Journal of Development Economics 153 (2021) 102747

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Short communication

Texts don't nudge: An adaptive trial to prevent the spread of COVID-19 in India[☆]

Girija Bahety [a], Sebastian Bauhoff [b,c,*], Dev Patel [d], James Potter [b]

[a] Tufts University, Department of Economics and the Fletcher School of Law and Diplomacy, United States of America
[b] Harvard TH Chan School of Public Health, United States of America
[c] Center for Global Development, United States of America
[d] Harvard University, Department of Economics, United States of America

ARTICLE INFO

Dataset link: https://doi.org/10.7910/DVN/0H5UGL

JEL classification:
D8
D83
D91
I1
I12
I15
O1

Keywords:
Information and communication
Health behaviors
Adaptive trial
COVID-19
SMS
India

ABSTRACT

We conduct an adaptive randomized controlled trial to evaluate the impact of a SMS-based information campaign on the adoption of social distancing and handwashing in rural Bihar, India, six months into the COVID-19 pandemic. We test 10 arms that vary in delivery timing and message framing, changing content to highlight gains or losses for either one's own family or community. We identify the optimal treatment separately for each targeted behavior by adaptively allocating shares across arms over 10 experimental rounds using exploration sampling. Based on phone surveys with nearly 4,000 households and using several elicitation methods, we do not find evidence of impact on knowledge or adoption of preventive health behavior, and our confidence intervals cannot rule out positive effects as large as 5.5 percentage points, or 16%. Our results suggest that SMS-based information campaigns may have limited efficacy after the initial phase of a pandemic.

1. Introduction

Preventive behaviors such as handwashing and social distancing are critical to containing the spread of infectious diseases like COVID-19, particularly in densely populated areas of developing countries with crowded living quarters and public spaces. As a pandemic unfolds, identifying ways to encourage the adoption of protective health behaviors in a timely, efficient, and cost-effective way is critical for public health (Van Bavel et al., 2020).

We examine the impact of text messages (or Short Message Service, SMS) on preventive health behavior through a multi-arm iterative randomized-controlled trial in rural India. Using a sample of phone numbers from birth registers at health centers in Saran district in the state of Bihar, we randomly sent some individuals four text messages over the course of two days. We ran two experiments in parallel during the first peak of the country's COVID-19 pandemic, between 17 August and 20 October 2020, one encouraging handwashing and the other social distancing. For each outcome – handwashing and social

☆ We are grateful to Suvita for sending the messages, especially Fiona Conlon and Nithya Nagarathinam, and to Dr. N. K. Sinha (State Immunization Officer) for support to conduct this study. Priyal Patil provided excellent research assistance. We thank Jenny Aker, Arielle Bernhardt, Emily Breza, Maximilian Kasy, Cynthia Kinnan, Gabriel Kreindler and Gautam Rao for helpful comments. This study underwent human subjects review at IFMR (IRB00007107), the Harvard TH Chan School of Public Health (IRB20-0706), Harvard University (IRB20-0791) and Tufts University (STUDY00000574). The analysis was pre-registered on the AEA social science registry as AEARCTR-0005780. Replication code and data can be found here. Funding for the project was provided by J-PAL South Asia at IFMR's Cash Transfers for Child Health Initiative. Bahety acknowledges summer research-funding support from the Economics Department at Tufts University, and Patel acknowledges support from a National Science Foundation Graduate Research Fellowship under grant DGE1745303. All errors are ours.
* Corresponding author.
E-mail addresses: girija.bahety@tufts.edu (G. Bahety), sbauhoff@hsph.harvard.edu (S. Bauhoff), devpatel@fas.harvard.edu (D. Patel), jpotter@g.harvard.edu (J. Potter).

https://doi.org/10.1016/j.jdeveco.2021.102747
Received 9 May 2021; Received in revised form 26 August 2021; Accepted 30 August 2021
Available online 25 September 2021

distancing – we test 10 treatment arms that vary across two dimensions that could influence treatment effectiveness: message frame and delivery timing. Informed by research in public health, psychology and behavioral economics, we consider five variants of message framing, changing content to highlight public gain or loss, private gain or loss, or neutral. We also varied the time of day when the messages were sent: either twice in the morning (7:00–8:00 a.m. and 10:00–11:00 a.m.) or once in the morning and once in the evening (7:00–8:00 a.m. and 6:00–7:00 p.m.).

Testing these large number of arms lends itself to an adaptive trial approach to efficiently recover the best policy. Following the exploration sampling algorithm of Kasy and Sautmann (2021), we reallocated treatment shares over the course of 10 rounds to identify the combination of timing and message framing that is most effective. This approach uses a modified Thompson sampling procedure to assign observations in each stage of the experiment to arms such that we achieve the best possible convergence rate to the optimal treatment. We conducted phone surveys with recipients three days after the first message was sent, along with surveys of control households who did not receive any messages. To mitigate concerns about experimenter demand, we measured preventive health behaviors first through an open-ended question and then via direct elicitation and a list experiment. We conduct inference using both the standard asymptotic approach and randomization statistical inference. For a sub-sample, we conducted phone surveys five days after the first message was sent to check for decay in treatment effects.

We find no evidence that any of our SMS-based information campaigns improve knowledge or adoption of social distancing and handwashing. This is true across several elicitation methods designed to address concerns of experimenter demand. We cannot, however, reject potentially meaningful treatment effects: Looking at all arms together for each target behavior, our confidence intervals allow us to reject direct impacts of 5.5 percentage points (p.p.) off of a control mean of 36% for adopting social distancing, and 5.6 p.p. off a base of 35% for adopting handwashing. We also find no evidence of indirect effects, e.g., of handwashing messages on social distancing behaviors, nor evidence of heterogeneous treatment effects by timing of the experiment round, literacy, and recall period, although our estimates for such impacts are imprecise.

Our study makes several contributions: First, our results are particularly policy-relevant to the new waves in COVID-19 in Spring 2021 which occur after extensive awareness about the disease. Second, our study builds on a growing number of studies in economics using adaptive approaches to allocate treatment shares in an experimental settings (Caria et al., 2020; Kasy and Teytelboym, 2020a, 2020b). We implement what is to the best of our knowledge, one of the first applications of the exploration sampling approach of (Kasy and Sautmann, 2021). Third, we contribute to a rich literature that has tested the potential for nudges and information to improve health behavior (e.g., Alatas et al., 2020; Bennear et al., 2013; Dupas, 2009; Madajewicz et al., 2007; Meredith et al., 2013). Fourth, we also add to research on optimal message framing by cross-randomizing gain or loss message framing with public or private framing and comparing it with a neutral message to understand their marginal impacts. Much of the evidence on information campaigns during COVID-19 from developed countries shows mixed results regarding the importance of these specific design features of messages (Jordan et al., 2020; Favero and Pedersen, 2020; Falco and Zaccagni, 2020). Fifth, we also examine whether the delivery timing impacts the efficacy of information or nudges (Kasy and Sautmann, 2021).

Finally, our study also adds to recent research using phone-based information campaigns to encourage preventive health behavior for COVID-19 in South Asia. Banerjee et al. (2020) randomly sent video links to households in West Bengal in May 2020 and measure both the direct and spillover effects on respondents who directly received the link or may have learned about it through their networks. They document large positive overall impacts on social distancing, handwashing, and hygiene behaviors. Our confidence intervals on handwashing include their point estimates. They do not find private or public gain framed messages to have differential effects and do not find impacts on knowledge of symptoms or precautions. In another study in Uttar Pradesh and Bangladesh in April 2020, Siddique et al. (2020) find significant impacts on both COVID-19 knowledge and behavior from phone calls plus sending messages and phone calls alone relative to just sending SMS. Armand et al. (2021) randomly sent a WhatsApp video or audio recordings of messages from doctors to the urban poor in Uttar Pradesh and find decreased probabilities of leaving the slum area but no effects on handwashing. In rural Bangladesh, Chowdhury et al. (2020) find that information campaigns on social distancing and hygiene measures improved preventive behaviors.

There are several potential explanations for the lack of impacts in our study. First, our study takes place several months into the pandemic, at a time when cases were spiking and after households across India had received COVID-19 messages and endured lockdowns for several months. At this stage, households may have already been well-informed or too fatigued to respond; they may also have faced higher opportunity costs of adopting preventive behaviors like social distancing after the prolonged economic disruption of the pandemic. Second, our intervention featured relatively plain SMS, without any celebrity or professional endorsements nor the video or audio components of other studies. Text messages are accessible even on basic phones, potentially expanding access in the context of only 24% smartphone penetration in India (Rajagopalan and Tabarrok, 2020). Although SMS-based information campaigns require literacy and may be relatively less engaging (Favero and Pedersen, 2020), they can be effective at changing health behaviors (Orr and King, 2015; Armanasco et al., 2017). Third, we focus on a different elicitation approach to measure compliance with health behavior. To reduce risk of experimenter demand, we asked respondents to list all actions they are taking to protect against the virus. By contrast, Siddique et al. (2020) directly ask respondents if they wash their hands or maintain distance. We do elicit second-order beliefs on community behaviors like in Banerjee et al. (2020) but find no evidence of treatment effects on reported community-level social distancing and handwashing in our experiment. Fourth, while Banerjee et al. (2020) randomize across communities, our experiment varied treatment at the individual level. If there are significant spillovers from our treatment like in Banerjee et al. (2020), they will attenuate our point estimates.

The remainder of this paper proceeds as follows: Section 2 describes the setting of the experiment, study sample, intervention, the adaptive trial design, and primary data collection. Section 3 describes the empirical strategy and main results, and Section 4 discusses implications for health messaging campaigns, especially in the context of pandemics.

2. Study context and design

As of August 17, 2020, India had recorded a cumulative total of 2.7 million confirmed COVID-19 cases (Roser et al., 2020) and true infection rates were an order of magnitude higher (Mohanan et al., 2021). We conducted our study in collaboration with the Bihar state government and our NGO partner, Suvita, during the initial height of the country's pandemic between August 17 and October 20, 2020. The experiment took place in Saran, a rural district in the western part of Bihar that resembles the state's overall socio-economic characteristics and pandemic experience.[1] New cases increased in early July and peaked

[1] The study design was pre-registered at the AEA Registry as AEARCTR-0005780 but for two additional health behaviors, a smaller sample size, and in the state of Maharashtra. We changed the study location and number of behaviors in response to local conditions, but all remaining aspects of the pre-registration still apply.

in early August, just before our study started (Appendix Figure A.1). Bihar imposed a full lockdown in mid-July and maintained a partial lockdown in August that remained effective during the first three weeks of our trial. Throughout the pandemic, public service announcements advocating hygiene and social distancing to combat COVID-19 were widely distributed via television, radio, newspapers, and text messages. We are unable to test how the relative timing of the pandemic impacted the effectiveness of our intervention, but we expect essentially everyone in our sample to have been exposed to some messaging already about the benefits of handwashing and social distancing. Although cell phone ownership is almost universal among households in Bihar, this may overstate the potential scope for SMS campaigns: according to the 2011 census, just 67% of women and 82% of men in Saran were literate.

Study sample. Our sample was recruited from a list of households who entered phone numbers into birth registries at health centers in 15 out of 20 blocks in Saran between August 2019 and February 2020.[2] Although the phone numbers come from birth registers, the subjects of our intervention and surveys are the users of the phones. Our sample of respondents is comparable to the population of Bihar on basic characteristics (Table A.1). However, our sample is younger than the average adult Bihari.

We randomly selected phone numbers within four strata based on block characteristics from the 2011 Census: above and below average literacy rate and above and below average proportion of Scheduled Castes and Scheduled Tribes (SC/ST) population. Table 1 shows the summary statistics and balance across treatment and control groups for key demographic characteristics in Panel A, SMS-related information in Panel B, and knowledge of symptoms and access to health care in Panel C. About three-quarters of the respondents were male with an average age of 31 years. Less than a third of the sample was unemployed, and most of those who worked did so in a manual job. Eighty-six percent of respondents can read SMS in Hindi, but 36% do not ever read text messages. Less than a third read SMS daily in the week prior to the interview. Knowledge of COVID-19 symptoms and practice of ante-natal care is balanced across treatment and control.

Intervention design. Our trial compares 10 message types varying in framing and timing to target social distancing and handwashing. Each treated phone number was sent four text messages in Hindi over the course of two days. We chose five different message frames based on principles from psychology and behavioral economics (Tversky and Kahneman, 1979, 1991; Van Bavel et al., 2020): neutral, public gain or loss, and private gain or loss. These message frames may appeal to different emotions, such as fear (by making the threat of pandemic salient) or prosocial motivation (by highlighting externalities of the preventive actions). Table A.2 shows the different messages by content framing for each behavior. The neutral messages give simple, directed advice: for social distancing, the neutral message states "Coronavirus is here. Outside the house, keep a distance of at least two arms from others". For handwashing, the neutral message is "Coronavirus is here. Before touching any food or touching your face, wash your hands with water and soap". In the public loss arms, appealing to both fear and prosocial motivation, the first sentence is replaced with "Coronavirus kills. Your action can put your community at risk of infection". In the public gain arms, the first sentence is instead replaced with "Save lives. Your action can protect your community from coronavirus". The private gain and loss arms are the same as the public gain or loss arms, except "community" is replaced with "family".

The delivery timing can also impact the efficacy of information or nudges. We expected most households to be less busy and more likely to read SMS if they were delivered to them either at the start of the day or later in the evening. Moreover, Kasy and Sautmann (2021) find success with sending messages in the morning. Hence, to explore this issue further, we vary the time of the day when messages were sent across treatment arms. In all arms, the first message was sent between 7:00 and 8:00 a.m. in the morning. In twice-morning arms, a second SMS was sent between 10:00 and 11:00 a.m., while in morning–evening arms, the second message was sent between 6:00 and 7:00 p.m. Overall, we create 10 treatment arms using each of the five framings for both delivery timings.

Experimental design. We randomly assigned our treatment sample to 10 rounds of treatment for each behavior. We implemented the "exploration sampling" procedure from Kasy and Sautmann (2021) to allocate sample to the different treatment arms over the course of the experiment. An adaptive approach is particularly appealing in this setting because it provides more statistical power in identifying optimal treatment over a large set of alternatives. The "exploration sampling" method uses a modified Thompson sampling procedure to iterate over the different messaging campaign attributes to identify those that are most effective in shaping reported behavior. In each phase of the experiment t, the probability that a unit is assigned to arm j is given by q_t^j, as defined in Eq. (1), where p_t^j is the posterior probability that arm j is optimal given outcomes up through period $t-1$.

$$q_t^j = \frac{p_t^j \cdot \left(1 - p_t^j\right)}{\sum_j p_t^j \cdot \left(1 - p_t^j\right)} \tag{1}$$

While traditional Thompson sampling procedures weight by p_t^j, this modification shifts weight towards the close competitors of the best performing arms. Indeed, exploration sampling is equivalent to Thompson sampling if the same treatment assignment is never assigned twice in a row. The key advantage of exploration sampling is that it achieves the best possible exponential rate of convergence subject to the constraint that in the limit, half of observations are assigned to the best treatment, thereby converging much faster than Thompson sampling or non-adaptive assignment. This particular approach does not have an explicit stopping rule, and thus we decided on 10 rounds based on the budget for conducting the surveys. We could have continued to run the experiment to obtain more precise estimates of the treatment effect for each arm. We find that the treatment shares assigned by the algorithm stabilize by the second half of the experiment as shown in Figure A.2, suggesting that 10 rounds were sufficient for identifying the optimal treatment. We used open-ended reported practice of either social distancing or handwashing as our main outcome for the corresponding target behavior to adapt shares (described further below).

Although we intended to begin with equal priors across all arms (and therefore equal shares), due to an initial coding mistake, allocations were not matched to the correct arms when the messages were sent for the first several weeks. In practice, this means that when the algorithm began being implemented correctly, some arms (randomly) had more observations upon which to form a prior about their effectiveness. This can be seen visually in Figure A.2 by the fact that all lines are not beginning at 10 percent in the first round. This error does not affect the validity of our treatment effect estimates. However, the error could inhibit our ability to identify the most effective arm because, due to the error, the initial shares were assigned randomly rather than optimally.[3] We see no systematic evidence of this concern in practice, as the algorithm settles on arms by about round 6, after which it consistently identifies the same arms as performing relatively better. Moreover, these arms are not systematically correlated with those arms which were randomly assigned more observations in round 1.

In addition to these treatment arms, we include a pure control group that received no message. This design choice allows us to both test

[2] According to the 2019–2020 National Family Health Survey (NFHS-5), 75% of births in rural Bihar take place at a health facility.

[3] We explicitly account for this mistake in our simulated treatment assignments while conducting randomization inference.

Table 1
Summary Statistics and Balance Tests.

	Control			Treatment			Difference	
	Mean	S.D.	N	Mean	S.D.	N	*Δ*	S.E.
Panel A: Demographics								
District - Saran	0.81	0.39	1,087	0.81	0.39	2,863	−0.00	0.014
Location - Town	0.06	0.24	1,079	0.06	0.24	2,845	0.00	0.009
Location - Rural Area	0.78	0.42	1,079	0.78	0.41	2,845	0.01	0.015
Age	30.66	11.07	1,047	31.10	11.10	2,794	0.42	0.402
Male	0.75	0.43	1,051	0.71	0.45	2,788	−0.04***	0.016
Finished secondary school	0.17	0.38	1,034	0.17	0.38	2,763	−0.00	0.014
More than secondary school	0.28	0.45	1,034	0.28	0.45	2,763	−0.00	0.016
Unemployed	0.28	0.45	1,026	0.31	0.46	2,746	0.03**	0.017
Manual job	0.34	0.47	1,026	0.32	0.46	2,746	−0.02	0.017
Scheduled Castes	0.18	0.38	768	0.20	0.40	2,142	0.02	0.016
Other Backward Classes	0.63	0.48	768	0.59	0.49	2,142	−0.04**	0.020
Hindu	0.92	0.27	789	0.90	0.30	2,179	−0.02*	0.012
Muslim	0.08	0.27	789	0.10	0.30	2,179	0.02	0.012
Own phone	0.92	0.27	872	0.93	0.26	2,350	0.00	0.011
Panel B: SMS-related								
Can read SMS in Hindi	0.86	0.35	862	0.86	0.35	2,340	−0.00	0.014
Trust information on SMS	0.96	0.21	90	0.92	0.26	702	−0.03	0.024
Did not read any SMS	0.38	0.49	784	0.35	0.48	2,143	−0.03	0.020
Read SMS daily	0.30	0.46	784	0.29	0.45	2,143	−0.02	0.019
Any SMS delivered (Admin)				0.72	0.45	2,779		
# SMS delivered (Admin)				2.72	1.77	2,779		
Panel C: Health								
Know symptom: Fever	0.77	0.42	1,034	0.75	0.43	2,761	−0.02	0.015
Know: Cough	0.79	0.41	1,034	0.76	0.43	2,761	−0.02	0.015
Received Antenatal Care	0.78	0.42	108	0.75	0.44	292	−0.03	0.048
Child immunized	0.92	0.27	439	0.90	0.30	1,221	−0.02	0.015
Joint significance F-test								
Panel A (*p-value*)	0.256							
Panel B (*p-value*)	0.204							
Panel C (*p-value*)	0.431							

Table 1 presents summary statistics for the control and pooled treatment group for key demographic characteristics in Panel A, SMS-related characteristics in Panel B and knowledge of COVID symptoms and access to routine health care in Panel C. The difference between treatment and control (*Δ*) is shown in Column 7 (controlling for strata) with standard errors (S.E.) in Column 8. We also report the *p*-value for joint significance tests for the variables in Panels A, B and C. Admin in Panel B refers to SMS delivery reports from the telecommunications provider. The statistics on ante-natal care and child immunization is reported only for those households which had a pregnant woman and a child under 1 year of age, respectively. * p<0.1; ** p<0.05; *** p<0.001.

the "behavioral" phrasings against a neutral framing as well as the efficacy of any SMS against none. Table 1 shows that the treatment and control groups are balanced on most demographic characteristics. The treatment group has more women, more unemployed respondents, and fewer households identifying as Other Backward Classes and Hindu. The bottom panel of the table shows that the joint test of significance for these covariates are not statistically significant. However, we control for gender, occupation, education, and age fixed effects in all treatment effect specifications.

Data collection. Three days after the first text message was sent, a team of enumerators called respondents over the phone.[4] If the phone number was not answered, then the enumerators repeatedly redialed after the full list was tried once. We find no evidence of overall differential response rates by treatment status (*p*-value = 0.567 for social distancing and 0.627 for handwashing) (results not shown).[5] A random subset of phone numbers were called five days later instead of three to test for potential decay effects. We had expected that if there had been treatment effects, they would fade at some point. Moreover, the five-day delay worked well with the survey schedule for the adaptive trial. We staggered the phone surveys for control, social distancing and handwashing samples to facilitate the updating of the treatment shares over the frequent iterations. Out of a total of 12,799 phone numbers called, we had a response rate of 34.7%, of which – conditional on answering the call – 8.9% did not consent to the interview and 0.62% were under the age of 18 years and were excluded from the sample. Of 3,964 eligible respondents who consented, 91.6% of respondents answered key outcome questions, and 74.9% completed all questions in the survey. We use all available answers for our analyses.[6] The survey covered basic respondent and household characteristics, phone usage behavior, risk perceptions, and knowledge and action regarding COVID-19 prevention.

Given concerns about experimenter demand, we elicited key preventive health behaviors using an open-ended (unprompted) question: "What are you doing to protect against the virus?". We classify compliance with social distancing and handwashing based on whether the

[4] There were public/national holidays on August 21, September 17 and October 2. We did not send SMS on holidays and the day before and made no calls on holiday and the following two days.

[5] We also find no difference in consent rates for the social distancing arm (*p*-value = 0.608), though those who received a handwashing message were 2.7 p.p. less likely to consent conditional on answering (*p*-value = 0.058) (results not shown). This could be suggestive evidence that the SMS had a discouraging effect on engagement with the research team.

[6] We made an average of 1.4 attempts per number, ranging from 1 to 8 attempts. Out of the phone numbers from which we did not get a response, 24% did not pick up despite multiple attempts, 22% were either unreachable or switched off and 17% were invalid numbers. The median completed interview lasted 17 min.

Table 2
First-Stage Results on Self-Reported Receipt of COVID-Related SMS.

	Any SMS	# SMS	SD SMS | Any SMS	SD SMS | Any SMS	HW SMS | Any SMS	HW SMS | Any SMS
Pooled treatment	0.286**	1.045**	0.038		0.122*	
	(0.026)	(0.137)	(0.061)		(0.050)	
Treatment - SD				0.136		
				(0.093)		
Treatment - HW						0.209**
						(0.059)
R^2	0.21	0.22	0.23	0.23	0.16	0.16
N	1,988	1,949	791	791	773	773
Control Mean	0.16	0.68	0.27	0.19	0.09	0.13
F-statistic	118.43	58.10	0.38	2.17	5.86	12.41

Table 2 shows the first stage results for four self-reported measures of receipt of any COVID-related SMS: any SMS, number of SMS received in Column 1 and 2, recall of social distancing in Column 3 and 4 and handwashing messages in Column 5 and 6, respectively. The last four measures are conditional on receiving any COVID-related SMS. The regressions include fixed effects for gender, occupation, education, age, target behavior, block, day of the week, round of the experiment, enumerator, and (random) order of the knowledge and action question for the key outcomes. Robust standard errors in parentheses. Asymptotic *p*-values are denoted by: * p<0.1; ** p<0.05; *** p<0.001.

respondent mentions each practice, respectively, and use this indicator to guide our adaptive trial.[7] We also elicited knowledge about preventive measures with a similar open-ended question. The order of the knowledge and practice questions was randomized across the respondents. We subsequently directly asked respondents whether they practice social distancing and handwashing.[8] We also conducted a list experiment to measure uptake of behaviors sensitive to social desirability bias by bundling (or veiling) the sensitive questions with two other innocuous statements (Chuang et al., 2020; Jamison et al., 2013; Karlan and Zinman, 2012). We asked respondents how many actions they did in the past two days: watching TV, speaking on the phone, and either social distancing or handwashing. Finally, we administered a randomly selected subset of three questions from the 13 statements in the Marlowe–Crowne Scale Social Desirability Scale (Form C) which assesses correlates to social desirability bias on measured self-reported outcomes (Crowne and Marlowe, 1960; Reynolds, 1982; Dhar et al., 2018). We used a subset of the full scale to reduce the survey length.

3. Empirical strategy and results

We estimate treatment effects with the ordinary least squares specification shown in Eq. (2):

$$Y_i = \alpha_i + \beta T_i + \gamma O_i + \mathbf{X}_i' \lambda + \varepsilon_i \tag{2}$$

where Y_i is the outcome for individual i, T_i indicates treatment for the target behavior (either social distancing or handwashing), O_i indicates treatment for the other behavior, and $\mathbf{X}_i$ is a vector of fixed effects including gender, occupation, education, age, target behavior, block, day of the week, round of the experiment, enumerator, and (random) order of the knowledge and action question for the key outcomes. Our results are robust to regressions specifications without any controls or with only strata fixed effects. We include both treatment groups in all specifications: for each *target* outcome or behavior – social distancing and handwashing – we include a separate treatment indicator for those who were treated for the *other* behavior (Muralidharan et al., 2019).

Table 3
ITT Results by Pooled Treatment.

	Distancing		Handwashing	
	Know	Act	Know	Act
Treatment - SD	−0.002	−0.003	−0.035	−0.051*
	(0.030)	(0.029)	(0.029)	(0.029)
	[0.957]	[0.928]	[0.206]	[0.056]
Treatment - HW	−0.001	0.018	0.034	0.002
	(0.028)	(0.027)	(0.027)	(0.028)
	[0.954]	[0.448]	[0.158]	[0.943]
Adjusted R^2	0.09	0.05	0.05	0.05
N	3,563	3,563	3,563	3,563
Control Mean	0.49	0.36	0.32	0.35

Table 3 shows the ITT results by pooled treatment for the four main outcomes. The regressions include fixed effects for gender, occupation, education, age, target behavior, block, day of the week, round of the experiment, enumerator, and (random) order of the knowledge and action question for the key outcomes. Robust standard errors are in parentheses and Fisher exact *p*-values are in square brackets. Asymptotic *p*-values are denoted by: * p<0.1; ** p<0.05; *** p<0.001.

This allows us to test for evidence of "attention" substitution – being reminded about one behavior takes attention away from another – or substitution in health practices—if social distancing more makes people feel like handwashing is less necessary, for instance. We use the same control group across both targeted behaviors. Unless otherwise noted, we conduct analysis on the sample of treated respondents who were reached three days after the first message was sent. We interviewed control group respondents throughout the study period, on alternate days.

Sample averages in adaptive trials are typically biased (Hadad et al., 2019), but under exploration sampling, this bias is negligible in large samples because assignment shares of sub-optimal treatments are bounded away from zero (Kasy and Sautmann, 2021). Thus, as long as the law of large numbers and central limit theorem apply, we can run standard t-tests ignoring the adaptivity. Because some of our treatment shares end up being quite small for some arms, in addition to asymptotic standard errors that are robust to heteroskedasticity, we also report exact *p*-values from randomization inference (using a two-tailed comparison). This inference approach is particularly appealing in heterogeneous treatment effect specifications that can be vulnerable to high leverage observations given the asymmetric treatment shares (Young, 2019). To conduct the randomization inference, we randomly allocate treatment status holding constant the observed distribution of outcomes. We recreate the full adaptive trial data-generating process to create the synthetic treatment allocations. Holding constant the initial strata, we randomly re-assign the initial treatment shares and then re-run the Bayesian process to generate

[7] For distancing, we use keeping two arms distance from others. For handwashing, we include both washing hands with water and washing hands with soap regularly as the main outcome.

[8] The prompted question for distancing asks "Have you come into close contact with anyone not in your household, that is within 2 arms distance or less? For example, when you went to meet someone in a group or for a meeting, get-together, or to go to the market or shopping". The corresponding question for handwashing asks "Have you washed hands with soap and running water, or used hand sanitizer?" Table 5 compares our definitions of outcomes to those used in related studies.

Table 4
Summary Statistics by Treatment Arm.

Behavior	Framing	Timing	N	μ_j	σ_j	p^j
Distancing	Neutral	2× morning	69	0.310	0.054	0.019
Distancing	Public gain	2× morning	81	0.325	0.051	0.027
Distancing	Public loss	2× morning	79	0.309	0.051	0.014
Distancing	Private gain	2× morning	14	0.125	0.080	0.003
Distancing	Private loss	2× morning	199	0.373	0.034	0.067
Distancing	Neutral	Morn./Even.	142	0.396	0.041	0.224
Distancing	Public gain	Morn./Even.	106	0.343	0.045	0.034
Distancing	Public loss	Morn./Even.	46	0.292	0.065	0.022
Distancing	Private gain	Morn./Even.	285	0.425	0.029	0.575
Distancing	Private loss	Morn./Even.	73	0.307	0.053	0.015
Handwashing	Neutral	2× morning	133	0.422	0.042	0.127
Handwashing	Public gain	2× morning	176	0.461	0.037	0.431
Handwashing	Public loss	2× morning	77	0.354	0.053	0.018
Handwashing	Private gain	2× morning	137	0.439	0.042	0.223
Handwashing	Private loss	2× morning	127	0.426	0.043	0.152
Handwashing	Neutral	Morn./Even.	113	0.348	0.044	0.005
Handwashing	Public gain	Morn./Even.	51	0.302	0.062	0.005
Handwashing	Public loss	Morn./Even.	114	0.379	0.045	0.025
Handwashing	Private gain	Morn./Even.	88	0.300	0.048	0.001
Handwashing	Private loss	Morn./Even.	103	0.362	0.047	0.013

Table 4 presents summary statistics by treatment arm. Column 1 denotes the target behavior. Column 2 denotes the framing of the message. Column 3 denotes the timing of when the SMS were sent. Column 4 lists the total number of treated respondents who were reached in a phone survey. Column 5 (μ_j) shows the mean outcome for each arm, and Column 6 (σ_j) presents the standard deviation. Column 7 (p^j) lists the posterior probability that each arm is the optimal arm at the conclusion of the experiment. For calculating posterior probabilities, the above samples were restricted to respondents who (1) consented to the interview, (2) were at least 18 years old, (3) were assigned to a 3-day recall period, and (4) had a recorded response for the outcome variable (total N=2,283).

future shares for each round. We do not adjust for multiple hypothesis testing, as our estimates are mostly not statistically significant even without this adjustment.

First-stage. We assess implementation fidelity by assessing whether the treatment message was delivered to the targeted recipients, as shown in Panel B of Table 1. Beginning on August 24th, 2020, a week after the experiment began, we received reports from the telecommunications provider on whether the message was delivered to the recipients' phones. Within the treatment group, on average, 72% of the respondents successfully received at least 1 message. On average and unconditional on receiving any SMS, the treatment group received 2.7 messages out of a total of four messages that were sent. Non-delivery and partial deliveries are likely due to phones being switched off or service interruptions. Almost all respondents in the treatment and control groups stated that they trusted the information in messages related to the Coronavirus.

We also compare treatment compliance by assessing self-reported measures of whether the respondent received any COVID-related SMS in the week prior to the survey and, conditional on having received any SMS, the number of SMS received and their recall of message content (Table 2). Column 1 shows that treated respondents were 28.6 p.p. more likely to report receiving any SMS related to COVID-19 in the previous week, off a base of 16% in the control group. Column 2 shows that the number of COVID-related messages the treatment group report receiving is about one. Treated households who received handwashing messages are 21 p.p. more likely to remember that specific guidance relative to control households who also received COVID content, as shown in column 6. The comparable effects for social distancing are noisier but still positive (13.6 p.p.), as shown in column 4. These results are consistent with our finding in Panel B of Table 1 where more than a third of the respondents across the treatment and control groups did not read any SMS at all in the week prior to the survey. For comparison, Banerjee et al. (2020) found an average viewing rate of 1.14% for their YouTube videos, while Armand et al. (2021) estimate that respondents on average listened to 19%–23% of WhatsApp messages.

Treatment effects on primary outcomes. Overall, we find no evidence that sending SMS increased uptake of social distancing and handwashing. First, we show results for treatment arms pooled together for each targeted behavior in Table 3 (results also shown in Appendix Figure A.3). Looking at the social distancing arm in the top row of the table, the observed treatment effects on both knowledge and uptake of social distancing are small, negative (a decrease in 0.2 and 0.3 p.p., off of a control mean of 49% and 36%, respectively), and not statistically significant based on either asymptotic or randomization exact p-values. Similarly for the handwashing arm in the bottom row of the table, the treatment effect on knowledge is 3.4 p.p. off of a control mean of 32%. The impact on uptake of handwashing is 0.2 p.p. off of a control mean of 35%. Both are statistically insignificant across both inference methods. Our 95% asymptotic confidence intervals are large enough such that we cannot rule out direct effects as large as 5.5 p.p. for each of our main behaviors. We find no systematic evidence of treatment effects from messages targeting one behavior on the other, although there is suggestive evidence of a negative effect of the social distancing messages on the uptake of handwashing (a decline of 5.1 p.p., statistically significant at the 10% level). However, this could be due to statistical chance. Using the randomized treatment assignment as an instrument, we find no evidence of positive treatment effects for either behavior, as shown in Table A.3. This is true both when using administrative delivery reports on text message receipt as the endogenous variable in a treatment-on-the-treated (TOT) specification (Panel A) or self-reported receipt of any COVID-related message in an instrumental variable (IV) specification (Panel B), suggesting that even among the "compliers" who did receive and recall the SMS, the content had no impact on uptake of preventive measures.

The null effects on practice of these behaviors may not be surprising given the lack of impact on stated awareness. We consider our intervention to be a nudge or reminder about handwashing and social distancing as opposed to providing new information. This aligns with the findings by Banerjee et al. (2020) that already in March 2020, many months before our experiment, respondents in West Bengal had heard about social distancing 20.2 times and washing hands 16.9 times in the previous two days alone. Thus, we interpret the outcome we refer to as "knowledge" as capturing some measure of awareness that we hope might be spurred by our text messages.

We evaluate the treatment effects by pooling treatment arms within the five frames in Table A.4 and within delivery timings in Table A.5. We find no systematic evidence of any impact of different framings or timings on behavior. Table A.6 presents treatment effects separately for each of the 10 treatment arms and shows no consistent effects. The few statistically significant point estimates across these specifications are likely due to chance.

Optimal message design. Taken together, the previous set of results suggests no consistent or compelling evidence that a particular framing or timing was especially effective in increasing preventive health behavior. The few statistically significant effects we document are largely not replicated for the other behavior and could simply be the artifact of statistical noise. We can more formally explore the optimal message design using the insights from the adaptive procedure. Comparing each of our treatment arms against one another, we calculate the posterior probability p^j that each arm is optimal. Despite the lack of treatment effects, the exploration sampling approach did converge towards a small number of specific framings and timings as shown in Figure A.2. We present the posterior probabilities in Table 4. For social distancing, the private gain framing messages sent once in the morning and once in the evening were optimal with probability 0.575. For handwashing, the public gain messages sent twice in the morning were optimal with probability 0.431. Disaggregating the treatment effects using Eq. (2) by 10 treatment arms in Table A.6 suggests that relative to the control mean of 36% for social distancing and 35% for handwashing, these arms stand out with statistically significant treatment effects among

Table 5
Comparison of Related Studies.

	Bahety et al. (2021)	Banerjee et al. (2020)	Siddique et al. (2020)	Armand et al. (2020)
Study design				
Setting	Bihar	West Bengal	Uttar Pradesh (& Bangladesh)	Uttar Pradesh
Intervention	Aug–Oct 2020	May 2020	Apr–May 2020	[e]
Sample size	3,563	1,883	1,680 (India)	3,991
Target behaviors	Distancing; Handwashing	Distancing; Hygiene	Distancing; Handwashing; Respiratory hygiene	Debunking fake news; COVID protocols[e]
Treatment arms	Ten per behavior[a]; Only SMS	Four[b]; Only SMS with link to a celebrity video	Three {SMS only[c], Phone calls only, SMS & Phone calls}; Once a month in April and May; No pure control arm	Four[d] {WhatsApp Doctor video & WhatsApp Doctor Audio}; Message debunking fake Bollywood news to control arm
Respondents	Registered at health clinic 6–12 months prior	Current and former village council members	Households previously surveyed by two local organizations	Census of households within mapped slum borders from 2017
Called after	3 and 5 days	2–14 days	1 month & 2 months	[e]
Treatment compliance	Received any COVID-related SMS; +28.5 p.p. (1.78 times more than control) (*p*-value <0.05)[f]	Video viewing rate; 1.14%[e]	[e]	Audio message listened (%); −16 p.p. for low &−12 p.p. for low incentive
First-order outcomes (respondent behavior)				
Distancing	*Open-ended:* Maintain 2 arms distance −0.3 p.p. (*p*-value=0.511)[f] *Direct:* Never came within 2 arms distance when meeting someone outside of household. +0.4 p.p. (*p*-value=0.525)	*Direct:* Number of non-household people who came within 2 arms distance in the last 2 days −1.47 (*p*-value=0.206) *Direct:* Went outside the village (yesterday & day before yesterday) −7.4 p.p. (*p*-value=0.026)	*Direct:* No close contact with outsiders at least on 3 separate days in the past week +85.3 p.p. (*p*-value=0.001) for calls-only; +94.8 p.p. (*p*-value=0.001) for calls+SMS[g]	*Direct:* Received visitors last week[e] −6 p.p.(*p*-value >0.1) for low; +2 p.p. (*p*-value >0.1) for high *Direct:* Left slum last week[e] −3 p.p.(*p*-value >0.1) for low; −8 p.p. (*p*-value <0.05) for high
Handwashing	*Open-ended:* Wash hands with water and/or soap +0.2 p.p. (*p*-value=0.362)[f] *Direct:* Always washed hands in the last two days +0.7 p.p. (*p*-value=0.283)	Not collected	*Direct:* Washed hands five times in a day at least on 3 separate days in the past week +80.2 p.p. (*p*-value=0.001) for calls-only; +92.6 p.p. (*p*-value=0.001) for calls+SMS[g]	*Open-ended:* Number of correct hand-washing practices reported[e] −0.07 (*p*-value >0.1) for low ; +0.07 (*p*-value >0.1) for high
Second-order outcomes (community behavior)				
Distancing	*Direct:* Typical community member maintained 2 arms distance −0.7 p.p. (*p*-value=0.531)	Not collected	Not collected	Not collected
Handwashing	*Direct:* Typical community member washed hands with soap and water or used hand sanitizer +2.2 p.p. (*p*-value=0.125)	*Direct:* Times a typical person in the village washed hands with soap upon return (%) +4.7 p.p. (*p*-value=0.044)	Not collected	Not collected

[a] 5 frames ×2 delivery times.
[b] Two motivation frames (externality+internality; internality-only) crossed with two ostracism frames (no ostracism; neutral).
[c] The SMS scripts have private gain framing.
[d] Cross-randomized video & audio message with high-incentive or low-incentive lottery.
[e] Detailed data not available.
[f] First-stage results from Table 2; treatment effects on targeted behavior from Table 3.
[g] Using Fisher exact *p*-values from their table A5. Estimates are relative to SMS-only arm, and only for India sample.

other modifications of the messages. It is unclear why the optimal message characteristics are so different between the two behaviors. One possibility is that the individuals on the margin for handwashing and social distancing are different in ways that impact which messages are more effective. The optimal odds could also still evolve if the experiment were to have continued for longer. We note, however, that after 10 rounds, even the best arms are not effective at changing behavior relative to no message. Overall, the results do not highlight a clear recommendation for a single SMS design for other campaigns.[9]

Experimenter demand effects. One challenge in measuring preventive health behavior in this setting is experimenter demand: respondents may report practicing social distancing or handwashing simply because they are aware they are expected to be doing so. For this reason, our primary outcomes use responses to the unprompted elicitation of behaviors or practices respondents are taking to prevent COVID-19, based on the idea that if people are not directly asked whether they are practicing a behavior, their answers will be more accurate. Table A.7 presents correlations across each measurement of our primary outcomes within the control group, and – consistent with experimenter demand – the direct elicitation approaches yield considerably higher rates of both social distancing and handwashing. For both outcomes, the correlations between our main measure and the other elicitation approaches are low, which we interpret as evidence that experimenter demand may be of particular concern. Perhaps most surprisingly, the

[9] This is true even before implementing potential adjustments to account for the Winner's curse in this type of experiment (Andrews et al., 2021; Banerjee et al., 2021).

correlation between second-order beliefs about one's community and one's own response to the unprompted elicitation is not significant for either behavior. We view this as evidence that the community questions like those asked in Banerjee et al. (2020) are potentially capturing meaningfully different variation than true individual practice of preventive health measures.

To explore these issues, we compare our preferred measure to alternative elicitation approaches in Table A.8. Within each behavior, the first two columns show effects on the open-ended question (our preferred outcome) and the direct elicitation measure, respectively. The third column presents outcomes from our list elicitation, in which we embedded the behavior of interest with additional statements about whether the respondent watched television yesterday or talked to a relative on the phone yesterday. Respondents answered how many of the statements (out of three) applied to them. Following Banerjee et al. (2020), the fourth column reports impacts on second-order beliefs of a typical community member's practice of social distancing and handwashing. For social distancing, we also report whether the respondent was with non-household members or at any other house other than her or his own at the time of the interview in columns 5 and 6.

Across all of these measures, there is no evidence of treatment effects. In Panel A, we evaluate the results for different measures for each of the outcomes by pooling treatment arms by targeted behavior. In Panel B of Table A.8, we test for treatment effect heterogeneity by a measure of social desirability bias using the Marlowe–Crowne scale. Because we only elicited a random subset of the full set of items for each respondent, we estimate a 1-parameter item response theory (IRT) model to aggregate across individuals onto a common scale and use this measure for heterogeneity analysis. The intuition behind this test is that if respondents report practicing the behavior because they believe that is what the enumerator wants to hear, then we should observe stronger treatment effects among those with a greater latent propensity to desire social approval. In our analysis, we find no systematic evidence of differential effects along this margin.[10]

Heterogeneity and spillovers. We explore three further dimensions of heterogeneity. First, we examine whether our treatment effects varied over the course of the study by dividing the experiment duration into three periods of approximately four weeks. Given that the exploration sampling was not correctly implemented in the first weeks of our experiment, we create three approximately equal groups by classifying the first round as the early period and compare treatment effects over middle rounds (rounds 2 to 5) and later rounds (rounds 6 to 10) in Table A.9. We do not find any evidence of differing treatment effects by periods of the experiment, though these effects are somewhat difficult to interpret for two reasons: first, there is endogenous change in our treatment as we allocate more shares over the course of the experiment to more effective arms, and second because the underlying disease environment and associated risks are changing at the same time. Second, low SMS-literacy could also attenuate treatment effects; per Table 1 about 86% of our respondents can read SMS in Hindi. We see no strong evidence that our treatments were more effective among this population in A.10, though the point estimates for handwashing are large and noisy. Third, we test for treatment effects on the 18.2% of our sample who were randomly assigned to be interviewed five days after the first message to test for decay of treatment effects. As shown in Table A.11, we find some evidence of decay of any potential baseline treatment effect: relative to those interviewed three days after receiving the first handwashing message, those surveyed five days later are about six percentage points less likely to report washing their hands. The point estimates are close to zero for social distancing.

[10] We do find statistically significant heterogeneous treatment effects in the list elicitation for social distancing, but this is difficult to attribute to experimenter demand because of the additional statements.

Additionally, we evaluate the effects of social distancing and handwashing outcomes on wearing protective masks and respiratory hygiene (covering mouth and nose while coughing or sneezing) in Table A.12. Overall, we find null effects for both outcomes. We also check for differences in risk perceptions of getting sick from or dying of COVID-19 and do not see any statistically significant differences between treatment and control group participants (Figure A.4). This suggests that the SMS did not cause people to become particularly more concerned about COVID-19 through increasing the salience of the disease.

4. Discussion

During a pandemic, effective communication is critical to encourage the take-up of preventive health behaviors that can help slow the spread of infections (Van Bavel et al., 2020). This is especially important in densely populated areas of developing countries with crowded living quarters and public spaces, and weak health systems.

We examine whether SMS-based information campaigns can be effective at encouraging the adoption of social distancing and handwashing, two key behaviors in preventing the spread of COVID-19. In our setting of rural Bihar, India, treatment participants are about 2.8 times (28.6 p.p.) more likely to receive a SMS related to COVID-19 than the control group. However, this first-stage does not translate into any meaningful impact on knowledge or uptake of social distancing and handwashing behavior. Based on estimated confidence intervals, we cannot rule out increases as large as 5.5–5.6 p.p. for knowledge and adoption of social distancing and 8.8 and 5.6 p.p. for handwashing knowledge and practice, respectively. Our main results are not directly comparable to those of similar experiments conducted during the COVID-19 pandemic in India. These studies work with different populations and only used direct elicitation (see Table 5 for a detailed comparison). Our point estimates for prompted questions are statistically insignificant and at most 0.4 p.p. for always maintaining two arms distance and 0.7 p.p. for washing hands. Banerjee et al. (2020) find that the combined direct and spillover effects of their video intervention decreased travel outside of villages by 7.4 p.p. (20%) and no significant effect on socially distanced interactions; they also find no statistically significant effects among the relatively small sample of respondents who were directly targeted by the experiment. Siddique et al. (2020) find that phone calls and phone calls paired with SMS increased knowledge of preventive behaviors by 53 p.p. (28%) and 85 p.p. (45%), respectively. Reported handwashing and avoiding contact increased by 80–95 p.p. compared to very low compliance in the control group that received only SMS. In terms of second order beliefs about typical community member complying with handwashing, we find an increase of 2.2 p.p. (3% increase relative to control mean) relative to 4.7 p.p. (7%) in Banerjee et al. (2020).

There are several substantive explanations for our null findings. First, our study takes place during an advanced stage of the pandemic when citizens might already be well-informed or too fatigued to respond to nudges. Indeed, in a small number of qualitative interviews we conducted towards the end of the study period, respondents indicated that they had been exposed to many information campaigns and had stopped abiding by these advisories. Similarly, participants may experience higher opportunity costs of adopting certain preventive behaviors, especially social distancing, after the prolonged economic disruption. Second, the majority of study participants do not have a high or very high risk perception of getting infected or dying from COVID-19 (Appendix Figure A.4). This may reduce their responsiveness to our information campaign. This perception could be a result of the low local reported infection rates (Appendix Figure A.1), even though true infections and deaths may be substantial and under-estimated (Mohanan et al., 2021). Third, SMS may not be a sufficiently engaging medium (Favero and Pedersen, 2020) and might convey too little information (Sadish et al., 2021). In contrast to text messages, speaking with a real person (Siddique et al., 2020) or watching a video featuring a well-known person (Banerjee et al., 2020) appear to be effective at changing behaviors. More generally, many of our respondents do not read SMS on a daily basis. However, these approaches require that recipients have smartphones and are willing to use network or internet bandwidth to download videos, pictures or audio files, or require more costly live operators to place calls. Finally, although most respondents indicated that they can read SMS in Hindi, literacy rates in our study area are low. Other modes of communication, such as phone calls or picture messages, may be more appropriate and effective for this population (Siddique et al., 2020). Finally, our relatively young sample may be less responsive to information if their perceived risk is low, even if they may be more comfortable with SMS messages.

Overall, information campaigns based on text messages have low marginal costs and potential to scale but may be ineffective at encouraging preventive behaviors, at least after the initial stage of a pandemic. Other approaches may have more impact and, ultimately, be more cost-effective.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and analysis code are available on the Harvard DataVerse, https://doi.org/10.7910/DVN/0H5UGL.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2021.102747.

References

Alatas, Vivi, Chandrasekhar, Arun G., Mobius, Markus, Olken, Benjamin A., Paladines, Cindy, 2020. Designing effective celebrity messaging: Results from a nationwide Twitter experiment on public health in Indonesia. NBER Working Paper w25589, National Bureau of Economic Research.

Andrews, Isaiah, Kitagawa, Toru, McCloskey, Adam, 2021. Inference on winners. NBER Working Paper w25456, National Bureau of Economic Research.

Armanasco, Ashleigh A., Miller, Yvette D., Fjeldsoe, Brianna S., Marshall, Alison L., 2017. Preventive health behavior change text message interventions: A meta-analysis. Am. J. Prev. Med. 52 (3), 391–402.

Armand, A., Augsburg, B., Bancalari, A., 2021. Coping With Covid-19 Measures in Informal Settlements. Final Report COVID-19-20077-IND-1, International Growth Center.

Banerjee, Abhijit, Alsan, Marcella, Breza, Emily, Chandrasekhar, Arun, Chowdhury, Abhijit, Duflo, Esther, Goldsmith-Pinkham, Paul, Olken, Benjamin, 2020. Messages on COVID-19 prevention in India increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities. NBER Working Paper w27496. National Bureau of Economic Research.

Banerjee, Abhijit, Chandrasekhar, Arun, Dalpath, Suresh, Duflo, Esther, Floretta, John, Jackson, Matthew O., Harini Kannan, Francine Loza, Sankar, Anirudh, Schrimpf, Anna, Shrestha, Maheshwor, 2021. Selecting the most effective nudge: Evidence from a large-scale experiment on immunization, Working Paper.

Bennear, Lori, Tarozzi, Alessandro, Pfaff, Alexander, Balasubramanya, Soumya, Ahmed, Kazi Matin, Van Geen, Alexander, 2013. Impact of a randomized controlled trial in arsenic risk communication on household water-source choices in Bangladesh. J. Environ. Econ. Manag. 65 (2), 225–240.

Caria, Stefano, Gordon, Grant, Kasy, Maximilian, Osman, Soha, Quinn, Simon, Teytlboym, Alex, 2020. An adaptive targeted field experiment: Job search assistance for refugees in Jordan. CESifo Working Paper w25589. CESifo.

Chowdhury, Shyamal, Schildberg-Hörisch, Hannah, Schneider, Sebastian O., Sutter, Matthias, 2020. Nudging or paying? Evaluating the effectiveness of measures to contain COVID-19 in rural Bangladesh in a randomized controlled trial. In: Presentation at the NBER Development Economics Program Meeting Fall 2020. National Bureau of Economic Research.

Chuang, Erica, Dupas, Pascaline, Huillery, Elise, Seban, Juliette, 2020. Sex, Lies, and measurement: Consistency tests for indirect response survey methods. J. Dev. Econ. 148, 102582.

Crowne, Douglas P., Marlowe, David, 1960. A new scale of social desirability independent of psychopathology. J. Consult. Psychol. 24 (4), 349.

Dhar, Diva, Jain, Tarun, Jayachandran, Seema, 2018. Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. NBER Working Paper w25331. National Bureau of Economic Research.

Dupas, Pascaline, 2009. What matters (and what does not) in households' decision to invest in malaria prevention? Amer. Econ. Rev. 99 (2), 224–230.

Falco, Paolo, Zaccagni, Sarah, 2020. Promoting social distancing in a pandemic: Beyond the good intentions. http://dx.doi.org/10.31219/osf.io/a2nys.

Favero, Nathan, Pedersen, Mogens Jin, 2020. How to encourage "togetherness by keeping apart" amid COVID-19? The ineffectiveness of prosocial and empathy appeals. J. Behav. Public Adm. 3 (2).

Hadad, Vitor, Hirshberg, David A, Zhan, Ruohan, Wager, Stefan, Athey, Susan, 2019. Confidence intervals for policy evaluation in adaptive experiments. Working Paper arXiv:1911.02768, arXiv.

Jamison, Julian C., Karlan, Dean, Raffler, Pia, 2013. Mixed method evaluation of a passive mhealth sexual information texting service in Uganda. NBER Working Paper w19107. National Bureau of Economic Research.

Jordan, J., Yoeli, Erez, Rand, David G., 2020. Don't get it or don't spread it? Comparing self-interested versus prosocial motivations for Covid-19 prevention behaviors. Working Paper yuq7x, PsyArXiv.

Karlan, Dean S., Zinman, Jonathan, 2012. List randomization for sensitive behavior: An application for measuring use of loan proceeds. J. Dev. Econ. 98 (1), 71–75.

Kasy, Maximilian, Sautmann, Anja, 2021. Adaptive treatment assignment in experiments for policy choice. Econometrica 80 (1), 113–132.

Kasy, Maximilian, Teytelboym, Alex, 2020a. Adaptive combinatorial allocation. Working Paper arXiv:2011.02330, arXiv.

Kasy, Maximilian, Teytelboym, Alex, 2020b. Adaptive targeted infectious disease testing. Oxford Review of Economic Policy 36 (Suppl 1), S77–S93.

Madajewicz, Malgosia, Pfaff, Alexander, van Geen, Alexander, Graziano, Joseph, Hussein, Iftikhar, Momotaj, Hasina, Sylvi, Roksana, Ahsan, Habibul, 2007. Can information alone change behavior? Response to arsenic contamination of groundwater in Bangladesh. J. Dev. Econ. 84 (2), 731–754.

Meredith, Jennifer, Robinson, Jonathan, Walker, Sarah, Wydick, Bruce, 2013. Keeping the doctor away: Experimental evidence on investment in preventative health products. J. Dev. Econ. 105, 196–210.

Mohanan, Manoj, Malani, Anup, Krishnan, Kaushik, Acharya, Anu, 2021. Prevalence of SARS-CoV-2 in Karnataka, India. JAMA e210332.

Muralidharan, Karthik, Romero, Mauricio, Wüthrich, Kaspar, 2019. Factorial designs, model selection, and (incorrect) inference in randomized experiments. NBER Working Papers w26562. National Bureau of Economic Research.

Orr, Jayne A., King, Roberrt J., 2015. Mobile phone SMS messages can enhance healthy behaviour: A meta-analysis of randomised controlled trials. Health Psychol. Rev. 9 (4), 397–416.

Rajagopalan, Shruti, Tabarrok, Alexander T., 2020. Pandemic policy in developing countries: Recommendations for India. Mercatus Special Edition Policy Brief, Mercatus Center, George Mason University.

Reynolds, William M., 1982. Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. J. Clin. Psychol. 38 (1), 119–125.

Roser, Max, Ritchie, Hannah, Ortiz-Ospina, Esteban, Hasell, Joe, 2020. Coronavirus pandemic (COVID-19). Database, Our World in Data.

Sadish, D., Adhvaryu, Achyuta, Nyshadham, Anant, 2021. (Mis)information and anxiety: Evidence from a randomized Covid-19 information campaign. Journal of Development Economics 152, 102699.

Siddique, Abu, Rahman, Tabassum, Pakrashi, Debayan, Ahmed, Firoz, Islam, Asad, 2020. Raising Covid-19 awareness in rural communities: A randomized experiment in Bangladesh and India. Munich Papers in Political Economy 9/2020. Technical University of Munich.

Tversky, Amos, Kahneman, Daniel, 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263–291.

Tversky, Amos, Kahneman, Daniel, 1991. Loss aversion in riskless choice: A reference-dependent model. Q. J. Econ. 106 (4), 1039–1061.

Van Bavel, Jay J, Baicker, Katherine, Boggio, Paulo S, Capraro, Valerio, Cichocka, Aleksandra, Cikara, Mina, Crockett, Molly J, Crum, Alia J, Douglas, Karen M, Druckman, James N, et al., 2020. Using social and behavioural science to support Covid-19 pandemic response. Nat. Hum. Behav. 4 (5), 460–471.

Young, Alwyn, 2019. Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Q. J. Econ. 134 (2), 557–598.