslim greeting, thus creating exogenous variation in religion concordance; third, we cross randomized households in both the doctor and control message groups into a lower-incentive lottery with a value of Rs. 2,500 (US\$32) or a higher-incentive lottery with a value of Rs. 5,000 (US\$64).

In line with the pre-analysis plan (Armand et al., 2020) and to obtain a standard level of statistical power, we estimate treatment effects up to the second level of randomization, focusing on the effect of the content of the message and its combination with either the religion of the sender or with the level of monetary incentives. For the latter, because the lottery amounts are both sizable, and therefore differential impacts are marginal, we present the results in Appendix D.5 and discuss them in Section 6 when relevant. Section 5 discusses potential threats of spillover effects deriving from the experimental design, and how we exploit household-level randomization to test for spillover effects.

4 Data

We draw on two data sources, summarized in this section: a panel survey of slum residents, and administrative data on the implementation of interventions. Appendix B offers detailed description of each variable, including the type (self-reported, elicited or from administrative records) and the round (baseline or follow-up), and elaborates on the ethical considerations related to data collection activities.

Primary panel data. We collected primary data among slum residents on households' experiences during the COVID-19 pandemic, such as their knowledge on how to prevent the virus, compliance with policies, as well as information on sources of information and trust, and beliefs. We collected a baseline survey in June–July 2020, reaching 3,991 households. Two waves of follow-up panel data were collected in October–November 2020, and December 2020–January 2021 (3.5 and 5.5 months after the baseline survey), reaching 3,816 households during the first follow-up and 3,906 during the second follow-up survey. To keep the time gap between the intervention and follow-up data collection similar across individuals, we split the sample in four batches determined by the operational capacity of the field team. In each batch, we interviewed households two weeks after sending the voice messages by conducting phone conversations. The sampled households that were not reachable at the time of the survey were replaced with replacement households randomly selected from the sampling frame described in Section 3.

Combining both follow-up surveys, we re-interviewed 87% of residents at least once, with a low implied attrition rate (13%) compared to phone surveys conducted in similar settings. Response rates are

typically around 50% in non-crisis contexts, while during crisis contexts this is expected to be lower. For instance, a study during the Ebola crisis was able to re-interview only 38% (Himelein et al., 2020). Attrition is orthogonal to treatment allocation, while being female and a dwelling owner significantly reduces attrition (Appendix C).

Appendix Table C1 presents descriptive statistics of the sample. Twenty-one percent of respondents are Muslim, almost 80% of respondents are male, mostly represented by the household head, with an average age of 40 years. More than 80% live in a strong dwelling with four other members, and 38% have a ration card (i.e., an official document giving access to the subsidized purchase of essential commodities). At the time of the baseline survey, 12% of respondents report that at least one member was having COVID-19 symptoms.

The primary outcome is the compliance with recommended practices to avoid spreading COVID-19, as highlighted in the doctor messages. We obtain information about behavior related to wearing face masks, hand-washing and physical distancing using multiple questions split in two modules: the “hand-washing” module and the “physical distancing” module, which included a question on the use of face masks. Individual questions are detailed in Appendix B. To guarantee both a high quality of information and a concise interview, each module was administered to a random subset of households only. We then aggregate individual variables using an index of z-scores following Kling et al. (2007), by first normalizing individual variables in standard deviations from the control group, and then averaging available information. This procedure allows to obtain a compliance index for all respondents that responded to one of the modules.

While compliance is based on self-reported behavior, it is important to highlight that information was collected two weeks after exposure to the interventions. This extended time period reduces concerns regarding experimenter demand effects (i.e., changes in behavior by experimental subjects due to cues about what constitutes appropriate behavior), as well as spurious priming effects (i.e., effects that dissipate within hours after the intervention and are only driven by the salience of the message, but not by a change in knowledge, attitudes or behavior). To further alleviate concerns about experimenter demand effects, we collected baseline data on social desirability using the Marlowe-Crowne Social Desirability Scale (Fischer and Fick, 1993), and discuss heterogeneity depending on the level of social desirability in Section 6.

We supplement information about compliance to recommended practices with beliefs over the efficacy of different ways to prevent infection from COVID-19. We asked respondents about the level of agreement with various recommended preventive practices (i.e., present in policy recommendations) and non-evidence-based preventive practices (i.e., not present in policy recommendations), all of them discussed in the doctor messages (Section 3). Evidence-based practices included wearing a face mask, hand-washing, and keeping physical distance. In contrast, non-evidence-based practices included the two most-common views collected at baseline on how to protect from the virus, which are also the ones that

the doctor message debunked: to rely on vegetarianism or on the Indian immune system.¹¹ The beliefs over the efficacy of recommended practices are strongly positively correlated with the compliance index, and the beliefs over the efficacy of non-evidence-based practices are negatively correlated with the compliance index, validating the index (Appendix D.3).

Finally, we measure how participants respond to misinformation about COVID-19. We gather information on *fact-checking*, a proxy for evidence-based behavior related to misinformation. Additionally, we introduce a novel survey instrument to elicit how participants respond when facing misinformation shared by other citizens. In line with the literature (see, for instance, Scheufele and Krause, 2019), we define misinformation as incorrect views based on faulty knowledge or understanding. We present respondents with two statements attributed to a third person living in UP, whom we refer to as the *interlocutor*, and we then elicit their level of agreement with each statement. Statements are presented in a random order during the interview to avoid question order bias.¹² The content of the statements was chosen to reflect common claims by the media, including some with significant religious salience. The first statement, “if you are vegetarian, you do not need to worry about the coronavirus”, carries specific religious salience since, in the context of India, vegetarianism is widely associated with the dominant ideology of Hinduism. The second statement, “if you are a good person, you do not need to worry about the coronavirus”, carries general religious salience, with the idea that religion helps becoming a good person.¹³

As agreement with misinformation is often associated with motivated thinking (i.e., the set of emotional biases leading individuals to agree with views based on desirability rather than evidence), agreement with these statements may vary based on the interlocutor’s identity. This aspect is crucial in our context, where religious tensions can blur the lines between misinformation agreement and group identity, often linked to religion (Tankard and Paluck, 2016; Nyhan, 2021). To investigate this, we choose the name of the interlocutor to signal different religious identities using 5 options: 1 male Muslim name, 1 female Muslim name, 1 male Hindu name, 1 female Hindu name, or a generic “people”. Names were selected using information on the most common names by religion from the census of slum residents (see Section 3). For each respondent, statements are randomly allocated to one of these 5 options. Because the list of statements is constant in the survey, but interlocutors vary in each interview, we can measure agreement with statements shared by citizens that are either *in-group* or *out-group* interlocutors, depending on whether the respondent shares the same religion signalled by the interlocutor. When two interlocutors

¹¹Baseline information for these variables is not available because the baseline questionnaire elicited practices through an open-ended question, rather than in levels of agreement with their efficacy.

¹²The exact script of the question reads as follows: “We have surveyed a few people from UP and we would like to hear if you agree with their opinion. Note that responses to the statements are a matter of opinion. There is no scientific evidence about their truthfulness. On a scale of 1 to 5, where 1 means you strongly disagree and 5 you strongly agree, how much do you agree or disagree with the following statements. [Interlocutor] says that [statement].”

¹³We elicit agreement with three further statements, which contain views about COVID-19 that are not necessarily based on facts or knowledge. We label these statements as *opinions*. Because opinions are harder to be influenced by information campaigns and by fact-checking (Walter and Salovich, 2021), we use them as placebo statements. Impacts on these variables are discussed in Section 6.3, while Appendix D.6 presents descriptive statistics.

fall in the same identity category, we average agreement with their individual statements.

Figure 1 provides descriptive statistics on these variables for the control group. Panel A focuses on the index of compliance with recommended practices and on respondents' levels of agreement with evidence- and non-evidence-based preventive practices over the course of study. Panel B focuses on fact-checking and respondents' levels of agreement with misinformation shared by in-group and out-group citizens.¹⁴ A few observations are worth highlighting. First, likely because some of the restrictions were removed in the follow-up (e.g., self-employed were allowed to work, and offices, super-market and entertainment industries re-opened), the average level of compliance with preventive practices reduces over time. At the same time, the level of agreement with evidence-based ways to protect from the virus (first three figures in Panel A) remains significantly higher than agreement with non-evidence-based practices (last two figures in Panel A). Moreover, the level of misinformation is noteworthy, as people on average neither agree nor disagree with misconceptions shared both by in-group and out-group citizens, and it increases slightly over time (last two figures in Panel B).¹⁵

Administrative data. The voice messages were sent to the whole sample in two rounds. For each round, we gather information about the delivery of voice messages, and about the duration and the share of the voice message that each user played. Descriptive statistics and differential effects of the interventions on the take-up of messages is discussed in Section 6.1.

5 Empirical approach

To assess treatment impacts we rely on post-baseline data, in line with the trial registry (Armand et al., 2020), and justified by having successfully created observationally-equivalent groups. Appendix Table C1 shows mean differences at baseline between the different treatment arms for respondent's characteristics. We find balance in terms of observable characteristics across groups allocated to the doctor and control message, as well as across Muslim and Hindu sender within the doctor message group.

The primary objective is to test different hypothesis on how the interventions translate into behavioral impacts, as discussed in Section 2. The first hypothesis is that the doctor message, which carries informational content related to COVID-19, impacts health-related behavior (as compared to the control message, which has no content related to preventive practices). The second hypothesis is that the doctor message with religion concordance between the sender and the receiver generates differential impacts as compared to the doctor message in which the religion of the sender is different from the one of the receiver. In the experimental design, there exists another hypothesis in which the control message with

¹⁴Appendix Figure A5 shows respondents' levels of agreement with each statement, distinguishing by whether the interlocutor is in- or out-group.

¹⁵We find differences by religion in beliefs for non-evidence-based practices as well as for misinformation (Appendix D.1). On average, Hindu respondents are significantly more likely to agree with non-evidence-based ways and with misinformation shared by other citizens, a difference that is mainly driven by beliefs about vegetarianism, the predominant diet among the Hindu population. Finally, agreement with misinformation tend to be relative constant over time and similar across different types of interlocutors.

religion concordance generates differential impacts as compared to a control message in which the religion of the sender is different from the one of the receiver. However, because the control message has no content related to preventive practices and it is not expected to impact health-related behavior, we expect no differential impact. We in fact treat this comparison as a placebo comparison and discuss it in Appendix D.2.

For the first hypothesis, we estimate the impact of the doctor message using the following specification:

$$Y_{ijt} = \beta_D \text{doctor}_i + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \quad (1)$$

where Y_{it} are outcomes of interest of respondent i in slum j at time t . The variable doctor_i is an indicator variable equal to 1 if the receiver i is in the doctor message treatment group, and 0 otherwise. \mathbf{X}_{ij} is a set of control variables, and δ_t are period-of-survey indicator variables. In the main analysis, \mathbf{X}_{ij} includes only the indicator variables for randomization strata. Adding more control variables selected with the post-double selection LASSO (PDSL) procedure (Belloni et al., 2013; Tibshirani, 1996) or controlling for the baseline value of the outcome variable (ANCOVA specification) does not affect the results; if any, precision improves (Appendix D.7). The error term ϵ_{it} is assumed to be clustered at the slum level, but results are robust to alternative assumptions about standard errors, such as clustering at the individual level.

For the second hypothesis, we estimate the role of religion concordance with the sender of the doctor message by restricting the sample to the doctor message group, therefore focusing on a group that received the same informational content, and estimating the following specification:

$$Y_{ijt} = \beta_C \text{concordance}_i + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \quad (2)$$

where concordance_i is an indicator variable equal to 1 if the receiver i was sent a message in which the sender and the receiver share the same religion, and 0 otherwise. The parameter β_C captures the differential effect of receiving a religion-concordant doctor message as compared to a religion-discordant doctor message.¹⁶ It is therefore testing whether religion concordance, as compared to discordance, creates differences in the effects of the doctor messages estimated in equation (1). We note that this approach complements the pre-specified one, which proposed an interacted model, imposing that the main effect of religion concordance (i.e., the effect of sending a message with concordance independently from the content) is the same in the doctor message and in the control group. Results using this approach are in line with the ones presented in the main text, but less precise for some outcomes (Appendix D.4). Because in the final design of the experiment, the content in control messages is very different and significantly shorter than the one in the doctor messages, and assuming homogeneity of the main effect of religion concordance reduces precision, our preferred strategy remains that of presenting the results

¹⁶Appendix D.5 provides estimates of the effect of a Hindu versus a Muslim greeting, independently from the religion of the recipient. We observe no effect for these comparisons.

using equation (1) and equation (2) separately, assuming that the main effect of religion concordance is heterogeneous in the doctor and the control messages. In line, religion concordance has differential effects over the take-up of interventions depending on the content of the message (see Section 6.1).

We estimate both equation 1 and equation 2 by pooling data from the two follow-up surveys together, therefore estimating the average impact in the follow-up period (i.e., assuming β_D and β_C are constant over time). When outcome variables are measured in close temporal proximity, this approach allows averaging out the noise in the outcome variables and increasing power (McKenzie, 2012). Appendix D.4 shows results for each follow-up survey separately. Appendix D.1 shows how estimates vary in sub-samples defined by pre-specified variables (religion of the respondent, and % of Muslim living in the slum), and by other relevant dimensions (caste, strength of religious identity, trust in the government, and social desirability), which we discuss in the next section.

Because not everybody listens to the message that is sent to (Section 6.1 provides details about treatment compliance), as standard in mass information campaigns, we supplement main estimates with IV estimates that consider the actual exposure to the interventions. Using administrative data to compute the share of each message that is effectively listened on the phone, we estimate versions of equation 1 and equation 2 in which the treatment indicator is multiplied by this share. This interaction term captures the actual (endogenous) exposure to each treatment. We instrument this variable with the treatment indicators $doctor_i$ and $concordance_i$, respectively, and estimate the equations using two-stage least squares (2SLS). The parameters β_D and β_C in equation 1 and equation 2 can be interpreted as ITT effects (i.e., they capture the effect of sending a message, independently from whether a person listens to it). Conversely, IV estimates can be interpreted as ToT effects, which inform about the magnitude of the effects in presence of full compliance. In light of the likely heterogeneity in the (potential) impacts of the intervention, ToT estimates corresponds to the local average treatment effects for participants that comply with the intervention (see, for instance, Imbens and Angrist, 1994).

Finally, for statistical inference, we supplement in each table standard inference for the ITT estimates of equation 1 and equation 2 with multiple hypothesis testing adjusting p -values for the significance of each individual coefficient in the table using the List et al. (2019) bootstrap-based procedure. To this end, we categorize hypotheses by grouping variables in three groups and present the results in Section 6. First, in Section 6.1, we test whether the interventions impacted the take-up of the messages. Second, in Section 6.2, we test whether the interventions changed compliance with recommended practices and belief over the efficacy of preventive behavior. Third, in Section 6.3, we focus on whether interventions influenced the response of study participants to misinformation. Finally, in Section 6.4, we verify whether estimates are influenced by potential threats from spillover effects.

6 Results

6.1 Take-up of the campaign

Table 1 shows estimates of the effect of the doctor message and of religion concordance in the doctor message on the probability to have picked up the call, on the share and on the duration (in minutes) of the message that is listened to. These variables are computed from administrative data. Heterogeneity in these effects by the round of messages and by religion, a pre-specified heterogeneity dimension, are reported in Appendix D.1.

We begin by focusing on the effect of sending doctor messages versus control messages. In Panel A, we estimate equation (1) using the full sample of respondents. On average, 38.1% of respondents in the control group picked up the call at least once. Conditional on having picked up the call, they listened to the message for 0.55 minutes (33 seconds) or 67.4% of the message. Sending a doctor message did not shift the probability of picking up the call, but significantly decreased the share of the message that is listened to by 24.6 percentage points, while increasing the duration of the message listened to by 0.30 minutes (18 seconds). This seemingly counter-intuitive results can be explained by the fact that the doctor’s message is longer than the control message. Although the doctor’s message keeps the respondent on the phone for an additional 18 seconds, this extended duration does not result in a higher proportion of the message being listened to. Panel A in Figure 2 highlights these differences separately for the first and second round of messages.

In Panel B, we focus on the introduction of social proximity with the sender in the doctor message and estimate treatment effects using equation (2). Since we do not include the control group in this estimation, the length of the message is the same across groups. On average, 37.7% of respondents that received a doctor message with an introduction from a different religion picked up the call and, conditional on having picked up the call, they listened to 39.8% of the message, corresponding to 0.79 minutes (47 seconds). The impact of religion concordance changes exposure to the doctor message significantly. For one, we find that the share of respondents that picked up the call is reduced by 2.9 percentage points (a decrease by 7.7% over the mean for messages with religion discordance). While ex-ante people would not know about the source of the call, and thus one would not expect any difference across treatment groups, we show in Appendix D.1 that this reduced probability of picking up the phone is driven by the second call. It is therefore likely, that some respondents recognized the number and decided not to pick up again.

Importantly, conditional on having picked up the call, religion concordance leads to a significantly larger exposure to the doctor message. The share of the message that is listened to increases by 5.3 percentage points, corresponding to an additional 0.12 minutes (7 seconds). These effects, corresponding to an increase by 13.3% and 15.3% over the means for messages with religion discordance, are specific to the doctor message. In fact, in Appendix D.2, we show that religion concordance in the control messages had no effect on the probability to take-up the call, nor in the share of the message that is listened. We

conclude that it is the combination of religion concordance with relevant informational content that is driving respondents to listen for longer to the information campaign.

We observe that it is a full shift in the distribution of listening time that is driving these results. This highlights the importance of not only the very first seconds of the call, when the sender introduces the message, but also the content that follows the introduction. Panel B in Figure 2 presents the distribution of the share of each message that is listened to by study participants in the presence of religion-concordance or -discordance for both the control group (left figure) and the doctor message group (right figure). While we do not observe any difference in the control message, we observe a difference in the distributions for the doctor message group.

6.2 Compliance with preventive practices and beliefs about their efficacy

We first focus on the effect of sending doctor messages versus control messages (Table 2) before turning to impacts of sending a doctor message that is religion-concordant (Table 3). Panel A presents ITT estimates and Panel B shows IV estimates of the effect of doctor messages on compliance with preventive practices and on the beliefs about their efficacy in fighting COVID-19. In column (1), we focus on compliance with recommended practices using the index that aggregates different indicators of preventive behavior (see Section 4). In columns (2)–(5), we focus on the respondents’ beliefs over the efficacy of different preventive practices, in columns (2)–(3) on recommended practices, and in columns (4)–(5) on non-evidence-based practices, such as relying on vegetarianism or on the Indian immunity to the virus. Sending the doctor messages increases significantly the compliance with recommended practices by 0.06 standard deviations relative to the control group percentage points. This effect is driven by increases in both hand-washing and in physical distancing (Appendix D.3), indicating that doctor messages were effective at promoting recommended practices to avoid contagion. This effect is homogeneous by the level of social desirability of the respondent (Appendix D.1), indicating that experimenter demand effects are not driving this result.

The increase in compliance with recommended practices is accompanied by changes in beliefs over the efficacy of evidence-based practices. We find a significant increase in agreement with using face masks and practicing hand-washing to protect against the virus by 0.6 percentage points (0.75% over the control mean), while agreement with social distancing also increased, though not significantly. This result may be influenced by the constraints of living in overcrowded spaces, as is the case in the slums where the study was conducted. We do not observe any effect on the beliefs over the efficacy of non-evidence based practices at conventional significance levels. Inference for these effects is robust to multiple hypothesis testing at standard confidence levels.¹⁷ Estimates increase significantly in magnitude when considering IV estimates (Panel B). Listening to the full doctor message increases compliance with recommended practices by 0.36 standard deviations, and by 4.1 percentage points in the level of agreement with using

¹⁷While the effect on beliefs over the efficacy of face masks and hand-washing are stronger in the first follow-up round, compliance with recommended practices is significantly affected in both rounds (Appendix D.4).

face masks and practicing hand-washing (corresponding to a 5.0% increase relative to the control mean). Overall, while the effects on compliance are large in magnitude, the effects on beliefs are either absent or relatively small.

Table 3 focuses on the effect of sending a doctor message that is religion-concordant, as compared to religion-discordant. Religion concordance in the doctor message increases the compliance with recommended practices by 9.4 percentage points. Again, inference for this effect is robust to multiple hypothesis testing and the effect is not driven by experimenter demand effects as it is homogeneous with respect to the level of social desirability of the respondent (Appendix D.1). Because this effect is almost double the estimate of the effect of the doctor message, it indicates that the efficacy of doctor messages in promoting compliance is almost uniquely driven by messages in which the sender and the receiver share the same religion. This result is confirmed by estimating the effects with an interaction model (Appendix D.4). This finding is possibly due to the fact that these receivers listen to a larger proportion of the message (Section 6.1), and/or associate a stronger importance to the message. IV estimates indicate a large magnitude of the effect when the whole message is listened by the receiver, leading to an increase in compliance with recommended practices by 0.57 standard deviations.

As compared to religion discordance, religion concordance in the doctor messages does not alter beliefs over the efficacy of recommended practices, but it does reduce agreement with non-evidence-based practices. We observe a reduction by 1.6 percentage points in the agreement with vegetarianism being a way to prevent contagion, an effect that is significant only at the 15% level when corrected for multiple hypothesis testing. The magnitude of this effect is larger as compared to the effect on beliefs of doctor messages alone, as it corresponds to a reduction in beliefs by 2.8% over the mean for religion-discordant doctor messages.

These effects are not driven by changes in risk perceptions about the contagion with COVID-19, which is unaffected by the interventions (Appendix D.6). In addition, the placebo test confirms that the effects of religion concordance are specific to the doctor message; i.e., similarly to the case of the take-up of messages, we observe no differential effect of religion concordance in the control group (Appendix D.2). Although the campaign influences behavior by shaping beliefs about the efficacy of recommended practices, we also find that it is largely ineffective in altering beliefs regarding non-evidence-based preventive practices. These unproven practices continue to persist among the study population.

6.3 Response to misinformation

Results on behavioral outcomes indicate that the informative content in information campaigns is more effective at shifting compliant behavior especially when there is social proximity between the sender and the receiver. In this section, we focus on whether the campaign was also effective at inoculating against misinformation.

We begin by studying whether sending doctor messages achieves this end as compared to sharing gossips in the control message. Table 4 presents the results. In column (1), fact-checking is measured as an

indicator variable equal to 1 if the respondent frequently checks the truthfulness of the information he/she shares or discusses, and zero otherwise. In columns (2)–(3), we focus on the level of agreement with misinformation shared by in-group citizens and by out-group citizens. The elicitation procedure for these outcomes is described in Section 4. Panel A presents ITT estimates, while Panel B provides IV estimates of ToT effects.

Sending messages from a trusted source, in this case doctors, crowd-out fact-checking.¹⁸ The inclination of respondents to verify information shared and discussed with family and friends decreases significantly by 2.3 percentage points (6.5% over the control mean), and remains significant after adjusting p-values for multiple hypothesis testing. This effect, which is driven by the second round of data collection (Appendix D.4), translates in a ToT estimate of 14.7 percentage points across survey rounds when the respondent listens to the full doctor message (or 41.8% over the control mean). Perhaps this crowding out happens because individuals, having heard the messages from doctors, feel more confident in dismissing misinformation. Reductions in fact checking following the doctor message are slightly higher, but not robust to multiple hypothesis testing, when respondents were incentivized with the higher lottery amount (Appendix D.5), potentially driven by participants paying closer attention to the campaign.

In terms of agreement with misinformation shared by other citizens, we observe that doctor messages do not impact this dimension when shared by in-group citizens, but it does reduce agreement when misinformation is shared by out-group citizens. Doctor messages lead to a significant reduction in agreement by 1.6 percentage points (an effect of 3.2% over the control mean) when the statement is made by citizen of a different religion. This effect is robust to multiple hypothesis testing and corresponds to a ToT estimate of 10.5 percentage points reduction when the doctor message is listened fully, corresponding to a reduction of by 21.3% over the control mean. Thanks to the design of the survey instrument, the content of the statements used to measure how respondents react to misinformation is orthogonal to the citizen sharing it being in- or out-group (i.e., statements are constant, while the citizen varies exogenously). These results highlight how measuring impacts on the response to misinformation requires considerations of social norms and group identity. The information campaign promoted by the doctor messages is effective at protecting against misinformation carrying little group identity (i.e., shared by citizens of a different religious group). However, doctor messages alone are ineffective at protecting against misinformation when carrying stronger group identity (i.e., shared by citizens of the same religious group of the respondent).

We then turn the attention to whether sending a doctor message with religion concordance or discordance impacts the response to misinformation. Table 5 provides estimates of these effects by restricting the sample to the recipients of the doctor message. Religion concordance does not introduce, on average, any significant differential effect for fact-checking nor agreement with misinformation shared by out-group citizens. However, religion concordance inoculates against misinformation shared by citizens with the

¹⁸The interventions have no effect on the level of trust. We do not find any effect in reported levels of trust in information shared by different groups, including doctors and health experts and other citizens of UP (Appendix D.6).

same identity. While the doctor message decreases agreement with misinformation shared by out-group citizens, it is only in presence of religion concordance that a doctor message influences also agreement with misinformation reported by in-group citizens. When the doctor message is introduced by a religion-concordant greeting, agreement with this type of misinformation is reduced by 2.3 percentage points as compared to a doctor message introduced by a religion-discordant greeting (corresponding to a decrease by 4.6% over the mean for religion-discordant messages). This effect is highly significant and robust to multiple hypothesis testing. The magnitude of the ToT estimate is a reduction of 15.0 percentage points in agreement after listening fully to a doctor message introduced by a religion-concordant greeting (corresponding to a decrease by 30.2% over the mean for religion-discordant messages). On the contrary, religion concordance does not shift further disagreement with misconceptions reported by out-group citizens as compared to the main effect provided the informative content.

Similar to the effects presented in Section 6.1 and Section 6.2, these effects are specific to the combination of religion concordance with a doctor message, as receiving the religion-concordant greeting with the control message does not affect agreement with any of the variables presented in Table 5 (Appendix D.2).

To verify whether these effects are specific to misinformation, we present a placebo test by estimating treatment effects on agreement with a different type of statement shared by citizens. We focus on opinions related to COVID-19 rather than misinformation because these are harder to be influenced via fact-checking (see, for instance, [Walter and Salovich, 2021](#)). Appendix D.6 show that doctor messages, with or without religion concordance, have no impact on opinions, independently from whether these are reported by an in-group interlocutor or an out-group interlocutor. This finding reinforces that the pattern of effects observed is specific to misinformation about COVID-19. It also suggests that the limited effectiveness of the information campaign in influencing beliefs over the efficacy of non-evidence-based practices might be related to non-factual opinions, which are more persistent and harder to be influenced by information campaigns.

In summary, protection against misinformation can be more effectively achieved through informative content shared by sources that are trusted. However, in order to fully safeguard against misinformation and break the connection between beliefs and group identity, we need to factor in social proximity in information campaigns. Only in the presence of religion concordance, the agreement with misinformation shared by both in- and out-group citizens is reduced by the informative content of the campaign. In the absence of concordance, the reduction in disagreement with misconceptions occurs solely when the misinformation originates from out-group citizens.

6.4 Information spillovers

Randomization into the experimental arms is conducted at the household level because the intervention is directed one-to-one through mobile phones, and we wanted to prevent informational spillovers within households. The interpretation of the estimates of treatment effects discussed in Sections 6.1–6.3 would

be affected by the presence of information spillovers (see, for example, [Vazquez-Bare, 2022](#)). Spillover effects are mitigated by the voice messages being automatic calls that cannot be forwarded or shared, but there remains the possibility of word-of-mouth information sharing, particularly within communities. While this study was not specifically designed to capture spillover effects in information campaigns, we test for their presence by leveraging variation in intervention exposure across slums, induced by the household-level randomization not stratified at slum level. The availability of precise geo-location on slum borders, as well as where each household resides, allows us to measure the share of households living in the same slum as the respondent that is allocated to the doctor message group, and, conditional on being allocated to the doctor message group, the share that also receives a religion-concordant message.¹⁹ By design, the probability of neighbors being in each of these groups is on average 0.5. However, household-level randomization allows for random variation in this probability across respondents. Figure 3 shows the distribution of these variables by whether the respondent took-up the intervention. The distributions confirm not only the random pattern of treatment allocation among neighbors, but also the similarity of this pattern across respondents that did and did not pick up the intervention call. Kolmogorov-Smirnov tests fail to reject the equality of the distributions along this dimension. Exploiting this variation, we estimate both equation (1) and equation (2) controlling for this measure of ‘neighbor treatment’. Rejecting the null hypothesis of a zero coefficient for this measure indicates the presence of information spillovers. Table 6 presents the results. The estimates of treatment effects discussed in Sections 6.1–6.3 are unaffected by controlling for the treatment allocation among neighbors. In addition, the effect of treatment allocation among neighbors is not statistically significantly different from zero for most of the outcomes, indicating a limited importance of community-level information sharing. These results highlight that, despite interventions having the potential to spread information across individuals, community-level spillovers do not play a central role, and we can interpret main results as consistent estimates of causal effects of the intervention campaign.

7 Conclusions

Ensuring that misinformation and fake news do not overshadow scientific evidence and mislead individuals is of global importance for public health. We demonstrate that information campaigns based on trusted sources like doctors and using low-cost interventions that rely on mobile phone technology can be effectively deployed to improve preventive behavior in low-income settings. In particular, we show that increasing social proximity between the campaign sender and the receiver can generate beneficial effects. This is evident not only in improved preventive behavior, but also in the recipient’s increased ability to effectively respond to misinformation. Religion concordance in the information campaign inoculated individuals against misinformation stated by in-group citizens.

These findings open new avenues for future research to explore both the effectiveness of information

¹⁹Results using the treatment allocation of the nearest neighbor are in line (Appendix D.8).

campaigns and the role of social proximity in decision-making. In particular, while this study focuses on religion as a dimension of social proximity, future research could delve into other social dimensions such as ethnicity, community ties, or socio-economic status. Understanding how various social factors influence information dissemination can guide more comprehensively the design of information campaigns.

Understanding how social proximity interacts with information campaigns and health-related behaviors offers opportunities for targeted policy interventions. Policymakers can leverage these insights to create more effective and culturally attuned campaigns, thereby enhancing public health outcomes across diverse communities. In particular, policymakers should consider incorporating social proximity into the design of information campaigns, ensuring that messages resonate based on the audience's social connections. At the same time, in light of the ongoing challenges posed by misinformation, policy interventions should aim not only at disseminating accurate information but also at effectively countering misinformation. This study underscores the potential of mobile-based campaigns as effective tools in low-income areas, offering scalable and cost-effective methods for widespread information dissemination.

References

- Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics* 115(3), 715–753.
- Al Jazeera (2020). Misinformation, fake news spark India coronavirus fears. Online; accessed on 2021-07-13.
- Alesina, A. and E. L. Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature* 43(3), 762–800.
- Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2), 211–236.
- Alsan, M. and S. Eichmeyer (2021). Experimental Evidence on the Effectiveness of Non-Experts for Improving Vaccine Demand. National Bureau of Economic Research (No. w28593).
- Alsan, M., O. Garrick, and G. Graziani (2019). Does diversity matter for health? Experimental evidence from Oakland. *American Economic Review* 109(12), 4071–4111.
- Alsan, M., S. Stantcheva, D. Yang, and D. Cutler (2020). Disparities in coronavirus 2019 reported incidence, knowledge, and behavior among us adults. *JAMA Network Open* 3(6), e2012403–e2012403.
- Armand, A., B. Augsburg, and A. Bancalari (2020). COVID-19, fake news and religious tensions: experimental evidence from India. AEA RCT Registry, October 08. <https://doi.org/10.1257/rct.6564-1.0>.

- Armand, A., B. Augsburg, A. Bancalari, and M. Ghatak (2023). Public Service Delivery, Exclusion and Externalities: Theory and Experimental Evidence from India. CEPR discussion paper DP18636, Centre for Economic Policy and Research.
- Atkin, D., E. Colson-Sihra, and M. Shayo (2021). How do we choose our identity? A revealed preference approach using food consumption. *Journal of Political Economy* 129(4), 1193–1251.
- Awasthi, P. (2020). UP seals off hotspots in 15 districts, including Lucknow, Noida to contain spread. *The Week*.
- Azrieli, Y., C. P. Chambers, and P. J. Healy (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy* 126(4), 1472–1503.
- Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review* 115(4), 1325–41.
- Baicker, Katherine, S. M. and J. Schwartzstein (2015). Behavioral Hazard in Health Insurance. *Quarterly Journal of Economics* 130(4), 1623–6.
- Banerjee, A., M. Alsan, E. Breza, A. G. Chandrasekhar, A. Chowdhury, E. Duflo, P. Goldsmith-Pinkham, and B. A. Olken (2020, July). 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. NBER Working Paper 27496, National Bureau of Economic Research.
- Banerjee, A., M. Faye, A. Krueger, P. Niehaus, and T. Suri (2020). Effects of a Universal Basic Income during the pandemic. Working Paper, UC San Diego.
- Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics* 182, 104123.
- Bavel, J. J., K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman, J. Drury, O. Dube, N. Ellemers, E. J. Finkel, J. H. Fowler, M. Gelfand, S. Han, S. A. Haslam, J. Jetten, S. Kitayama, D. Mobbs, L. E. Napper, D. J. Packer, G. Pennycook, E. Peters, R. E. Petty, D. G. Rand, S. D. Reicher, S. Schnall, A. Shariff, L. J. Skitka, S. S. Smith, C. R. Sunstein, N. Tabri, J. A. Tucker, S. van der Linden, P. van Lange, K. A. Weeden, M. J. Wohl, J. Zaki, S. R. Zion, and R. Willer (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* 4(5), 460–471.
- Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019). Unity in diversity? How intergroup contact can foster nation building. *American Economic Review* 109(11), 3978–4025.
- Belloni, A., V. Chernozhukov, and C. Hansen (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies* 81(2), 608–650.
- Bhalotra, S., I. Clots-Figueras, G. Cassan, and L. Iyer (2014). Religion, politician identity and develop-

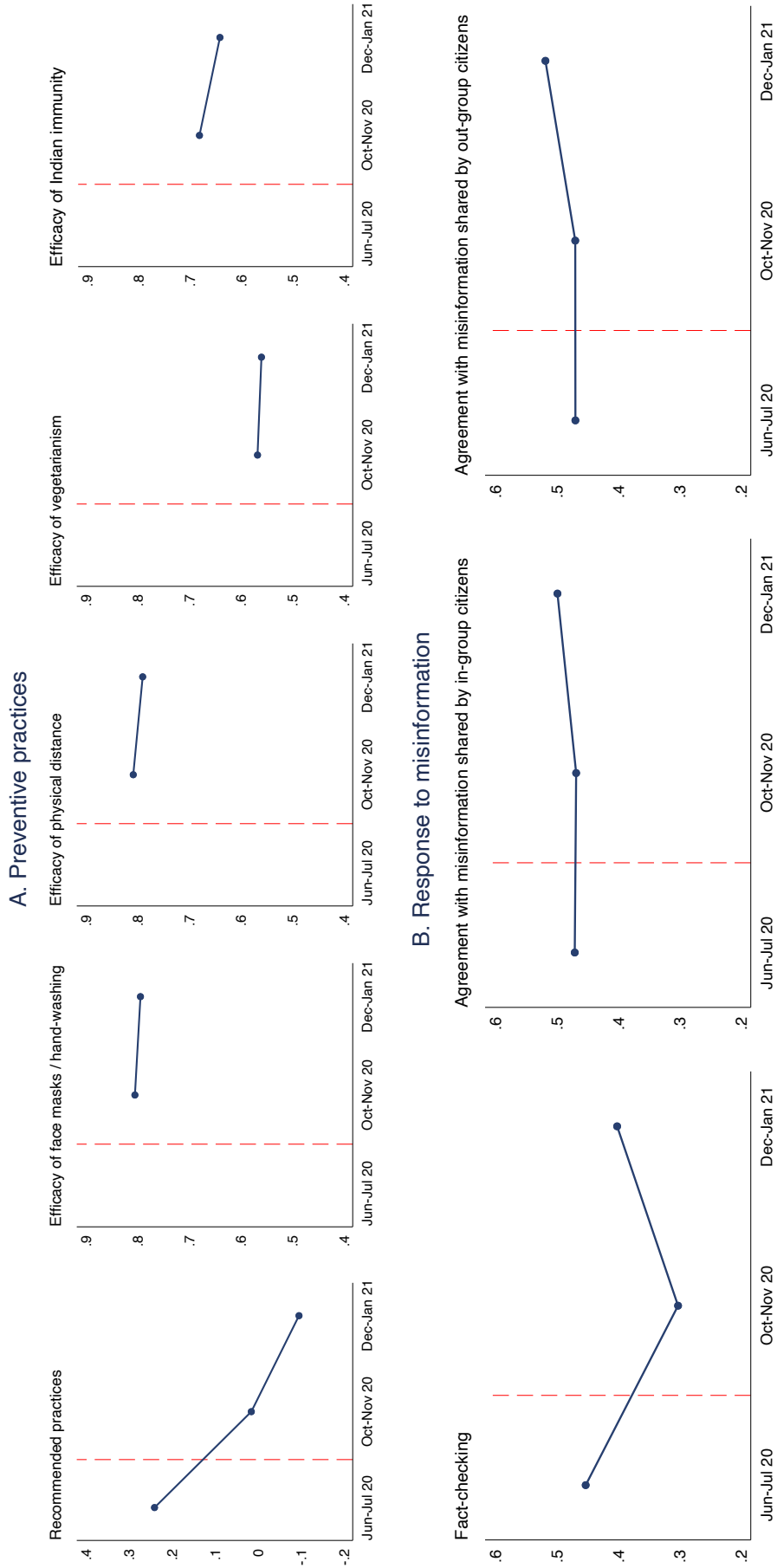
- ment outcomes: Evidence from india. *Journal of Economic Behavior & Organization* 104, 4–17.
- Bicchieri, C., E. Dimant, S. Gaechter, and D. Nosenzo (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior* 132, 59–72.
- Bowles, J., H. Larreguy, and S. Liu (2020). Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe. *PloS one* 15(10), e0240005.
- Breza, E., F. C. Stanford, M. Alsan, B. Alsan, A. Banerjee, A. G. Chandrasekhar, S. Eichmeyer, T. Glushko, P. Goldsmith-Pinkham, K. Holland, and Others (2021). Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial. *Nature Medicine*, 1–7.
- Bursztyjn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2023). Opinions as Facts. *Review of Economic Studies* forthcoming.
- Chen, Y. and S. X. Li (2000). Group Identity and Social Preferences. *American Economic review* 99(1), 431–457.
- DellaVigna, S. and E. Kaplan (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics* 122(3), 1187–1234.
- Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics* 3(1), 425–449.
- Enikolopov, R., A. Makarin, and M. Petrova (2020). Social Media and Protest Participation: Evidence from Russia. *Econometrica* 88(4), 1479–1514.
- Facebook’s Crowd Tangle Team (2020). CrowdTangle.
- Fischer, D. G. and C. Fick (1993). Measuring social desirability: Short forms of the marlowe-crowne social desirability scale. *Educational and Psychological Measurement* 53(2), 417–424.
- Fisman, R., D. Paravisini, and V. Vig (2017). Cultural proximity and loan outcomes. *American Economic Review* 107(2), 457–92.
- Fisman, R., A. Sarkar, J. Skrastins, and V. Vig (2020). Experience of communal conflicts and intergroup lending. *Journal of Political Economy* 128(9), 3346–3375.
- Government of India (2011). Census of India 2011.
- Government of India (2021). Orders prescribing lockdown for containment of COVID-19 Epidemic in the country and guidelines for phased re-opening (unlocks). <https://www.mha.gov.in/sites/default/files/PRNationalLockdown26032020.pdf>.
- Greenwood, B. N., S. Carnahan, and L. Huang (2018). Patient–physician gender concordance and increased mortality among female heart attack patients. *Proceedings of the National Academy of Sciences of the United States of America* 115(34), 8569–8574.

- Greenwood, B. N., R. R. Hardeman, L. Huang, and A. Sojourner (2020). Physician-patient racial concordance and disparities in birthing mortality for newborns. *Proceedings of the National Academy of Sciences of the United States of America* 117(35), 21194–21200.
- Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of economic literature* 61(1), 3–40.
- Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review* 101(4), 709–725.
- Hill, A., D. Jones, and L. Woodworth (2020). Physician-Patient Race-Match and Patient Outcomes. SSRN Working Paper.
- Himelein, K., S. Eckman, and D. McKenzie (2020). Mobile Phone Surveys for Understanding COVID-19 Impacts: Part II Response, Quality, and Questions. Development Impact Blog.
- Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2), 467–475.
- Iyer, S. (2016). The new economics of religion. *Journal of Economic Literature* 54(2), 395–411.
- Jha, S. (2013). Trade, institutions, and ethnic tolerance: Evidence from south asia. *American political Science review* 107(4), 806–832.
- Khan, A., S. Nasim, M. I. Shaukat, and A. Stegmann (2021). Building Trust in the State with Information: Evidence from Urban Punjab. *Journal of Public Economics*.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- Kremer, M., G. Rao, and F. Schilbach (2019). Behavioral development economics. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 2, pp. 345–458. Elsevier.
- Lazer, D. M., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain (2018). The science of fake news: Addressing fake news requires a multidisciplinary effort. *Science* 359(6380), 1094–1096.
- Lilford, R., O. Oyeboode, D. Satterthwaite, G. Melendez-Torres, Y.-F. Chen, B. Mberu, and et al. (2017). Improving the health and welfare of people who live in slums. *Lancet* 389, 559–70.
- List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics* 22(4), 773–793.
- Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review* 111(6), 1807–1844.
- Manacorda, M. and A. Tesei (2020). Liberation Technology: Mobile Phones and Political Mobilization

- in Africa. *Econometrica* 88(2), 533–567.
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics* 99(2), 210–221.
- Mitra, A. and D. Ray (2014). Implications of an economic theory of conflict: Hindu-Muslim violence in India. *Journal of Political Economy* 122(4), 719–765.
- Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science* 369(6505), 866–870.
- Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences* 118(15), e1912440117.
- O’Keefe, D. J. (2016). *Persuasion: Theory and Research* (3 ed.). Thousand Oaks, CA: SAGE Publications, Inc.
- Patel, A. (2020). Preventing COVID-19 Amid Public Health and Urban Planning Failures in Slums of Indian Cities. *World Medical and Health Policy* 12(3), 266–273.
- Pauly, M. V. and F. E. Blavin (2008). Moral Hazard in Insurance, Value-Based Cost Sharing, and the Benefits of Blissful Ignorance. *Journal of Health Economics* 27(6), 1407–17.
- Philpott, D. (2007). Explaining the Political Ambivalence of Religion. *American Political Science Review* 101(3), 505–525.
- Reserve Bank of India (2019). State-wise Poverty Rate. Technical report, National Sample Survey Organisation, NITI Aayog, Delhi, India.
- Sadish, D., A. Adhvaryu, and A. Nyshadham (2021). (Mis)information and anxiety: Evidence from a randomized Covid-19 information campaign. *Journal of Development Economics* 152(102699).
- Scheufele, D. A. and N. M. Krause (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America* 116(16), 7662–7669.
- Siddique, A., T. Rahman, D. Pakrashi, A. Islam, and F. Ahmed (2022, 03). Raising Health Awareness in Rural Communities: A Randomized Experiment in Bangladesh and India. *The Review of Economics and Statistics*, 1–45.
- Sinding Bentzen, J. (2019). Acts of God? Religiosity and natural disasters across subnational world districts. *The Economic Journal* 129(622), 2295–2321.
- Solís Arce, J. S., S. S. Warren, N. F. Meriggi, A. Scacco, N. McMurry, M. Voors, G. Syunyaev, A. A. Malik, S. Aboutajdine, O. Adejo, et al. (2021). COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nature medicine* 27(8), 1385–1394.
- Statista (2021). Main sources of news across India in 2019, by age.
- Tabellini, G. (2008). The Scope of Cooperation: Values and Incentives. *Quarterly Journal of Eco-*

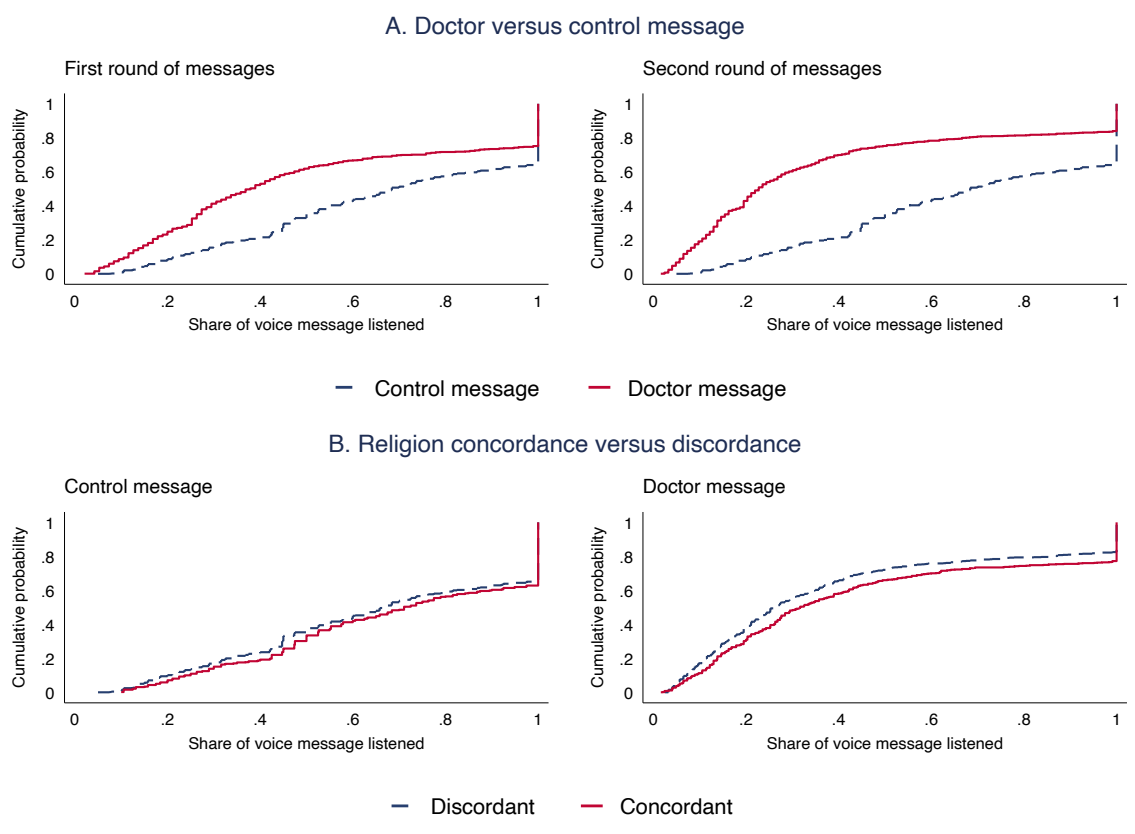
- nomics* 123(3), 905–50.
- Tankard, M. E. and E. L. Paluck (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review* 10(1), 181–211.
- The Guardian (2020). Coronavirus conspiracy theories targeting Muslims spread in India. The Guardian.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–288.
- Torres, C., L. Ogbu-Nwobodo, M. Alsan, F. C. Stanford, A. Banerjee, E. Breza, A. G. Chandrasekhar, S. Eichmeyer, M. Karnani, T. Loisel, P. Goldsmith-Pinkham, B. A. Olken, P. L. Vautrey, E. Warner, and E. Duflo (2021). Effect of Physician-Delivered COVID-19 Public Health Messages and Messages Acknowledging Racial Inequity on Black and White Adults’ Knowledge, Beliefs, and Practices Related to COVID-19: A Randomized Clinical Trial. *JAMA Network Open* 4(7), e2117115–e2117115.
- United Nations (2019). World Urbanization Prospects: The 2018 Revision. United Nations, Department of Economic and Social Affairs, Population Division, New York, US.
- Van der Linden, S., A. Leiserowitz, S. Rosenthal, and E. Maibach (2017). Inoculating the Public against Misinformation about Climate Change. *Global Challenges* 1(2), 1600008.
- Vazquez-Bare, G. (2022). Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*.
- Walter, N. and N. A. Salovich (2021). Unchecked vs. uncheckable: How opinion-based claims can impede corrections of misinformation. *Mass Communication and Society* 24(4), 500–526.
- Wasdani, K. P. and A. Prasad (2020). The impossibility of social distancing among the urban poor: the case of an Indian slum in the times of COVID-19. *Local Environment* 25(5), 414–418.
- World Health Organisation (2020). COVID-19 Response. Seventy-third world health assembly (a73.conf./1 rev. 1), World Health Organization.
- Yasir, S. (2020). India is scapegoating Muslims for the spread of the Coronavirus. *Foreign Policy*.
- Yom-Tov, E., J. Shembekar, S. Barclay, and P. Muennig (2018). The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. *NPJ Digital Medicine* (24).
- Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual Review of Economics* 12, 415–438.

Figure 1: Preventive practices and response to misinformation (control group)



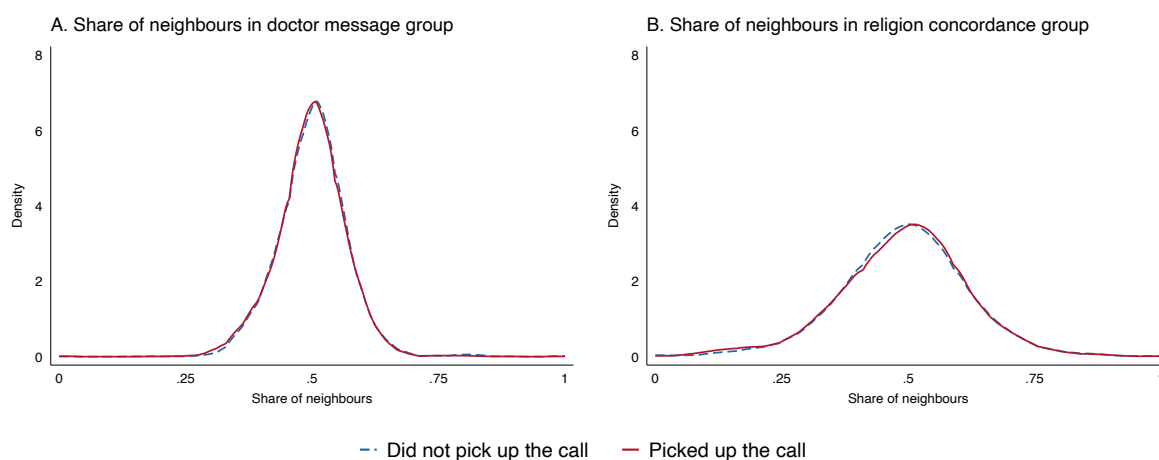
Notes. Each figure shows the average of outcome variables measured at different points in time. The vertical line separates the baseline measurement from the follow-up measurements. In both panels, the sample is restricted to the control group that did not receive the doctor message. Further details about these variables are provided in Section 4. Variables are defined in Appendix B.

Figure 2: Share of voice messages listened by study participants



Note. The figures show the share of the messages listened by study participants, conditional on having picked up the call. Information is based on administrative data from the intervention. Panel A includes the full sample separated by round of intervention, panel B restricts the sample to the control message group in the left figure and to the doctor message group in the right figure. Treatment effects on the take-up of messages are reported in Table 1. The duration of the call can be larger than the duration of the recorded message if the receiver spends time to reply to the question at the end of the message.

Figure 3: Treatment allocation among neighbours, by respondent's group



Notes. The figures show the distribution of the share of households living in the same slum of the respondent that are allocated to the doctor message group (Panel A) or to the religion concordance group (Panel B), depending on whether the respondent picked up or did not pick up the intervention call. In panel B, the sample is restricted to the doctor message group. Distributions are estimated non-parametrically using kernel density estimation, assuming an Epanechnikov kernel function with a bandwidth of 0.02 in Panel A and 0.04 in Panel B. The p-values of Kolmogorov-Smirnov tests of equality of distributions are 0.18 in Panel A and 0.75 in Panel B.

Table 1: Treatment effects on the take-up of messages

	Picked up (1)	% listened (2)	Duration (minutes) (3)
A. Full sample			
Doctor message	-0.016 (0.013) [0.21, 0.21]	-0.246 (0.014) [0.00, 0.00]	0.302 (0.027) [0.00, 0.00]
Mean (control message)	0.381	0.674	0.551
Observations	7700	2873	2873
B. Sample restricted to doctor message group			
Religion concordance	-0.029 (0.016) [0.09, 0.09]	0.053 (0.021) [0.01, 0.02]	0.121 (0.047) [0.01, 0.02]
Mean (religion discordance)	0.377	0.398	0.790
Observations	3851	1406	1406

Notes. Estimates based on OLS regressions using equation (1) in Panel A and equation (2) restricting the sample to participants allocated to the doctor message in Panel B (see Section 5). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in each panel. The dependent variables are: in column (1) *Picked up* is an indicator variable equal to 1 if the respondent picked up the call in any of the two rounds of interventions, and 0 otherwise; in column (2) *% listened* is the share of the message that is listened, conditional on having picked up; in column (3) *Duration (minutes)* is the duration of the call, conditional on having picked up. Note that the doctor messages have different duration as compared to the control messages (see Section 3). All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 2: Preventive practices: doctor versus control message

	Compliance		Beliefs over the efficacy of...		
	Recommended practices (1)	Face masks / hand-washing (2)	Physical distancing (3)	Vegetarianism (4)	Indian immunity (5)
A. OLS					
Doctor message	0.057 (0.022) [0.01, 0.06]	0.006 (0.003) [0.01, 0.06]	0.005 (0.004) [0.20, 0.36]	0.002 (0.006) [0.80, 0.80]	-0.007 (0.005) [0.13, 0.34]
B. IV					
% listened · doctor message	0.358 (0.135) [0.01]	0.041 (0.016) [0.01]	0.031 (0.024) [0.19]	0.010 (0.038) [0.80]	-0.047 (0.030) [0.13]
Mean (control message)	-0.041	0.799	0.799	0.563	0.661
Observations	5079	7700	7698	7692	7697

Notes. Estimates based on OLS regressions using equation (1) in Panel A and on 2SLS regressions in Panel B (see Section 5). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. In Panel A, the first value is from individual testing, the second is adjusted for testing that each treatment is jointly different from zero for all outcomes presented in the panel. The dependent variables are: in column (1) *Recommended practices* is an index capturing adherence to WHO's recommendations to protect from infection, built using the procedure of Kling et al. (2007) described in Section 4; column (2) *Face masks and hand-washing* concerns the average level of agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer; column (3) *Physical distancing* concerns keeping physical distance with other people; column (4) *Vegetarianism* concerns the level of agreement with relying on eating a vegetarian diet; column (5) *Indian immunity* concerns the level of agreement with relying on the Indian immune system. The level of agreement in columns (1)–(4) is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 3: Preventive practices: the effect of religion concordance in the doctor message

	Compliance		Beliefs over the efficacy of...		
	Recommended practices (1)	Face masks / hand-washing (2)	Recommended practices Physical distancing (3)	Non-evidence-based practices Vegetarianism (4)	Indian immunity (5)
A. OLS					
Religion concordance	0.094 (0.032) [0.00 , 0.02]	-0.004 (0.004) [0.34 , 0.55]	-0.006 (0.004) [0.20 , 0.46]	-0.016 (0.008) [0.04 , 0.15]	-0.001 (0.007) [0.85 , 0.86]
B. IV					
% listened · religion concordance	0.572 (0.198) [0.00]	-0.023 (0.023) [0.33]	-0.035 (0.027) [0.20]	-0.102 (0.051) [0.04]	-0.008 (0.042) [0.85]
Mean (religion discordance)	-0.035	0.807	0.806	0.571	0.654
Observations	2519	3851	3849	3846	3849

Notes. The sample is restricted to respondents in the doctor message group. Estimates based on OLS regressions using equation (1) in Panel A and on 2SLS regressions in Panel B (see Section 5). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. In Panels A and C, the first value is from individual testing, the second is adjusted for testing that each treatment is jointly different from zero for all outcomes presented in each panel. The dependent variables are: in column (1) *Recommended practices* is an index capturing adherence to WHO’s recommendations to protect from infection, built using the procedure of Kling et al. (2007) described in Section 4; column (2) *Face masks and hand-washing* concerns the average agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer; column (3) *Physical distancing* concerns keeping physical distance with other people; column (4) *Vegetarianism* concerns the average agreement with relying on eating a vegetarian diet; column (5) *Indian immunity* concerns the average agreement with relying on the Indian immune system. The level of agreement in columns (1)–(4) is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 4: Response to misinformation: doctor versus control message

	Fact-checking	Agreement with misinformation shared by...	
	(1)	In-group citizens (2)	Out-group citizens (3)
A. OLS			
Doctor message	-0.023 (0.010) [0.03 , 0.06]	0.004 (0.007) [0.56 , 0.55]	-0.016 (0.006) [0.01 , 0.01]
B. IV			
% listened · doctor message	-0.147 (0.066) [0.03]	0.027 (0.046) [0.55]	-0.105 (0.037) [0.01]
Mean (control message)	0.352	0.483	0.494
Observations	7700	5182	6704

Notes. Estimates based on OLS regressions using equation (1) in Panel A and on 2SLS regressions in Panel B (see Section 5). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. In Panel A, the value first is from individual testing, the second is adjusted for testing that each treatment is jointly different from zero for all outcomes presented in the panel. The dependent variables are: in column (1) *Fact-checking* is an indicator variable equal to 1 if the respondent always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; in columns (2)–(3) *Agreement with misinformation shared by [...]* is the average level of agreement with statements including incorrect views based on faulty knowledge or understanding, where 0 refers to strongly disagree and 1 refers to strongly agree. In column (2), the outcome variables include only statements from an interlocutor with the same religion of the respondent. In column (3), the outcome variables include only statements from an interlocutor with a religion different from the one of the respondent or from the generic term “people”. Individual statements and categorization are described in Appendix A.1. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 5: Response to misinformation: the effect of religion concordance in the doctor message

	Fact-checking	Agreement with misinformation shared by...	
	(1)	In-group citizens	Out-group citizens
	(1)	(2)	(3)
A. OLS			
Religion concordance	0.006 (0.015) [0.69 , 0.69]	-0.023 (0.009) [0.01 , 0.03]	0.006 (0.008) [0.46 , 0.72]
B. IV			
% listened · religion concordance	0.037 (0.091) [0.69]	-0.150 (0.059) [0.01]	0.037 (0.051) [0.46]
Mean (religion discordance)	0.326	0.497	0.476
Observations	3851	2587	3340

Notes. The sample is restricted to respondents in the doctor message group. Estimates based on OLS regressions using equation (1) in Panel A and on 2SLS regressions in Panel B (see Section 5). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. In Panel A, the first value is from individual testing, the second is adjusted for testing that each treatment is jointly different from zero for all outcomes presented in the panel. The dependent variables are: in column (1) *Fact-checking* is an indicator variable equal to 1 if the respondent always or very frequently check the truthfulness of information shared or discussed with family and friends, and 0 otherwise; in columns (2)–(3) *Agreement with misinformation shared by [...]* is the average level of agreement with statements including incorrect views based on faulty knowledge or understanding, where 0 refers to strongly disagree and 1 refers to strongly agree. In column (2), the outcome variables include only statements from an interlocutor with the same religion of the respondent. In column (3), the outcome variables include only statements from an interlocutor with a religion different from the one of the respondent or from the generic term “people”. Individual statements and categorization are described in Appendix A.1. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table 6: Spillover effects of the interventions

	Compliance		Beliefs over the efficacy of ...			Fact-checking		Agreement with misinformation shared by...	
	Recommended practices (1)	Face masks / hand-washing (2)	Recommended practices (3)	Vegetarianism (4)	Indian immunity (5)	In-group citizens (6)	In-group citizens (7)	Out-group citizens (8)	
A. Full sample									
Doctor message	0.062 (0.022) [0.01, 0.03]	0.008 (0.003) [0.01, 0.03]	0.006 (0.004) [0.18, 0.45]	0.001 (0.007) [0.84, 0.84]	-0.006 (0.005) [0.27, 0.48]	-0.022 (0.011) [0.04, 0.09]	0.005 (0.008) [0.49, 0.48]	-0.013 (0.006) [0.04, 0.13]	
Doctor message (% neighbours)	0.281 (0.201) [0.16, 0.50]	0.067 (0.043) [0.11, 0.41]	0.048 (0.053) [0.36, 0.61]	-0.006 (0.082) [0.94, 0.95]	0.075 (0.070) [0.28, 0.60]	0.016 (0.146) [0.91, 0.91]	0.052 (0.089) [0.56, 0.82]	0.137 (0.088) [0.12, 0.33]	
Mean (control message)	-0.041	0.799	0.799	0.563	0.661	0.352	0.483	0.494	
Observations	5079	7700	7698	7692	7697	7700	5182	6704	
B. Sample restricted to doctor message group									
Religion concordance	0.100 (0.033) [0.00, 0.01]	-0.004 (0.004) [0.36, 0.60]	-0.005 (0.005) [0.24, 0.55]	-0.017 (0.008) [0.04, 0.15]	-0.002 (0.007) [0.83, 0.84]	0.005 (0.015) [0.74, 0.76]	-0.025 (0.009) [0.01, 0.02]	0.004 (0.008) [0.60, 0.85]	
Religion concordance (% neighbours)	0.214 (0.150) [0.16, 0.52]	-0.004 (0.029) [0.89, 0.99]	0.003 (0.035) [0.92, 0.93]	-0.039 (0.047) [0.40, 0.83]	-0.011 (0.051) [0.84, 1.00]	-0.040 (0.081) [0.63, 0.62]	-0.086 (0.052) [0.10, 0.27]	-0.065 (0.052) [0.21, 0.39]	
Mean (religion discordance)	-0.035	0.807	0.806	0.571	0.654	0.326	0.497	0.476	
Observations	2519	3851	3849	3846	3849	3851	2587	3340	

Notes. Estimates based on OLS regressions using equation (1) in Panel A and equation (2) restricted to the doctor message group in Panel B (see Section 5). % neighbours is the share of households living in the same slum of the respondent that are allocated to the correspondent group. Standard errors clustered at the slum level are reported in parentheses. P -values are presented in brackets. The first value is from individual testing, the second is adjusted for testing that each variable is jointly different from zero for all outcomes grouped according to Table 2 and Table 4. Dependent variables in columns (1)–(5) are defined in Table 2, while dependent variables in columns (6)–(8) are defined in Table 4. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

ONLINE APPENDIX

Social Proximity and Misinformation:

Experimental Evidence from a Mobile Phone-Based Campaign in India

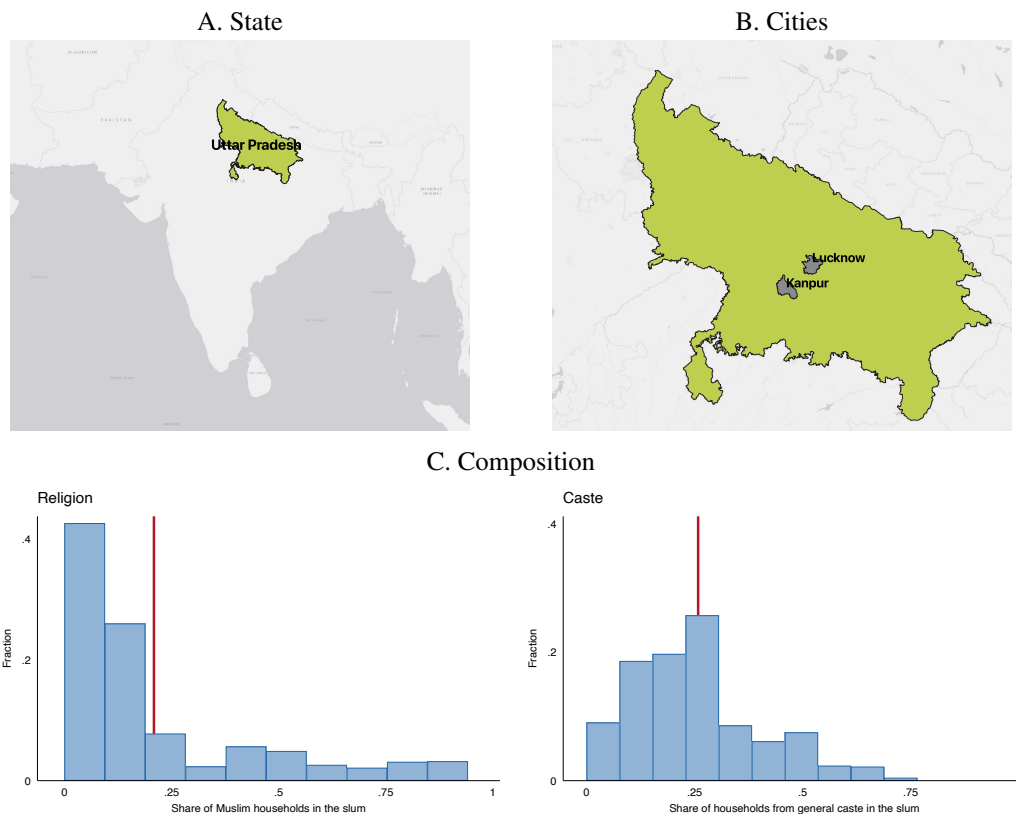
Alex Armand, Britta Augsburg, Antonella Bancalari and Kalyan Kumar Kameshwara

A Study area and timeline	2
A.1 Preventive practices and misinformation: additional descriptives	3
A.2 Intervention content	6
B Variable definition	6
B.1 Research methods and ethical concerns	9
C Study population, balance and attrition	11
D Additional analysis	12
D.1 Heterogeneous treatment effects	12
D.2 Effect of religion concordance in the control message	12
D.3 Effect on compliance, by component	19
D.4 Alternative specifications	19
D.5 The Effect of the Hindu introduction and of higher incentives	24
D.6 Other outcomes	24
D.7 Estimates with controls using post-double selection LASSO and ANCOVA	28
D.8 Additional evidence on spillovers	28

A Study area and timeline

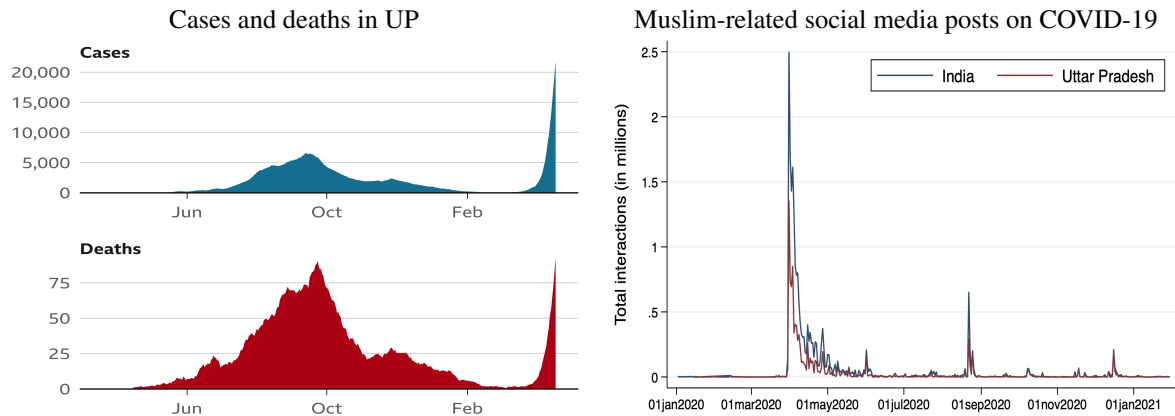
Panel A and B of Figure A1 illustrate the geographic location of the study area. Panel C shows the distribution of the share of the Muslim population at slum level in the study area. Panel A in Figure A2 reports the time series of the number of COVID-19 cases and deaths in UP from the beginning of 2020 until April 2021 (see Figure A3 for a comparison with the timeline of the study). Panel B focuses on trends in social media interactions (Facebook and Facebook-related media) targeting and blaming the Muslim population for the spread of the virus.

Figure A1: Study location and religious/caste composition



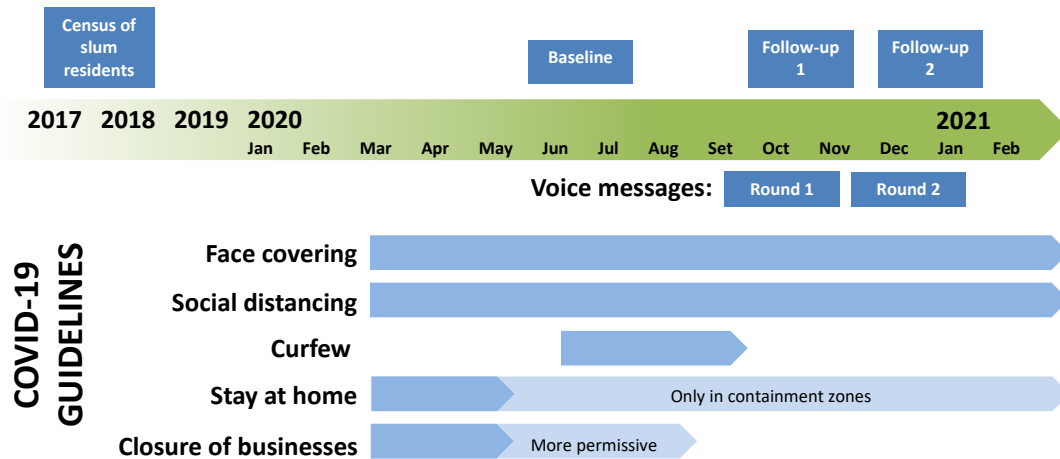
Notes. Panel A shows the location of the state of UP, while Panel B show the location of Lucknow and Kanpur in the state (basemap source: Esri). In Panel B, the Muslim and the general caste population is computed at slum level. The vertical lines indicate the sample mean.

Figure A2: COVID-19 cases and deaths in UP and misinformation in social media



Notes. In Panel A, the figure shows the reported number of cases and deaths from the beginning of 2020 until April 2021 using rolling seven-day averages. The source of data is the Indian Ministry of Health and Family Welfare. Graphic elaboration produced by BBC (<https://www.bbc.com/news/world-asia-india-56799303>). In Panel B, the data shows the evolution over time of Muslim-related social media posts about COVID-19 spread between January 2020 and February 2021. The vertical axis depicts the total number of times a Facebook post created on a given date is liked, shared or commented upon. Data is computed from Facebook's Crowd Tangle Team (2020). We select the following keywords (both in Hindi and in Latin transliteration): Corona_Jihad, CoronaJihad, Corona Jihad, Tablighi, Tablighi jamat, Tablighijamat, Tablighi_jamat, jihadivirus, Muslim virus, Nizamuddin Markaz. These keywords were the most-commonly used to spread misinformation linking the Muslim religion with COVID-19.

Figure A3: Study timeline and comparison with COVID-19 guidelines in UP

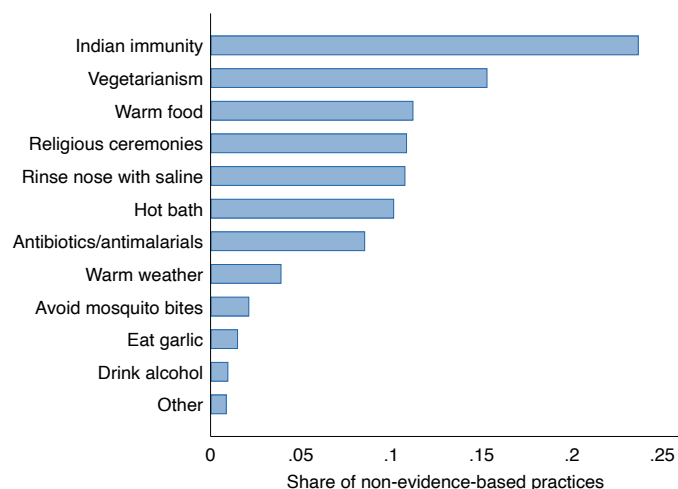


Notes. Guidelines are compiled from official sources (Government of India, 2021; Awasthi, 2020). Lucknow and Kanpur were included in the red zone in May 2020. Red zones are the areas with high coronavirus cases and high doubling rate in the previous 21 days. The first phase of the closure of businesses included all businesses apart from essential shops and services, while the second more permissive phase allowed the re-opening of the following activities: shopping malls, religious places, hotels and restaurants in June 2020 (unlock phases 1 and 2); gyms and yoga centers in August 2020 (unlock phase 3); entertainment, sport, political, academic and social functions and gatherings with a limited number of participants in September 2020 (unlock phases 4, 5 and 6). Curfews were first characterized by night curfews from 9pm to 5am in June and July 2020, and then to weekend curfews until September 2020. Local authorities had the power to impose curfews based on local conditions.

A.1 Preventive practices and misinformation: additional descriptives

Figure A4 presents the most-commonly reported misconceptions about protecting against COVID-19. Table A1 estimates how likely individuals in the baseline sample are to identify misinformation, based on their individual characteristics.. Figure A5 shows average levels of agreement with statements including misinformation shared by other citizens, restricting the sample to the control group and distinguishing by whether the citizen is in-group or out-group.

Figure A4: Non-evidence-based preventive practices, at baseline



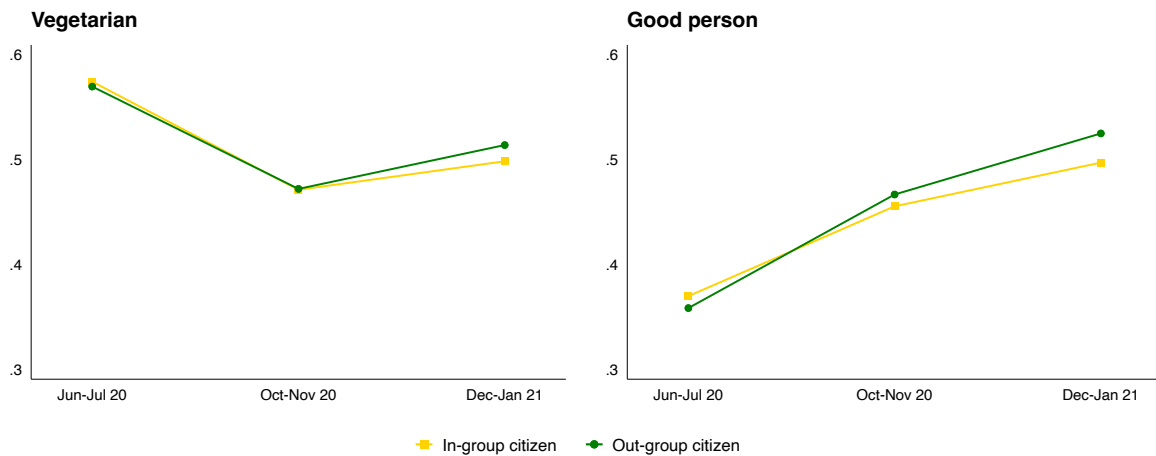
Notes. Respondents were asked about what, according to their opinion, would help in protecting them, or their family, from getting coronavirus. The questions were open-ended and responses were categorized into evidence-based and non-evidence-based preventive practices. We present the share of each non-evidence-based practice out of all non-evidence-based practices reported by the respondent. The sample is restricted to baseline observations and to respondents that reported at least one non-evidence-based practice.

Table A1: Baseline correlates of preventive practices

	Number of reported preventive practices		At least 1 non-evidence-based practice (3)
	Evidence-based (1)	Non-evidence-based (2)	
Respondent is male	-0.03 (0.08)	-0.03 (0.03)	-0.01 (0.02)
Head is male	0.11 (0.07)	0.05 (0.04)	0.02 (0.02)
Respondent is Muslim	-0.08 (0.10)	-0.04 (0.03)	-0.01 (0.02)
Caste: General	0.24*** (0.08)	0.05 (0.04)	0.02 (0.02)
Age	-0.01** (0.00)	-0.00 (0.00)	-0.00 (0.00)
Household members	0.01 (0.01)	0.01 (0.01)	0.00 (0.00)
Share females	-0.03 (0.09)	0.06 (0.06)	0.03 (0.03)
No children	0.03 (0.06)	-0.01 (0.03)	-0.01 (0.02)
Own dwelling	0.13 (0.10)	0.01 (0.03)	0.01 (0.02)
BPL ration card	-0.06 (0.07)	-0.01 (0.02)	-0.00 (0.01)
Member with COVID-19	0.23 (0.18)	0.13** (0.05)	0.11*** (0.03)
COVID-19 symptoms known	-0.03 (0.09)	-0.07* (0.04)	-0.05*** (0.02)
Slums	142	142	142
Households	3,975	3,975	3,975

Notes. The dependent variables are: in column (1) *Number of evidence-based preventive practices* is the number of practices reported by the respondent that are evidence-based; (2) *Number of non-evidence-based preventive practices* is the number of practices reported by the respondent that are non-evidence-based; column (3) *At least 1 non-evidence-based practice* is an indicator equal to one if the respondent reported at least 1 non-evidence-based preventive practice, and 0 otherwise. All specifications include strata (city and managed by main provider) variables as controls. Standard errors clustered at the slum level are presented in parenthesis.

Figure A5: Agreement with misinformation shared by other citizens, by statement



Notes. Each figure shows the average level of agreement in the control group with the following statements: *vegetarian* “if you are vegetarian, you do not need to worry about the coronavirus”; *good person* “if you are a good person, you do not need to worry about the coronavirus”. *In-group* averages only statements in which the respondent and the interlocutor assigned to the statement share the same religion. *Out-group* averages only statements in which the respondent and the interlocutor assigned with the statement do not share the same religion or the statement is associated with the generic “people”. Each outcome is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree. Details about the survey instrument is described in Section 4.

A.2 Intervention content

The **control message** consisted of gossip about popular actresses of Bollywood (uninformative message). The **doctor message** treatment involved two informative messages sent to the study participants twice during the study period. Although these treatment messages had a similar structure, they each addressed a different topic. The script of the messages reads as follows:

Introduction (included in both control and doctor message)

Sender: Greeting! ['namaste' or 'salam alaykum', according to randomization] I am a resident of UP and like me, you might also be confused about information shared on social media. If this is the case, then the following messages might be helpful for you. After watching this video, if you answer the question correctly, then you can get a chance to win the lottery of up to Rs. [high or low amount, according to randomization] in the form of mobile recharge.

First round of the doctor message

Sender: So, let's listen to what the renowned doctors have to say about this question: Is it correct that being a vegetarian or eating only a vegetarian diet fully protects from contracting the virus? *Doctor 1:* No, this misconception is spread inside the society, there is no such thing. You can see that people all over the world are non-vegetarians or vegetarians and everyone is getting infected. *Doctor 2:* Yes, it is true that vegetarian food is good food and healthy food. It also increases some immunity. But it is a misconception that if we take vegetarian food then there is no need to do other measures and we will not be infected from Corona. *Doctor 3:* The most important thing to avoid coronavirus is to use masks, social distance, wash hands frequently with soap, use of sanitizer.

Second round of the doctor message

Sender: So, let's listen to what the renowned doctors have to say about this question: Is it correct that we Indians need not worry about the coronavirus because our immune system is quite strong? *Doctor 1:* This is a myth. It can lead to false beliefs among people that they we will not get the disease. Please do not live with this false belief. In fact, the Indian population has contracted many diseases in the past. Please look at how many people are contracting the virus: the number of people getting the disease is increasing in the country and the world. *Doctor 2:* Coronavirus is a threat to the entire human civilization today. Do not stay under the misconception that we are immune to the virus. We need to be careful, protect ourselves from the virus, and follow the guidelines set by the government. *Doctor 3:* Maintain physical distance, use face mask and sanitizer and take nutritious diet. All these things are being emphasized, so keep doing all these. Avoid fake news and the confusion that is being spread, and follow all these things.

All rounds of the doctor message

Sender: We thank the doctors. Now, things are clear for me and hopefully for you too. If you have understood the message, please spread it to others. If each of us makes this contribution, we can save a lot of lives together. To enter the lottery, you would have to answer the following question correctly: "Can we Indians be carefree and not worry about coronavirus because our immune system is very strong?" (first round of doctor message) / "When eating pure vegetarian, you cannot get coronavirus." (second round of doctor message). Press 1 for true or 2 for false.

B Variable definition

Variable	Description	Type (round)
Respondent's characteristics		
Gender	Indicator variable equal to 1 for male respondents, and 0 otherwise.	Self-report (BL)
Head is male	Indicator variable equal to 1 if household head is male, and 0 otherwise.	Self-report (BL)
Muslim	Indicator variable equal to 1 if respondent is Muslim, and 0 otherwise.	Self-report (BL)
Incomplete primary education	Indicator variable equal to 1 if household head did not complete primary education, and 0 otherwise.	Self-report (BL)
Caste: general	Indicator variable equal to 1 if respondent belongs to General caste, and 0 otherwise (other backward caste, scheduled caste, or scheduled tribe).	Self-report (BL)
Share females	Number of women in the household.	Self-report (BL)
No children	Indicator variable equal to 1 if household has no children (less than five years old), and 0 if household has children.	Self-report (BL)
BPL ration card	Indicator variable equal to 1 if household possess a below poverty line ration card, and 0 if it does not.	Self-report (BL)
Own dwelling	Indicator variable equal to 1 if the dwelling is owned, and 0 otherwise.	Self-report (BL)
Own latrine	Indicator variable equal to 1 if the latrine is owned, and 0 otherwise.	Self-report (BL)
Member with COVID-19	Indicator variable equal to 1 if any household member has tested positive with COVID-19, and 0 otherwise.	Self-report (BL+FU)
COVID-19 symptoms known	Indicator variable equal to 1 if any household member has COVID-19 symptoms, and 0 otherwise.	Self-report (BL+FU)
Intervention		
Doctor message	Indicator variable equal to 1 if the receiver is in the doctor message treatment group, and 0 otherwise.	Records (FU)
Duration (minutes)	Duration of the call, reported in minutes. It is coded as missing for the respondents that did not pick up the call.	Records (FU)
Picked up	Indicator variable equal to 1 if the respondent picked up the call in any of the two rounds of interventions, and 0 otherwise.	Records (FU)
Religion concordance	Indicator variable equal to 1 if the receiver received a message in which the sender and the receiver shares the same religion, and 0 otherwise.	Records (FU)
% listened	Proportion of the audio message that is listened by the respondent. In IV regressions it is coded as 0 for the respondents that did not pick up the call. In treatment compliance regressions, it is coded as missing for the respondents that did not pick up the call.	Records (FU)
Outcomes		
Recommended practices	Index capturing adherence to WHO's recommendations to protect from infection, built using the procedure of Kling et al. (2007) (Section 4). Individual components of the index includes the following variables: <i>wore face mask</i> is an indicator variable equal to 1 if the respondent wears a face mask when leaving the house, and 0 otherwise; <i>washed hands frequently</i> is an indicator variable equal to 1 if the respondent indicates at least 3 moments (which corresponds to the within-sample median value) in which he/she washed hands the day before the interview, and 0 otherwise; <i>did not leave slum</i> is an indicator variable equal to 1 if the respondent did not leave the slum the week previous to the interview, and 0 otherwise; <i>did not receive a visit</i> is an indicator variable equal to 1 if the respondent did not leave receive a visit from a person living outside the slum the week previous to the interview, and 0 otherwise; <i>did not meet anybody</i> is an indicator variable equal to 1 if the respondent did not meet anybody from outside the slum the day before the interview, and 0 otherwise.	Self-report (BL+FU)

(continued on next page)

Variable	Description	Type (round)
Face masks / hand-washing	Respondent's level of (average) agreement with wearing face masks in crowded places and washing hands with soap more frequently and for longer to protect themselves against COVID-19. Agreement is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.	Self-report (FU)
Physical distancing	Respondent's level of agreement with keeping physical distance with other people to protect themselves against COVID-19. Agreement is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.	Self-report (FU)
Vegetarianism	Respondent's level of agreement with eating a vegetarian diet to protect themselves against COVID-19, measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.	Self-report (FU)
Indian immunity	Respondent's level of agreement with relying on the Indian immune system to protect themselves against COVID-19. Agreement is measured using a re-scaled likert scale where 0 refers to strongly disagree and 1 refers to strongly agree.	Self-report (FU)
Fact-checking	Indicator variable equal to 1 if the respondent always or very frequently check the truthfulness of information shared or discussed with family and friends, 0 otherwise.	Self-report (BL+FU)
Agreement with misinformation	Average level of agreement with statements including incorrect views based on faulty knowledge or understanding, where 0 refers to strongly disagree and 1 refers to strongly agree. Misinformation shared by <i>in-group</i> includes statements from an interlocutor with the same religion of the respondent. Misinformation shared by <i>out-group</i> includes statements from an interlocutor with a religion different from the one of the respondent or from the generic term "people".	Elicited (BL+FU)
Appendix outcomes and other heterogeneity dimensions		
Contagion extremely unlikely	Indicator variable equal to 1 if the event that someone in the household to become ill from coronavirus is extremely unlikely, and 0 otherwise.	Self-report (BL+FU)
Opinions	Average level of agreement with statements reporting public views concerning opinions. Statements are aggregated by averaging responses using a re-scaled likert scale in which 0 refers to strongly disagree and 1 refers to strongly agree. Opinions shared by <i>in-group</i> includes statements from an interlocutor with the same religion of the respondent. Opinions shared by <i>out-group</i> includes statements from an interlocutor with a religion different from the one of the respondent or from the generic term "people". Individual statements are described in Appendix D.6.	Elicited (BL+FU)
Average trust in government	High trust is an indicator equal 1 if the average trust in the government in the slum is below the median of the sample distribution, and 0 otherwise. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.	Self-report (BL)
Muslim share of the slum	Share of Muslim households in the slum. High % of Muslim households in the slum is an indicator variable equal to 1 if the share of Muslim households in the slum is below the median of the sample distribution, and 0 otherwise.	Self-report (BL)
Risk of contagion	Risk of someone in the household to become ill from coronavirus, with 0 indicating extremely unlikely and 1 indicating extremely likely.	Self-report (BL+FU)
Strength of religious identity	Indicator variable equal to 1 if the respondent strongly agree or agree to the statements "My religious faith/philosophy of life has a pronounced impact on my daily life" and "When I take important decisions, my religious faith/philosophy of life plays a considerable role", and 0 otherwise.	Self-report (FU)

(continued on next page)

Variable	Description	Type (round)
Social desirability	High social desirability is an indicator equal to 1 if social desirability is below the median of the sample distribution, and 0 otherwise. Social desirability is measured using the short version of the Marlowe–Crowne Social Desirability Scale (MC–SDS).	Self-report (BL)
Trust in government	Indicator variable equal to 1 if respondent trusts or strongly trusts information shared by government officials, and 0 otherwise.	Self-report (BL)
Trust in information	Respondent’s level of trust in the information shared by different individuals, including <i>doctors and health experts</i> and <i>other citizens</i> (which includes people from UP and by people from UP of other religions). Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust.	Self-report (BL+FU)

B.1 Research methods and ethical concerns

Participants of the study were selected based on data previously-collected by the research team. Participants were interviewed as part of a separate field experiment, completed in January 2020, for which we obtained ethical approval from the *UCL Research Ethics Committee* (ref. 2168/012). Informed consent was secured for both participation in the original study and for potential contact in future survey rounds and related research. All participants were above 18 years old and provided written consent. For the current study, which focuses on the same population, we received separate ethics approval from the *LSE Research Ethics Review* (ref. 1132).

Due to the limitations imposed by the COVID-19 pandemic, data collection was conducted through mobile phone interviews. To ensure the autonomy and well-being of participants, we obtained their voluntary and informed consent orally at the start of each interview. Participation in the survey was entirely voluntary, and no monetary compensation was offered to the respondents. The consent form script, translated into Hindi, is as follows:

Hello. My name is [NAME] and I work with Morsel Research and Development on a research project called “COVID-19 Spread in Informal Settlements” and funded by the London School of Economics (LSE). Researchers at the LSE, Institute for Fiscal Studies in the United Kingdom and the Nova School of Business and Economics in Portugal are interested in collecting information to assess slum dwellers’ response to the COVID-19 pandemic. We are not affiliated to the government. Results from this research will be shared with policymakers and academics. However, they will not get any information about each participant, including names. We would like to interview you for approximately ten minutes. All the information you provide remains confidential and can be accessed only by selected members of the research team. You have the right to decline your participation or withdraw from the study at any time without the need to explain yourself and your decision will carry no consequences. Should you take part in the study, you agree that we can contact you again in the future to collect more information related to this study, at which point you will again be able to choose whether to participate or not. Please let me know if you have any questions at this point. We have just sent you a text message with contact information, should you have any queries about this study and your interview going forward. Please confirm that you have understood the information just provided and that you were given the opportunity to clarify any doubts or questions. [If respondent says ‘Yes’ proceed with the survey].

Respondents were informed that they could ask questions about the study at any time before, during

and after the interview. In line with this, following the informed consent, a text message was sent to participants providing a contact number for any study-related inquiries. Additionally, to support the well-being of participants, the text message provided also information on how to contact the COVID-19 helpline for issues and questions related to the pandemic.

To ensure confidentiality of participants' responses, we assured them that only anonymous data would be analyzed. We implemented several measures to maintain this confidentiality. Firstly, interviewers used headphones to avoid that responses could be overheard by anyone. Secondly, interviewers were trained not to view or share any information about respondents other than for what was strictly required for the purposes of data collection. To guarantee this condition, the project used Computer-Assisted Personal Interviewing (CAPI) to collect data, a well-established system designed specifically with the needs of confidentiality and data-security in mind, including, for example, single log-in and access to data available only during the interview. Thirdly, all collected data were encrypted and stored on a secured server. For added security, data backups are maintained on an off-site machine stored securely with a third-party company. Network access to these servers is strictly limited to technical support staff. In terms of the questionnaire's content, we did not identify any issues related to causing stress among participants. Recognizing that the target population is a vulnerable group, we took special care in how we framed the questions. This included conducting a pilot survey where respondents could give feedback on the types of questions, wording, and interview length. During this pilot phase, we did not receive any reports of issues from the participants.

Regarding the interventions, our approach was strictly non-political and focused on providing participants with content based on scientific evidence. Importantly, the project did not involve deceiving respondents in any way. It is worth noting that exposure to different religions is a common aspect in the target population, as evidenced by the religious diversity in the sample. Moreover, our methodology was carefully designed to avoid communicating any discriminatory messages related to specific religions.

C Study population, balance and attrition

Tables C1 reports descriptive statistics for observable characteristics of the respondent and the household and of outcome variables. Table C2 reports correlates of attrition.

Table C1: Respondents' characteristics and attrition

	Control mean	Full sample Difference with doctor message treatment	N	Sample restricted to doctor message Muslim sender (mean)	Difference with Hindu sender	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. Respondent characteristics						
Respondent is male	0.79 [0.41]	-0.00 (0.01)	3983	0.78 [0.41]	0.02 (0.02)	1996
Head is male	0.82 [0.39]	-0.00 (0.01)	3983	0.81 [0.39]	-0.00 (0.02)	1996
Respondent is Muslim	0.21 [0.41]	-0.00 (0.00)	3983	0.23 [0.42]	0.01 (0.01)	1996
Caste: General	0.16 [0.36]	0.01 (0.01)	3983	0.16 [0.36]	0.02 (0.02)	1996
Age	39.77 [11.41]	-0.50 (0.38)	3983	39.34 [11.59]	-0.15 (0.47)	1996
Household members	5.10 [1.96]	0.04 (0.06)	3983	5.20 [1.97]	-0.10 (0.09)	1996
Share females	0.35 [0.16]	-0.01 (0.01)	3983	0.34 [0.16]	0.01 (0.01)	1996
No children	0.72 [0.45]	-0.02 (0.01)	3983	0.71 [0.45]	-0.02 (0.02)	1996
Own dwelling	0.73 [0.44]	-0.00 (0.01)	3983	0.73 [0.44]	0.00 (0.02)	1996
Own latrine	0.61 [0.49]	0.00 (0.02)	3977	0.61 [0.49]	0.02 (0.02)	1995
BPL ration card	0.38 [0.49]	-0.01 (0.02)	3983	0.38 [0.49]	-0.01 (0.02)	1996
Member with COVID-19	0.12 [0.32]	0.01 (0.01)	3983	0.14 [0.34]	-0.01 (0.01)	1996
COVID-19 symptoms known	1.60 [0.66]	-0.03 (0.02)	3975	1.58 [0.66]	-0.02 (0.02)	1991
B. Attrition						
Attrition BL-any FU	0.13 [0.34]	-0.01 (0.01)	3983	0.14 [0.34]	-0.01 (0.01)	1996
Attrition BL-FU1	0.28 [0.45]	0.00 (0.01)	3983	0.29 [0.45]	-0.00 (0.02)	1996
Attrition BL-FU2	0.24 [0.42]	-0.00 (0.01)	3983	0.23 [0.42]	-0.00 (0.02)	1996

Notes. Column (1) reports the mean and standard deviation of the each variable for the control group in the doctor message treatment, while column (2) shows the difference to this mean of those who were sent the doctor messages. Column (3) reports the joint sample size. Columns (4)–(6) report the same information comparing those that were sent the message with a Muslim sender to those that were sent a message with a Hindu sender, hence restricting the sample to the doctor message treatment group.

Table C2: Correlates of attrition

<i>Sample:</i>	Dependent variable: attrition indicator			
	Full sample		Restricted to doctor message group	
	(1)	(2)	(3)	(4)
Doctor message	-0.01 (0.01)	-0.01 (0.01)		
Doctor message x Muslim		0.01 (0.02)		
Religion concordance			-0.01 (0.01)	-0.02 (0.02)
Religion concordance x Muslim				0.04 (0.04)
Respondent is male	0.03** (0.01)	0.03** (0.01)	0.04* (0.02)	0.04* (0.02)
Head is male	-0.04** (0.02)	-0.04** (0.02)	-0.04* (0.02)	-0.04* (0.02)
Respondent is Muslim	0.04** (0.02)	0.03 (0.02)	0.04* (0.03)	0.02 (0.03)
Caste: General	-0.02* (0.01)	-0.02* (0.01)	-0.03 (0.02)	-0.03 (0.02)
Age	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Household members	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Share females	0.00 (0.02)	-0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)
No children	-0.01 (0.01)	-0.01 (0.01)	0.00 (0.02)	0.00 (0.02)
Own dwelling	-0.03** (0.02)	-0.03** (0.02)	-0.04* (0.02)	-0.04* (0.02)
BPL ration card	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)	0.01 (0.02)
Member with COVID-19	-0.02 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)
COVID-19 symptoms known	0.02** (0.01)	0.02** (0.01)	0.02 (0.01)	0.02* (0.01)
Sample	All	All	Doctor message	Doctor message
Attrition Rate	0.13	0.13	0.13	0.13
Slums	142	142	142	142
Households	3,975	3,975	1,991	1,991

Notes. Attrition indicator is an indicator variable equal to 1 if a households was neither re-interviewed in follow-up 1 or follow-up 2, and 0 otherwise. Standard errors are clustered at the slum level and presented in parenthesis. Columns (1)–(2) are for the full sample, columns (3)–(4) restrict the sample to the doctor message treatment group.

D Additional analysis

D.1 Heterogeneous treatment effects

Figure D1 shows the average of outcome variables separate for Hindu and Muslim respondents. Figure D2 shows estimates of treatment effects separately by the round of messages sent and by the religion of the respondent. Figures D3–D5 report estimates of heterogeneous treatment effects of the doctor message using equation (1) (Panel A), and of religion concordance in the doctor message using equation (2) restricted to the doctor message group (Panel B).

D.2 Effect of religion concordance in the control message

Tables D1, D2 and D3 present estimates of the effects on the take-up of messages, on preventive practices and on response to misinformation, respectively. Estimates are produced restricting the sample to recipients of the control message and estimating equation (2).

Table D1: Take-up of the control message and religion concordance

	Picked up	% listened	Duration (minutes)
	(1)	(2)	(3)
Religion concordance	-0.004 (0.018) [0.844, 0.844]	0.024 (0.019) [0.212, 0.373]	0.031 (0.024) [0.197, 0.416]
Mean (control message)	0.387	0.659	0.533
Observations	3849	1467	1467

Notes. Estimates based on OLS regressions using equation (2) restricted to respondents in the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in the table. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D2: Preventive practices: the effect of religion concordance in the control message

	Compliance		Beliefs over the efficacy of...		
	Recommended practices	Recommended practices		Non-evidence-based practices	
		Face masks / hand-washing	Physical distancing	Vegetarianism	Indian immunity
	(1)	(2)	(3)	(4)	(5)
Religion concordance	-0.044 (0.031) [0.166 , 0.540]	-0.002 (0.004) [0.622 , 0.626]	-0.007 (0.004) [0.126 , 0.506]	-0.008 (0.008) [0.306 , 0.665]	-0.004 (0.007) [0.519 , 0.769]
Mean (religion discordance)	-0.018	0.800	0.802	0.567	0.663
Observations	2560	3849	3849	3846	3848

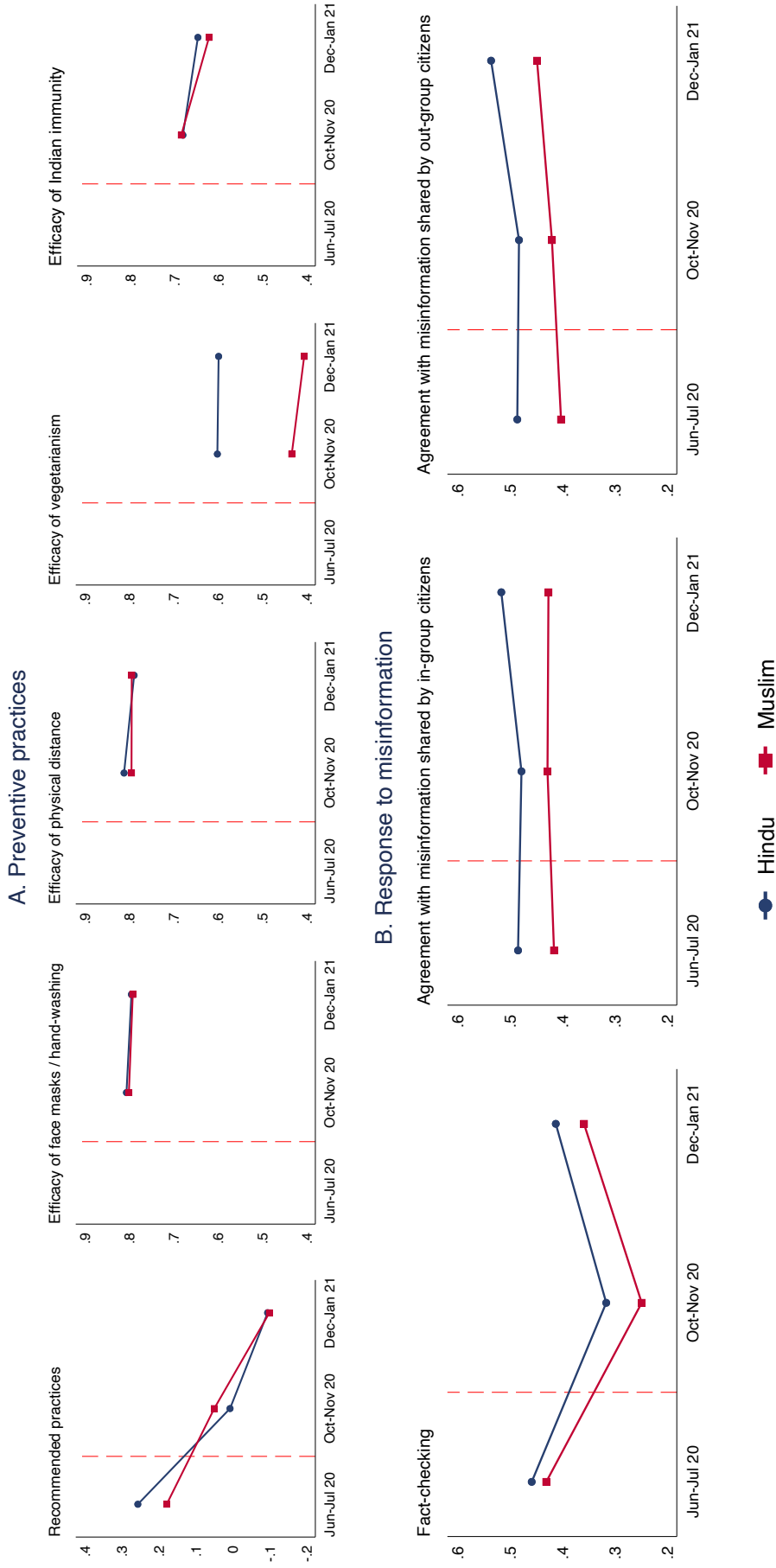
Notes. Estimates based on OLS regressions using equation (2) restricted to respondents in the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. In Panels A and C, the first value is from individual testing, the second is adjusted for testing that each treatment is jointly different from zero for all outcomes presented in each panel. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D3: Response to misinformation: the effect of religion concordance in the control message

	Fact-checking	Agreement with misinformation shared by...	
	(1)	In-group citizens	Out-group citizens
		(2)	(3)
A. OLS			
Religion concordance	0.017 (0.017) [0.321 , 0.701]	0.001 (0.012) [0.952 , 0.952]	0.002 (0.008) [0.848 , 0.978]
Mean (control message)	0.345	0.483	0.493
Observations	3849	2595	3364

Notes. Estimates based on OLS regressions using equation (2) restricted to respondents in the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. In Panels A and C, the first value is from individual testing, the second is adjusted for testing that each treatment is jointly different from zero for all outcomes presented in each panel. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

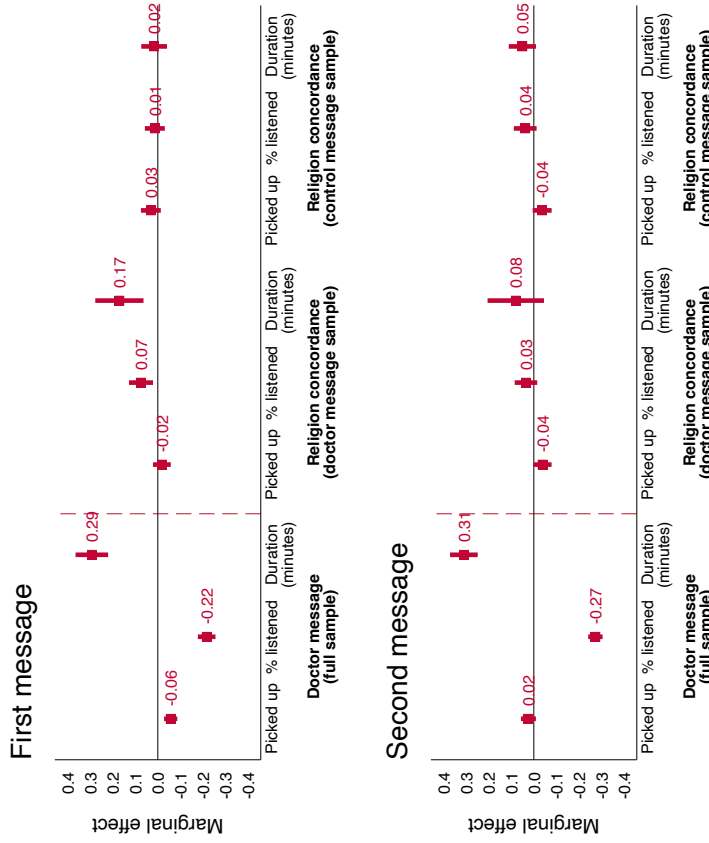
Figure D1: Preventive practices and response to misinformation, by respondent's religion



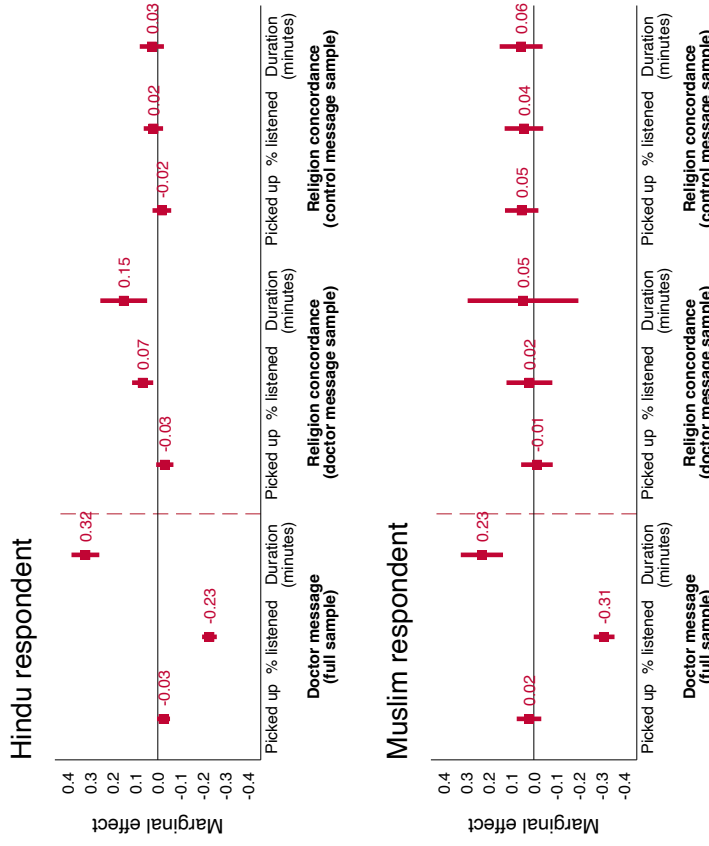
Notes: Each figure shows the average of outcome variables measured at different points in time. *Hindu* restricts the sample to Hindu respondents, and *Muslim* restricts the sample to Muslim respondents. The vertical line separates the baseline measurement from the follow-up measurements. In all panels, the sample is restricted to the control group. Further details about these variables are provided in Section 4. Variables are defined in Appendix B.

Figure D2: Take-up of messages, by sub-group

A. By timing of the message

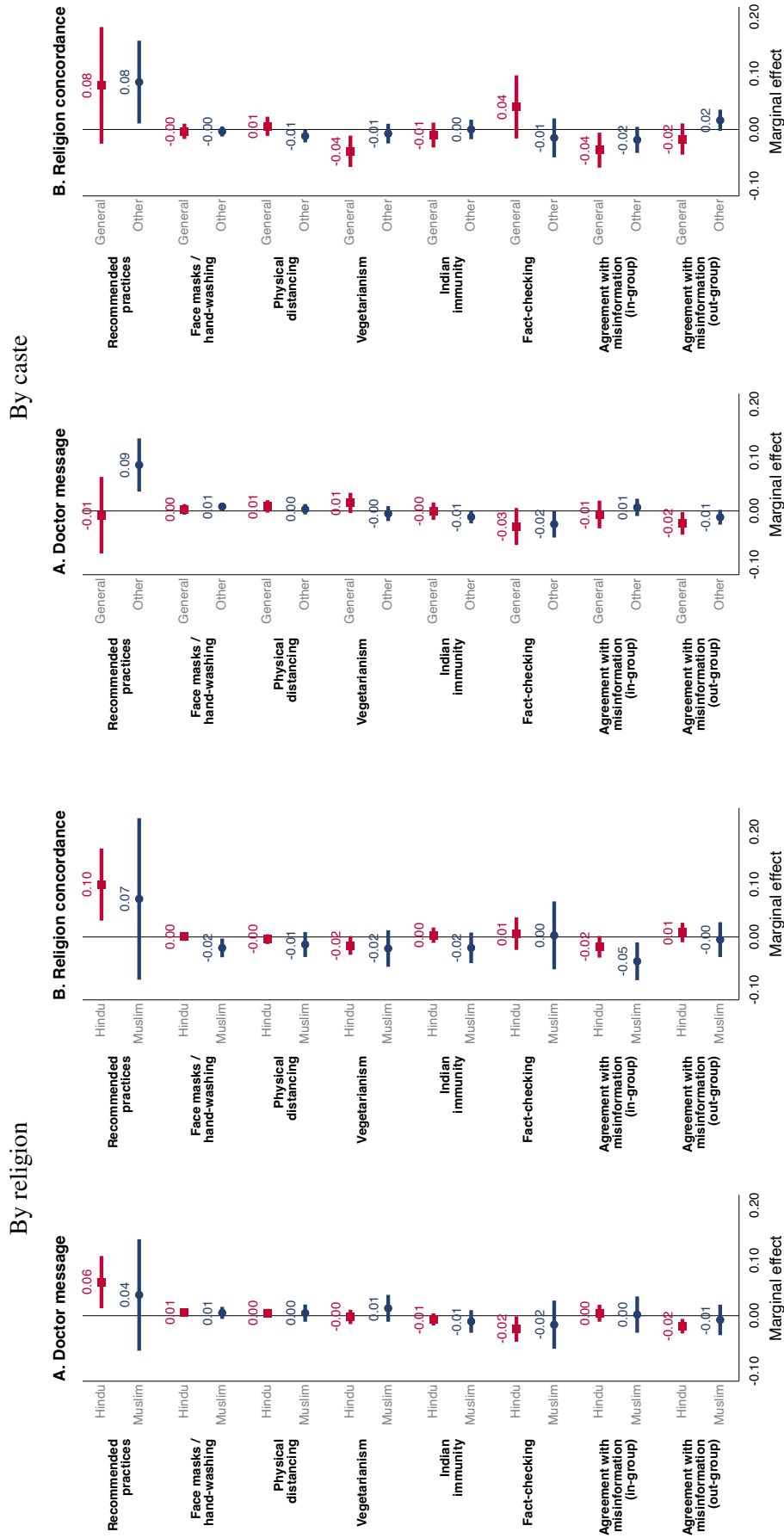


B. By respondent's religion



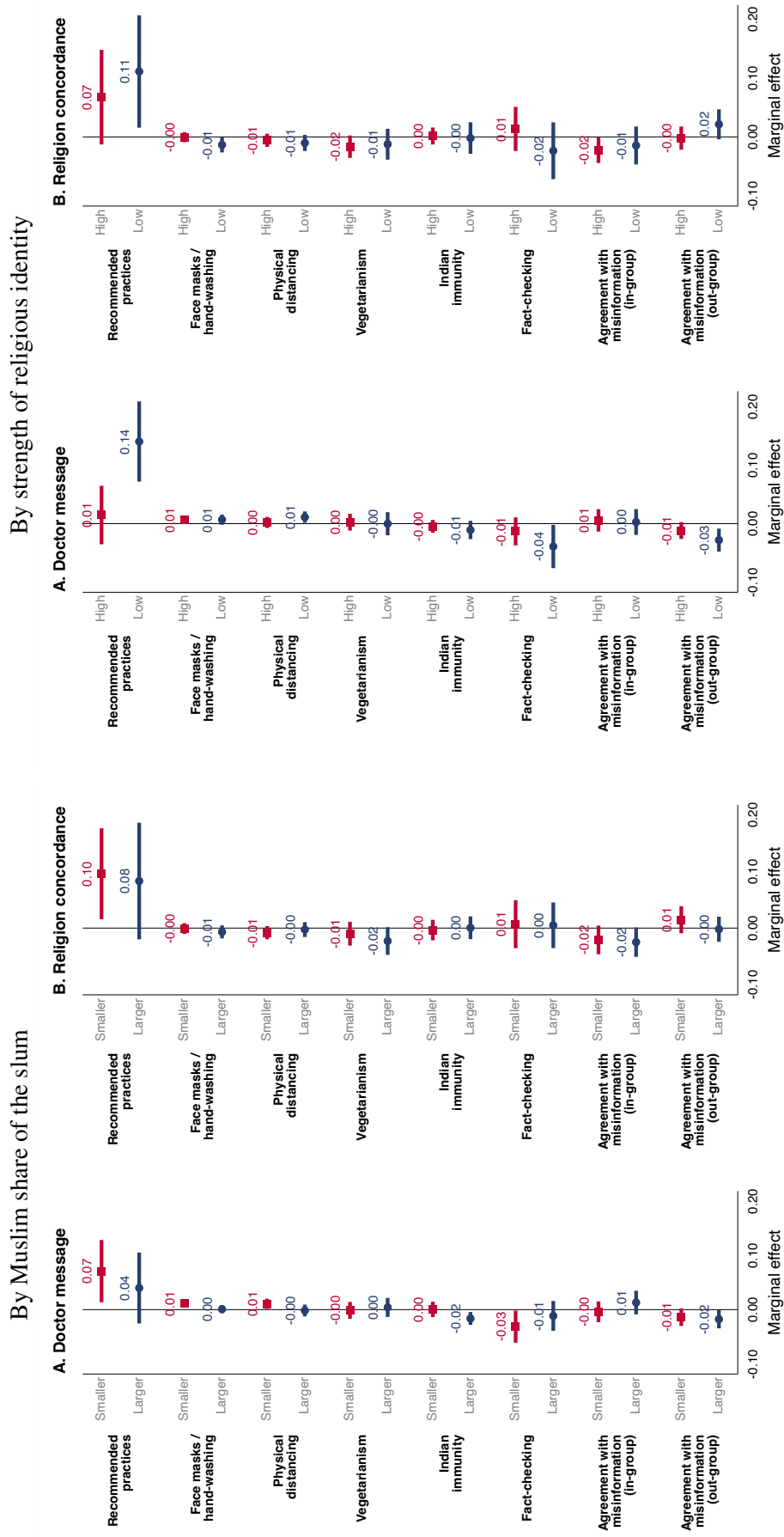
Note. The figure shows the effect of the doctor message and of religion concordance on compliance with treatments. Estimates based on OLS regressions using equation (1) and aggregating information in both rounds of data collection. Standard errors are clustered at the slum level, confidence intervals reported at 95% level. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent). *Picked up* is an indicator variable equal to 1 if the respondent picked up the call in any of the two rounds of interventions, and 0 otherwise. *% listened* is the share of the message that is listened, conditional on having picked up. The distributions of the share of each message that is listened by study participants (conditional on having picked up the call) is presented in Figure 2. *Duration (minutes)* is the duration of the call, conditional on having picked up. Note that the doctor messages have different duration as compared to the control messages.

Figure D3: Heterogeneous effects by respondent's religion and caste



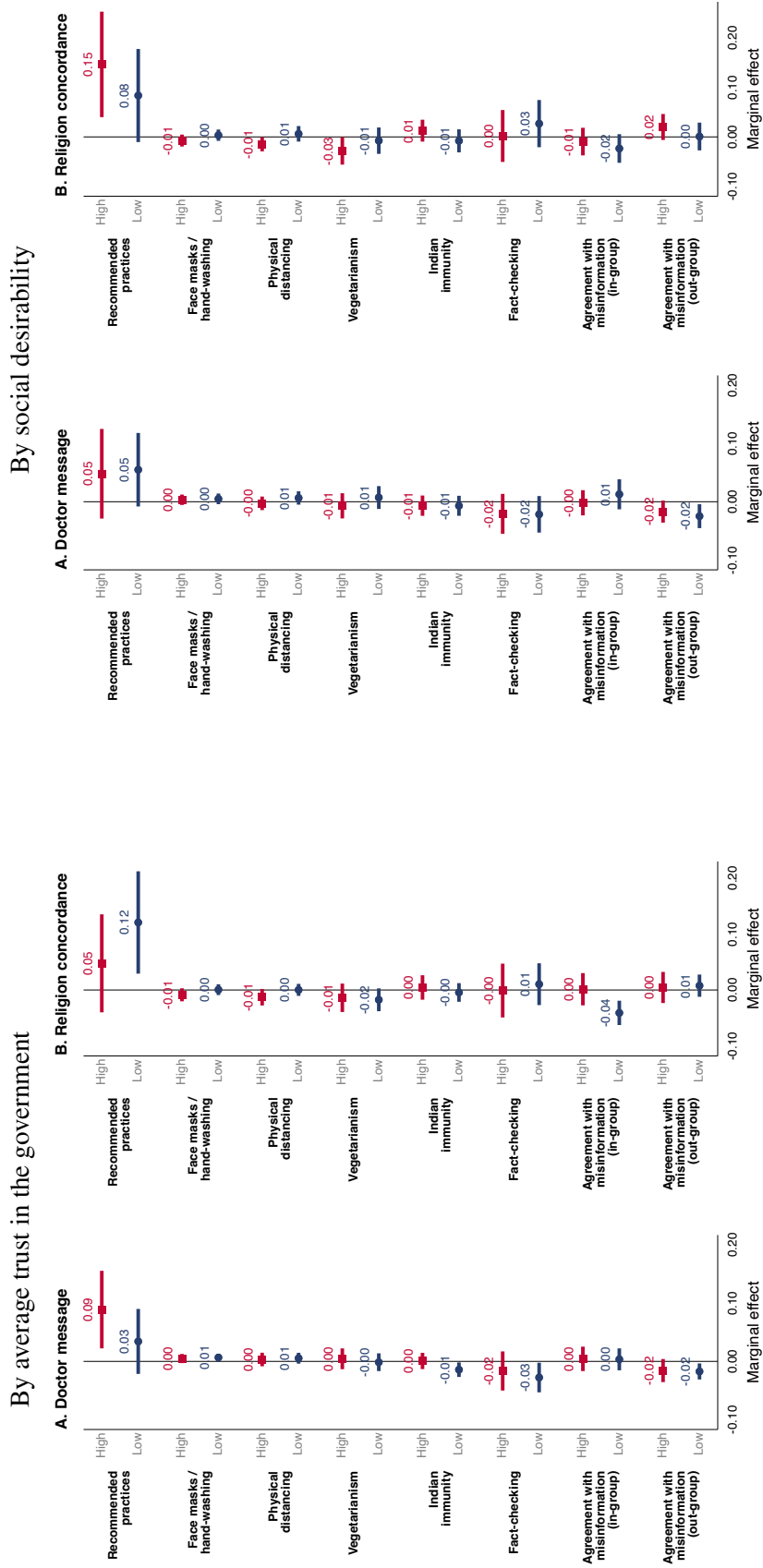
Notes. Heterogeneity by religion y is based on an indicator variable equal to 1 if the respondent is of Muslim religion, and 0 otherwise. Heterogeneity by caste is based on an indicator variable equal to 1 if the respondent is of a general caste, and 0 if the respondent is of other backward caste, scheduled caste, or scheduled tribe. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix B.

Figure D4: Heterogeneous effects by religious features



Notes. Heterogeneity by Muslim share of the slum is based on an indicator variable equal to 1 if the share of Muslim households in the slum is below the median of the sample distribution, and 0 otherwise. Heterogeneity by strength of religious identity is based on an indicator variable equal to 1 if the respondent's strength of religious identity is below the median of the sample distribution, and 0 otherwise. Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix B.

Figure D5: Heterogeneous effects by other dimensions



Notes. Heterogeneity by average trust in the government is based on an indicator variable equal to 1 if the average trust in the government in the slum is below the median of the sample distribution, and 0 otherwise. Trust is measured using a re-scaled likert scale where 0 refers to strongly distrust and 1 refers to strongly trust. Heterogeneity by social desirability is based on an indicator variable equal to 1 if social desirability is below the median of the sample distribution, and 0 otherwise. Social desirability is measured using the short version of the Marlowe-Crowne Social Desirability Scale (MC-SDS). Estimates based on OLS regressions using equation (1) and restricting the sample to the categories reported in column. Panel A presents estimates of the effect of the doctor message and includes all participants in the sample. Panel B presents estimates of the effect of religion concordance and restricts the sample to participants allocated to the doctor message. Standard errors are clustered at the slum level. Confidence intervals at 95% level. Variables are defined in Appendix B.

D.3 Effect on compliance, by component

Table D4 shows treatment effects on individual indicators of compliance with recommended practices and with indices capturing compliance by sub-category (face masks and hand-washing and physical distancing). Similar to the overall measure of compliance, these indices are computed using the procedure of Kling et al. (2007). The number of observations can vary because questionnaire modules were implemented in different random sub-samples to limit the duration of the interview. Table D5 shows (conditional) correlations between the overall measure of compliance and each outcome variable studied in the main text.

Table D4: Compliance with recommended practices

	Face masks / hand-washing			Physical distancing			
	z-score index	Wore face mask	Washed hands frequently	z-score index	Did not leave slum	Did not receive a visit	Did not meet anybody
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Full sample							
Doctor message	0.039 (0.025) [0.124]	0.000 (0.017) [0.988]	0.030 (0.018) [0.089]	0.037 (0.025) [0.137]	0.014 (0.011) [0.214]	0.019 (0.018) [0.277]	0.012 (0.011) [0.259]
Mean (control message)	0.043	0.698	0.763	-0.219	0.078	0.242	0.080
Observations	5079	2554	2525	2554	2554	2554	2546
B. Sample restricted to doctor message group							
Religion concordance	0.096 (0.039) [0.016]	0.026 (0.027) [0.339]	0.058 (0.023) [0.013]	0.035 (0.035) [0.321]	0.005 (0.017) [0.764]	0.012 (0.024) [0.605]	0.024 (0.016) [0.139]
Mean (religion discordance)	0.033	0.687	0.763	-0.205	0.089	0.251	0.076
Observations	2519	1229	1290	1229	1229	1229	1226

Notes. Estimates based on OLS regressions using equation (1) in Panel A and on equation (2) restricting the sample to the doctor message group in Panel B. Standard errors clustered at the slum level are reported in parentheses. P -values are presented in brackets. Dependent variables in column (1) and (4) are z -score indices computed for each sub-category indicated in the table's heading using the procedure of Kling et al. (2007). Other dependent variables are detailed in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

D.4 Alternative specifications

In this section, we estimate treatment effects using alternative specifications to equation (1) and equation (2). First, in Table D6, we present estimates of treatment effects with an interaction model. In Panel A, to capture the effect on the content of the message and of religion concordance, we estimate the following specification using the full sample:

$$Y_{ijt} = \beta_D \text{doctor}_i + \beta_C \text{concordance}_i + \beta_{DC} \text{doctor}_i \cdot \text{concordance}_i + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \quad (3)$$

where doctor_i is an indicator variable equal to 1 if the receiver i is in the doctor message treatment group, and 0 otherwise, and concordance_i is an indicator variable equal to 1 if the receiver i was sent a message in which the sender and the receiver share the same religion, and 0 otherwise. \mathbf{X}_{ij} is a set of indicator variables for randomization strata, and δ_t are period-of-survey indicator variables. The error term ϵ_{it} is assumed to be clustered at the slum level. In Panel B, to capture the effect of the content of the

Table D5: Correlation between compliance and other outcomes

	Dependent variable: Compliance with recommended practices							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Face masks and hand-washing	0.522 (0.077) [0.000]							0.540 (0.130) [0.000]
Physical distancing		0.416 (0.066) [0.000]						0.198 (0.112) [0.080]
Vegetarianism			-0.134 (0.047) [0.005]					-0.021 (0.070) [0.765]
Indian immunity				-0.112 (0.056) [0.048]				-0.100 (0.076) [0.194]
Fact-checking					0.111 (0.026) [0.000]			0.132 (0.032) [0.000]
Agreement with misinformation (in-group)						-0.339 (0.060) [0.000]		-0.251 (0.059) [0.000]
Agreement with misinformation (out-group)							-0.343 (0.057) [0.000]	-0.213 (0.066) [0.002]
Mean (control message)	-0.041	-0.041	-0.042	-0.042	-0.041	-0.024	-0.027	-0.000
Observations	5079	5079	5073	5076	5079	3583	4607	3106

Notes. Estimates based on OLS regressions using equation (1). Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent and independent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

message and of monetary incentives, we follow a similar approach to equation 3 but using an indicator variable equal to 1 if the receiver i is offered a higher financial incentive, and 0 otherwise.

Second, in Tables D7–D8, we provide estimates using the specifications of Tables 2 and 4, but estimating treatment effects separately for each follow-up measurement.

Table D6: Treatment effects estimated using an interacted model (PAP version)

	Compliance			Beliefs over the efficacy of ...					Fact-checking		Agreement with misinformation shared by...	
	Recommended practices (1)	Face masks / hand-washing (2)	Physical distancing (3)	Vegetarianism (4)	Indian immunity (5)	(6)	(7)	(8)	In-group citizens	Out-group citizens		
A. Content and religion concordance												
Doctor message	-0.013 (0.030) [0.658]	0.007 (0.004) [0.045]	0.004 (0.005) [0.375]	0.006 (0.009) [0.500]	-0.009 (0.006) [0.174]	-0.018 (0.016) [0.252]	0.016 (0.011) [0.131]	-0.018 (0.008) [0.027]				
x religion concordance	0.139 (0.044) [0.002]	-0.002 (0.005) [0.697]	0.001 (0.006) [0.891]	-0.009 (0.011) [0.431]	0.003 (0.009) [0.771]	-0.010 (0.022) [0.646]	-0.024 (0.015) [0.099]	0.005 (0.011) [0.665]				
Religion concordance	-0.046 (0.031) [0.147]	-0.002 (0.004) [0.628]	-0.007 (0.004) [0.129]	-0.008 (0.008) [0.319]	-0.004 (0.007) [0.515]	0.017 (0.017) [0.325]	0.001 (0.012) [0.910]	0.001 (0.008) [0.877]				
Observations	5079	7700	7698	7692	7697	7700	5182	6704				
B. Content and monetary incentives												
Doctor message	0.044 (0.032) [0.175]	0.008 (0.003) [0.021]	0.006 (0.005) [0.269]	0.000 (0.009) [0.986]	-0.001 (0.007) [0.928]	-0.018 (0.015) [0.229]	0.013 (0.011) [0.220]	-0.013 (0.009) [0.149]				
x high incentive	0.026 (0.047) [0.584]	-0.003 (0.005) [0.522]	-0.002 (0.006) [0.751]	0.003 (0.011) [0.814]	-0.013 (0.010) [0.166]	-0.011 (0.019) [0.574]	-0.018 (0.016) [0.252]	-0.006 (0.012) [0.598]				
High incentive	0.000 (0.032) [0.991]	0.003 (0.004) [0.459]	-0.000 (0.004) [0.954]	-0.009 (0.008) [0.266]	0.009 (0.007) [0.178]	-0.019 (0.016) [0.222]	0.013 (0.012) [0.273]	-0.003 (0.008) [0.757]				
Observations	5079	7700	7698	7692	7697	7700	5182	6704				

Notes. Estimates based on OLS regressions using equation (3), including the baseline value of the dependent variable (when available). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D7: Preventive practices, estimates by survey round

	Beliefs over the efficacy of ...										
	Compliance			Recommended practices			Non-evidence-based practices			Immune system	
	Recommended practices		Face masks and hand-washing	Physical distancing		Vegetarianism	Vegetarianism		Immune system		
<i>Follow-up measurement:</i>	FU1 (1)	FU2 (2)	FU1 (3)	FU2 (4)	FU1 (5)	FU2 (6)	FU1 (7)	FU2 (8)	FU1 (9)	FU2 (10)	
A. Full sample											
Doctor message	0.053 (0.029) [0.074]	0.060 (0.029) [0.042]	0.007 (0.004) [0.059]	0.005 (0.004) [0.145]	0.007 (0.005) [0.184]	0.003 (0.005) [0.609]	0.014 (0.009) [0.109]	0.014 (0.009) [0.143]	-0.011 (0.007) [0.143]	-0.007 (0.007) [0.292]	-0.007 (0.006) [0.277]
Mean (control message)	0.015	-0.096	0.805	0.794	0.808	0.790	0.567	0.560	0.560	0.681	0.641
Observations	2496	2583	3801	3899	3800	3898	3797	3797	3895	3799	3898
B. Sample restricted to doctor message group											
Religion concordance	0.129 (0.044) [0.004]	0.060 (0.041) [0.151]	-0.005 (0.006) [0.362]	-0.002 (0.005) [0.720]	-0.005 (0.006) [0.433]	-0.006 (0.007) [0.376]	-0.021 (0.012) [0.089]	-0.021 (0.012) [0.089]	-0.012 (0.011) [0.249]	-0.001 (0.011) [0.942]	-0.002 (0.009) [0.840]
Mean (religion discordance)	-0.001	-0.068	0.815	0.800	0.818	0.795	0.590	0.553	0.553	0.675	0.634
Observations	1240	1279	1898	1953	1897	1952	1896	1896	1950	1897	1952

Notes. Estimates based on OLS regressions using equation (1) in Panel A and equation (2) restricted to the sample of respondents in the doctor message group in Panel B. *FU1* restricts the sample to data collected in the first follow-up survey (October–November 2020). *FU2* restricts the sample to data collected in the second follow-up survey (December 2020–January 2021). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D8: Response to misinformation, estimates by survey round

<i>Follow-up measurement:</i>	Fact-checking		Agreement with misinformation shared by...			
	FU1 (1)	FU2 (2)	In-group citizens		Out-group citizens	
	FU1 (1)	FU2 (2)	FU1 (3)	FU2 (4)	FU1 (5)	FU2 (6)
A. Full sample						
Doctor message	-0.008 (0.015) [0.576]	-0.037 (0.015) [0.013]	-0.010 (0.010) [0.341]	0.017 (0.010) [0.076]	-0.014 (0.008) [0.090]	-0.018 (0.007) [0.012]
Mean (control message)	0.302	0.401	0.468	0.498	0.470	0.518
Observations	3801	3899	2551	2631	3298	3406
B. Sample restricted to doctor message group						
Religion concordance	-0.000 (0.021) [0.997]	0.011 (0.022) [0.602]	-0.022 (0.013) [0.093]	-0.025 (0.013) [0.053]	0.003 (0.011) [0.804]	0.009 (0.012) [0.478]
Mean (religion discordance)	0.291	0.360	0.467	0.526	0.454	0.497
Observations	1898	1953	1267	1320	1643	1697

Notes. Estimates based on OLS regressions using equation (1) in Panel A and equation (2) restricted to the sample of respondents in the doctor message group in Panel B. *FU1* restricts the sample to data collected in the first follow-up survey (October–November 2020). *FU2* restricts the sample to data collected in the second follow-up survey (December 2020–January 2021). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

D.5 The Effect of the Hindu introduction and of higher incentives

Panel 1 in Table D9 shows estimates of the effect of the Hindu greeting at the beginning of the message estimated using the following specification restricted to either the doctor message group (Panel A) or the control group (panel B) on the outcomes studied in the main text:

$$Y_{ijt} = \beta_H Hindu_i + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \quad (4)$$

where $Hindu_i$ is an indicator variable equal to 1 if the message sent to receiver i is introduced by a Hindu greeting, and 0 if introduced by a Muslim greeting. Panel 2 shows estimates of the effect of offering a higher monetary incentive estimated using the following specification restricted to either the doctor message group (Panel A) or the control group (panel B) on the outcomes studied in the main text:

$$Y_{ijt} = \beta_L Higher_i + \alpha \mathbf{X}_{ij} + \delta_t + \epsilon_{ijt} \quad (5)$$

where $higher_i$ is an indicator variable equal to 1 if the receiver i is offered a higher financial incentive, and 0 if offered a lower financial incentives. In both equations, the remaining terms are in line with equation (1).

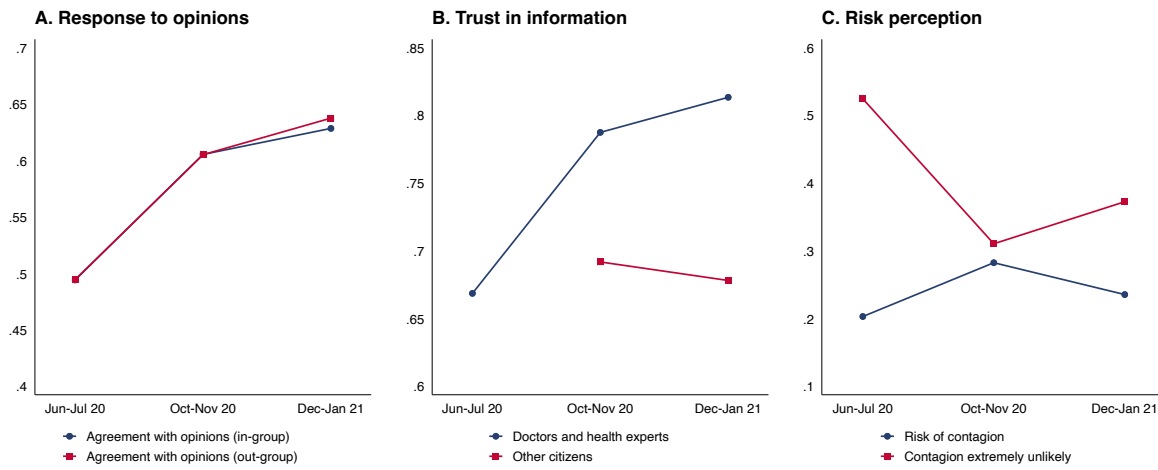
D.6 Other outcomes

We focus in this section on other outcome variables. First, during the interview, we also asked about agreement with 3 statements presenting views that are not necessarily based on facts or knowledge (or *opinions*), and are therefore harder to be influenced by information campaigns and by fact-checking. The first opinion, “religious gatherings should be allowed”, is particularly relevant in the study context due to the early outbreak linked to the Islamic missionary movement *Tablighi Jamaat*, which led to Islamophobic reactions across media. The second opinion, “unity and brotherhood will help us fight the coronavirus”, is connected with Islam and with the Hindu nationalist party BJP. The third opinion, “the virus was created in a laboratory”, is related to theories about the creation of the virus, which often lead to conspiracy theories targeting Muslims in India ([The Guardian, 2020](#)). We exploit the random allocation of the interlocutor and measure agreement with opinions shared by citizens that are in-group or out-group relative to the respondent. Panel A in Figure D6 shows the average level of agreement with these opinions, while columns (1)–(2) in Table D10 provides estimates of treatment effects.

Second, we focus on trust in information shared by different people, including doctors and health experts, and other citizens. Panel B in Figure D6 shows the average level of trust, while columns (3)–(4) in Table D10 present estimates of treatment effects. Finally, we focus on risk perceptions. We measure this dimension using the following question “Do you think it is possible that someone from your household might at some point get sick with the coronavirus?”, 0 indicating that it is extremely unlikely and 1 it is extremely likely. Panel C in Figure D6 shows the average level of perceived risk of contagion and the

share of respondents reporting it is extremely unlikely. Columns (5)–(6) in Table D10 present estimates of treatment effects.

Figure D6: Average agreement with citizen’s opinions, trust in information and risk perception



Note. Average levels are measured using a re-scaled likert scale where 0 refers to low agreement/risk and 1 refers to high agreement/risk. Variables are defined in Appendix B.

Table D9: Preventive practices and response to misinformation: the effect of Hindu introductions and higher incentives

	Compliance		Beliefs over the efficacy of ...		Fact-checking		Agreement with	
	Recommended practices (1)	Face masks / hand-washing (2)	Physical distancing (3)	Vegetarianism (4)	Non-evidence-based practices (5)	In-group citizens (6)	Out-group citizens (7)	misinformation shared by... citizens (8)
I. EFFECT OF HINDU INTRODUCTION								
1A. Sample restricted to doctor message group								
Hindu greeting	0.065 (0.030) [0.03, 0.16]	0.004 (0.004) [0.23, 0.66]	-0.001 (0.005) [0.87, 0.87]	-0.009 (0.008) [0.24, 0.57]	0.007 (0.006) [0.27, 0.47]	0.004 (0.014) [0.75, 0.76]	-0.006 (0.009) [0.53, 0.78]	0.007 (0.008) [0.36, 0.72]
Mean (lower incentive)	-0.020	0.804	0.804	0.571	0.650	0.328	0.490	0.475
Observations	2519	3851	3849	3846	3849	3851	2587	3340
1B. Sample restricted to control group								
Hindu greeting	-0.025 (0.030) [0.40, 0.80]	-0.002 (0.004) [0.59, 0.84]	-0.007 (0.004) [0.12, 0.45]	0.010 (0.008) [0.21, 0.62]	-0.003 (0.007) [0.63, 0.63]	0.007 (0.016) [0.68, 0.70]	0.004 (0.011) [0.70, 0.90]	0.000 (0.008) [0.98, 0.98]
Mean (lower incentive)	-0.029	0.800	0.802	0.558	0.663	0.349	0.480	0.495
Observations	2560	3849	3849	3846	3848	3849	2595	3364
2. EFFECT OF HIGHER INCENTIVES								
2A. Sample restricted to doctor message group								
Higher incentive	0.025 (0.031) [0.43, 0.93]	-0.000 (0.003) [0.99, 0.99]	-0.002 (0.004) [0.61, 0.84]	-0.006 (0.008) [0.47, 0.92]	-0.005 (0.007) [0.50, 0.86]	-0.030 (0.016) [0.06, 0.16]	-0.006 (0.011) [0.61, 0.61]	-0.008 (0.009) [0.38, 0.61]
Mean (lower incentive)	0.003	0.806	0.805	0.567	0.656	0.344	0.488	0.482
Observations	2519	3851	3849	3846	3849	3851	2587	3340
2B. Sample restricted to control group								
Higher incentive	0.001 (0.032) [0.97, 0.97]	0.003 (0.004) [0.43, 0.78]	-0.000 (0.004) [0.96, 1.00]	-0.009 (0.008) [0.27, 0.69]	0.009 (0.007) [0.17, 0.56]	-0.019 (0.016) [0.23, 0.56]	0.013 (0.012) [0.29, 0.47]	-0.003 (0.008) [0.74, 0.74]
Mean (lower incentive)	-0.040	0.798	0.798	0.567	0.657	0.361	0.477	0.496
Observations	2560	3849	3849	3846	3848	3849	2595	3364

Notes. Estimates based on equation (4) in Panel 1 and on equation (5) in Panel 2. The samples are restricted to participants allocated to the doctor message (Panels 1A and 2A), or to participants allocated to the control group (Panels 1B and 2B). Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets, the first from individual testing, the second adjusting for testing that each treatment is jointly different from zero for all outcomes presented in each panel. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D10: Agreement with citizen's opinions, trust in information and risk perception

	Agreement with opinions shared by...		Trust in information shared by...		Risk perception	
	In-group citizens	Out-group citizens	Doctors and health experts	Other citizens	Risk of contagion	Contagion extremely unlikely
	(1)	(2)	(3)	(4)	(5)	(6)
A. Full sample						
Doctor message	0.007 (0.005) [0.180]	-0.005 (0.005) [0.277]	0.002 (0.003) [0.449]	-0.002 (0.004) [0.552]	0.002 (0.005) [0.693]	0.005 (0.010) [0.652]
Mean (control message)	0.617	0.622	0.801	0.685	0.259	0.342
Observations	6704	7700	7700	7700	7700	7700
B. Sample restricted to doctor message group						
Religion concordance	-0.000 (0.008) [0.958]	0.003 (0.007) [0.644]	-0.001 (0.006) [0.813]	-0.002 (0.005) [0.740]	0.006 (0.007) [0.440]	-0.005 (0.015) [0.727]
Mean (religion discordance)	0.625	0.615	0.804	0.683	0.257	0.351
Observations	3340	3851	3851	3851	3851	3851

Notes. Estimates based on OLS regressions using equation (1) in Panel A and on equation (2) restricting the sample to the doctor message group in Panel B. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

D.7 Estimates with controls using post-double selection LASSO and ANCOVA

Table D11 presents estimates of treatment effects of the doctor message and of the religion concordance treatments using ANCOVA specifications (i.e., using equation (1) and controlling for the baseline value of the dependent variable), while Table D12 provides estimates of treatment effects using the specification defined in equation (1) and including control variables selected with the post-double selection LASSO (PDSL) procedure Belloni et al. (2013); Tibshirani (1996). In the latter, the set of potential control variables include the following observable characteristics (all continuous variables are also included in their squared term and are standardized): individual characteristics described in Table C1; the slum-level averages of individual characteristics; the baseline value of outcome variables presented in Tables 2-4. Additional information about outcome variables is provided in Appendix B. In order to have the same sample size of estimates as in the main tables, missing values are replaced by the value 0 and an indicator variable equal to 1 if the observation had a missing value is introduced in the list of available controls. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D11: Compliance and response to misinformation, ANCOVA estimates

	Compliance	Fact-checking	Agreement with misinformation shared by...	
	Recommended practices (1)	(2)	In-group citizens (3)	Out-group citizens (4)
A. Full sample				
Doctor message	0.056 (0.021) [0.010]	-0.023 (0.010) [0.028]	0.004 (0.007) [0.548]	-0.016 (0.006) [0.006]
Mean (control message)	-0.041	0.352	0.483	0.494
Observations	5079	7700	5182	6704
B. Sample restricted to doctor message group				
Religion concordance	0.095 (0.032) [0.004]	0.006 (0.015) [0.678]	-0.023 (0.009) [0.010]	0.006 (0.008) [0.442]
Mean (religion discordance)	-0.035	0.326	0.497	0.476
Observations	2519	3851	2587	3340

Notes. Estimates based on OLS regressions using equation (1) and controlling for the baseline value of the dependent variable. When the dependent variable is missing at baseline, we impute it with the slum-level average value of the dependent variable at baseline. Panel B restricts the sample to participants allocated to the doctor message. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

D.8 Additional evidence on spillovers

Table D13 replicates estimates in Table 6 but using the allocation to treatments of the respondents' nearest neighbour as a measure for spillover.

Table D12: Preventive practices and response to misinformation, including controls (PDSL procedure)

	Compliance		Beliefs over the efficacy of ...			Fact-checking		Agreement with misinformation shared by...	
	Recommended practices (1)	Face masks / hand-washing (2)	Physical distancing (3)	Vegetarianism (4)	Indian immunity (5)	(6)	In-group citizens (7)	Out-group citizens (8)	
A. Full sample									
Doctor message	0.051 (0.022) [0.020]	0.007 (0.003) [0.009]	0.006 (0.004) [0.117]	0.002 (0.006) [0.733]	-0.006 (0.005) [0.165]	-0.024 (0.010) [0.019]	0.002 (0.007) [0.732]	-0.018 (0.006) [0.002]	
Observations	5079	7700	7698	7692	7697	7700	5182	6704	
B. Sample restricted to doctor message group									
Religion concordance	0.092 (0.032) [0.004]	-0.003 (0.004) [0.408]	-0.005 (0.004) [0.240]	-0.016 (0.008) [0.054]	-0.001 (0.007) [0.849]	0.004 (0.014) [0.790]	-0.024 (0.009) [0.007]	0.005 (0.008) [0.560]	
Observations	2519	3851	3849	3846	3849	3851	2587	3340	
C. Sample restricted to control group									
Religion concordance	-0.049 (0.031) [0.117]	-0.002 (0.004) [0.544]	-0.006 (0.004) [0.128]	-0.008 (0.008) [0.302]	-0.004 (0.006) [0.516]	0.018 (0.017) [0.279]	0.001 (0.011) [0.924]	0.003 (0.008) [0.742]	
Observations	2560	3849	3849	3846	3848	3849	2595	3364	

Notes. Estimates based on equation (1) including control variables using the PDSL procedure (Belloni et al., 2013; Tibshirani, 1996). Panel B restricts the sample to participants allocated to the doctor message, Panel C restricts the sample to participants allocated to the control group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables are defined in Appendix B. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).

Table D13: Preventive practices and response to misinformation, controlling for spillovers

	Compliance			Beliefs over the efficacy of ...			Fact-checking		Agreement with misinformation shared by...	
	Recommended practices (1)	Face masks / hand-washing (2)	Physical distancing (3)	Vegetarianism (4)	Indian immunity (5)	(6)	In-group citizens (7)	Out-group citizens (8)		
A. Full sample										
Doctor message	0.056 (0.021) [0.009]	0.006 (0.003) [0.011]	0.005 (0.004) [0.193]	0.001 (0.006) [0.805]	-0.007 (0.005) [0.124]	-0.023 (0.010) [0.027]	0.004 (0.007) [0.576]	-0.016 (0.006) [0.004]		
Doctor message (nearest neighbour)	-0.007 (0.023) [0.755]	0.002 (0.003) [0.478]	-0.002 (0.004) [0.588]	-0.006 (0.006) [0.331]	0.003 (0.005) [0.509]	0.013 (0.011) [0.238]	-0.010 (0.007) [0.179]	0.003 (0.006) [0.657]		
Mean (control message)	-0.041	0.799	0.799	0.563	0.661	0.352	0.483	0.494		
Observations	5079	7700	7698	7692	7697	7700	5182	6704		
B. Sample restricted to doctor message group										
Religion concordance	0.153 (0.041) [0.000]	-0.001 (0.005) [0.808]	-0.007 (0.006) [0.256]	-0.026 (0.012) [0.031]	0.001 (0.009) [0.873]	0.015 (0.020) [0.442]	-0.027 (0.013) [0.046]	-0.010 (0.012) [0.405]		
Religion concordance (nearest neighbour)	0.050 (0.039) [0.201]	0.001 (0.006) [0.918]	-0.006 (0.007) [0.386]	-0.009 (0.011) [0.395]	0.011 (0.010) [0.282]	-0.028 (0.020) [0.171]	-0.011 (0.014) [0.442]	-0.007 (0.012) [0.569]		
Mean (religion discordance)	-0.070	0.808	0.808	0.567	0.655	0.326	0.493	0.484		
Observations	1520	2319	2317	2315	2319	2319	1543	2018		

Notes: Estimates based on OLS regressions using equation (1) in Panel A and equation (2) restricted to the doctor message group in Panel B. *Nearest neighbour* is an indicator variable equal to 1 if the nearest neighbour is allocated to the correspondent group. Standard errors clustered at the slum level are reported in parentheses. *P*-values are presented in brackets. Dependent variables in columns (1)–(5) are defined in Table 2, while dependent variables in columns (6)–(8) are defined in Table 4. All specifications include indicator variables for data collection rounds, and strata indicators (city and religion of respondent).