Stanford | King Center on Global Development

Normalizing Community Mask-wearing: a Cluster Randomized Trial in Bangladesh

Jason Abaluck
Laura H Kwong
Ashley Styczynski
Ashraful Haque
Md. Alamgir Kabir
Ellen Bates-Jefferys
Emily Crawford
Jade Benjamin-Chung
Salim Benhachmi
Shabib Raihan
Shadman Rahman
Neeti Zaman
Peter J. Winch
Md. Maqsud Hossain
Hasan Mahmud Reza
Stephen P. Luby
Ahmed Mushfiq Mobarak

April, 2021
Working Paper No. 1086

John A. and Cynthia Fry Gunn Building | kingcenter.stanford.edu

NBER WORKING PAPER SERIES

NORMALIZING COMMUNITY MASK-WEARING: A CLUSTER RANDOMIZED TRIAL IN BANGLADESH

Jason Abaluck
Laura H Kwong
Ashley Styczynski
Ashraful Haque
Md. Alamgir Kabir
Ellen Bates-Jefferys
Emily Crawford
Jade Benjamin-Chung
Salim Benhachmi
Shabib Raihan
Shadman Rahman
Neeti Zaman
Peter J. Winch
Md. Maqsud Hossain
Hasan Mahmud Reza
Stephen P. Luby
Ahmed Mushfiq Mobarak

Working Paper 28734
http://www.nber.org/papers/w28734

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2021

This project was funded with a grant from Givewell. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Jason Abaluck, Laura H Kwong, Ashley Styczynski, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, Jade Benjamin-Chung, Salim Benhachmi, Shabib Raihan, Shadman Rahman, Neeti Zaman, Peter J. Winch, Md. Maqsud Hossain, Hasan Mahmud Reza, Stephen P. Luby, and Ahmed Mushfiq Mobarak. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to

Normalizing Community Mask-Wearing: A Cluster Randomized Trial in Bangladesh
Jason Abaluck, Laura H Kwong, Ashley Styczynski, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, Jade Benjamin-Chung, Salim Benhachmi, Shabib Raihan, Shadman Rahman, Neeti Zaman, Peter J. Winch, Md. Maqsud Hossain, Hasan Mahmud Reza, Stephen P. Luby, and Ahmed Mushfiq Mobarak
NBER Working Paper No. 28734
April 2021
JEL No. I12,I14,I15

ABSTRACT

A randomized-trial of community-level mask promotion in rural Bangladesh during COVID-19 shows that the intervention tripled mask usage and is a cost-effective means of promoting public health.

Jason Abaluck
Yale School of Management
Box 208200
New Haven, CT 06520-8200
and NBER
jason.abaluck@yale.edu

Laura H Kwong
Woods Institute for the Environment
Stanford University
Palo Alto, CA 94305
lakwong@stanford.edu

Ashley Styczynski
Division of Infectious Diseases
and Geographic Medicine
Stanford University
Palo Alto, CA 94034
astyczyn@stanford.edu

Ashraful Haque
Innovations for Poverty Action Bangladesh
Dhaka, Bangladesh
mahaque@poverty-action.org

Md. Alamgir Kabir
Innovations for Poverty Action Bangladesh
Dhaka, Bangladesh
Evanston, IL 60208
mdkabir@poverty-action.org

Ellen Bates-Jefferys
Innovations for Poverty Action
Evanston, IL 60208
Ebates-jefferys@poverty-action.org

Emily Crawford
Yale School of Management
Yale University
New Haven, CT 06520-8200
emily.crawford@yale.edu

Jade Benjamin-Chung
Department of Epidemiology &
Population Health
Stanford University
jadebc@stanford.edu

Salim Benhachmi
Yale University
New Haven, CT 06511
salim.benhachmi@yale.edu

Shabib Raihan
Innovations for Poverty Action Bangladesh
Dhaka, Bangladesh
sraihan@poverty-action.org

Shadman Rahman
Innovations for Poverty Action Bangladesh
Dhaka, Bangladesh
srahman3@poverty-action.org

Neeti Zaman
Innovations for Poverty Action Bangladesh
Dhaka, Bangladesh
nzaman@poverty-action.org

Peter J. Winch
Division of Social and Behavioral
Interventions
Johns Hopkins Bloomberg School of
Public Health
Baltimore, MD 21205
pwinch@jhu.edu

Md. Maqsud Hossain
North South University
Dhaka, Bangladesh
muhammad.maqsud@northsouth.edu

Hasan Mahmud Reza
North South University
Dhaka, Bangladesh
hasan.reza@northsouth.edu

Stephen P. Luby
Division of Infectious Diseases and
Geographic Medicine
Stanford University
Palo Alto, CA 94034
sluby@stanford.edu

Ahmed Mushfiq Mobarak
Yale School of Management
135 Prospect Street
Box 208200
New Haven, CT 06520
and NBER
ahmed.mobarak@yale.edu

Abstract

Background: A growing body of scientific evidence suggests that face masks can slow the spread of COVID-19 and save lives, but mask usage remains low across many parts of the world, and strategies to increase mask usage remain untested and unclear.

Methods: We conducted a cluster-randomized trial of community-level mask promotion in rural Bangladesh involving 341,830 adults in 600 villages. We employed a series of strategies to promote mask usage, including free household distribution of surgical or cloth masks, distribution and promotion at markets and mosques, mask advocacy by Imams during Friday prayers, role modeling by local leaders, promoters periodically monitoring passers-by and reminding people to put on masks, village police accompanying those mask promoters, providing monetary rewards or certificates to villages if mask-wearing rate improves, public signaling of mask-wearing via signage, text message reminders, messaging emphasizing either altruistic or self-protection motives for mask-wearing, and extracting verbal commitments from households. The primary objective was to assess which of these interventions would increase proper (covering nose and mouth) wearing of face masks, and secondarily, whether mask promotion unintentionally creates moral hazard and decreases social distancing. This analysis is part of larger study evaluating the effect of mask-wearing on transmission of SARS-CoV-2.

Results: There were 64,937 households in the intervention group and 64,183 households in the control group; study recruitment has ended. In the control group, proper mask-wearing was practiced by 13% of those observed across the study period. Free distribution of masks along with role modeling by community leaders produced only small increases in mask usage during pilot interventions. Adding periodic monitoring by mask promoters to remind people in streets and public places to put on the masks we provided increased proper mask-wearing by 29.0 percentage points (95% CI: 26.7% - 31.3%). This tripling of mask usage was sustained over all 10 weeks of surveillance, which includes a period *after* intervention activities ended. Physical distancing, measured as the fraction of individuals at least one arm's length apart, also increased by 5.2 percentage points (95% CI: 4.2%-6.3%). Beyond the core intervention package comprised of free distribution and promotion at households/mosques/markets, leader endorsements plus periodic monitoring and reminders, several elements had *no additional ef-*

fect on mask wearing, including: text reminders, public signage commitments, monetary or non-monetary incentives, altruistic messaging or verbal commitments, or village police accompanying the mask promoters (the last not cross-randomized, but assessed in panel data). No adverse events were reported during the study period.

Conclusions: Our intervention demonstrates a scalable and cost-effective method to promote mask adoption and save lives, and identifies a precise combination of intervention activities that were necessary. Comparisons between pilots shows that free mask distribution alone is not sufficient to increase mask-wearing, but adding periodic monitoring in public places to remind people to wear the distributed masks had large effects on behavior. The absence of any further effect of the village police suggests that the operative mechanism is not any threat of formal legal sanctions, but shame and people's aversion to a light informal social sanction. The persistence of effects for 10 weeks and after the end of the active intervention period, as well as increases in physical distancing, all point to changes in social norms as a key driver of behavior change. Our cross-randomizations suggest that improved mask-wearing norms can be achieved without incentives that require costly monitoring, that aesthetic design choices and colors can influence mask-wearing, and that surgical masks with a substantially higher filtration efficiency can be a cost-effective alternative to cloth masks (1/3 the cost) and are equally or more likely to be worn. Implementing these interventions – including distribution of free masks, and the information campaign, reminders, encouragement – cost \$2.30-\$3.75 per villager, or between \$8 and \$13 per person adopting a mask. Combined with existing estimates of the efficacy of masks in preventing COVID-19 deaths, this implies that the intervention cost \$28,000-\$66,000 per life saved. Beyond reducing the transmission of COVID-19, mask distribution is likely to be a cost-effective strategy to prevent future respiratory disease outbreaks.

Trial registration: ClinicalTrials.gov Identifier: NCT04630054 Funding: GiveWell

1 Introduction

As of March 2021, the COVID-19 pandemic has taken the lives of more than 2.8 million people. A growing body of scientific evidence suggests that face masks can slow the spread of the disease and save lives [1]. Laboratory studies of influenza and other coronaviruses establish the potential efficacy of face masks as a means of source control, blocking particles emitted by infected individuals [2]. Since a substantial share of coronavirus transmission stems from asymptomatic or presymptomatic individuals, the efficacy of masks for source control suggests benefits from *universal* mask-wearing rather than mask-wearing among only those with symptoms [3]. Randomized trials conducted in hospitals also provide evidence that surgical masks can protect the wearer against respiratory diseases. While evidence from community mask-wearing is inconclusive, this is largely attributable to low compliance rates in these studies [4]. A number of observational studies suggest that masks are effective at slowing community spread of COVID-19. For example, countries with mask mandates or mask-wearing norms have had lower infection rates [3, 5], states which mandate mask use subsequently experience declines in case growth rates [6], and model simulations indicate that mask mandates reduce the weekly growth of cases by at least 10%, preventing 19-47% of deaths [7]. As a result of this growing body of evidence, over 40% of the world's population live in countries where mask-wearing is mandated in public areas, and another 40% live in countries where universal mask norms prevail absent a legal mandate [8].

Despite this evidence, theoretical and practical concerns regarding mask adoption remain. The World Health Organization declined to recommend mask adoption until June 2020, citing, in part, concerns that masks would create a false sense of security [9]. Critics of face mask use argued that those who wore masks would engage in compensating behaviors, such as failing to physically distance from others, resulting in a net increase in transmission [10]. A second major concern was that universal mask wearing would lead to hoarding of medical masks, making them unavailable for healthcare workers. The US Surgeon General explicitly discouraged mask-wearing early in the pandemic on these grounds [11].

In many countries, mask use has become politicized, with mask use seen as signalling affiliation with a broader set of political views. This raises questions about how exactly to increase community-wide mask-wearing: is it sufficient to increase accessibility, or correct misinformation and counter anti-mask attitudes (perhaps with encouragement from leaders and role models), or does this need to be supplemented by informal social sanctions, or even threats of legal punishment?

These questions are especially pressing in lower income countries where governments may lack adequate resources to enforce mask mandates. In Bangladesh, a quarter of those observed in public areas in June 2020 wore masks, and only a fifth wore masks *properly* (covering both the nose and mouth), despite a nationwide mask mandate in effect at the time. While vaccines may constrain the spread of COVID-19 in the long-term, it is unlikely that a substantial fraction of the population in low- and middle-income countries will have access to vaccines before the end of 2021 [12]. Uncovering scalable and cost-effective means of promoting mask-wearing, as well as measuring the impact of these interventions on proper, consistent mask-wearing and physical distancing behavior, is thus of first-order policy importance, especially for developing countries.

We employ a series of strategies over multiple rounds of piloting and surveillance to identify the core elements that are necessary for large-scale mask adoption. The iterative research process made clear that mask distribution, coupled with information provision and even role-modeling is unlikely to be sufficient to generate consistent mask-wearing. Adding active mask promotion and periodic monitoring on top of distribution, information, and role-modeling produced much larger increases in mask-wearing.

The main intervention package we ultimately settled on – and test using a cluster-randomized controlled trial covering 341,830 adults in 600 villages in rural Bangladesh – combined mask distribution with encouragement to wear masks at households, mosques, markets, and other public places, communication about the value of mask-wearing, periodic monitoring and reminders, and role-modeling by public officials and community leaders. The selection of elements in this package was informed by both our pilot results and research in public health, psychology [13, 14, 15], economics [16, 17, 18], marketing [19, 20, 21], and other social sciences [22] on product promotion and dissemination strategies.

Entire villages were selected to receive that intervention package or serve as controls. The clustered village-level randomization was important for two reasons. First, unlike technologies with primarily private benefits, mask adoption is likely to yield especially large benefits at the community-level. Second, mask adoption by some may influence mask adoption by others because mask-wearing is immediately visible to other members of the community [23].[1]

The intervention we report here is part of a larger study designed to evaluate the impact of mask wearing on COVID-19 transmission (ClinicalTrials ID NCT04630054). In this paper we report on the first-stage outcome: the effect of the intervention on the prevalence of mask-wearing.[2]

Mask-wearing was assessed through direct observation in public locations including mosques, markets, the main entrance roads to villages, and tea-stalls. Surveillance staff noted whether residents were wearing any mask or face covering, whether the mask was one distributed by our project (and if so, the color), and whether the mask was worn over both the mouth and nose. The mask distribution and promotion was conducted by the Bangladeshi NGO GreenVoice, a grassroots organization with a network of volunteers across the country. Household surveys and surveillance were performed independently by Innovations for Poverty Action (IPA). To minimize the likelihood that village residents would perceive that their mask-wearing behavior was being observed, surveillance staff were separate from mask promoters and wore no identifying apparel. The Bangladesh Directorate General of Health Services under the Ministry of Health, North-South University in Dhaka, and Aspire to Innovate (a2i), an information and data-focused organization within the Bangladesh government, partnered in the study design and discussions and reviewed protocols.

Our full intervention was designed to last 8 weeks in each village (with 10 weeks of surveillance). The intervention started in different villages at different times, rolling out over a 6-week period in 7 batches (with control and intervention villages paired within each batch). Given the

[1]This design was also necessary to properly assess the full impact of masks on infections, including preventing *transmission* of the virus. Individual-level randomization would identify only whether masks protect wearers.

[2]Results for the second-stage outcome, symptomatic seropositivity, will not be available until the serosurveys are completed in June. To predict the cost-per-life-saved from our mask distribution now, we combine the effect of our interventions on mask take-up with current best estimates available from the literature on the efficacy of masks in reducing disease transmission. This prediction will be made more precise once the serosurveys are completed.

time-sensitivity of the subject matter, in this paper we report results from observation conducted during the first 10 weeks of the trial, providing between 4 and 10 weeks of surveillance data for each village. Our study design and results are immediately relevant for Bangladesh's plans for larger-scale distribution of 80 million masks across all rural areas. In fact, we have started devising specific plans with a consortium of government, NGO, and development partners in Bangladesh to scale up the strategies that appear most effective. Results are also relevant for mask dissemination and promotion campaigns planned in other countries. The intervention package that proves effective in our trial would be feasible to implement in a similar fashion throughout South Asia and in other world regions. Beyond masks, the dissemination strategies we employ could be applied to encourage the adoption of other health behaviors and technologies that citizens in low-income countries have also resisted adopting, including water purification technologies [24], sanitation [23, 25], deworming pills [26], cleaner-burning cookstoves [27], and agricultural technologies [28, 29].

2 Background and Context

Bangladesh is a densely populated country with 165 million inhabitants. A serosurvey conducted in July 2020 found 45% of Dhaka residents had antibodies against SARS-CoV-2. This suggests a 0.55% case detection rate based on reported cases, implying that the true infection rate may be two orders of magnitude higher than reported [30, 31]. By January 21, 2021, there were 529,687 reported COVID-19 cases in Bangladesh.[3] The number of daily cases has surged seven-fold between January 21 and March 21 2021 to reach 3500 per day, nearing the June 2020 peak in the country.

Between April and June 2020, our team and others conducted several surveys in Bangladesh to quantify mask-wearing behavior. The evolution of mask use over time in Bangladesh is discussed

[3]http://dashboard.dghs.gov.bd/webportal/pages/covid19.php. Naively extrapolating the case detection rate in Dhaka, up to 58% of the population may have been infected by January 21st, 2021. Similar estimates have been observed in India [32, 33].

in greater detail in [34]. In Bangladesh, the government strongly recommended mask use in early April 2020. This policy was initially accompanied by consistent public messaging, as well as attempts by police and NGOs to confront those who were seen in public without masks. When we surveyed respondents at the end of April 2020, over 80% self-reported wearing a mask and 97% self-reported owning a mask. Anecdotal reports from surveyors suggested that mask-use was pervasive. The Bangladeshi government formally mandated mask use in late May 2020 and threatened to fine those who did not comply, although enforcement was weak to non-existent, especially in rural areas. Anecdotally, mask-wearing was substantially lower than indicated by our self-reported surveys. To investigate, we conducted surveillance studies throughout public areas in Bangladesh in two waves. The first wave of surveillance took place between May 21 and May 25, 2020 in 1,441 places in 52 districts. About 51% out of more than 152,000 individuals we observed were wearing a mask. The second wave of surveillance was conducted between June 19 and June 22, 2020 in the same 1,441 locations, and we found that mask wearing dropped to 26%, with 20% wearing masks that covered their mouth and nose. These observations, coupled with the increasing case load, motivated the interventions we implemented to increase mask use.

3 Interventions and Data Collection

3.1 Sampling frame and strategy

Innovations for Poverty Action (IPA) Bangladesh selected 1,000 rural and peri-urban unions out of 4,500 unions in Bangladesh as a sampling frame, using data on population and infrastructure. We excluded Dhaka district, due to high initial seropositivity, and three hill districts, due to the logistical challenges in accessing the region. We also dropped remote coastal districts where population density is low. The final sampling frame of 1000 unions were located in 40 different districts (*zillas*) (out of 64) and 144 sub-district (*upazilas*) (out of 485).

We used a pairwise randomization to select 300 intervention and 300 control unions within the same upazila/sub-district. This randomization procedure, described in detail in Appendix B, was designed to pair unions that were similar in terms of COVID-19 case data (which was limited), population size, and population density. Each union consists of roughly 80,000 people, or around 80 villages. In each union, we selected a single village. To do so, we identified the largest market and the village within which the market is located and demarcated this territory as the intervention unit (during this scoping process, surveyors were blinded to whether the union was an intervention or control union). We ensured that no two markets/villages were within 2 km of each other to minimize the risk of spillover effects. Within each village, adults from every household were eligible to participate in the study.

The intervention was rolled out in waves, with between 14 and 59 village-pairs grouped in each wave based on geographic proximity. Paired control and treatment villages were always included in the same wave. The first wave was rolled out on 17-18 November 2020 and the last wave was rolled out on 5-6 January 2021. In this paper, we report results from 10 weeks of observation in wave 1, 4 weeks of observation in the last wave 7, and an intermediate number for other waves.

IPA staff travelled to many villages that had low mask uptake in the first five weeks of the study and found that in these villages local leaders were not very engaged in supporting mask promotion. Hence, we retrained mask promotion staff part-way through the intervention to work more closely with local leaders and set specific milestones for that partnership.

3.2 Intervention Materials and Activities

Our entire intervention, including the masks, did not require any unusual or specialized skills or equipment; it was designed to be easily adopted by other NGOs or government agencies and required minimal monitoring.

In focus groups conducted prior to the study, participants said they preferred cloth over surgical masks because they perceived surgical masks to be single-use only and cloth masks to be more durable. Focus group participants also provided feedback on different cloth masks designs and sizes. Both types of masks were manufactured in Bangladesh. The cloth mask had an exterior layer of 100% non-woven polypropylene (70 grams/square meter [gsm]), two interior layers of

60% cotton / 40% polyester interlocking knit (190 gsm), an elastic loop that goes around the head above and below the ears, and a nose bridge. The surgical mask had three layers of 100% non-woven polypropylene (the exterior and interiors are spunbond and the middle layer is meltblown), elastic ear loops, and a nose bridge. The filtration efficiency was 37% (standard deviation [SD] = 6%) for the cloth masks, and 95% (SD = 1%) for the surgical masks (manuscript forthcoming).[4] The filtration efficiency of the surgical masks after washing them 10 times with bar soap and water was 76% (manuscript forthcoming). Surgical masks were outfitted with a sticker that had a logo of a mask with an outline of the Bangladeshi flag and a phrase in Bengali that noted the mask could be washed and reused. The project masks were woven by and procured from local Bangladeshi garment factories within 6 weeks after ordering. The relatively large scale of our bulk order allowed us to negotiate mask prices of $0.50 per cloth mask and $0.13 per surgical mask.

To emphasize the importance of mask-wearing, we prepared a brief video of notable public figures discussing why, how, and when to wear a mask. The video was shown to each household during the mask distribution visit and featured the Honorable Prime Minister of Bangladesh Sheikh Hasina, the head of the Imam Training Academy, and the national cricket star Shakib Al Hasan. During the distribution visit, households also received a brochure based on WHO materials depicting proper mask-wearing.

We implemented a basic set of interventions in all treatment villages, and cross-randomize additional intervention elements in randomly chosen subsets of treatment villages to investigate whether those have any additional impact on mask usage. The basic intervention package consists of five main elements:

1. One-time mask distribution and promotion at households

2. Mask distribution in markets on 3-6 days per week

[4]The filtration efficiency test was conducted using a Fluke 985 particle counter that has a volumetric sampling rate of 2.83 liters per minute. The measurement was taken of particles 0.3–0.5 µm in diameter flowing through the material with a face velocity of 8.5 cm/s. In our internal testing, we found that cloth masks with a external layers made of Pellon 931 polyester fusible interface ironed onto interlocking knit with a middle layer of interlocking knit could achieve a 60% filtration efficiency. Upon discussions with the manufacturers, we learned that those materials could not be procured. Using materials that were available, the highest filtration efficiency possible was 37%.

3. Mask distribution at mosques on three Fridays during the first four weeks of the intervention

4. Mask promotion in public spaces and markets where non-mask wearers were encouraged to wear masks (weekly or biweekly)

5. Role-modeling and advocacy by local leaders, including imams discussing the importance of mask-wearing at Friday prayers using a scripted speech provided by the research team

Details on each of these elements can be found in the intervention protocol, available at https://osf.io/vzdh6/.

3.3 Cross-randomization of behavior change communication and incentives

Village-level Cross-randomizations Our intervention had four village-level cross-randomizations. All intervention villages were assigned to either the treatment or the control group of each of these four randomizations. These village-level randomizations were:

1. Randomization of treated villages to either cloth or surgical masks.

2. Randomization of treated villages to no incentive, monetary incentive of 190 USD, or non-monetary incentive. We announced that the monetary reward or the certificate would be awarded if village-level mask wearing among adults exceeded 75% 8-weeks after the intervention started.

3. Randomization of treated villages to public commitment (providing households signage and asking them to place signage on doors that declares they are a mask-wearing household), or not. The signage was meant to encourage formation of social norms through public signalling.

4. Randomization of treated villages to 0% or 100% of households receiving twice-weekly text message reminders about the importance of mask wearing.

Household-level Cross-randomizations We had three household-level cross-randomizations. In any one village, only one of these household randomizations was operative. Recall that our data collection protocols relied on passive observation at the village-level, which implied that we cannot directly observe household-specific mask-wearing behavior. We therefore varied the color of the masks distributed to the household based on its cross-randomization status. Surveillance staff recorded the color of masks worn by community members, which allowed us to infer the effect of the household-level randomization. In order to avoid conflating the effect of the household-specific treatment with the effect of the mask color, we also randomized which color corresponded to which intervention status across villages (this way a specific color was not fully coincident with a specific treatment). The household-level randomizations were:

1. Households were randomized to receive messages emphasizing either altruism or self-protection.

2. Households were randomized to receive twice-weekly text reminders or not. As mentioned above, the text message saturation was randomly varied to 0%, 50%, or 100% of all villagers receiving texts, and in the 50% villages, the specific households that received the texts was also random.

3. Households were randomized to making a verbal commitment to be a mask-wearing household (all adults in the household promise to wear a mask when they are outside and around other people) or not. This experiment was conducted in a third set of villages where there was no public signage commitment.

This randomization procedure is described in further detail in Appendix C. Participants, mask promoters, and mask surveillance staff were not blinded as intervention materials were distinct. The pre-specified analyses and sample exclusions were made by analysts blinded to the treatment assignment. The intervention protocol, pre-specified analysis plan, and CONSORT checklist are available at https://osf.io/vzdh6/.

Conceptual Basis for Chosen Interventions Our goal was to choose interventions that had a reasonable chance of persuading rural Bangladeshis to wear masks. We consulted literature in public health, development and behavioral economics, and marketing, to identify some of the most promising strategies. An extensive literature identifies price and access as a key deterrent to adoption of welfare-improving products, and especially of technologies that produce positive health externalities, such as face masks [26, 16]. Household distribution of free face-masks therefore formed the core part of our strategy. Inspired by large literature in marketing and economics on the role of opinion leaders in new product diffusion, we additionally emphasized a partnership with community leaders in mask distribution [20, 35].

The additional village- and household-level interventions we experimented with were also inspired by insights from marketing, public health, development, and behavioral economics. For example, masks are a visible good where social norms are expected to be important so we consulted the literature documenting peer effects in product adoption [36, 37, 38]. We experimented with incentives because it is unclear whether extrinsic rewards crowd out intrinsic motivation [39, 40, 41]. We test whether soft commitment devices encourage targets to follow through with actual behaviour change [42, 43], whether public displays can promote social norms [22], whether an altruistic framing inspires people more or less than self-interest [44], whether social image concerns and signaling can lead to higher compliance [45, 17], and whether regular reminders are a useful tool to ensure adoption [18].

3.4 Surveillance Strategies

Mask-wearing and physical distancing were measured through direct observation. Surveillance staff recorded the mask-wearing behavior of all adults observed during surveillance periods; observations were not limited to adults from enrolled households.[5] We defined proper mask-wearing

[5] On December 30th (after 5 weeks of surveillance in wave 1), it was clarified that surveillance staff should only record mask-wearing behavior of people who appear to be 18 years or older. Prior to this, some surveyors included children (especially older children) in their counts. Since the same staff member conducted surveillance in paired intervention and control villages, this change affected the treatment and control groups equally.

as wearing either a project mask or an alternative face-covering over the mouth and nose. Surveillance staff recorded a person as practicing physical distancing if s/he was at least one arm's length away from the nearest person. This is consistent with the WHO guideline that defines physical distancing as one meter of separation.[6] For the sake of assessing risk compensation, this assumes that all nearby people are from different households. Surveillance was conducted using a standard protocol that suggested when surveillance staff should observe the mask use and physical distancing practices of community members; surveillance staff were encouraged to adjust their locations and timing across days to observe as many people as possible. Direct surveillance was conducted at baseline and then once per week on weeks 1, 2, 4, 6, 8, and 10 after the intervention. Each village was observed on two alternating days of the week. Across all villages, observations took place on all seven days of the week, with observation in 150 villages occurring on Friday to over-sample days when mosques were most crowded. Observations generally took place from 9am to 7pm.

3.5 Piloting Interventions

IPA implemented two pilots – Pilot 1 from July 22 to 31 and Pilot 2 from August 13 to 26, 2020. The objective of the pilots was to mimic some of the major aspects of the main experiment to identify implementation challenges. Each of the pilots was conducted in 10 unions that were not part of the project areas. We describe these pilots in more detail below, as we use the difference between the pilots to better understand which elements of our full intervention were essential. The same sizes in our pilots were large enough for reasonably precise statistical inference.

We also conducted focus group discussions and in-depth interviews with village residents, community leaders, religious leaders, and political leaders to elicit opinions on how to maximize the effectiveness of the intervention.

[6] https://www.who.int/westernpacific/emergencies/covid-19/information/physical-distancing. Accessed January, 30 2021.

4 Results

Our analysis follows our pre-registered analysis plan (ClinicalTrials ID NCT04630054) except where indicated. Our pre-registered plan called for an 8-week intervention (with 10 weeks of surveillance). We are here reporting results from 4-10 weeks of surveillance data in each village due to the urgency of the question and the clarity of our findings. Our primary outcome is the proportion of individuals observed to be properly wearing masks. The study overall was powered around the outcome of symptomatic seropositivity, which requires much larger sample sizes to identify precisely. As a result, we are highly over-powered to identify the effects of treatment on mask wearing.

4.1 Balance

While our stratification procedure should have achieved balance with respect to variables observed at the time of randomization, we assessed the balance of these variables two months after intervention assignment but prior to the start of our intervention. This assessment was not pre-registered. For each characteristic, we report the results of a t-test comparing the two groups. This t-test parallels our primary specification. For the balance tests, we replaced the dependent variable of mask-wearing with several baseline characteristics, and used analytic weights proportional to the number of adults recorded in the baseline household survey as well as heteroskedasticity robust standard errors. We also report a bottom-line F-test of the joint significance of any intervention-control differences.

In Table A1 we present summary statistics for the villages in our sample separately for control and intervention villages, including number of households, baseline mask wearing (assessed via observation), and baseline respiratory symptoms reported. Of the four variables we tested, only one was significantly different between the control and intervention groups at the 10% level and the F-test failed to reject balance.

4.2 Primary Analyses

Mask-Wearing The first column of Table 1 reports coefficients from a regression of mask wearing on a constant, an intervention indicator (based on the assigned groups), baseline mask wearing, the baseline symptom rate, and indicators for each control-intervention pair.[7] Mask-wearing in intervention villages increased by 29.0 percentage points. If we omit all covariates (except the fixed effects necessary due to our stratification procedure), our point-estimate is identical. These estimates are quite precise, with standard errors of about 1 percentage point. Considering only surveillance conducted when no mask distribution was taking place, we measure a 28.1 percentage point increase in mask use relative to a baseline level of 13.3 percent.

We also run this analysis separately in mosques, markets, and other locations such as tea stalls, the entrance of restaurants, and the main road in the village. The increase in mask use was largest in mosques (37.6 percentage points), while in all other locations it was 25-29 percentage points.

Physical Distancing Contrary to concerns that mask-wearing would promote risk compensation, we find that our intervention increases distancing behavior. In Table 2, we report identical specifications to Table 1, but with physical distancing as the dependent variable. Evidently, protective behaviors like mask-wearing and social-distancing are complements rather than substitutes: endorsing mask-wearing and informing people about its importance encourages rural Bangladeshis to take the pandemic threat more seriously and engage in another form of self-protection.

While we find increases in physical distancing of 5.2 percentage points pooling across all locations, there was substantial heterogeneity across locations. In markets, individuals become substantially more likely to physically distance (7.4 percentage points). In mosques, we observed no change. Group prayer rituals are completely inelastic: there was no physical distancing practiced in any mosque, in either treatment or control villages.

[7] More details of our statistical methods and standard error construction are available in Appendix D

Table 1: Surveilled Rates of Proper Mask Wearing

	Full	No Active Promotion	Mosques	Markets	Other Locations
No Baseline Controls					
Intervention Coefficient	0.290***	0.281***	0.376***	0.290***	0.253***
	(0.012)	(0.012)	(0.016)	(0.012)	(0.012)
Average Mask Wearing Rate in Paired Control Villages§	0.133	0.133	0.123	0.120	0.146
With Baseline Controls					
Intervention Coefficient	0.290***	0.280***	0.376***	0.289***	0.253***
	(0.012)	(0.012)	(0.017)	(0.012)	(0.012)
N Villages	572	572	570	570	568

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regressions also include an indicator for each control-intervention pair. The regressions ”with baseline controls” include controls for baseline rates of mask wearing and baseline symptom rates.
Baseline Symptom Rate is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO-definition of a probable COVID-19 case. This is defined as any of the following:
(a) fever and cough;
(b) any three of the following (fever, cough, general weakness/fatigue, headache, muscle aches, sore throat, coryza [nasal congestion or runny nose], dyspnoea [shortness of breath or difficulty breathing], anorexia [loss of appetite]/nausea/vomiting, diarrhoea, altered mental status;
(c) anosmia [loss of smell] and ageusia [loss of taste].
We assume that (1) all reported symptoms were acute onset, (2) all people live or work in an area with high risk of transmission of virus and (3) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster.
§We report the mean rate of proper mask wearing among the control villages after the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
”Other Locations” include the Tea Stall, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
The analysis excludes 14 villages and their village-pairs in the full sample, 15 villages and their pairs in the mosque and market sub-samples, and 16 villages and their pairs in the other location sub-sample because the observation data has not yet been submitted by our surveyors.

Table 2: Surveilled Rates of Physical Distancing

	Full	No Active Promotion	Mosques	Markets	Other Locations
			No Baseline Controls		
Intervention Coefficient	0.052***	0.057***	0.000	0.074***	0.070***
	(0.005)	(0.005)	(0.000)	(0.007)	(0.006)
Average Distancing Rate in Paired Control Villages§	0.241	0.253	0.000	0.291	0.311
			With Baseline Controls		
Intervention Coefficient	0.052***	0.058***	0.000	0.075***	0.071***
	(0.005)	(0.005)	(0.000)	(0.007)	(0.006)
N villages	572	572	570	570	568

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regressions also include an indicator for each control-intervention pair. The regressions ”with baseline controls” include controls for baseline rates of social distancing and baseline symptom rates.
Baseline Symptom Rate is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO-definition of a probable COVID-19 case. This is defined as any of the following:
(a) fever and cough;
(b) any three of the following (fever, cough, general weakness/fatigue, headache, muscle aches, sore throat, coryza [nasal congestion or runny nose], dyspnoea [shortness of breath or difficulty breathing], anorexia [loss of appetite]/nausea/vomiting, diarrhoea, altered mental status;
(c) anosmia [loss of smell] and ageusia [loss of taste].
We assume that (1) all reported symptoms were acute onset, (2) all people live or work in an area with high risk of transmission of virus and (3) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster.
§We report the mean rate of physical distancing among the control villages after the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
”Other Locations” include the Tea Stall, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
The analysis excludes 14 villages and their village-pairs in the full sample, 15 villages and their pairs in the mosque and market sub-samples, and 16 villages and their pairs in the other location sub-sample because the observation data has not yet been submitted by our surveyors.

5 Mechanisms

Our intervention combines multiple distinct elements: we provide people with free masks; we provide information about why mask-wearing is important; we conduct mask promotion in the form of monitors encouraging people to wear masks and stopping non-mask-wearing individuals on roads and public places to remind them about the importance of masks; we partner with local public officials to encourage mask-wearing at mosques and markets; and in some villages, we provide a variety of reminders and commitment devices as well as incentives for village leaders. In this section, we attempt to decompose which elements were most critical to increase mask use. We first report results from several cross-randomizations, and then we report non-randomized evidence based on changes over time as our intervention details changed between the rounds of piloting, launch of the full project, and thereafter.

5.1 Village-level Cross-randomizations

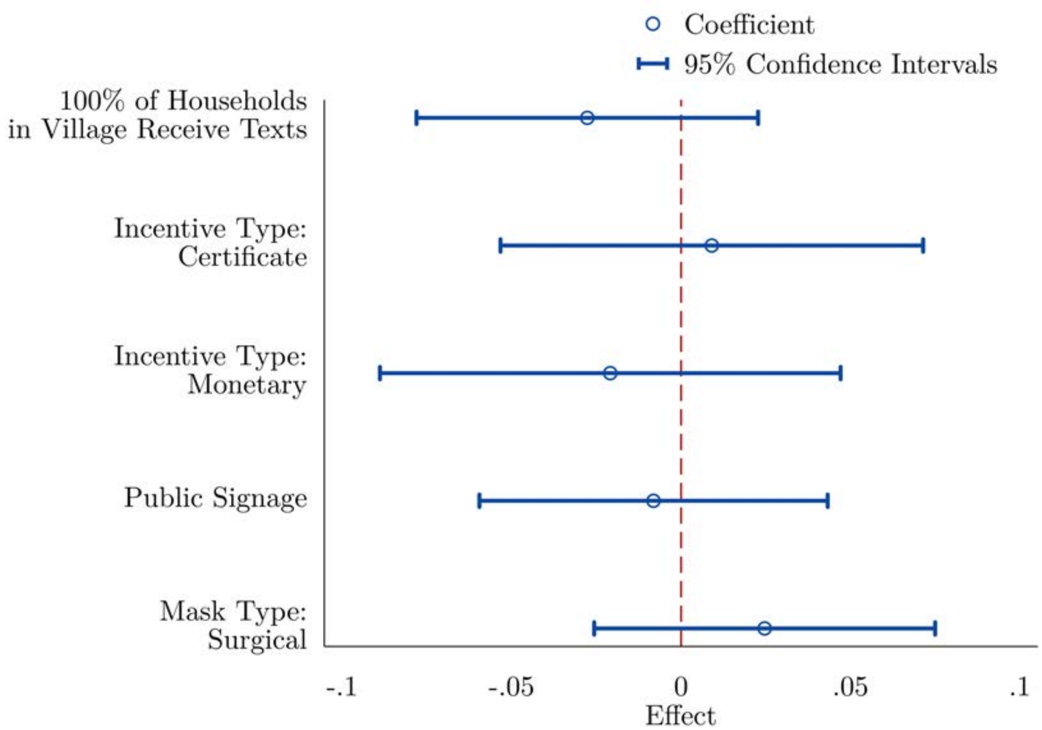
Results from the same regression specification as our primary analysis, adding indicators for each village-level cross-randomization are reported in Figure 1 and Table A7. *None* of the village-level cross-randomizations had any statistically significant impact on mask-wearing behavior, beyond our basic intervention package. These null effects are fairly precise (with standard errors ranging from 2.6-3.5 percentage points). Text message reminders, incentives for village-leaders, or explicit commitment signals explain little of the mask increase we document.

5.2 Household-level Cross-randomizations

We analyzed the effects of household-specific randomized treatments (e.g. verbal commitments or not) by regressing the probability of wearing a mask color corresponding to the treatment on indicators for each household-level randomization, as well as controls for color and surgical masks (recall that the mask-color corresponding to treatment varied across villages).

Results of the household-level cross-randomizations are reported in Figure 2 and Table A8.

Figure 1: Village-Level Cross Randomizations



The figures corresponds to the regressions in A7, upper panel, among the full surveillance data.
Villages were assigned to the treatment or control arms of one of the following four village-level randomizations:
Texts: 0% or 100% of households in a village receive text reminders on the importance of mask wearing;
Incentives: Villages either received no incentive, a certificate, or a monetary reward for meeting a mask-wearing threshold,
Public Signage: All or none of the households in a village are asked to publicly declare they are a mask wearing households;
Mask Type: Villages receive either a cloth or surgical mask.
For a more detailed description of the village-level cross randomizations, see Section 3.3.

Once again, we saw no significant effects of any of the household-level cross-randomizations: compared to self-protection messaging alone, altruistic messaging had no greater impact on mask-wearing, and twice-weekly text messages and a verbal commitment had no significant effects. Surgical masks were perhaps slightly *more* likely than cloth masks to be worn, although the difference is significant at only the 10% level and we saw no significant difference in the village-level randomization of surgical vs. cloth masks.

We did see an impact of mask color on mask adoption. In villages where surgical masks were distributed, blue surgical masks were 2.9 percentage points more likely than green surgical masks to be observed. In villages where cloth masks were distributed, violet masks were 5.8 percentage points more likely than red masks to be observed.

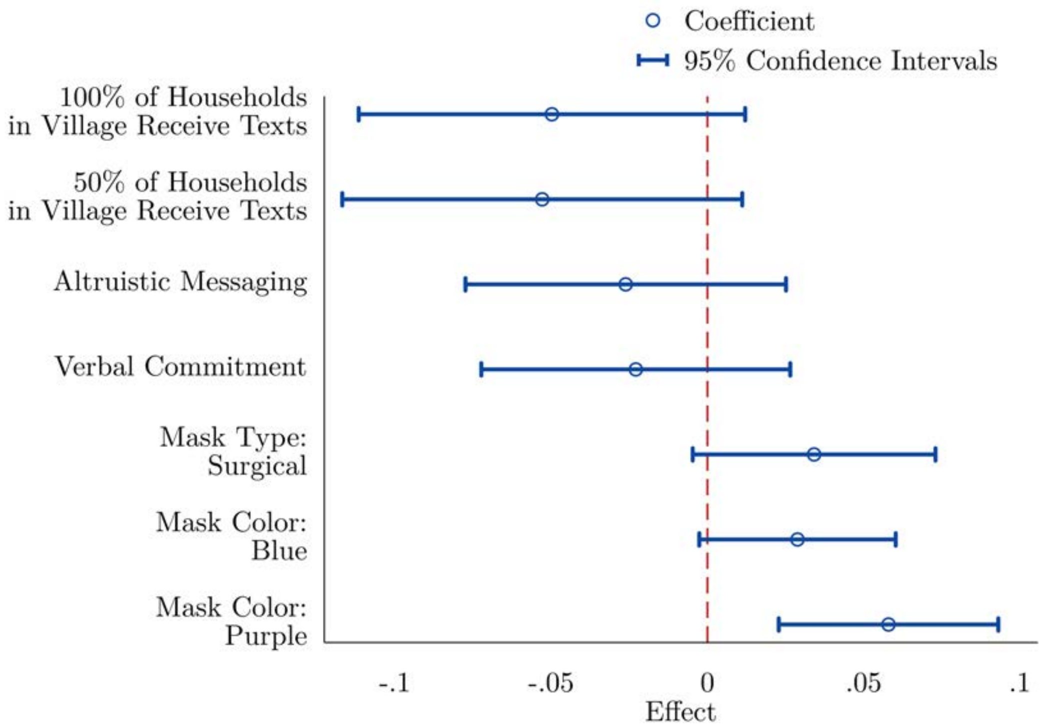
5.3 Mask Promotion

As noted above, we ran two pilots prior to launching the full project. Both pilots were conducted in Naogaon and Joypurhat districts, but in different unions. While the unions were not selected at random, there was no systematic difference in the selection process between the two pilots. In both cases, unions were selected based on convenience and proximity to existing Greenvoice personnel.

Both pilots included elements 1, 2, 3, and 5 enumerated in Section 3.2: masks were distributed at households, markets, and mosques, and there was role-modeling and advocacy by local leaders, including Imams. The second pilot added to these elements explicit mask promotion: mask promoters patrolled public areas a few times a week and asked those not wearing masks to put on a mask. The full intervention also included mask promotion.

The comparison between the two pilots is thus instructive about the impact of active mask promotion. This comparison is shown in Table A5. The difference is striking. The first pilot increased mask-use by 10.9 percentage points (insignificantly different from zero). The second pilot, which included mask promotion, increased mask-use by 28.4 percentage points, comparable to the 29.0 percentage points we see several months later in our full intervention. The presence of mask promotion appears to be crucial for the success of our intervention.

Figure 2: Household-Level Cross Randomizations



The figure corresponds to the regression presented in Table A8.
Villages were assigned to the treatment or control arms of one of the following four village-level randomizations:
Texts: 0%, 50% of 100% of households in a village receieve text reminders on the importance of mask wearing;
Messaging: Households receive messaging emphasizing the altruistic or self-protective benefits of mask wearing;
Verbal Commitment: Households were asked to verbally commit to mask wearing;
Mask Colors: Surgical masks distributed to households were blue or green. Cloth masks distributed to households were violet or red.
For a more detailed description of the household-level cross-randomizations, see Section 3.3.

5.4 Police Enforcement

Our intervention did not vary the likelihood that individuals would face explicit legal sanction if they do not wear masks. During the period of our intervention, mask-wearing was legally required in public areas in Bangladesh, but there was little formal enforcement of this norm.

This raises the question of whether our intervention encouraged mask-use due to the *social* sanction associated with being asked to wear a mask or due to a perceived increase in the threat of legal sanction. In some villages in our intervention, we actively sought the involvement of Chawkidar, or village police, to assist our mask promoters. While the Chawkidar did not arrest people who refused to wear mask, their presence may have suggested an increased likelihood of formal punishment. The involvement of Chawkidar was a later addition, so in all such villages, we have several weeks of surveillance prior to Chawkidar involvement. In Table A6, we run a panel regression of mask-wearing in each village-week on a post-Chawkidar dummy variable, as well as fixed effects for calendar weeks and weeks since the start of the intervention (in the second column, we repeat the same specification using only the 215 villages where Chawkidar were involved in mask promotion). In both specifications, we find no evidence that the threat of formal sanction by the presence of Chawkidar increased mask-wearing, beyond the effect produced by our ordinary mask promoters. This suggests that the implicit social sanction from the mask promoter, or the awkwardness or embarrassment of that conversation was the key mechanism at play in increasing mask usage.

5.5 Persistence of Effects Over Time

In Table A2, we report estimates of our primary specification separately by week of surveillance. In the top panel, we include all villages with surveillance at each respective week. There are fewer villages for each successive week because of the staged roll-out. In the bottom panel, we report the same regression in each week, but only for the 30 villages for which we have at least 10 weeks of data. Week 10 is especially interesting, as it is two weeks after the end of mask promotion.

We find no evidence that the impact of the intervention attenuates over time. In the 30 villages for which we have 10 weeks of surveillance, the point estimates are slightly smaller in week 10 (24.1 percentage points) than week 1 (27.2 percentage points), although this difference is not statistically significant. This is consistent with social norms around mask-wearing taking hold, where adoption by some in the community has a demonstration effect encouraging subsequent adoption by others. If mask usage was driven by a "novelty factor" associated with the our mask promotion campaign, we would have instead expected some attenuation over the course of the 8 weeks of intervention. The point estimates of the impact of intervention by week for the panel of 30 villages for which we have data in all weeks are plotted in Figure A2.

5.6 Subgroup Analyses

We also considered how the impact of our intervention differed between subgroups. In Table A4, we conducted our main analyses on both mask-wearing and physical distancing separately by gender, as well as by whether baseline mask-wearing was above or below the median. In the gender results, we drop surveillance observations for mosques where only men are observed (hence the lower average increases due to the intervention). We found that the intervention increased mask-wearing more for men than for women. This may be due to our mask promoters being predominantly men, although we do not have the variation to test this.

We also found a larger increase in mask usage in villages with below-median baseline mask-wearing (where mask wearing increased from 8.5 percent to 43.4 percent after the intervention), than those with above-median initial mask-wearing (where the increase was from 17.6 percent to 42.5 percent after the intervention).

6 Cost-Benefit Calculations

In Appendix E, we calculated that after subtracting surveillance costs, our intervention cost $12.36 for each person induced to regularly wear a cloth mask and $8.56 for each person to regularly wear a surgical mask. In Table 3, we report estimates of the implied cost per life saved under different assumptions about the effectiveness of masks.[8] The 1- month rows show the dollars per life saved during a month of our intervention, while the 4-month rows show the dollars per life saved extrapolating four months out; notably, we do not assume continued mask wearing beyond one month. Rather, infections prevented during the one month of the intervention propagate into infections prevented in future months. We estimate that after four months, the interventions we conducted cost between $28,000 and $66,000 per life saved, depending on mask effectiveness. Importantly, this does not account for reductions of morbidity associated with hospitalization or other complications of COVID-19. For context, [46] estimate that the value of a statistical life is $205,000 in Bangladesh, implying that our intervention was 4-10 times more cost-effective than what the typical Bangladeshi would be willing to pay to reduce mortality risk. This cost-effectiveness calculation was not pre-specified.

While we did not directly assess harms in this study, there could be opportunity costs resulting from discomfort with increased mask-wearing, adverse health effects such as dermatitis or headaches, or impaired communication. Additionally, disposable surgical masks, even when re-used, can have negative environmental impacts through the generation of non-degradable microplastic material. No adverse events were reported during the study period.

7 Discussion

We present results from a cluster-randomized controlled trial of a scalable intervention designed to increase mask-wearing. Our estimates suggest that mask-wearing increased by 29.0 percentage points, corresponding to an estimated 50,947 additional adults wearing masks in intervention villages. Combining mask distribution, role-modeling and active mask promotion – rather than mask distribution and role-modeling alone – seems critical to achieving the full effect. The effects of our intervention on mask-wearing persist even after active mask promotion was discontinued. The estimated cost per life saved of the intervention after four months was $28,000-$66,000 taking into

[8]Many of the assumptions we made about infection fatality rates and other parameters, and citations justifying those assumptions, are detailed in Table A3

Table 3: Estimates of Cost per Life Saved (Dollars)

Mask Effectiveness	Intervention	Cloth	Surgical
30% Effectiveness			
1 month	266,392	335,068	232,054
4 months	66,598	83,767	58,013
50% Effectiveness			
1 month	159,830	201,035	124,315
4 months	39,955	50,256	34,805
70% Effectiveness			
1 month	114,168	143,601	99,452
4 months	28,542	35,900	24,863

The table reports the estimated dollar cost per life-saved of our intervention under different assumptions about the effectiveness of masks in preventing COVID, after 1 and 4 months. These numbers are computed by multiplying the cost per mask adopter computed in Appendix E by the number needed to treat computed in Table A3.

account the estimated mortality impact, but not reductions in morbidity. This is between 4 and 10 times lower than estimates of the value of a statistical life in Bangladesh [46], and therefore a 'very good buy' for policymakers. Even then, we think that many cost elements can be brought down further in any 'at-scale implmentation'. This is because some of our information campaigns and promotion activities had to be individualized for the purposes of conducting a trial with a control group, whereas at scale the government could use mass media and social media based dissemination strategies more cost-effectively.

While critics of mask mandates suggest that mask-wearing may result in people engaging in riskier behavior, we found no evidence of risk compensation due to increased mask-wearing. In fact, we found that our intervention increased the likelihood of physical distancing, presumably because individuals participating in the intervention took the threat of COVID-19 more seriously. These findings should be interpreted with caution, as these behavioral responses may be context-dependent. For example, it is possible that our specific interventions induced more physical distancing in Bangladesh, where mask promotion is taken to signal the seriousness of the threat, but that mask promotion would have opposite effects in regions where it is viewed as political over-reach. Our results demonstrate that in this one context, the behavioral response to masks enhances the health benefits generated by the intervention.

We found evidence that aesthetics matter, even for a decision of this nature with potentially serious health consequences. In our specific setting, we found that blue masks were preferred to green and violet to red. This was in contrast to our expectations that green and red masks would be more popular because they are the national colors. Specific colors are likely context and culturally dependent, but they highlight the fact that aesthetic taste was important for mask-wearing. In designing effective mask promotion policies, it may be prudent to first host focus group sessions to evaluate preferences, or to offer a variety of masks of different colors and designs to increase take-up.

One limitation in this study was that due to the distinct appearance of project-associated masks and elevated mask-wearing in intervention villages, it was impossible to blind surveillance staff to study arm assignment. Even though surveillance staff were plain-clothed and were instructed to remain discreet, community members may have recognized that they were being observed and changed their behavior. Although control villages were at least 2 km from intervention villages, adults from control villages may have come to intervention villages to receive masks, reducing the apparent impact of the intervention.

Our results suggest a cocktail of core elements essential to promote mask-adoption: mask distribution and role-modeling, combined with mask promotion, leads to large and sustained increases in mask use. Our results also highlight many factors that appear inessential: we find no evidence that public commitments, village-level incentives, text messages, altruistic messaging, or verbal commitments change mask-wearing behavior. These results do not necessarily imply that these approaches are not worth trying in other contexts: for example, the cost of sending texts was minimal, and perhaps an alternative implementation of these strategies would prove effective. What our results do suggest is that an effective intervention is possible without these enhancements. Some of these enhancements could be costly to implement, such as the village-level incentive. Assigning performance-based incentives requires monitoring, which can be expensive and cumbersome for the government bureaucracy, even if the value of the incentive itself is small. The RCT results teach us that such a costly element can be safely avoided in Bangladesh.

Additionally, we found evidence that surgical masks were as likely to be adopted as sturdier-appearing cloth masks (perhaps slightly more likely). This is an important finding because surgical masks have higher filtration efficiency than cloth masks and can be purchased for as little as $0.07 per mask in Bangladesh.[9] In initial project development discussions with officials from the Bangladesh Directorate General of Health Services and public health experts in the country, we were warned that surgical masks would likely prove to be unpopular or ineffective because they were viewed as a disposable, lower-quality, cheaper product among the population. While we distributed 40% more surgical masks, presumably because they were more frequently disposed of, we nonetheless find that the surgical mask treatment arm was the more cost-effective strategy. In Table 3, we report that an intervention consisting only of surgical masks would cost between $25,000 and $58,000 per life saved. At scale, this efficiency could be further improved. For example, if the information about washability on our mask sticker could be conveyed through other means, the cost of our surgical masks would be almost halved.

Even if our intervention turns out to be only 20% as effective in other parts of the world, it would be an extremely cost-efficient means of improving public health both in the current pandemic and in future respiratory disease outbreaks. Further, the product promotion strategies we tested may be more generally useful to overcome the stubborn resistance we have experienced in our efforts to promote other health and welfare-improving products and behaviors, including vaccines, toilets, stoves, weather insurance, new agricultural techniques, and rational-response behaviors like migration [47].

[9] The surgical masks we procured cost $0.13 instead of $0.07 due to the inclusion of a sticker on the mask with a written message that indicated the mask was washable and reusable.

References

[1] J. Howard, *et al.*, *Proceedings of the National Academy of Sciences* **118**, e2014564118 (2021).

[2] N. H. Leung, *et al.*, *Nature Medicine* **26**, 676 (2020).

[3] J. Abaluck, *et al.*, *SSRN* (2020).

[4] C. R. MacIntyre, A. A. Chughtai, *The bmj* **350**, h694 (2015).

[5] C. T. Leffler, *et al.*, *American Journal of Tropical Medicine and Hygiene* **103**, 2400 (2020).

[6] W. Lyu, G. L. Wehby, *Health Affairs* **39**, 1419 (2020).

[7] V. Chernozhukov, H. Kasaha, P. Schrimpf, *Journal of Econometrics* **220**, 23 (2021).

[8] H. Ludwig, W. Keating, P. Zdimal, P. Navratil, P. Hausenblass, What countries require masks in public or recommend masks? (2020). Accessed on 4 Jan 2021.

[9] T. A. Ghebreyesus, WHO Director-General's opening remarks at the media briefing on COVID-19 - 5 June 2020 (2020). Publisher: WHO.

[10] L. Brosseau, M. Sietsema, Commentary: Masks-for-all for COVID-19 not based on sound data (2020). Publisher: Center for Infectious Disease Research and Practice (CIRDRAP). University of Minnesota.

[11] N. Ibrahim, Did the US Surgeon General Recommend the Public Not Wear Masks?

[12] A. Mullard, How COVID vaccines are being divvied up around the world (2020).

[13] D. Kahneman, D. T. Miller, *Psychological review* **93**, 136 (1986).

[14] J. Jordan, E. Yoeli, D. Rand (2020).

[15] R. B. Cialdini, N. J. Goldstein, *Annu. Rev. Psychol.* **55**, 591 (2004).

[16] Jameel Poverty Action Lab (J-PAL), The price is wrong: Charging small fees dramatically reduces access to important products for the poor, *Tech. rep.* (2011). Publisher: Innovations for Poverty Action.

[17] A. Karing, Social signaling and childhood immunization: A field experiment in Sierra Leone, *Tech. rep.* (2018). Publisher: Innovations for Poverty Action.

[18] D. Karlan, M. McConnell, S. Mullainathan, J. Zinman, *Management Science* **62**, 3393 (2016).

[19] N. J. Goldstein, R. B. Cialdini, V. Griskevicius, *Journal of consumer Research* **35**, 472 (2008).

[20] G. Miller, A. M. Mobarak, *Marketing Science* **34** (2014).

[21] P. Manchanda, Y. Xie, N. Youn, *Marketing Science* **27**, 961 (2008).

[22] C. Bicchieri, *Norms in the wild: How to diagnose, measure, and change social norms* (Oxford University Press, 2016).

[23] R. Guiteras, J. Levinsohn, A. M. Mobarak, *Science* **348**, 903 (2015).

[24] N. Ashraf, J. Berry, J. M. Shapiro, *American Economic Review* **100**, 2383 (2010).

[25] S. R. Patil, *et al.*, *PLOS Medicine* **11**, e1001709 (2014).

[26] M. Kremer, E. Miguel, *Quarterly Journal of Economics* **122**, 1007 (2007).

[27] A. M. Mobarak, P. Dwivedi, R. Bailis, L. Hildemann, G. Miller, *Proceedings of the National Academy of Sciences* **109**, 10815 (2012).

[28] E. Duflo, M. Kremer, J. Robinson, *American Economic Review* **101**, 2350 (2011).

[29] A. BenYishay, A. M. Mobarak, *The Review of Economic Studies* **86**, 976 (2018).

[30] ICDDR,B, The IEDCR and partners share insights on the prevalence, seroprevalence and genomic epidemiology of COVID-19 in Dhaka city (2020).

[31] Coronavirus COVID-19 dashboard, 2020 (2021). Publisher: Government of Bangladesh Directorate General of Health Services.

[32] A. Malani, *et al.*, *The Lancet Global Health* **9**, E110 (2020).

[33] M. Mohanan, A. Malani, K. Krishnan, A. Acharya, *medRxiv* (2020).

[34] J. Abaluck, A. M. Mobarak, Getting all Bangladeshis to wear masks (2020). Publisher: WhiteBoard Magazine.

[35] P. S. Van Eck, W. Jager, P. S. Leeflang, *Journal of Product Innovation Management* **28**, 187 (2011).

[36] E. Oster, R. Thornton, *Journal of the European Economic Association* **10**, 1263 (2012).

[37] H. Allcott, *Journal of public Economics* **95**, 1082 (2011).

[38] R. Guiteras, J. Levinsohn, A. M. Mobarak, Demand estimation with strategic complementarities: Sanitation in Bangladesh (2019). Publisher: CEPR Discussion Paper No. DP13498.

[39] N. Ashraf, O. Bandiera, K. Jack, *Journal of Public Economics* **120**, 1 (2014).

[40] R. Chetty, E. Saez, L. Sandor, *Journal of Economic Perspectives* **28**, 169 (2014).

[41] D. Ariely, A. Bracha, S. Meier, *American Economic Review* **99**, 544 (2009).

[42] G. Bryan, D. Karlan, S. Nelson, *Annual Review of Economics* **2**, 671 (2010).

[43] J. Luoto, D. Levine, J. Albert, S. Luby, *Journal of Development Economics* **110**, 13 (2014).

[44] N. Ashraf, O. Bandiera, E. Davenport, S. S. Lee, *American Economic Review* **110**, 1355 (2020).

[45] L. Bursztyn, R. Jensen, *Annual Review of Economics* **9**, 131 (2017).

[46] W. K. Viscusi, C. J. Masterman, *Journal of Benefit-Cost Analysis* **8**, 226 (2017).

[47] G. Bryan, S. Chowdhury, A. M. Mobarak, *Econometrica* **82**, 1671 (2014).

8 Conflict of Interest

The funder had no role in the study design, interpretation of results, or decision to publish.

9 Research Ethics Approvals

Our study protocols were reviewed and approved by the Yale University Institutional Review Board (Protocol ID: 2000028482), and by the Bangladesh Medical Research Council National Research Ethics Committee (IRB registration number: 330 26 08 2020). We also received separate administrative approval from the Bangladesh Ministry of Health and Family Welfare. Study protocols and plans were also discussed with public health experts at the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) and a2i, an innovation agency within the Bangladesh government (https://a2i.gov.bd/), prior to implementation.

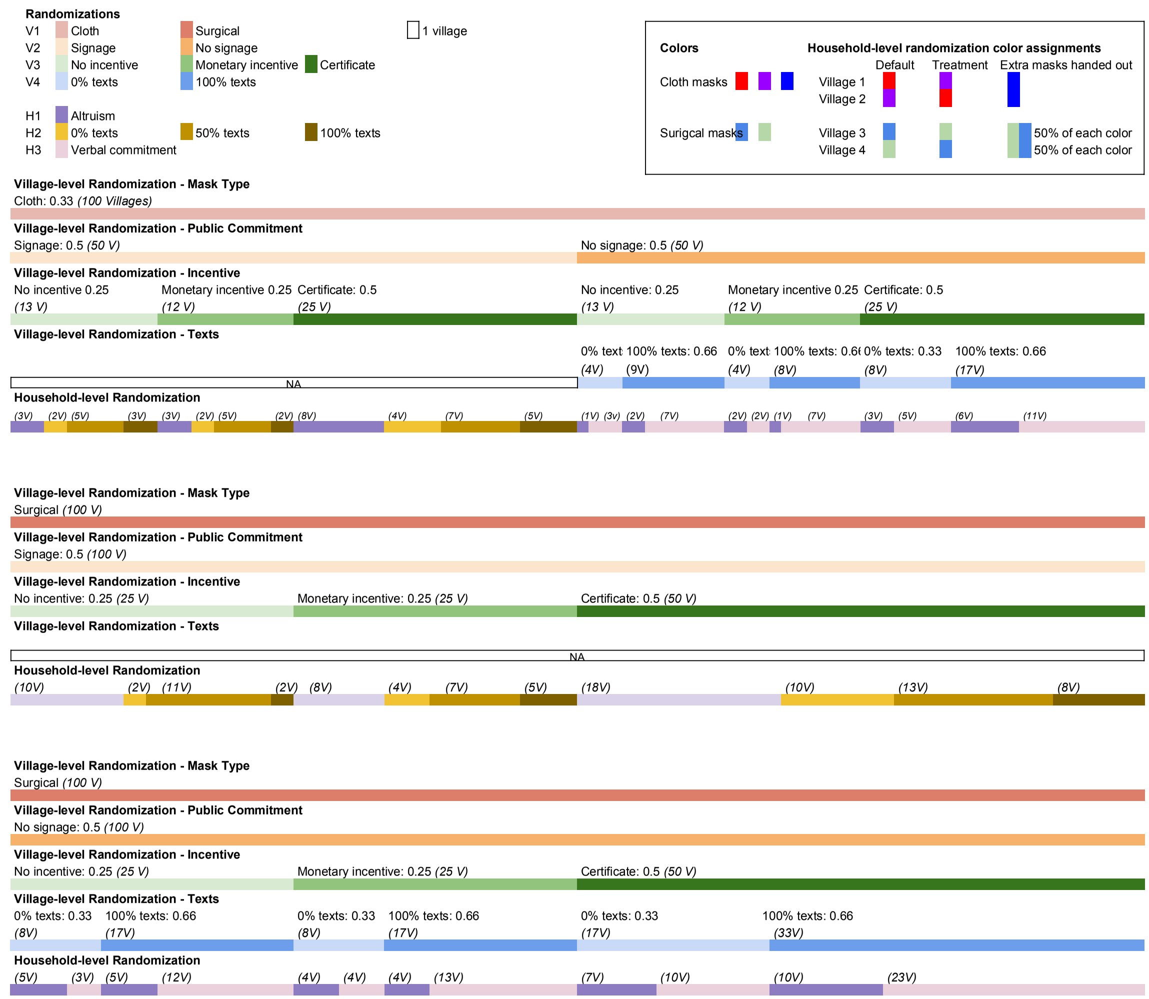
10 Acknowledgements

Thanks to Givewell for their generous funding, Dr. Sabrina Flora, additional director general of the Directorate General of Health Services in Bangladesh, James Snowden and Karen Levy for ongoing support and encouragement. Thanks to WHO Chief Scientist Soumya Swaminathan for her encouragement to conduct this trial. Thanks to Anir Chowdhury, policy advisor to the Bangladesh government and Dr. Shams El Arifeen of icddr,b for establishing connections with the Bangladesh Directorate General of Health Services in the Ministry of Health and Family Welfare. Thanks to Asif Saleh, Executive Director of BRAC, for help with implementation. Thanks also to Bhavani Prathap Kasina and Stacey Daves-Ohlin for their advice and assistance, Tom Schmidt for help in shipping, Arnab Bhattacharya of the Tata Institute of Fundamental Research and Shailabh Kumar of Stanford University for conducting filtration efficiency testing of the study masks and evaluation of the impact of washing on surgical mask performance, Yasser Choudhury from Katex for assistance with the surgical masks, Anisur Rahman from Standard Group for assistance with

the cloth masks, GreenVoice, and many employees at IPA Bangladesh for assistance throughout this project.

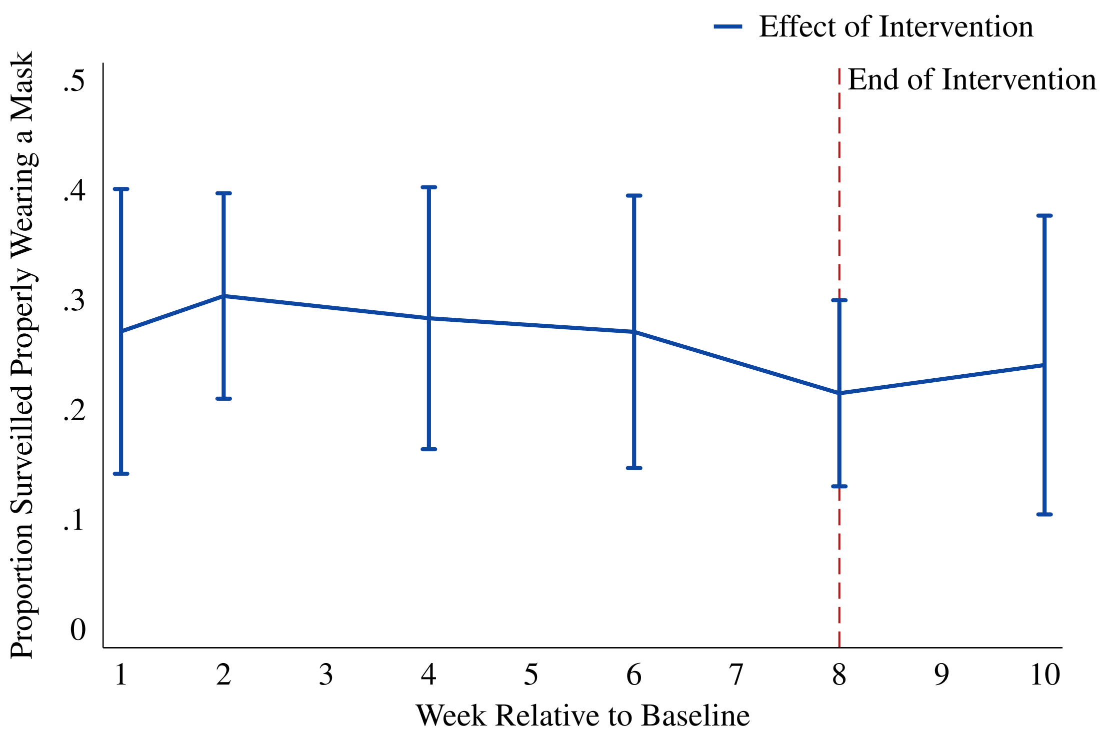
A List of Supplementary Materials

Figure A1: Schematic of Cross-Randomizations



Notes: Each box represents one village and each color represents a village-level or household-level randomization. Different tones of the same hue represent different possible realizations for each randomization. The "Colors" box in the upper right exemplifies the color of masks used to denote households that received the default or intervention condition of the household-level randomization.

Figure A2: Persistence of Mask Wearing



The figure corresponds to the regressions presented in Table A2, second panel. We present the effect of the intervention separately across weeks 1, 2, 4, 6, and 10 after the baseline observation with 95% confidence intervals. The analysis is run across a panel of 30 villages with observation through the entirety of the study. Week 10 observation occurs after mask promotions have ceased.

Table A1: Balance Tests

Coefficient	Influenza-like Illness	WHO-Defined Probable COVID-19	Baseline Mask Wearing Rate	Number of Households
Summary Statistics				
Intervention	0.008	0.027	0.123	227
Control	0.009	0.025	0.125	224
Balance Tests				
Intervention Coefficient	–0.001*	0.001	0.001	2.723
	(0.001)	(0.002)	(0.005)	(3.962)
N villages	572	572	572	572
F		1.84		
Joint-Test Prob $>$ *F*		0.1014		

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
We define influenza-like illness as fever and cough.
We classify a WHO-Defined Probable COVID-19 case as any of the following
(a) fever and cough;
(b) three or more of the following symptoms (fever, cough, general weakness/fatigue, headache, myalgia, sore throat, coryza, dyspnea, anorexia/nausea/vomiting, diarrhea, altered mental status);
(c) loss of taste or smell.
The baseline rate of mask wearing was measured through observation over a 1 week period, defined as the rate of those observed who wear a mask or face covering that covers the nose and mouth.
Household count was assessed in a scoping visit prior to the intervention.
The analysis excludes 14 villages and their village-pairs in the full sample, because the observation data has not yet been submitted by our surveyors.

Table A2: Persistence of Mask Wearing

	Week from Baseline Observation					
	1	2	4	6	8	10
	Batches 1-7					
Intervention Coefficient	0.303***	0.289***	0.301***	0.290***	0.256***	0.241***
	(0.015)	(0.015)	(0.015)	(0.018)	(0.023)	(0.069)
N villages	498	498	496	436	196	30
	Batches 1-3 (At Least 10 Weeks of Observation)					
Intervention Coefficient	0.272***	0.304***	0.284***	0.271***	0.215***	0.241***
	(0.066)	(0.048)	(0.061)	(0.063)	(0.043)	(0.069)
N villages	30	30	30	30	30	30

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regressions also include an indicator for each control-intervention pair. The regressions ”with baseline controls” include controls for baseline rates of mask wearing and baseline symptom rates.
Baseline Symptom Rate is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO-definition of a probable COVID-19 case. This is defined as any of the following:
(a) fever and cough;
(b) any three of the following (fever, cough, general weakness/fatigue, headache, muscle aches, sore throat, coryza [nasal congestion or runny nose], dyspnoea [shortness of breath or difficulty breathing], anorexia [loss of appetite]/nausea/vomiting, diarrhoea, altered mental status;
(c) anosmia [loss of smell] and ageusia [loss of taste].
We assume that (1) all reported symptoms were acute onset, (2) all people live or work in an area with high risk of transmission of virus and (3) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster.
”Other Locations” include the Tea Stall, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
This analysis was limited to villages within the first 3 batches (between November 15, 2020 and X), and estimates separate intervention effects for 1, 2, 4, 6, 8, and 10 weeks after baseline observation.
The second panel estimates the same effect, but only among a consistent group of 30 villages that have at least 10 weeks of observation. The 10th week of observation occurs after all active promotion of mask wearing has ceased.

Table A3: Calculation of Number Needed to Treat

Mask Effectiveness & Duration of Use	Bangladesh Population (2020)*	Anticipated Rate of Seroconversions	Estimated IFR for Bangladesh	Anticipated Deaths	Projected Deaths Given Mask Effectiveness	ARR	NNT
30 % Effectiveness							
1 month	164,689,383	0.058	0.00212	20,250	14,175	3.69E-05	27,109
4 months	164,689,383	0.232	0.00212	81,001	56,701	1.48E-04	6,777
50 % Effectiveness							
1 month	164,689,383	0.058	0.00212	20,250	10,125	6.15E-05	16,265
4 months	164,689,383	0.232	0.00212	81,001	40,500	2.46E-04	4,066
70 % Effectiveness							
1 month	164,689,383	0.058	0.00212	20,250	6,075	8.6E-05	11,618
4 months	164,689,383	0.232	0.00212	81,001	24,300	3.4E-04	2,905

IFR = Infection Fatality Rate; ARR = Absolute Risk Reduction; NNT = Number Needed to Treat;
* worldometers.info

To determine the impact of the intervention in reducing mortality from COVID-19, we first identified some key parameters. We determined the risk of SARS-CoV-2 infection based on a prior seroprevalence study conducted in Dhaka. COVID-19 was first introduced into Bangladesh in March 2020 with widespread community transmission observed in urban centers by April. 45% of the population in Dhaka had been exposed to SARS-CoV-2 by the end of July 2020 (1). When compared with reported cases in Dhaka for the same time period, this translates into a case detection rate of 0.55% (2). The number of reported cases has been relatively steady in Bangladesh since the onset of the pandemic. If we extrapolate a 0.55% case detection rate to the entire country, we estimate that up to 58% of the population of Bangladesh may have been exposed to SARS-CoV-2 by the end of January 2021. This translates to a seroconversion rate of approximately 5.8% per month.

The estimated infection fatality rate (IFR) for Bangladesh, accounting for population structure, is 0.212% (3). We can use this and the estimated number of seroconversions per month to calculate the anticipated deaths per month from COVID-19. We can then estimate the reduction in deaths for each month following the intervention. We can convert this to the number needed to treat (NNT) during the intervention to prevent one death for each month after the intervention by taking the inverse of the absolute risk reduction (ARR) across a range of estimates of mask effectiveness for reducing COVID-19 (4-9).

References 1. https://www.icddrb.org/quick-links/press-releases?id=97task=view 2. http://dashboard.dghs.gov.bd/webportal/pages/covid19.php 3. O'Driscoll, M., Ribeiro Dos Santos, G., Wang, L. et al. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature (2020). https://doi.org/10.1038/s41586-020-2918-0 4. Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early COVID-19 pandemic in the US. J Econometrics. 2021;220(1):23-62. https://www.sciencedirect.com/science/article/pii/S0304407620303468?via5. Worby C, Chang H. Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic. Nature Communications. 2020;11. https://www.nature.com/articles/s41467-020-17922-x 6. Eikenberry SE, et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infectious Disease Modeling. 2020;5:293-308. https://doi.org/10.1016/j.idm.2020.04.001 7. Chu D, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet. 2020;395(10242):1973-87. https://doi.org/10.1016/S0140-6736(20)31142-9 8. Doung-Ngern P, Suphanchaimat R, Panjangampatthana A, et al. Case-Control Study of Use of Personal Protective Measures and Risk for Severe Acute Respiratory Syndrome Coronavirus 2 Infection, Thailand. Emerg Infect Dis. 2020;26(11).10.3201/eid2611.203003. https://www.ncbi.nlm.nih.gov/pubmed/32931726 9. Payne DC, Smith-Jeffcoat SE, Nowak G, et al. SARS-CoV-2 Infections and Serologic Responses from a Sample of U.S. Navy Service Members – USS Theodore Roosevelt, April 2020. MMWR Morb Mortal Wkly Rep. 2020;69(23):714-721.10.15585/mmwr.mm6923e4. https://www.ncbi.nlm.nih.gov/pubmed/32525850

Table A4: Subgroup Analyses

	Female Only	Male Only	Above Median	Below Median
	No Baseline Controls			
Intervention Coefficient	0.225***	0.272***	0.249***	0.349***
	(0.014)	(0.013)	(0.019)	(0.022)
Average Control Mask Wearing Rate§	0.310	0.116	0.176	0.085
	With Baseline Controls			
Intervention Coefficient	0.224***	0.272***	0.250***	0.353***
	(0.014)	(0.013)	(0.019)	(0.021)
N villages	568	568	200	202

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regressions also include an indicator for each control-intervention pair. The baseline control regressions include controls for baseline rates of mask wearing and baseline symptom rates. For the gender subgroup analyses, the baseline symptom rate and baseline mask wearing rate was defined across all individuals, not just those among females and males, respectively.
Baseline Symptom Rate is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO-definition of a probable COVID-19 case. This is defined as any of the following:
(a) fever and cough;
(b) any three of the following (fever, cough, general weakness/fatigue, headache, muscle aches, sore throat, coryza [nasal congestion or runny nose], dyspnoea [shortness of breath or difficulty breathing], anorexia [loss of appetite]/nausea/vomiting, diarrhoea, altered mental status;
(c) anosmia [loss of smell] and ageusia [loss of taste].
We assume that (1) all reported symptoms were acute onset, (2) all people live or work in an area with high risk of transmission of virus and (3) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster.
The sex-specific subgroup samples excludes 2 unions because of lack of data. The above-median and below-median samples includes 85 singleton observations which were dropped.

Table A5: Pilot Analyses

	Main Intervention	Pilot 1	Pilot 2
	No Baseline Controls		
Intervention Effect	0.290***	0.109	0.284
	(0.012)	[-0.161, 0.320]	[0.081, 0.408]
Average Control Mask Wearing Rate§	0.1329	0.129	0.095
	With Baseline Controls		
Intervention Effect	0.290***	0.096	0.341
	(0.012)	[-0.126, 0.315]	[0.135, 0.509]
N villages	572	10	10

Standard errors are in parentheses. Confidence intervals are in brackets, computed using wild bootstrap.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
The first column reports the results of our main intervention; equivalent to the results in Table 1, using full surveillance data.
§We report the mean rate of mask wearing among the control villages after the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.

Table A6: Chawkidar Intervention

	All Treatment Villages	All Villages that Ever Had Chawkidar Intervention
Post-Chawkidar Intervention Coefficient	0.011	0.026
	(0.016)	(0.034)
N village-surveillance weeks	1,325	953

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regressions include an indicator for each village, week of observation, and calendar week.

Table A7: Village-Level Cross Randomizations

Coefficient	Full	No Active Promotion	Mosques	Markets	Other Locations
No Baseline Controls					
Mask Type (Surgical)	0.025	0.025	0.056	0.015	0.017
	(0.026)	(0.025)	(0.036)	(0.027)	(0.025)
Commitment w/ Signage	–0.008	–0.005	–0.016	–0.006	–0.006
	(0.026)	(0.026)	(0.034)	(0.027)	(0.026)
Incentive Type					
Monetary	–0.021	–0.023	0.012	–0.032	–0.026
	(0.035)	(0.034)	(0.046)	(0.035)	(0.036)
Certificate	0.009	0.007	0.028	0.010	–0.001
	(0.032)	(0.032)	(0.041)	(0.032)	(0.033)
100% Text	–0.028	–0.023	–0.043	–0.023	–0.017
	(0.026)	(0.025)	(0.033)	(0.027)	(0.026)
With Baseline Controls					
Mask Type (Surgical)	0.025	0.024	0.055	0.014	0.018
	(0.026)	(0.025)	(0.035)	(0.027)	(0.025)
Commitment w/ Signage	–0.005	–0.001	–0.018	–0.003	–0.003
	(0.026)	(0.026)	(0.034)	(0.027)	(0.026)
Incentive Type					
Monetary	–0.020	–0.023	0.012	–0.031	–0.026
	(0.034)	(0.034)	(0.045)	(0.034)	(0.036)
Certificate	0.013	0.010	0.032	0.013	0.002
	(0.031)	(0.031)	(0.040)	(0.031)	(0.032)
100% Text	–0.026	–0.021	–0.037	–0.022	–0.017
	(0.026)	(0.025)	(0.033)	(0.027)	(0.026)
N villages	286	286	285	285	284

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
The baseline control regressions include controls for baseline rates of mask wearing and baseline symptom rates.
Baseline Symptom Rate is defined as the rate of surveyed individuals in a village who report symptoms coinciding with the WHO-definition of a probable COVID-19 case. This is defined as any of the following:
(a) fever and cough;
(b) any three of the following (fever, cough, general weakness/fatigue, headache, muscle aches, sore throat, coryza [nasal congestion or runny nose], dyspnoea [shortness of breath or difficulty breathing], anorexia [loss of appetite]/nausea/vomiting, diarrhoea, altered mental status;
(c) anosmia [loss of smell] and ageusia [loss of taste].
We assume that (1) all reported symptoms were acute onset, (2) all people live or work in an area with high risk of transmission of virus and (3) all people have been a contact of a probable or confirmed case of COVID-19 or are linked to a COVID-19 cluster.
”Other Locations” include the Tea Stall, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
The analysis excludes 14 villages in the full sample, 15 villages in the mosque and market sub-samples, and 16 villages in the other location sub-sample because the observation data has not yet been submitted by our surveyors.

Table A8: Household-Level Cross Randomizations

Coefficient	Full
Household-Level Text Randomization	
50% of Households in Village	–0.050
	(0.032)
100% of Households in Village	–0.053
	(0.033)
Altruistic Messages	–0.026
	(0.026)
Verbal Commitment	–0.023
	(0.025)
Mask Type (Surgical)	0.034*
	(0.020)
Mask Color	
Blue	0.029*
	(0.016)
Violet	0.058***
	(0.018)
N villages	286

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
Surgical masks distributed to households were blue or green. Cloth masks distributed to households were violet or red.
The analysis excludes 14 villages and their pairs because the observation data has not yet been submitted by our surveyors.

B Pairwise Randomization Procedure

Villages were assigned to strata as follows:

1. We began with 1,000 villages in 1,000 separate unions to ensure sufficient geographic distance to prevent spillovers (Bangladesh is divided into about 4,600 separate unions).

2. We collected these unions into "Units", defined as the intersection of upazila x (above/below) median population x case trajectory, where above/below median population was a 0-1 indicator for whether the union had above-median population for that upazila and case trajectory takes the values -1, 0, 1 depending on whether the cases per 1,000 are decreasing, flat or increasing. We assessed cases per person using data provided to us from the Bangladeshi government for the periods June 27th-July 10th and July 11th-July 24th, 2020.

3. If a unit contained an odd number of unions, we randomly dropped one union.

4. We then sort unions by "cases per person" based on the July 11th-July 24th data, and create pairs of unions. We randomly kept 300 such pairs.

5. We randomly assigned one union in each pair to be the intervention union.

6. We then tested for balance with respect to cases, cases per population, and density.

7. Finally, we repeated this entire procedure 50 times, selecting the seed that minimized the maximum of the absolute value of the balance tests with respect to case trajectory and cases per person.

C Cross-Randomization Procedure

Unions were assigned to village-level cross-randomizations as follows:

1. We began with the 300 union-pairs (600 villages total) identified in the pairwise randomization procedure, and limited to only the villages in the intervention group.

2. Using a random number generator, we ordered the villages, and assigned the first 1/3 of the intervention villages to be distributed cloth masks and 2/3 to be distributed surgical masks.

3. Within the mask-type randomization, we randomly reordered the unions, then assigned the first 1/2 of villages to hang signage on their door as a visual commitment to mask-wearing, and 1/2 of villages to not have signage on their door.

4. Within the previous two randomizations, we randomly assigned 1/4 of villages to receive no incentive, 1/4 to receive a monetary award, and 1/2 to receive a certificate incentive. If there was an odd-number of villages within this randomization, then we broke the difference by rounding the number of villages in the randomization to the nearest whole number.

5. In villages without signage, we randomly ordered the villages and assigned the first 2/3 to receive texts encouraging mask-wearing, and the remaining 1/3 receive no such messages. If the number of villages was not divisible by thirds, then we broke the difference by rounding the number of villages to the nearest whole number.

Unions were assigned to household-level cross-randomizations using the following procedure. Note that each village may have only one household-level randomization.

1. In villages with the signage randomization, we assigned 2/3 of villages to receive messages emphasizing the self-protection benefits of masks, and the remaining 1/3 to receive altruistic messages about the benefits of mask wearing in addition to the self-protection messages. If the number of villages was not divisible by thirds, we broke the difference by rounding to the nearest whole number.

2. In villages with the signage randomization and no household-level altruism randomization (and by definition, no village-level text message randomization), we assigned 1/4 of villages to receive no household-level text-message randomization, 1/2 of villages to have 50% of their households receive text-message reminders, and the remaining 1/4 of villages to have 100% of their households receive texts.

3. In villages without the signage randomization, we assigned 2/3 of villages to receive messages emphasizing the self-protection benefits of masks, and the remaining 1/3 to receive messages emphasizing the altruistic reasons to wear masks in addition to the self-protection messages.

4. In the villages without the signage randomization and no household-level altruism randomization, we asked households to make a verbal commitment to be a mask-wearing household.

D Statistical Analysis

This section describes details of our statistical analyses.

Mask-Wearing We created a data set with an observation for each village j. We defined proper mask use as anyone wearing either a project mask or an alternative face-covering that covered their mouth and nose. We considered two definitions of the proportion of observed individuals wearing masks (p_j). In our primary specification, we defined p_j using all observed adults. In a secondary specification, we considered adults observed only in locations where we there was not simultaneous mask distribution. The purpose of this second specification was to investigate separately whether the intervention increased mask-wearing in places where we did not have promoters on site.

Our goal was to estimate the impact of the intervention on the probability of mask-wearing, defined as $\psi_1 = E_x[E(p_j|T_j = 1, x_j) - E(p_j|T_j = 0, x_j)]$ where T_j is an indicator for whether a village was treated and x_j is a vector of the village-level covariates, including the prevalence of baseline mask-wearing in each village (constructed analogously to p_j), baseline respiratory symptom rates, and indicators for each pair of villages from our pairwise stratification method.

We estimated this equation at the village-level with an ordinary least squares regression, using analytic weights proportional to the number of observed individuals (the denominator of p_j) and heteroskedastic-robust standard errors. In this specification, the dependent variable is p_j, the independent variable of interest was T_j, and controls were included for the x_j covariates.

Physical Distancing Using analogous methods, we estimated the impact of the intervention on the probability that wearing a mask influenced physical distancing (being within one arm's length of any other person at the time of observation).

D.1 Estimating Effects of Village-level Cross-randomizations

We analyze all four village level cross-randomizations jointly via a linear regression:

$$E(p_j|T_j, x_j, D_k) = \beta T_j + \sum_k D_k \delta_k + x_j \gamma \tag{1}$$

where $D_k = 1$ if the village has been assigned to the intervention group of the village-level cross-randomization denoted by letter k, and 0 otherwise. This specification is otherwise identical to our estimating equation for the impact of intervention on mask-wearing, with the addition of the D_k terms.

D.2 Estimating Effects of Household-level Cross-randomizations

To evaluate the effect of household-level cross-randomizations, we constructed a regression with an observation for each *village* where we ask whether masks of the color representing the treatment were more commonplace than masks of the color representing the control. In each village, we computed Δ_j, the difference in the fraction of individuals wearing treatment mask colors vs. control mask colors. We alternated across villages which color corresponds to intervention, so we can control directly for whether specific colors are more popular (denote these by d_{jc}; $d_{jc} = 1$ if treated masks in village j are color c). We index the various household randomizations by m. Our estimate for each household randomization will be α_{0m}, given by:

$$E(\Delta_j|d_{jc}) = \alpha_{0m} + \sum_c \alpha_c d_{jc} + surgical_j \tag{2}$$

α_{0m} tells us how much more likely individuals are to wear masks of the treated color than masks of the control color. $surgical_j$ is, as its name implies, a dummy for whether surgical masks were distributed in village j. We estimate this equation at the village-level by ordinary least squares, using analytic weights proportional to the number of observed individuals (the denominator of Δ_j) and heteroskedasticity-robust standard errors.

E Intervention Cost and Benefit Estimates

The average person-day of staff time in our intervention cost \$20 of wages plus \$0.50 of communication costs. All management salaries, benefits, support, internal monitoring, and equipment costs \$71,696. We exclude these from the below calculation as they will vary from setting to setting. As reported in the main text, we estimate that we induced 51,125 people to regularly wear masks, or 172 people per intervention village.

Costs per village The main fixed costs of the intervention (as opposed to costs that vary over days):

- Masks for initial household distribution (3 masks per household), (\$0.13 per surgical mask and \$0.50 per cloth masks), 370,643 cloth masks, and 924,849 surgical masks
- Staffing for initial household distribution (4 person-days per village)
- 1 person-day of training per village
- PPE for staff: \$70 per village
- Media costs: \$100 per village
- Other transportation and materials costs: \$30 per village

This amounts to fixed costs of: \$302.50 per village for non-mask materials, \$275.10 worth of cloth masks per village, and \$88.90 of surgical masks per village. We estimate that we induced 594 x

29% = 172 people per village to wear masks, which amounts to $3.36 per adult induced to wear a mask in cloth mask villages, and $2.28 per adult in surgical mask villages.

Costs per village-day of intervention The main costs paid per day of the intervention:

- 1,295,492 masks distributed over an average of 29 days per village. Of these, there were 370,643 cloth masks distributed (129 cloth masks per day per village) and 924,849 surgical masks distributed (160 surgical masks per day per village).

- 14 person-days per week per village in week 1, 8 person-days per week per village in week 2, 6 person-days per village in weeks 3, 4 and 5, and 4 person-days per week per village thereafter.

Over the first four weeks of our intervention, this amounts to mask supply costs of $64.50 per village-day for cloth masks and $20.8 per village-day for surgical masks. The promotion costs were $23 per village-day. Dividing by the number of people induced to wear masks per village (172), we obtain costs of $0.51 per person-day in cloth mask villages and $0.25 per person-day in surgical mask villages.

In Table A3, we compute the benefits of mask-wearing in terms of reductions in mortality risk as a function of duration of mask-wearing and mask-effectiveness. Based on the anticipated rate of seroconversions and estimates of the number of infections and deaths per seroconversions, we estimate the absolute reduction in deaths and thus the number needed to treat to avert one death.

For the one-month of the intervention, the number needed to treat to prevent one death ranges from 11,618 to 27,109. Our estimates above suggest that the total cost of our intervention per person induced to wear a mask for a month was: $3.36 + $0.51 x 30 = $18.66 in cloth mask villages and $2.28 + $0.25 x 30 = $9.78 in surgical mask villages. Depending on the estimate, this translates to between $144,000 and $355,000 per life saved in cloth mask villages, and between $99,000 and $232,000 per life saved in surgical mask villages.

However, these estimates improve substantially as we extend the duration of our window to consider the next several months after the intervention. Even assuming that no one continues to wear masks after the intervention, the intervention will become increasingly cost-effective as the infections prevented in the first month fail to propagate into future infections. This reduces the number needed to treat to prevent one death by a factor of 4, and improves the cost effectiveness accordingly. After a 4-month window, the cost-effectiveness of the intervention for cloth masks was between $36,000 and $88,750 per life saved, and for surgical masks was between $24,750 and $58,000 per life saved.

F Impact of Mask Use on COVID-19

F.1 Methods

Sample size calculation Sample size calculations were based on the primary outcome of symptomatic SARS-CoV-2 infections. We anticipated that we could directly survey up to 500 people per village per wave across 600 villages, or 300,000 people. We solved for the clinically detectable effect size (delta) as a function of number of individuals per intervention arm (m = 150,000), the desired significance level (5%), the desired power level (80%), the expected rates of respiratory symptoms in the intervention and control groups, the number of individuals per cluster (n = 500), and the intracluster correlation (rho = 0.0022) using the following equation[10]:

$$\delta = ((Z_{1-\alpha/2} + Z_{1-\beta}) * \sqrt{((P_1 * (1 - P_1) + P_2 * (1 - P_2)))} * \sqrt{(1 + (\rho * n - \rho)))} / \sqrt{(m)} \qquad (3)$$

We estimated the intracluster coefficient using our baseline data of respiratory symptoms from a nation-wide telephone and variation across districts in our survey, typically between 50 and 150 surveyed members per upazila. These power calculations are conservative because they ignore the possibility of controlling for baseline rates from our initial survey and of stratifying, both of which could lead to substantial increases in power.

The below table demonstrates the minimum detectable effect under different assumptions about the fraction of the population infected with SARS-CoV-2 over our sample period. The first column gives possible (true) SARS-CoV-2 infection rates in the population. The second column gives the resulting rates of respiratory disease we expect to observe in our survey assuming 2% rates with no COVID and that 50% of COVID patients are symptomatic. Column 3 reports the minimum detectable effect as an absolute reduction in respiratory symptoms, and Column 4 the implied effectiveness of masks in reducing respiratory disease, computed by dividing column 3 by column 2. Column 5 computes a "per mask" detectable effect assuming that we increase mask use by 30%

[10] See equation 5 from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4521133/

Table A9: Sample Size Calculation Parameters

Cumulative COVID-19 Incidence	Resp Disease	Detectable Effect	Total Impact of Masks	Impact Per Mask (30 percentage point increase)
5%	4.50%	0.0033	7.34%	24.47%
10%	7.00%	0.0041	5.81%	19.37%
15%	9.50%	0.0047	4.92%	16.40%
20%	12.00%	0.0052	4.31%	14.37%

(column 4 divided by 0.3).

Blood sample collection We collected capillary blood samples on a cohort of 7,500 randomly-selected participants at baseline and endline. We additionally collected blood samples from participants who reported symptoms consistent with the WHO COVID-19 case definition for suspected or probable cases during the study period.[11] Blood samples were obtained by puncture with a 20-Gauge safety lancet to the third or fourth digit. 500 microliters of blood were collected into Microtainer® capillary blood collection serum separator tubes (BD, Franklin Lakes, NJ, Cat 365967). Blood samples were transported on ice and stored at -20°C until testing.

SARS-CoV-2 testing Blood samples were tested for the presence of IgG antibodies against SARS-CoV-2 using the SCoV-2 Detect™ IgG ELISA kit (InBios, Seattle, Washington, Cat COVE-G). This assay detects IgG antibodies against the spike protein subunit (S1) of SARS-CoV-2. The assays were performed according to the manufacturer's instructions. Briefly, serum samples were diluted 1:100 with sample dilution buffer. 50 microliters of diluted specimens were added to the SCoV-2 antigen-coated microtiter strip plates. After one hour of incubation at 37°C, the plate was washed six times with wash buffer, and conjugate solution was added to each well. The plate

[11] WHO COVID-19 Case Definition https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance$_{Case_Definition}-2020.2$

was incubated for another 30 minutes at 37°C and washed six times with wash buffer. 75 microliters of liquid TMB substrate were added to all wells followed by 20 minutes of incubation in the dark at room temperature before the reaction was stopped. The absorbance was read on a microplate reader at 450nm (GloMax® Microplate Reader, Promega Corporation, Madison, WI, Cat GM3000). After calibration according to positive, negative, and cut-off controls, the immunological status ratio (ISR) was calculated as the ratio of optical density divided by the cut-off value. Samples were considered positive if the ISR value was determined to be at least 1.1. Samples with an ISR value 0.9 or below were considered negative. Samples with equivocal ISR values were retested in duplicate, and resulting ISR values were averaged.

We tested x% of samples collected from individuals who reported COVID-like symptoms during the intervention period from both intervention and control groups (estimated 15,000 blood samples). In addition, all endline blood samples from the cohort of 7,500 were tested. For any of these individuals with positive endline results for IgG antibodies against SARS-CoV-2, the baseline blood samples were also tested (estimated 7,500 endline, 3,750 baseline blood samples).

Analysis Our primary outcome is symptomatic seropositivity, defined as reporting symptoms consistent with COVID-19 during the study period and a positive endline SARS-CoV-2 IgG antibody test. We will construct a dataset with an observation for each surveyed household (i indexes households and j villages). In each village, define $Y_{ij} = 1$ if the highest risk individual in each household is 1) reporting either a. dry cough and fever and b. either fatigue, lack of taste/smell or shortness of breath in the last month at either the fifth week or ninth week telephone survey and 2) are seropositive in our blood test at endline. If either of these conditions fail to hold, $Y_{ij} = 0$. To assess seropositivity, we will test all individuals who are symptomatic in either our 5-week or 9-week household survey.

Our goal will be to estimate the impact of the intervention on seropositivity, defined as: $\psi_0 = E_x[E(Y_{ij}|T_j = 1, x_j) - E(Y_{ij}|T_j = 0, x_j)]$ where T_j is an indicator for whether a village was treated and x_j are village-level covariates including baseline mask-use in each village (constructed as described below) and baseline respiratory symptom based on reported symptoms, as well as indicators for each pair of villages from our pairwise stratification method. In an auxiliary specification, we will also include fixed effects for each observation staff member.

We will estimate this parameter using a generalized linear model with a normal family (and identity link), clustering at the village-level.[12] The dependent variable is Y_{ij}, the independent variable of interest is T_j, and controls will be included for the x_j covariates.

To estimate the overall impact of masks on seroconversions, we have chosen a random cohort of 7,500 individuals from which we collected blood at baseline. The results of baseline antibody testing in this cohort will allow us to estimate baseline symptomatic seropositivity. By differencing the seropositivity rate among symptomatic individuals at baseline and endline, we can compute symptomatic seroconversions. We can determine the fraction of seroconversions prevented by dividing ψ_0 by the overall rate of symptomatic seroconversions.

[12] Hardin, James W and Hilbe, Joseph, Generalized linear models and extensions, Stata press,2007.