Countering Misinformation Via WhatsApp: Evidence from the COVID-19 Pandemic in Zimbabwe

Jeremy Bowles, Horacio Larreguy, and Shelley Liu

CID Faculty Working Paper No. 380
May 2020

Working Papers
Center for International Development
at Harvard University

Countering Misinformation Via WhatsApp: Evidence from the COVID-19 Pandemic in Zimbabwe

Jeremy Bowles,[1] Horacio Larreguy,[2] Shelley Liu[3]

[1]Department of Government, Harvard University.
jbowles@g.harvard.edu.
[2]Department of Government, Harvard University.
hlarreguy@fas.harvard.edu.
[3]Department of Government, Harvard University.
shelleyxliu@g.harvard.edu.

We examine how information from trusted social media sources can shape knowledge and behavior when misinformation and mistrust are widespread. In the con-text of the COVID-19 pandemic in Zimbabwe, we partnered with a trusted civil society organization to randomize the timing of the dissemination of messages aimed at targeting misinformation about the COVID-19 virus to 27,000 newslet-ter WhatsApp subscribers. We examine how exposure to these messages affects individual beliefs about how to deal with the COVID-19 virus and preventative behavior. The results show that social media messaging from trusted sources may have substantively large effects on not only individual knowledge but also ultimately on related behavior.

One Sentence Summary: WhatsApp messages from a trusted source increase Zimbabwean citizens' knowledge and preventative behavior regarding COVID-19.

WhatsApp has become a central source of information for individuals in the Global South (*1*). In particular, since in sub-Saharan Africa traditional media reach is low and mobile data costs to access the internet are high, WhatsApp has become a low-cost "one-stop-shop" (*1, 2*). Unfortunately, social media platforms are also highly susceptible to misinformation due to low cost of access, virality of posts, individuals' trust in their social network, and the high cost of fact-checking (*3–6*). Amidst the COVID-19 pandemic, as had been the case with the 2014-2015 Ebola epidemic (*7*) and the 2015-2016 Zika epidemic (*8*), WhatsApp has exacerbated this misinformation problem and muddied public knowledge about the virus throughout the Global South (*9–11*). We study whether trusted sources of information can also leverage WhatsApp's ubiquity to combat misinformation and related potentially harmful behavior. Specifically, we examine the effectiveness of WhatsApp messages from a trusted civil society organization (CSO) in Zimbabwe aimed at targeting misinformation in the context of the COVID-19 pandemic. Contrary to mixed results from the Global North on the dissemination of health-related misinformation (*12–15*), we find that WhatsApp messaging against misinformation from a trusted source can increase both knowledge about COVID-19 and also preventative behavior.

Zimbabwean citizens rely heavily on WhatsApp to access and share information due to prohibitive data costs and the anonymity that WhatsApp affords. As a result, the social network accounts for close to half of all internet traffic in Zimbabwe, far more than competing platforms (*16*). During the study period, the COVID-19 virus had reached Zimbabwe, and the government had just imposed a national lockdown to limit the spread of the virus. Already, across various social media platforms, and particularly through WhatsApp, posts with misinformation about virus transmission and cures had gone viral. Further, due to the low official infection rates, many questioned the necessity of preventative measures (*17*). Misinformation about the virus and low trust in the government threatened the likelihood of lockdown compliance in the country.

To combat this problem, we partnered with two organizations over a two-week period. First, we partnered with Internews, an international non-governmental organization (NGO) operating in Zimbabwe. Internews focuses on training and supporting independent media across the world to help provide people with trustworthy and high-quality information. Our second partner, which implemented the study, is Kubatana, a trusted online media civil society organization (CSO) that was formed in 2001. Kubatana primarily shares information with its subscribers on issues relating to civil and human rights in Zimbabwe through its email, Facebook, Twitter, and WhatsApp channels. The organization began using WhatsApp as a method of distribution in 2013. Today, it has over 27,000 WhatsApp subscribers from across the country divided roughly across 133 WhatsApp broadcast lists. These lists were created based on the month and year of subscription, and contain up to 256 subscribers per list.

Each week, our two partner organizations jointly crafted a short WhatsApp message (SM1). In the first week, the message explained COVID-19's rates of transmission and emphasized the importance of social distancing to lower them. In the second week, the message debunked a viral piece of misinformation on fake cures for COVID-19. Kubatana disseminated the messages in English, Shona, and Ndebele, which are the three main languages in Zimbabwe, through its WhatsApp broadcast lists. In addition, the organization maintained its usual publishing and activity schedule.

To evaluate their effect, we randomized the timing of these messages at the WhatsApp broadcast list level. Subscribers in broadcast lists assigned to the *treatment* condition in a given week were sent the message on Monday, while subscribers in broadcast lists assigned to the *control* condition were sent the message on Saturday. Between these two days of the week, Kubatana sent two additional WhatsApp messages to its subscribers. First, between Tuesday and Wednesday, it sent its weekly newsletter. Second, on Thursday, it distributed a short survey designed to test treatment effects on 1) knowledge of the information disseminated in the messages, and 2) behavior relating to social distancing. Respondents were given the option of responding to the survey either directly through WhatsApp or through Qualtrics. Notably, Kubatana disseminated both the messages and survey without sharing broadcast list information with us, to avoid potential reputational costs in a context where anonymity is highly valued. As we discuss later, this did not affect our results.

This research design has three advantages. First, by randomizing the timing of each message rather than the dissemination itself, all WhatsApp subscribers eventually received important information regardless of their treatment condition. Second, by having Kubatana's weekly newsletter in between the WhatsApp message to treated broadcast lists, we reduced the likelihood that survey respondents would scroll back to a previous message to search for the correct answer. Third, by allowing respondents to respond through WhatsApp, we maximized the response rate. In line with our expectation due to the mobile data costs in Zimbabwe, the survey response rate was four times higher through WhatsApp than through Qualtrics.

The survey sample comprises 868 respondents over two weeks, with 585 (2% response rate) from the first week and 283 (1% response rate) from the second week. These response rates are similar to those of other studies where survey respondents are recruited through social media in sub-Saharan Africa (*18*). 55% of our survey respondents are male and 76% live in urban localities, aligning with evidence from nationally-representative surveys, which estimate that 59% of frequent social media users in Zimbabwe are male and 69% live in urban areas (*19*). Descriptively, a substantial share of respondents report believing in fake cures that have prominently spread through social media. 30% of respondents believe that drinking hot water will cure the virus and 25% believe that inhaling steam will. Table 1 provides descriptive statistics relating to the sample.

We evaluate outcomes relating to *knowledge* and *behavior*. We measured knowledge using a standardized index, or z-score, of responses to factual questions that relate to the message sent in a given week. Directly asking about preventative behavior likely induces social desirability bias. Each week, we thus measured behavior using a list experiment. Respondents were given a list of activities and asked how many they have performed in the past three days. Some respondents received a *short* experimental list with four non-sensitive activities, while others received a *long* experimental list that also included a sensitive activity—visiting a friend or family member outside of their homes during the mandated nationwide COVID-19 lockdown period—indicating that they have not engaged in social distancing. Random assignment of respondents to a *short* or *long* experimental list was performed at the WhatsApp broadcast list level. A comparison of the reported number of activities, across respondents assigned to 'short' and 'long' experimental lists within a treatment condition, provides an unbiased measure of the prevalence of the sensitive activity among the respondents assigned to the treatment condition.

Each week, to assign each WhatsApp broadcast list to a treatment condition, we initially blocked broadcast lists into groups of four, grouping lists which had been created around the same time together. Then, within each block, we randomly assigned one list to each of the four possible combinations of treatment conditions and experimental list length. In Table 2, we show that survey response rates and respondent characteristics are balanced across treatment conditions.

We estimate treatment effects on *knowledge* by regressing the z-score index onto a treatment indicator. We estimate treatment effects on *behavior* by regressing the number of activities reported in the list experiment onto a treatment indicator, a long experimental list indicator, and the interaction between the two. We provide specifications with and without controlling for respondent covariates. We include week fixed effects and either randomization block fixed effects or, more demandingly, WhatsApp broadcast list fixed effects. Standard errors are clustered at the level of the WhatsApp broadcast list-week throughout. Further, we explore subgroup treatment effects by splitting our sample across gender, urbanity and week of the intervention. We provide additional information on estimation in SM4.

First, we examine the effects of treatment assignment on respondent knowledge about the information delivered. Figure 1 plots the treatment effects using different permutations of our specifications. The results suggest substantively large effects of the WhatsApp messages on individual knowledge. In the baseline specification with randomization block fixed effects, respondents assigned to a treated WhatsApp broadcast list in a given week report factual knowledge 0.26σ greater than respondents assigned to a control list ($p < 0.001$). Treatment effects are slightly larger in the specification with WhatsApp broadcast list fixed effects at 0.45σ ($p < 0.001$). These correspond to roughly 7 percentage points, or 12% increase, in correct responses. Across specifications, results are unchanged by the addition of respondent covariates.

Second, we examine treatment effects on respondents' preventative behavior. Figure 2 plots the treatment effects using different permutations of our specifications. In the baseline specification, *among respondents assigned to the control condition*, 37% ($p < 0.001$) did not comply with social distancing. However, *among respondents assigned to the treatment condition*, this behavior drops to 7% ($p = 0.47$). The difference between these effects is statistically significantly different ($p < 0.05$), implying that the WhatsApp messages changed related behavior. Estimated treatment effects are again slightly larger when using WhatsApp broadcast list fixed effects, and are robust to the addition of respondent covariates. The magnitudes of these treatment effects are comparable to those from other studies seeking to facilitate healthy behavior in the Global South (*20*). Importantly, due to the use of a list experiment, these treatment effects on behavior cannot be explained by respondents scrolling back to a previous message to search for the correct answer, and thus also bolster confidence in the effects of treatment assignment on knowledge.

Lastly, we examine subgroup treatment effects on the two outcomes in Figures 3 and 4 based on gender, rurality, and week of intervention. We find relatively uniformly estimated effects across subgroups. While statistically insignificant, treatment effects on knowledge among women are greater than among men ($p = 0.25$), while effects on behavior are not different between women and men ($p = 0.85$). SM6 provides a full set of tables to support the figures.

In sum, our results indicate encouraging positive changes in knowledge and behavior. While WhatsApp has been identified as a platform through which misinformation easily spreads, we show that trusted CSOs can also leverage WhatsApp's reach to successfully get individuals to reassess their misconceptions and correct related behavior. This effect is roughly similar across the urban-rural as well as the gender divide, highlighting the power of WhatsApp messages from a trusted source to counter misinformation. These findings, then, stress the potential of CSOs in sub-Saharan Africa to fight misinformation. They further highlight the similar role that other

WhatsApp newspapers in the region might play (e.g., The Continent in South Africa and 263Chat in Zimbabwe).

The study's context and findings contribute to recent work on the effectiveness of messages to correct misinformation across a variety of issues ranging from health to politics (*12, 15, 21*). These studies present mixed findings, and are particularly negative with respect to vaccination campaigns (*13, 14*). However, most them provide evidence from lab and online experiments in the Global North, while far fewer studies take place in the Global South. Similarly, there is a dearth of field experimental evidence in this context, which is likely to be most informative for scaling up related policies (*22, 23*). Our positive findings from a field experiment in Zimbabwe suggest that there are especially high returns to correcting misinformation, especially surrounding ongoing health crises where people are uncertain and seeking information (*7, 24, 25*).

Further, we confirm the important role that trusted sources play, particularly in confusing informational situations such as health crises (*26*), and in an authoritarian context where trust in information might be low (*27*). Existing scholarship emphasizes the importance of how information is framed (*28*), and the credibility of the information source for the recipient (*29*). During the COVID-19 pandemic, the identification and dissemination of correct information represent an important challenge. While fact-checking can contribute to a source's credibility (*30*), particularly during emergency situations, it might be outpaced by the spread of misinformation through social media (*31, 32*). We show that a trusted source of information can use the same social media channels to disseminate information that both combats misinformation and changes related behavior.

Future research should consider how best to integrate WhatsApp messages aimed at targeting misinformation into a CSOs' ongoing programming, while at the same time highlighting their relative importance. During the study, Kubatana's WhatsApp messaging increased threefold, from one WhatsApp message a week. Even after two weeks, the organization reported four unsubscribers—a number that, while low, is highly unusual for it. Moreover, in the second week, there was a 50% drop in survey responses relative to the first week. Additional work on identifying how to maximize the benefits of such messaging without inducing disengagement will be of great importance for devising a sustainable way to counter misinformation in the Global South.

References

1. The Economist, How Whatsapp is used and misused in Africa (2019).

2. N. Newman, R. Fletcher, A. Kalogeropoulos, D. A. L. Levy, R. K. Nielsen, Digital news report 2017. *Reuters Institute* (2017).

3. M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, W. Quattrociocchi, The spreading of misinformation online. *Proceedings of the National Academy of Sciences* **113**, 554–559 (2016).

4. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 US presidential election. *Science* **363**, 374–378 (2019).

5. D. M. J. Lazer, 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, J. L. Zittrain, The science of fake news. *Science* **359**, 1094–1096 (2018).

6. J. Zarocostas, How to fight an infodemic. *The Lancet* **395**, 676 (2020).

7. L. Spinney, In Congo, fighting a virus and a groundswell of fake news. *Science* **363**, 213–214 (2019).

8. L. Bode, E. K. Vraga, See something, say something: Correction of global health misinformation on social media. *Health communication* **33**, 1131–1140 (2018).

9. AFP, How to spot COVID-19 misinformation on Whatsapp (2020).

10. Al Jazeera, Misinformation, fake news spark India coronavirus fears (2020).

11. R. J. Limaye, M. Sauer, J. Ali, J. Bernstein, B. Wahl, A. Barnhill, A. Labrique, Building trust while influencing online COVID-19 content in the social media world. *The Lancet Digital Health* (2020).

12. M.-p. S. Chan, C. R. Jones, K. Hall Jamieson, D. Albarracín, Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci* **28**, 1531–1546 (2017).

13. B. Nyhan, J. Reifler, S. Richey, G. L. Freed, Effective messages in vaccine promotion: a randomized trial. *Pediatrics* **133**, e835–e842 (2014).

14. S. Pluviano, C. Watt, S. Della Sala, Misinformation lingers in memory: failure of three pro-vaccination strategies. *PLoS One* **12** (2017).

15. N. Walter, S. T. Murphy, How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs* **85**, 423–441 (2018).

16. Quartz Africa, Nearly half of all internet traffic in Zimbabwe goes to Whatsapp (2017).

17. The Guardian, 'we will starve': Zimbabwe's poor full of misgiving over COVID-19 lockdown (2020).

18. K. Hoffmann, F. Rampazzo, L. R. Rosenzweig, Online surveys and digital demography in the developing world: Facebook users in Kenya (2020).

19. Afrobarometer, Zimbabwe round 7 data (2018).

20. E. Grover, M. K. Hossain, S. Uddin, M. Venkatesh, P. K. Ram, R. Dreibelbis, Comparing the behavioural impact of a nudge-based handwashing intervention to high-intensity hygiene education: a cluster-randomised trial in rural bangladesh. *Tropical Medicine & International Health* **23**, 10–25 (2018).

21. B. Nyhan, J. Reifler, When corrections fail: The persistence of political misperceptions. *Political Behavior* **32**, 303–330 (2010).

22. J. N. Druckman, D. P. Green, J. H. Kuklinski, A. Lupia, *Cambridge Handbook of Experimental Political Science* (Cambridge University Press, 2011).

23. A. V. Banerjee, E. Duflo, *Handbook of Field Experiments* (North-Holland, 2017).

24. J. M. Carey, V. Chi, D. J. Flynn, B. Nyhan, T. Zeitzoff, The effects of corrective information about disease epidemics and outbreaks: Evidence from Zika and yellow fever in Brazil. *Science Advances* **6** (2020).

25. S. O. Oyeyemi, E. Gabarron, R. Wynn, Ebola, Twitter, and misinformation: a dangerous combination? *BMJ* **349** (2014).

26. P. Vinck, P. N. Pham, K. K. Bindu, J. Bedford, E. J. Nilles, Institutional trust and misinformation in the response to the 2018–19 Ebola outbreak in North Kivu, DR Congo: a population-based survey. *The Lancet Infectious Diseases* **19**, 529–536 (2019).

27. H. Huang, A war of (mis)information: The political effects of rumors and rumor rebuttals in an authoritarian country. *British Journal of Political Science* **47**, 283–311 (2017).

28. A. Gesser-Edelsburg, A. Diamant, R. Hijazi, G. S. Mesch, Correcting misinformation by health organizations during measles outbreaks: A controlled experiment. *PloS one* **13** (2018).

29. J. Cone, K. Flaharty, M. J. Ferguson, Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences* **116**, 9802–9807 (2019).

30. A. Agadjanian, N. Bakhru, V. Chi, D. Greenberg, B. Hollander, A. Hurt, J. Kind, R. Lu, A. Ma, B. Nyhan, *et al.*, Counting the Pinocchios: The effect of summary fact-checking data on perceived accuracy and favorability of politicians. *Research & Politics* **6**, 2053168019870351 (2019).

31. Quartz Africa, Facebook is launching fact-checking tools in Africa–but Whatsapp is its real problem (2018).

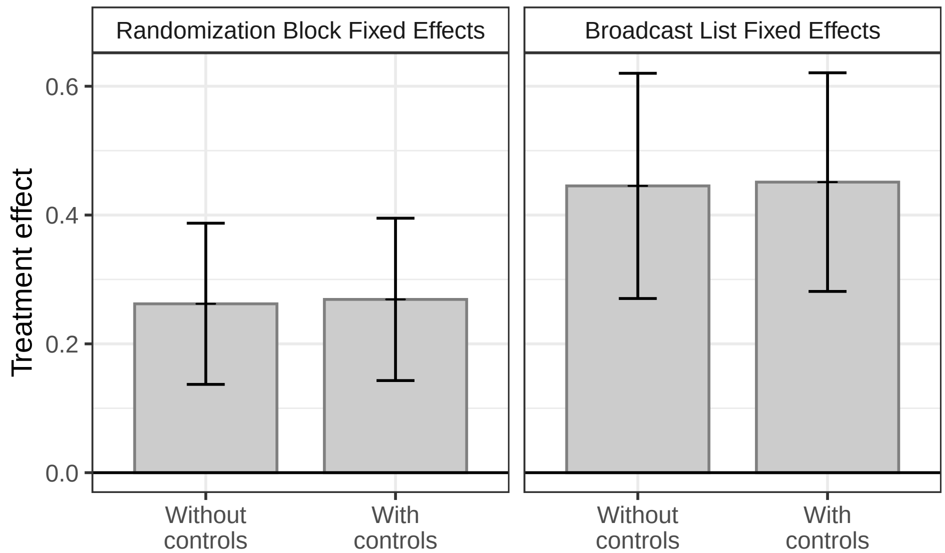
32. Quartz Africa, Whatsapp is limiting message forwarding as coronavirus misinformation takes hold in Africa (2020).

Acknowledgements

With thanks to our partnering organizations, Internews and Kubatana, for their cooperation. IRB exemption granted by Harvard Committee on the Use of Human Subjects (IRB20-0602). **Funding:** No funding support received. **Author contributions:** All authors contributed equally to research stages. **Conflict of interest:** None declared. **Data and materials availability:** All data is available in the manuscript or the supplementary materials.

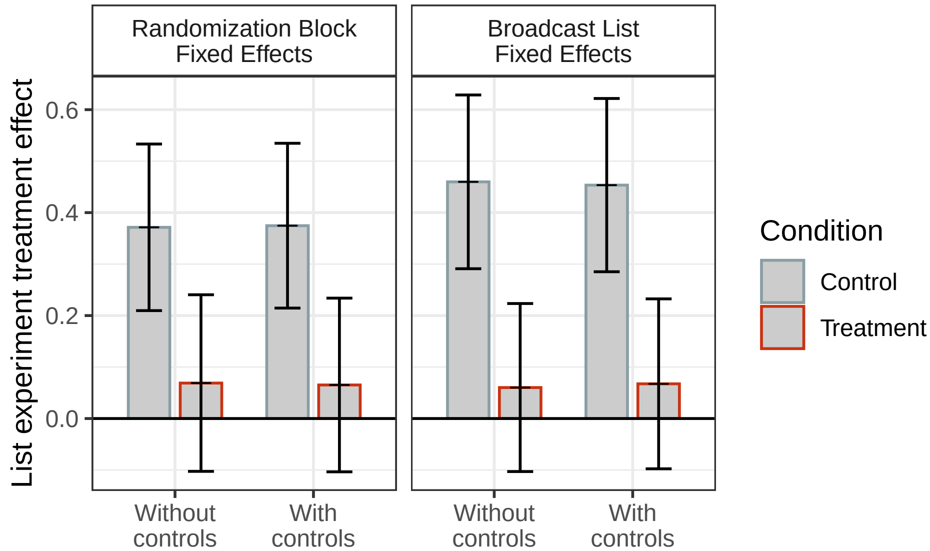
Figures

Figure 1: Treatment effects on knowledge



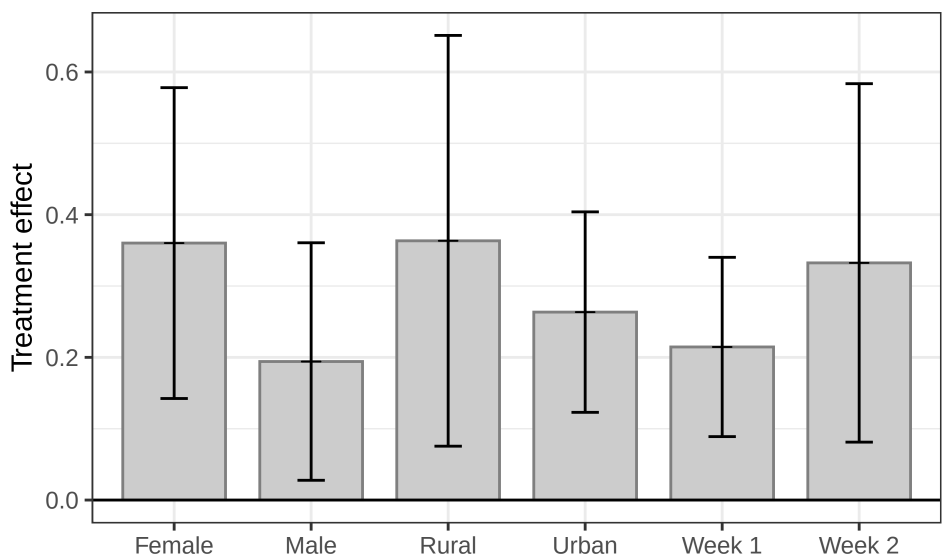
Estimates of the treatment effect of WhatsApp messages on a standardized index of responses to factual questions that relate to the messages sent. 95% confidence intervals plotted. All specifications include week fixed effects. Standard errors clustered at the week-broadcast list level.

Figure 2: Treatment effects on behavior



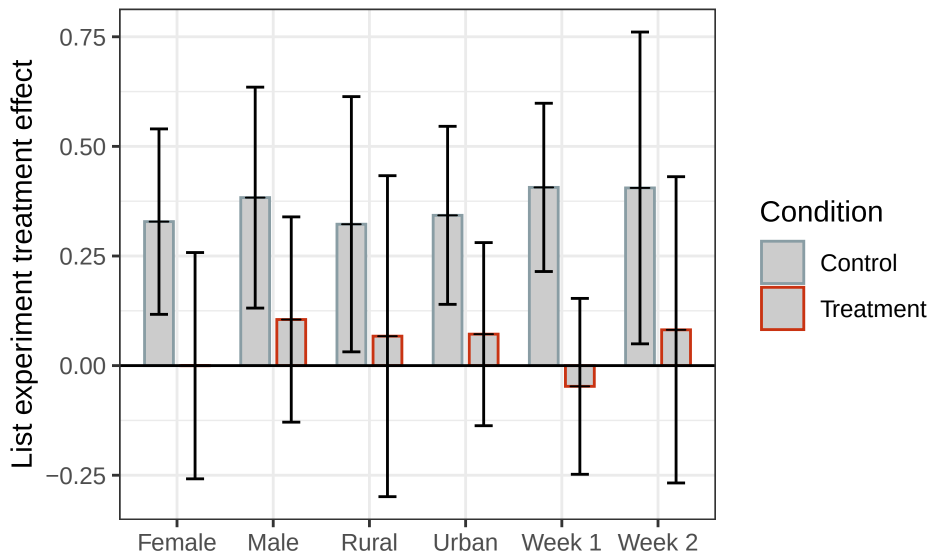
Estimates of the treatment effect of WhatsApp messages on behavior measured through a list experiment. 95% confidence intervals plotted. All specifications include week fixed effects. Standard errors clustered at the week-broadcast list level.

Figure 3: Subgroup treatment effects on knowledge



Estimates of the treatment effect of WhatsApp messages on a standardized index of responses to factual questions that relate to the messages sent. 95% confidence intervals plotted. All specifications include randomization block fixed effects and (apart from by-week estimates) week fixed effects. Standard errors clustered at the week-broadcast list level.

Figure 4: Subgroup treatment effects on behavior



Estimates of the treatment effect of WhatsApp messages on behavior measured through a list experiment. 95% confidence intervals plotted. All specifications include randomization block fixed effects and (apart from by-week estimates) week fixed effects. Standard errors clustered at the week-broadcast list level.

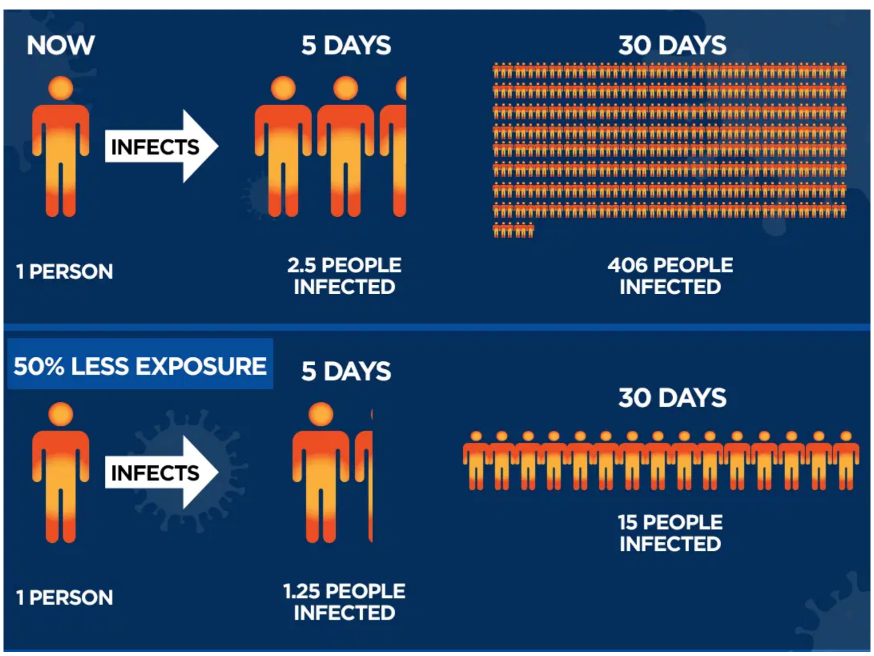
Supplementary Materials

Materials and Methods

1 Messages

Week 1:

With only 9 confirmed cases in Zimbabwe, and given the hardship lockdown imposes on people, many are questioning whether a 21 day lockdown is necessary, and what government's plan is in the longer term. But, it is possible to have the Coronavirus and not show any symptoms. At least 25% of people who have Coronavirus never show symptoms. This means you could catch it from someone who does not know they are sick, and you could unknowingly pass it on to other people, without even realising you were carrying it. This graphic visually demonstrates how physical distancing can help to contain the spread Covid-19.



Week 2:

Social media features a lot of false information about Coronavirus. One myth encourages people to breathe steam or drink hot water to kill Coronavirus. Importantly, **neither breathing hot steam nor drinking hot water kills the virus**. There is no miracle cure and researchers are doing their best to find something quickly, but it will take time. The best recommendations to avoid getting sick and to stop you spreading the virus are to:

- practise **physical distancing**
- **hand wash** thoroughly and frequently (with soap on your hands for 20 seconds)
- **wash surfaces regularly and well**, ideally with bleach or other disinfectant

You can read more here: `https://bit.ly/34rG14b`

2 Coding Decisions

Main treatment variables:

Treatment: Coded as "1" if the WhatsApp broadcast list is assigned to the treatment condition. Treatment assignment varies each week.

Long experimental list: Coded as "1" if the WhatsApp broadcast list is assigned to the long experimental list. Experimental list assignment varies each week.

Main outcome variables:

Knowledge: In **week 1**, there are two questions that test respondent knowledge of the treatment messaging (see Week 1, Q4 and Q5 in Section 3 for exact wording). We code whether the respondent selected the correct responses, and *Knowledge* is coded as the standardized index, or z-score, of these two variables. In **week 2**, there is one question that tests respondent knowledge. The question allows for multiple options, meaning that there are four potentially correct responses (see Week 2, Q4 in Section 3 for exact wording). We code whether the respondent answered correctly for each option, and *Knowledge* is coded as the z-score of all four options.

Behavior: Coded based on how many activities on the experimental list that respondents received they mention that they participated in the last three days. The short experimental list had four options and responses are coded from "0" to "4", while the long list has five options and responses are therefore coded from "0" to "5".

Other variables:

Qualtrics: Coded as "1" if the individual responded through the Qualtrics link and "0" if the individual responded directly through WhatsApp.

Urban: Coded as "1" if the individual responded to living in the following districts: Harare, Bulawayo, Chitungwiza, Mutare, Gweru, Chinhoye, Masvingo, Kwekwe, Kadoma, and Norton.

Female: Coded as "1" if the individual indicated that they were female.

Months subscribed: The number of months that the WhatsApp broadcast list has been active, counting backward from April 2020.

WhatsApp broadcast list response rate %: The number of responses per week from a WhatsApp broadcast list, divided by the total number of individuals in that list.

3 Surveys

Week 1

Hello! Researchers from Harvard University are helping Kubatana to assess the impact of the messages we share. Please could you answer the **5 short questions** in their survey? The survey will take you **less than three minutes** to complete, and your answers will be anonymous. To participate, you need to be over 18. You can read the questions below and reply us directly on WhatsApp, OR you can fill in their survey online here:

1. Where are you located? [Indicate your city or district.]
2. What is your gender?
   (a) Female
   (b) Male
3. In the last 3 days, **HOW MANY** of the following activities did you perform? [Indicate the **TOTAL NUMBER** of activities, not the actual activities]
   - Watched TV or listened to the radio
   - Spoke to friends or family on the phone or WhatsApp
   - **Visited a friend or family member**
   - Went grocery shopping
   - Received or earned money

   Answer: [Indicate the **TOTAL NUMBER** of activities from 0 to 5]
4. To the best of your knowledge, approximately, how many people infected with CORONAVIRUS never show symptoms? [Choose a single response.]
   (a) 0%
   (b) **25%**
   (c) 50%
   (d) 75%
   (e) Do not know
5. To the best of your knowledge, if people implement physical distancing by cutting their exposure to others in half, how will this change the spread of the virus? [Choose a single response.]
   (a) Physical distancing makes no difference.
   (b) Half as many people will be infected.

(c) A quarter as many people will be infected.

(d) **Physical distancing will almost eliminate the spread of the virus.**

(e) Do not know

Week 2

Hello! Thank you everyone for responding to our survey last week. This week again, researchers from Harvard University are helping Kubatana to assess the impact of the messages we share. Please could you answer the **5 short questions** in their survey? The survey will take you **less than three minutes** to complete, and your answers will be **anonymous**. To participate, you need to be over 18. You can read the questions below and reply us **by noon on Sunday** directly on WhatsApp, OR you can fill in their survey online here:

1. Where are you located? [Indicate your city or district.]
2. What is your gender?

   (a) Female

   (b) Male

3. In the last 3 days, **HOW MANY** of the following activities did you perform? [Indicate the **TOTAL NUMBER** of activities, not the actual activities]

   - Watched TV or listened to the radio
   - Spoke to friends or family on the phone or WhatsApp
   - **Visited a friend or family member**
   - Went grocery shopping
   - Received or earned money

   Answer: [Indicate the **TOTAL NUMBER** of activities from 0 to 5]

4. To the best of your knowledge, which of the following strategies most effectively deal with CORONAVIRUS? [Choose **ALL RESPONSES** that you think apply.]

   - Drinking hot water
   - Eating garlic, ginger, lemon and herbs[1]
   - **Hand washing with soap**
   - Inhaling hot steam
   - **Washing surfaces with bleach or other disinfectant**
   - None of these

[1] This information was not part of the messaging, and is thus not included in the coding for Knowledge.

5. To the best of your knowledge, what do you think will happen with new CORONAVIRUS cases when Zimbabwe's lockdown is over? [Choose a single response.][2]

    (a) There will be no new CORONAVIRUS cases.
    (b) The number of new CORONAVIRUS cases per day will decrease.
    (c) The number of new CORONAVIRUS cases per day will remain the same.
    (d) **The number of new CORONAVIRUS cases per day will increase.**
    (e) Do not know

[2]This information was not part of the messaging, and is thus not included in the coding for Knowledge.

4 Estimation

We estimate effects on knowledge using Equation (1):

$$y_{ilw} = \beta T_{lw} + \mu_w + \eta_b + \epsilon_{ilw}, \tag{1}$$

where outcome y for respondent i in broadcast list l in week w is regressed onto the treatment indicator T for WhatsApp broadcast list l in week w and week fixed effects. We additionally include either randomization block fixed effects η_b or, more demandingly, WhatsApp broadcast list fixed effects η_l. We cluster standard errors at the WhatsApp broadcast list-week level. β in Equation (1) estimates the causal effect of a WhatsApp message on knowledge.

We estimate effects on behavior using Equation (2):

$$y_{ilw} = \beta_1 T_{lw} + \beta_2 L_{lw} + \beta_3 (T_{lw} \times L_{lw}) + \mu_w + \eta_b + \epsilon_{ilw}, \tag{2}$$

where outcome y for respondent i in WhatsApp broadcast list l in week w is regressed onto the treatment indicator T for broadcast list l in week w, the list experiment indicator L for broadcast list l in week w, and the interaction of the two. We additionally include either randomization block fixed effects η_b or, more demandingly, WhatsApp broadcast list fixed effects η_l. Standard errors are clustered at the broadcast list-week level. β_1 in Equation (2) estimates treatment effects on the number of activities reported among respondents receiving the short experimental list; β_2 estimates the effect of receiving the long experimental list on the number of activities reported among those assigned to control, and β_3 estimates how the number of activities reported among respondents receiving the long experimental list varies between those assigned to the treatment as opposed to the control condition. β_3, therefore, estimates the causal effect of a WhatsApp message on behavior.

Tables

5 Summary Statistics and Balance

Table 1: Summary Statistics

	Obs.	Mean	SD	Min	Max
Main treatment variables:					
Treatment	868	0.52	0.50	0.00	1.00
Long list in list experiment	868	0.49	0.50	0.00	1.00
Main outcome variables:					
Knowledge	864	0.01	1.00	-2.03	1.85
Behavior	861	2.64	0.90	0.00	5.00
Correct response to knowledge questions:					
Week 1:					
25% of infected are symptomless	583	0.36	0.48	0.00	1.00
Distancing cuts infection rates almost entirely	570	0.83	0.38	0.00	1.00
Week 2:					
Drinking hot water helps	283	0.30	0.46	0.00	1.00
Hand washing with soap helps	283	0.70	0.46	0.00	1.00
Inhaling hot steam helps	283	0.25	0.43	0.00	1.00
Washing surface with disinfectant helps	283	0.56	0.50	0.00	1.00
Other variables:					
Qualtrics	868	0.18	0.39	0.00	1.00
Urban	868	0.76	0.43	0.00	1.00
Female	868	0.45	0.50	0.00	1.00
Months subscribed	868	20.63	19.41	1.00	76.00
WhatsApp broadcast list response rate (%)	868	0.02	0.02	0.00	0.17

Table 2: Balance

	Qualtrics	Urban	Female	List Time	Response Rate
Panel A:					
Treatment	0.01	0.03	0.00	-0.25	-0.00
	(0.03)	(0.02)	(0.03)	(0.15)	(0.00)
Panel B:					
Treatment	-0.00	0.04	0.02	0.00	-0.00
	(0.03)	(0.03)	(0.04)	(.)	(0.00)
Clusters	197	197	197	197	197
Observations	868	868	868	868	868

All specifications include week of intervention fixed effects. Panel A also includes randomization block fixed effects, while Panel B includes WhatsApp broadcast list fixed effects instead. Standard errors are clustered at week-list level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6 Regression Tables

Table 3: Knowledge

	All		Female		Male		Urban		Rural	
	No controls	Controls	No controls	Controls	No controls	Controls	No controls	Controls	No controls	Controls
Panel A:										
Treatment	0.26***	0.27***	0.36***	0.36***	0.19**	0.21**	0.26***	0.26***	0.36**	0.39***
	(0.06)	(0.06)	(0.11)	(0.11)	(0.08)	(0.08)	(0.07)	(0.07)	(0.15)	(0.15)
Panel B:										
Treatment	0.45***	0.45***	0.49***	0.49***	0.55***	0.54***	0.45***	0.46***	0.62**	0.62**
	(0.09)	(0.09)	(0.16)	(0.16)	(0.14)	(0.14)	(0.12)	(0.12)	(0.29)	(0.29)
Clusters	197	197	140	140	164	164	172	172	115	115
Observations	864	864	393	393	471	471	656	656	208	208

All specifications include week of intervention fixed effects. Panel A also includes randomization block fixed effects, while Panel B includes WhatsApp broadcast list fixed effects instead. Controls are indicators for Qualtrics response, urban, and female respondents. Standard errors are clustered at week-list level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Behavior

	All		Female		Male		Urban		Rural	
	No controls	Controls	No controls	Controls	No controls	Controls	No controls	Controls	No controls	Controls
Panel A:										
Treatment	0.32***	0.32***	0.35***	0.35***	0.30**	0.29**	0.33***	0.34***	0.26	0.26
	(0.08)	(0.07)	(0.10)	(0.10)	(0.12)	(0.12)	(0.09)	(0.09)	(0.18)	(0.18)
Long	0.37***	0.37***	0.33***	0.34***	0.38***	0.38***	0.34***	0.35***	0.32**	0.34**
	(0.08)	(0.08)	(0.11)	(0.11)	(0.13)	(0.13)	(0.10)	(0.10)	(0.15)	(0.15)
Treatment $\times$ Long	-0.30**	-0.31**	-0.33*	-0.34*	-0.28	-0.27	-0.27*	-0.28*	-0.26	-0.27
	(0.13)	(0.12)	(0.18)	(0.18)	(0.19)	(0.18)	(0.15)	(0.15)	(0.25)	(0.25)
α(Long + T $\times$ Long $\neq$ 0)	0.43	0.45	1.00	0.99	0.38	0.35	0.50	0.54	0.72	0.70
Panel B:										
Treatment	0.29***	0.28***	0.30*	0.30*	0.21*	0.19	0.30**	0.30**	0.22	0.20
	(0.09)	(0.09)	(0.17)	(0.17)	(0.12)	(0.13)	(0.12)	(0.12)	(0.26)	(0.27)
Long	0.46***	0.45***	0.52***	0.52***	0.48***	0.46***	0.47***	0.46***	0.31	0.29
	(0.09)	(0.09)	(0.12)	(0.12)	(0.15)	(0.15)	(0.12)	(0.12)	(0.23)	(0.24)
Treatment $\times$ Long	-0.40***	-0.39***	-0.39**	-0.39**	-0.43**	-0.41**	-0.35**	-0.34**	-0.35	-0.31
	(0.12)	(0.12)	(0.19)	(0.19)	(0.19)	(0.20)	(0.15)	(0.15)	(0.29)	(0.29)
α(Long + T$\times$Long $\neq$ 0)	0.47	0.42	0.34	0.34	0.69	0.68	0.19	0.19	0.86	0.91
Clusters	197	197	140	140	165	165	172	172	115	115
Observations	861	861	390	390	471	471	655	655	206	206

All specifications include week of intervention fixed effects. Panel A also includes randomization block fixed effects, while Panel B includes WhatsApp broadcast list fixed effects instead. Controls are indicators for Qualtrics response, urban, and female respondents. α(Long + Treatment$\times$Long $\neq$ 0) provides the p-value of the joint hypothesis that Long + Treatment$\times$Long $\neq$ 0. Standard errors are clustered at week-list level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Outcomes by week

	Week 1		Week 2	
	No controls	Controls	No controls	Controls
Panel A:				
Treatment	0.21***	0.21***	0.33**	0.37***
	(0.06)	(0.07)	(0.13)	(0.12)
Panel B:				
Treatment	0.37***	0.38***	0.36**	0.34**
	(0.10)	(0.10)	(0.15)	(0.14)
Long	0.41***	0.41***	0.41**	0.40**
	(0.10)	(0.09)	(0.18)	(0.18)
Treatment $\times$ Long	-0.45***	-0.46***	-0.32	-0.30
	(0.15)	(0.14)	(0.29)	(0.29)
α(Long + T$\times$Long $\neq 0$)	0.65	0.58	0.65	0.56
Clusters	110	110	87	87
Observations	581	581	280	280

All specifications include week of intervention fixed effects. Panel A also include randomization block fixed effects, while Panel B includes WhatsApp broadcast list fixed effects instead. Controls are indicators for qualtrics, urban, and female respondents. α(Long + Treatment$\times$Long $\neq 0$) provides the p-value of the joint hypothesis that Long + Treatment$\times$Long $\neq 0$. Standard errors are clustered at week-list level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.