# The Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh

Jason Abaluck[1,*,†], Laura H Kwong[2,3,†], Ashley Styczynski[4,†]

Ashraful Haque[5], Md. Alamgir Kabir[5], Ellen Bates-Jefferys[6]

Emily Crawford[1], Jade Benjamin-Chung[7], Shabib Raihan[5]

Shadman Rahman[5], Salim Benhachmi[8], Neeti Zaman[5]

Peter J. Winch[9], Maqsud Hossain[10], Hasan Mahmud Reza[11],

Abdullah All Jaber[10], Shawkee Gulshan Momen[10],

Aura Rahman[10], Faika Laz Banti[10], Tahrima Saiha Huq[10],

Stephen P. Luby[2,4,‡], Ahmed Mushfiq Mobarak[1,12,*,‡]

November 8, 2021

**Summary**: We ran a randomized trial of mask promotion in Bangladesh; the intervention increased mask-use and reduced symptomatic SARS-CoV-2 infections.

1. Yale School of Management, Yale University; New Haven, CT, USA. 2. Woods Institute for the Environment, Stanford University; Stanford, CA, USA. 3. Division of Environmental Health Sciences, University of California Berkeley; Berkeley, CA, USA. 4. Division of Infectious Diseases and Geographic Medicine, Stanford University; Stanford, CA, USA. 5. Innovations for Poverty Action Bangladesh; Dhaka, Bangladesh. 6. Innovations for Poverty Action; Evanston, IL, USA. 7. Department of Epidemiology and Population Health, School of Medicine, Stanford University. 8. Yale Research Initiative on Innovation and Scale, Yale University; New Haven, CT, USA. 9. Social and Behavioral Interventions Program, Johns Hopkins Bloomberg School of Public Health; Baltimore, MD, USA. 10. NGRI, North South University; Dhaka, Bangladesh. 11. Department of Pharmaceutical Sciences, North South University; Dhaka, Bangladesh. 12. Deakin University; Melbourne, Australia.

† denotes co-first author, ‡ denotes co-last author.

***Corresponding authors. Email: jason.abaluck@yale.edu and ahmed.mobarak@yale.edu**

## Abstract

We conducted a cluster-randomized trial to measure the effect of community-level mask distribution and promotion on symptomatic SARS-CoV-2 infections in rural Bangladesh from November 2020 to April 2021 (N = 600 villages, N = 342,183 adults). We cross-randomized mask type (cloth vs. surgical) and promotion strategies at the village and household level. Proper mask-wearing increased from 13.3% in the control group to 42.3% in the intervention arm (adjusted percentage point difference = 0.29 [0.26, 0.31]). The intervention reduced symptomatic seroprevalence (adjusted prevalence ratio (aPR) = 0.91 [0.82, 1.00]), especially among adults 60+ years in villages where surgical masks were distributed (aPR = 0.65 [0.45, 0.85]). Mask distribution and promotion was a scalable and effective method to reduce symptomatic SARS-CoV-2 infections.

**Trial registration**: ClinicalTrials.gov Identifier: NCT04630054

**Funding**: GiveWell.org

# 1 Introduction

As of September 2021, the COVID-19 pandemic has taken the lives of more than 4.7 million people. Inspired by the growing body of scientific evidence that face masks have the potential to slow the spread of the disease and save lives (*1–10*), we conducted a cluster-randomized controlled trial covering 342,183 adults in 600 villages in rural Bangladesh with the dual goals of (a) identifying strategies to increase community-wide mask-wearing, and (b) tracking changes in symptomatic SARS-CoV-2 infections as a result of our intervention. While vaccines may constrain the spread of SARS-CoV-2 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 (*11*). Developing scalable and effective means of combating COVID-19 is thus of first-order policy importance.

The World Health Organization declined to recommend mask adoption until June 2020, citing the lack of evidence from community-based randomized-controlled trials, as well as concerns that mask-wearing would create a false sense of security (*12*). Critics argued those who wore masks would engage in compensating behaviors, such as failing to physically distance from others, resulting in a net increase in transmission (*13*). We directly test this hypothesis by measuring physical distancing.

We designed our trial to encourage *universal* mask-wearing at the community level, rather than mask-wearing among only those with symptoms. We encouraged even healthy individuals to wear masks since a substantial share of COVID-19 transmission stems from asymptomatic or pre-symptomatic individuals (*14*), and masks may protect healthy wearers by reducing the inhalation of aerosols or droplets (*15–17*).

After piloting, we settled on a core intervention package that combined household mask distribution with communication about the value of mask-wearing, mask promotion and in-person reminders at mosques, markets, and other public places, and role-modeling by public officials and community leaders. We also tested several other strategies in sub-samples, such as asking people to make a verbal commitment, creating opportunities for social signaling, text messages, and providing village-level incentives to increase mask-wearing. The selection of strategies to

test was informed by both our pilot results and research in public health, psychology (*18–20*), economics (*21–23*), marketing (*24–26*), and other social sciences (*27*) on product promotion and dissemination strategies. We tested many different strategies because it was difficult to predict in advance which ones would lead to persistent increases in mask-wearing. Prediction studies we conducted with policymakers and public health experts at the World Health Organization, India's National Council of Applied Economic Research, and the World Bank suggest that even these experts with influence over policy design could not easily predict which specific strategies would prove most effective in our trial.

We powered our intervention around the primary outcome of symptomatic seroprevalence. During our study, we collected survey data on the prevalence of WHO-defined COVID-19 symptoms from all available study participants, and then collected blood samples at endline from those who reported symptoms anytime during the 8-week study. Our trial is therefore designed to track the fraction of individuals who are *both* symptomatic and seropositive. We chose this as our primary outcome because (a) the goal of public health policy is ultimately to prevent symptomatic infections (even if preventing asymptomatic infections is instrumentally important in achieving that goal) and (b) symptomatic individuals are far more likely to be seropositive so powering for this outcome required conducting an order of magnitude fewer costly blood tests. As secondary outcomes, we also report the effects of our intervention on WHO-defined symptoms for probable COVID-19 and mask-wearing.

Bangladesh is a densely populated country with 165 million inhabitants; reported infections reached 15,000 per day in during our study period, but reported cases and deaths are likely underestimated by 1-2 orders of magnitude (*28–32*). The evolution of mask use over time in Bangladesh is discussed in greater detail in (*33*). In Bangladesh, the government strongly recommended mask use from early April 2020. In an April 2020 telephone survey, over 80% of respondents self-reported wearing a mask and 97% self-reported owning a mask. 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. During in-person

surveillance between May 21-25, 2020 in 1,441 places in 52 districts, we observed 51% of approximately 152,000 individuals wearing a mask. In another wave of surveillance was conducted between June 19-22, 2020 in the same 1,441 locations, and mask-wearing dropped to 26%, with 20% wearing masks that covered their mouth and nose and 6% wearing masks improperly. An August 2020 phone survey in rural Kenya found that while 88% of respondents claim to wear masks in public, direct observation revealed that only 10% actually did (*34*). These observations suggest that mask promotion interventions could be useful in rural areas of low- and middle-income countries (LMIC), home to several billion people at risk for COVID-19.

# 2 Results

Our analysis followed our preregistered analysis plan (https://osf.io/vzdh6/) except where indicated. Our primary outcome was symptomatic seroprevalence for SARS-CoV-2. We also analyzed the impact of our intervention on mask-wearing, physical distancing, social distancing, and COVID-like symptoms. No adverse events were reported during the study period.

## 2.1 Sample Selection

Tables S1 and S2 summarizes sample selection for our analysis. We initially approved 134,050 households, of which 125,053 provided baseline information. From these 125,049 households, we collected baseline information from 342,183 individuals. Of these, 336,010 (98%) provided symptom data at week 5 and/or 9. Of these, 27,160 (8.0%) reported COVID-like symptoms during the 9-weeks since the study began. We attempted to collect blood samples from all symptomatic individuals. Of these, 10,790 (39.7%) consented to have blood collected (40.2% in the treatment group and 39.3% in the control group; $p = 0.24$). We show in Table S3 that consent rates are about 40% across men and women and among adults of different age groups in both treatment and control villages.

As such, the sample of individuals for whom we have symptom data is much larger than the

sample for whom we have serology data. We tested 9,512 (88.2%) of the collected blood samples to determine seroprevalence for SARS-CoV-2 IgG antibodies. Untested samples (<12%) either lacked sufficient quantity for our test or could not be matched to individuals from our sample because of a barcode scanning error. In our primary outcome analysis, we drop individuals for whom we are missing symptom data or who did not consent to blood sample collection. For the analyses where symptomatic status is the outcome, we report results using both this smaller sample, as well as the larger sample of all individuals who provided symptom data. In the baseline, we collected blood samples from a random sample of individuals (n = 10,085), of whom 339 had COVID-like symptoms. We use these to check balance with respect to baseline symptomatic seropositivity (as well as baseline symptomatic status).

Of the 600 villages initially recruited for the study, the analysis sample excludes 4 villages where interventions could not be performed due to lack of local government cooperation. We exclude an additional 11 villages and their village-pairs because we did not observe them in the baseline period prior to the intervention, and 1 village and its pair for lack of observational data throughout the intervention period, for a total analysis sample of 572 villages.

## 2.2 Primary Analyses

**Balance** While our stratification procedure should have achieved balance with respect to variables observed at the time of randomization, given the many possible opportunities for errors in implementation, we confirm in Appendix L that our control and treatment villages are balanced with respect to our primary outcome variables. This assessment was not preregistered. We investigate several other covariates and find a few small imbalances. We check whether these affect the main results we report in this paper. For example, we find more 18-30 year olds in the treatment group than in control, perhaps because households reported teenagers as 18 in order to receive more masks; our results are robust to dropping this age range.

**First and Second Stage Outcomes** We first report results in Table 1 on the effects of our core intervention package described in section 4.3 (free mask distribution, communication about the value of mask-wearing, mask promotion and in-person reminders in public places, and role-modeling by community leaders) on "first-stage" behavioral outcomes that may have changed due to the mask promotion intervention - rates of mask-wearing and physical distancing. Both were measured through direct observation, with surveillance procedures detailed in section 4.6 and Appendix G. We follow this up in Tables 2 and 3 with the primary epidemiological outcomes that may have changed secondary to those behavioral responses.

**Mask-Wearing** The first column in the top panel 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. More details of our statistical methods and standard error construction are available in Appendix K. Mask-wearing was 13.3% in control villages and 42.3% in treatment villages. Our regression adjusted estimate is an increase of 28.8 percentage points (95% CI: 0.26, 0.31). If we omit all covariates (except fixed effects for the strata within which we randomized), our point-estimate is identical (Table S5). Considering only surveillance conducted when no mask distribution was taking place, mask-wearing increased 27.9 percentage points, from 13.4% in control villages to 41.3% in intervention villages (regression adjusted estimate: 0.28, 95% CI: 0.26, 0.30). We also run our 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-wearing was largest in mosques (37.0 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 did not find evidence that our intervention undermines distancing behavior. In the second panel of Table 1, we report identical specifications to the first panel, but with physical distancing as the dependent variable. In control villages 24.1% of observed individuals practiced physical distancing compared to 29.2% in intervention villages, an increase of 5.1% (a regression adjusted estimate

Table 1: Mask-Wearing and Physical Distancing, Controlling for Baseline Variables

| | Full | No Active Promotion | Mosques | Markets | Other Locations | Surgical Mask Villages | Cloth Mask Villages |
|---|---|---|---|---|---|---|---|
| | | | *Proper Mask-Wearing* | | | | |
| Intervention Coefficient | 0.288*** | 0.279*** | 0.370*** | 0.287*** | 0.251*** | 0.301*** | 0.256*** |
| | (0.012) | (0.011) | (0.016) | (0.012) | (0.012) | (0.015) | (0.019) |
| | | | *Physical Distancing* | | | | |
| Intervention Coefficient | 0.051*** | 0.056*** | 0.000 | 0.074*** | 0.068*** | 0.054*** | 0.044*** |
| | (0.005) | (0.005) | (0.000) | (0.007) | (0.006) | (0.006) | (0.011) |
| N villages | 572 | 572 | 570 | 570 | 568 | 380 | 192 |

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 control-intervention pair and baseline symptom rates. The analyses in the top panel control for baseline rates of proper mask wearing, and the analyses in the bottom panel control for baseline rates of physical distancing.
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. 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.
"No Active Promotion" refers to any time that surveillance was conducted while promotion was not actively occurring (regardless of the week of the intervention). This excludes surveillance during the Friday Jumma Prayers in the mosque, when promoters were present and actively encouraged mask wearing.
"Other Locations" include tea stalls, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
"Surgical Villages" refer to all treatment villages which received surgical masks as part of the intervention, and their control pairs. "Cloth Villages" refer to all treatment villages which received cloth masks as part of the intervention, and their control pairs. The surgical and cloth sub-samples include surveillance from all available locations, equivalent to the to the column labeled "Full", but run separately for each subgroup.
Of the 572 villages included in the analyses sample, we exclude an additional village and its pair in the mosque and market sub-samples, and two villages and their pairs in the other location sub-sample because we did not observe them in the baseline period prior to the intervention. There are 190 treatment villages which received surgical masks as part of the intervention and 96 treatment villages which received cloth masks.

of 0.05 [95% CI: 0.04,0.06]). Evidently, protective behaviors like mask-wearing and physical-distancing are complements rather than substitutes: endorsing mask-wearing and informing people about its importance encouraged rural Bangladeshis to take the pandemic more seriously and engage in another form of self-protection. The increases in physical distancing were similar in cloth and surgical mask villages.

Physical distancing increased 5.1 percentage points overall but there was substantial heterogeneity across locations. In markets, individuals become 7.4 percentage points more likely to physically distance. In contrast, there was no physical distancing practiced in any mosque, in either treatment or control villages, probably as a result of the strong religious norm of standing shoulder-to-shoulder when praying.

**Social Distancing** It is possible that physical distancing increases because our intervention results in fewer total people being present in public spaces. If socializing increased in the intervention group, but only among risk-conscious people, then we might see physical distancing increase despite people engaging in overall riskier behavior. To assess this, as well as to assess directly if the intervention increased socializing, we study the effects of our intervention on the total number of people observed at public locations. While surveillance staff were not able to count everyone in busy public areas, the total number of people they were able to observe gives some indication of the crowd size. We find no difference in the number of people observed in public areas between the treatment and control groups overall (Table S6). The social distancing analysis was not pre-registered, although the specification exactly parallels our analysis of physical distancing.

**Symptomatic Seroprevalence** Among the 336,010 participants who completed symptom surveys, 27,160 (8.1%) reported experiencing COVID-like illnesses during the study period. More participants in the control villages reported incident COVID-like illnesses (n=13,853, 8.6%) compared with participants in the intervention villages (n=13,307, 7.6%). Over one-third (39.7%) of symptomatic participants agreed to blood collection. Omitting symptomatic participants who did not consent to blood collection, symptomatic seroprevalence was 0.76% in control villages and

0.68% in the intervention villages. Because the fractions we are reporting omit non-consenters from the numerator but not the denominator, it is likely that the true rates of symptomatic seroprevalence are substantially higher (perhaps by 2.5 times, if non-consenters have similar seroprevalence to consenters).

In Table 2 (and Table S7), we report results from a regression of symptomatic seroprevalence on a treatment indicator, clustering at the village level and controlling for fixed effects for each pair of control-treatment villages. In the tables, we report results with and without additional controls for baseline symptoms and mask-wearing rates. In Table S7, we report results from our pre-specified linear model and in Table 2 we report results from a generalized linear model with a Poisson family and log-link function. Here we discuss the latter results (which are in units of relative risk); the linear model implies results of an almost identical magnitude.

The results in all specifications are the same: we estimate a roughly 9% decline in symptomatic seroprevalence in the treatment group (adjusted prevalence ratio (aPR) = 0.91 [0.82, 1.00]) for a 29 percentage point increase in mask wearing over 8 weeks.[1] In the second column of Tables 2 and S7, we split our results by mask type (surgical vs. cloth). We find clear evidence that surgical masks lead to a relative reduction in symptomatic seroprevalence of 11.1% (aPR = 0.89 [0.78,1.00]; control prevalence = 0.81%; treatment prevalence = 0.72%). Although the point estimates for cloth masks suggests that they reduce risk, the confidence limits include both an effect size similar to surgical masks and no effect at all. (aPR = 0.94 [0.78,1.10]; control: 0.67%; treatment: 0.61%).

In Appendix N, we investigate the robustness of these results to alternative methods of dealing with missing data from non-consenters. In the main text, following our pre-specified analysis plan, we drop non-consenting symptomatic individuals. If we instead impute seropositivity for symptomatic non-consenters based on the population average seropositivity among symptomatic individuals, our pooled estimate of the impact of masking becomes larger and more precise. Notably,

[1] The confidence interval reported in the text corresponds to the specification in Table 2 with baseline controls (hence, "adjusted" prevalence ratio). To check robustness to the type of clustering, in panels S3a and S3b of Figure S3, we show the histogram of effect sizes arising from "randomization inference" if we randomly reassign treatment within each pair of villages and then estimate our primary specification. When doing so, we find that our estimated effect size is smaller than 7.0% of the simulated estimates with controls and 7.4% of the simulated estimates without controls (these are the corresponding p-values of the randomization inference $t$-test).

with this alternative imputation, we find effects for both cloth and surgical masks on symptomatic seroprevalence.

Not all symptomatic seroprevalence is necessarily a result of infections occurring during our intervention; individuals may have pre-existing SARS-CoV-2 infections and then become symptomatic (perhaps caused by an infection other than SARS-CoV-2). In Appendix I, we show that if either: a) masks have the same proportional impact on COVID and non-COVID symptoms or b) all symptomatic seropositivity is caused by infections during our intervention, then the percentage decline in symptomatic seroprevalence will exactly equal the decline in symptomatic seroconversions. More generally, the relationship between the two quantities depends on whether masks have a greater impact on COVID or non-COVID symptoms, as well as the proportion of symptomatic seropositivity that is a result of infections pre-existing at baseline.

**WHO COVID-19 Symptoms** In Tables 3 and S8, we report results from the same specifications with WHO-defined COVID-19 symptomatic status as the outcome. This is defined as any of following:

- Fever and Cough;
- 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;
- Anosmia [loss of smell] and ageusia [loss of taste].

We find clear evidence that the intervention reduced symptoms: we estimate a reduction of 11.6% (aPR = 0.88 [0.83,0.93]; control: 8.60%; treatment: = 7.63%). Additionally, when we look separately by cloth and surgical masks, we find that the intervention led to a reduction in COVID-like symptoms under either mask type ($p = 0.000$ for surgical, $p = 0.066$ for cloth), but the effect size in surgical mask villages was 30-80% larger depending on the specification. In Table S9, we run the same specifications using the smaller sample used in our symptomatic seroprevalence regression (i.e. those who consented to give blood). In this sample we continue to find an effect overall and an effect for surgical masks, but see no statistically significant effect for cloth masks.

Table 2: Symptomatic Seroprevalence, Expressed in Prevalence Ratios

| | Intervention Effect | Intervention Effect by Mask Type |
|---|---|---|
| *No Baseline Controls* | | |
| Intervention Prevalence Ratio | 0.905** [0.815, 0.995] | |
| Intervention Prevalence Ratio for Surgical Mask Villages | | 0.894* [0.782, 1.007] |
| Intervention Prevalence Ratio for Cloth Mask Villages | | 0.925 [0.766, 1.083] |
| Average Symptomatic-Seroprevalence Rate in Paired Control Villages§ | 0.0076 | 0.0076 |
| *With Baseline Controls* | | |
| Intervention Prevalence Ratio | 0.905** [0.815, 0.995] | |
| Intervention Prevalence Ratio for Surgical Mask Villages | | 0.889** [0.780, 0.997] |
| Intervention Prevalence Ratio for Cloth Mask Villages | | 0.942 [0.781, 1.103] |
| N individuals | 304,726 | 304,726 |
| N villages | 572 | 572 |

Confidence Intervals are in brackets.
*** 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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood. The regressions excludes an additional 17,377 individuals in 34 villages because there are 0 people who are symptomatic-seropositive in their village pairs.

## 2.3 Mechanisms for Increasing Mask-Wearing

Our core intervention package combined multiple distinct elements: we provided people with free masks and information about the importance of mask-wearing; we had mask promoters reinforce by stopping individuals in public places who were not wearing masks and reminding them, and we partnered with local leaders to encourage mask-wearing at mosques and markets. Additionally, in some villages we provided a variety of reminders, commitment devices, and incentives for village leaders. In Appendix J, we attempt to disentangle the role played by these different elements in encouraging mask use.

We find no evidence that any of our village-level or household-level treatments, other than mask color, impacted mask-wearing. For mask-color, we see marginally significant differences, small in magnitude. In surgical mask villages, blue masks were more likely to be observed than green (adjusted percentage point difference = 0.03, [-0.00,0.06]), and in cloth mask villages, red more likely than purple (adjusted percentage point difference = -0.02, [-0.04,-0.00]). Text message reminders, incentives for village-leaders, or explicit commitment signals explain little of the observed increase in mask-wearing. 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. We saw no significant difference in the rates of mask-wearing in the village-level randomization of surgical vs. cloth masks.

We do find non-experimental evidence that in-person mask promotion and reinforcement is a crucial part of our intervention. Our first pilot contained all elements of our intervention except in-person reinforcement. Our second pilot (one week later) and the full intervention (several months later) added in-person reinforcement. Under the assumption that treatment effects would otherwise be constant over time, we find that in-person reinforcement accounts for 19.2 percentage points of our effect (regression adjusted estimate 0.19 [-0.33,-0.05]), or 65% of the total effect size. In Table S10, we show that this difference is statistically significant whether or not we include baseline controls. This was not a pre-specified analysis.

Table 3: WHO-Defined COVID-19 Symptoms, Expressed in Prevalence Ratios

| | Intervention Effect | Intervention Effect by Mask Type |
|---|---|---|
| | *No Baseline Controls* | |
| Intervention Prevalence Ratio | 0.885***<br>[0.834, 0.934] | |
| Intervention Prevalence Ratio for Surgical Mask Villages | | 0.865***<br>[0.803, 0.928] |
| Intervention Prevalence Ratio for Cloth Mask Villages | | 0.922*<br>[0.838, 1.005] |
| Average Symptomatic Rate Rate in Paired Control Villages§ | 0.0860 | 0.0860 |
| | *With Baseline Controls* | |
| Intervention Prevalence Ratio | 0.884***<br>[0.834, 0.934] | |
| Intervention Prevalence Ratio for Surgical Mask Villages | | 0.874***<br>[0.809, 0.939] |
| Intervention Prevalence Ratio for Cloth Mask Villages | | 0.907**<br>[0.823, 0.991] |
| N individuals | 321,948 | 321,948 |
| N villages | 572 | 572 |

Confidence Intervals are in brackets.
*** 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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic status at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

## 2.4 Persistence of Effects over Time

In Appendix M, we present results on mask-wearing after our intervention ended. Even though the door-to-door free mask distribution occurred in the first week only, there was almost no attenuation of mask-wearing over the initial 10 weeks of surveillance. Notably, mask-wearing remained comparably elevated in the treatment group during the two weeks we continued surveillance after the end of all intervention activities in the village. 3-4 months later, mask-wearing waned, but remained 10 percentage point higher in treatment regions.

## 2.5 Subgroup Analyses

We also considered how the impact of our intervention differed between subgroups.

**Mask-Wearing by Gender** In Table S11, we analyze the impact of our intervention on mask-wearing and physical distancing separately by gender, as well as by whether baseline mask-wearing was above or below the median. Gender was recorded in 65% of observations; age was not recorded during the direct observation surveillance of mask-wearing in public places, and thus we do not conduct an age-stratified assessment. In the gender results, we drop surveillance observations for mosques because in rural Bangladesh it is rare for women to attend mosque. We found that the intervention increased mask-wearing by 27.1 percentage points for men ([0.25,0.30]) and 22.5 percentage points for women ([0.20,0.25]). Although we do not have the variation to test this, the gendered difference in effect size may be because our mask promoters were predominantly men, or because the mask-wearing rate in control villages was so much higher for women (31% for women vs. 12% for men). We intentionally hired predominantly men because most staff interactions would be with men. Men constituted 88.2% of all observed adults.

We also found a larger increase in mask-wearing in villages with below-median baseline mask-wearing (where mask-wearing increased from 8.7% to 41.9% at endline) than those with above-median baseline mask-wearing (where the increase was from 17.5% to 42.6%).

**Symptomatic Seroprevalence by Age** In Tables 4 and S12, we report results from our primary specification separately by age. Table S12 reports our preregistered specification, a linear model run separately for each decade of age, pooling cloth and surgical villages. Table 4 synthesizes these results, collapsing by categories of <40, 40-50, 50-60 and 60+, reporting results as a relative risk reduction, and showing results separately for surgical and cloth masks. We generally find that the impact of the intervention is concentrated among individuals over age 50. In surgical mask villages, we observe a 22.8% decline in symptomatic seroprevalence among individuals aged 50-60 (adjusted prevalence ratio of 0.77 [0.60,0.95]) and a 35.3% decline among individuals aged 60+ in our baseline specification ($p = 0.000$) (adjusted prevalence ratio of 0.65 [0.45, 0.85]). For cloth masks, we find an insignificant (5%) reduction overall, but some evidence of a reduction in symptomatic seroprevalence among 40-49 year olds; we investigate more deeply in Appendix N, and find that the age gradient appears to be sensitive to how we deal with missing values. In the second panel of Table 4, we report results where we impute the population average seroprevalence among all non-consenters rather than dropping them. This alternative approach yields more precise overall estimates, and suggests that both cloth and surgical masks have greater impacts on symptomatic seroprevalence at older ages, although the impact of surgical masks among age 60+ is smaller than in our baseline specification. Ex ante, it is not obvious to us which imputation method should be preferred, although the second approach makes our results less sensitive to differential consent rates that we observe in some waves of our intervention, as discussed in Appendix N.

**WHO COVID-19 Symptoms by Age** In Tables S13 and S14 (the latter our preregistered specification), we perform the same analysis using the larger sample of individuals who reported symptom information. In this sample, we continue to find larger effects at older ages, although the differences are not as stark as for the symptomatic seroprevalence outcome. In Table S15, we show that the age gradient is steeper for surgical masks.

**Biological Outcomes by Gender** In Appendix N and Table S28, we show results for symptoms and symptomatic seropositivity by gender. We see a similar pattern to the cloth and surgical results: we see significant effects for both genders for symptoms and symptomatic seroposivity when we impute seropositivity at the average value for non-consenters. If we instead drop non-consenters, the symptomatic seropositivity estimates for men become less precise and are no longer significantly different from zero, while the estimates for women remain unchanged.

**Additional Preregistered Specifications** In Appendix P, we discuss additional preregistered specifications not reported in the text, either because they were substantially underpowered given the available data or because data on required variables was unavailable. We also discuss ways in which trial implementation deviated from our pre-registered protocol, such as switching from exclusively phone surveys to household visits at weeks 5 and 9 in order to increase response rates.

## 2.6 Intervention Cost and Benefit Estimates

In Appendix Q, we assess the costs of implementing our intervention relative to the health benefits, specifically focusing on our ongoing efforts to implement this same intervention at scale in Bangladesh. We consider a range of possible estimates for excess deaths from COVID-19 from May 1, 2021 - September 1, 2021, and we assume that our age-specific impacts on symptomatic seroprevalence will lead to proportional to reductions in mortality. We estimate that a scaled version of our intervention being implemented in Bangladesh will cost about $1.50 per person, and between $10K and $52K per life saved, depending which estimate we use for excess deaths.

Table 4: Symptomatic Seroprevalence by Age Groups and Mask Type, Expressed in Prevalence Ratios

| | All | < 40 Y.O. | 40-49 Y.O. | 50-59 Y.O. | ≥ 60 Y.O. |
|---|---|---|---|---|---|
| *Pre-Registered Sample: Drop Individuals Without Blood Draws* | | | | | |
| Intervention Prevalence Ratio for Surgical Mask Villages | 0.889** [0.780, 0.997] | 0.967 [0.834, 1.100] | 1.009 [0.817, 1.200] | 0.772** [0.595, 0.949] | 0.647*** [0.448, 0.845] |
| Intervention Prevalence Ratio for Cloth Mask Villages | 0.942 [0.781, 1.103] | 1.058 [0.870, 1.247] | 0.713** [0.459, 0.967] | 0.838 [0.524, 1.153] | 1.084 [0.769, 1.399] |
| Avg. Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0076 | 0.0055 | 0.0095 | 0.0108 | 0.0104 |
| N Individuals | 287,349 | 146,306 | 35,839 | 24,086 | 27,943 |
| N Villages | 538 | 480 | 384 | 348 | 360 |
| *Imputing Symptomatic-Seroprevalence for Missing Blood Draws* | | | | | |
| Intervention Prevalence Ratio for Surgical Mask Villages | 0.873*** [0.801, 0.945] | 0.917* [0.829, 1.005] | 0.975 [0.862, 1.088] | 0.815*** [0.688, 0.942] | 0.701*** [0.577, 0.824] |
| Intervention Prevalence Ratio for Cloth Mask Villages | 0.890** [0.787, 0.993] | 0.861*** [0.758, 0.965] | 0.838** [0.678, 0.998] | 1.153 [0.970, 1.336] | 0.792** [0.601, 0.983] |
| Avg. Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0189 | 0.0152 | 0.0226 | 0.0229 | 0.0251 |
| N Individuals | 321,383 | 177,708 | 51,676 | 37,340 | 43,431 |
| N Villages | 570 | 566 | 528 | 504 | 534 |

Confidence Intervals are in brackets.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regression include an indicator for each control-intervention pair. The 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis in the top panel utilizes the pre-registered sample, equivalent to Table 2; it includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.
The analysis in the bottom panel replicates the regressions in the top panel, but imputes the seropositivity of individuals for who we did not draw blood. For symptomatic individuals we did not draw blood from, we simulate their symptomatic-seroprevalence status by using the average rate of conditional seropositivity among all symptomatic individuals. This analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

# 3 Discussion

We present results from a cluster-randomized controlled trial of a scalable intervention designed to increase mask-wearing and reduce COVID-19. Our estimates suggest that mask-wearing increased by 28.8 percentage points, corresponding to an estimated 51,357 additional adults wearing masks in intervention villages, and this effect was persistent even after active mask promotion was discontinued. The intervention led to a 9.5% reduction in symptomatic SARS-CoV-2 seroprevalence (which corresponds to a 105 fewer symptomatic seropositives) and an 11.6% reduction in the prevalence of COVID-like symptoms, corresponding to 1,541 fewer people reporting these symptoms. If we assume that non-consenting symptomatic individuals were seropositive at the same rate as consenting symptomatic individuals, the total estimated symptomatic seropositives prevented would be 354. The effects were substantially larger (and more precisely estimated) in communities where we distributed surgical masks, consistent with their greater filtration efficiency measured in the laboratory (manuscript forthcoming). In villages randomized to receive surgical masks, the relative reduction in symptomatic seroprevalence was 11% overall, 23% among individuals aged 50-60, and 35% among those over 60 in preferred specifications.

We found clear evidence that surgical masks are effective in reducing symptomatic seroprevalence of SARS-CoV-2. While cloth masks clearly reduce symptoms, we find less clear evidence of their impact on symptomatic SARS-CoV-2 infections, with the statistical significance depending on whether we impute missing values for non-consenting adults. The number of cloth mask villages (100) was half that for surgical masks (200), meaning that our results tend to be less precise. Additionally, we found evidence that surgical masks were no less likely to be adopted than cloth masks. Surgical masks have higher filtration efficiency, are cheaper, are consistently worn, and are better supported by our evidence as tools to reduce COVID-19.

Our results should *not* be taken to imply that mask-wearing can prevent only 10% of COVID-19 cases, let alone 10% of COVID-19 mortality. Our intervention induced 29 more people out of every 100 to wear masks, with 42% of people wearing masks in total. The total impact with near-universal masking–perhaps achievable with alternative strategies or stricter enforcement–may

be several times larger than our 10% estimate. Additionally, the intervention reduced symptomatic seroprevalence more when surgical masks were used, and even more for the highest-risk individuals in our sample (23% for ages 50-60 and 35% for ages 60+). These numbers likely give a better sense of the impact of our intervention on severe morbidity and mortality, since most of the disease burden of the COVID-19 pandemic is borne by the elderly. Where achievable, universal mask adoption is likely to have still larger impacts.

There are several possible theories for why we might observe a larger reduction in COVID-19 cases for older adults. We did not directly measure age during surveillance, but mask-wearing could have increased more for older adults.A second theory is that older adults are more susceptible to infections at viral loads preventable by masks. A third theory is that older adults have fewer social connections, so that reducing transmission through any one connection is more likely to prevent infection by severing all transmissible routes. A fourth theory is that people exercised more care and were more likely to wear masks when proximate to the elderly.

We identified a combination of core intervention elements that were effective in increasing mask-wearing in rural Bangladesh: mask distribution and role-modeling, combined with mask promotion, leads to large and sustained increases in mask use. Results from our pilots suggest that combining mask distribution, role-modeling, and active mask promotion – rather than mask distribution and role-modeling alone – seems critical to achieving the full effect. Our trial 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. The null results on our cross-randomizations do not necessarily imply that these approaches are not worth trying in other contexts, but they teach us that large, persistent increases in mask-wearing are possible without these elements.

Prediction studies we conducted with policymakers and public health experts at the World Health Organization and the World Bank prior to presentations of the study results suggests that our results are informative for policy design. The majority of respondents in the prediction studies anticipated that text messages, verbal commitments, and incentives would increase mask-wearing,

when in reality, we estimated fairly precise null effects, and poll respondents believed that in-person mask promotion would have no additional effect, whereas the evidence from our pilots suggests it is essential (for additional detail see Appendix R).

Our intervention design is immediately relevant for Bangladesh's plans for larger-scale distribution of masks across all rural areas. The Bangladesh Directorate-General of health has assigned the study team and the NGO *BRAC* the responsibility to scale up the strategies that were proven most effective in this trial to reach 81 million people (*35*). At the time of writing, we are implementing this program in the 37 districts prioritized by the government based on SARS-CoV-2 test positivity rates. Our results are also relevant for mask dissemination and promotion campaigns planned in other countries and settings which face similar challenges in ensuring mask usage as a result of limited reach and enforcement capacity. The mask promotion model described in this paper was subsequently adopted by governments and other implementers in Pakistan (*36*), India (*37*), and Nepal (*38*). The intervention package would be feasible to implement in a similar fashion in other world regions as well. Beyond face masks, the conceptual underpinning of our strategies could be applied to encourage the adoption of other health behaviors and technologies, in particular those easily observable by others outside the household, such as purchase and consumption of food, alcohol, and tobacco products in stores, restaurants, or other public spaces (*39*), hand washing and infection control in healthcare facilities (*40–42*), hygiene interventions in childcare and school settings (*43, 44*), improved sanitation (*45, 46*), or vaccination drives (*47*).

While critics of mask mandates suggest that individuals who wear masks are more likely to engage in high-risk behaviors (*48*), we found no evidence of risk compensation as a result of increased mask-wearing. In fact, we found that our intervention slightly increased the likelihood of physical distancing, presumably because individuals participating in the intervention took the threat of COVID-19 more seriously. These findings are consistent with other behaviors including seat belt use (*49*) or immunization (*50*) where risk compensation–even if present–is not sufficient to outweigh direct effects.

The intervention may have influenced rates of COVID-19 by increasing mask use and/or phys-

ical distancing and/or other risk prevention behaviors. Three factors suggest that the direct impact of masks is the most likely explanation for our documented health impacts. First, in Appendix O, we analyze cross-sectionally the relationship between our biological outcomes and both mask-wearing and physical distancing. We find that symptoms and symptomatic seropositivity are negatively correlated with mask-wearing but not with physical distancing after controlling for mask-wearing. This analysis uses variation in observational rather than solely experimental data, and should therefore be interpreted with caution, as discussed in the appendix. Second, we see no change in physical distancing in the highest risk environment in our study, typically crowded indoor mosques. However, women do not typically go to mosques in rural Bangladesh and their symptomatic seropositivity decreased by just as much as men, so outdoor transmission or transmission in settings we do not observe directly may be important. Third, our study complements a large body of laboratory and quasi-experimental evidence that masks have a direct effect on SARS-CoV-2 transmission (*1*).

We estimate that a scaled version of our intervention being implemented in Bangladesh will cost between $10K and $52K per life saved, depending on what fraction of excess deaths are attributable to COVID-19. This is considerably lower than the value of a statistical life in Bangladesh ($205,000, (*51*)) and under severe outbreaks, is comparable to the most cost-efficient humanitarian programs at scale (e.g. distributing insecticide nets to prevent malaria costs $9,200 per life saved (*52*)). This estimate includes only mortality impacts but not morbidity, and greater cost-efficiency is possible if our intervention can be streamlined to further isolate the essential components. The vast majority of our costs were the personnel costs for mask-promoters: if we consider only the costs of mask production, these numbers would be 20x lower. Thus, the overall cost to save a life in countries where mask-mandates can be enforced at minimal cost with existing infrastructure may be substantially lower than our estimates above.

**Study Limitations** Our study has several limitations. The distinct appearance of project-associated masks and elevated mask-wearing in intervention villages made it impossible to blind surveillance

staff to study arm assignment. However, staff were not informed about the exact purpose of the study. Even though surveillance staff were plain-clothed and were instructed to remain discreet, community members could have recognized that they were being observed and changed their behavior. Additionally, survey respondents could have changed their likelihood of reporting symptoms in places where mask-wearing was more widespread. If respondents were more cognizant of symptoms in mask-wearing areas, this may bias us towards underestimating the impact of masks; if respondents in mask-wearing areas were less concerned with mild symptoms and thus were less likely to recall them, this might bias us towards overestimating the impact of masks. While we confirm that blood consent rates are not significantly different in the treatment and control group and are comparable across all demographic groups, we cannot rule out that the composition of consenters differed between the treatment and control groups. The slightly higher point estimate for consent in the treatment group biases us away from finding an effect, since it raises symptomatic seroprevalence in the treatment group. 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. While we did not directly assess harms in this study, there could be costs resulting from discomfort with increased mask-wearing, adverse health effects such as dermatitis or headaches, or impaired communication.

Because the study was powered to detect differences in symptomatic seroprevalence, we cannot distinguish whether masks work by making symptoms less severe (through a reduced viral load at transmission) or by reducing new infections. We selected the WHO case definition of COVID-19 for its sensitivity, though its limited specificity may imply that the impact of masks on symptoms comes partly from non-SARS-CoV-2 respiratory infections. If masks reduce COVID-19 by reducing symptoms (for a given number of infections), they could help ease the morbidity and mortality resulting from a given number of SARS-CoV-2 infections. If masks reduce infections, they may reduce the total number of infections over the long-term by buying more time to increase the fraction of the population vaccinated. At the time of the study, the predominant circulating SARS-CoV-2 strain was B.1.1.7 (alpha) (*53*). The impacts of the delta variant on the number of

infections prevented by a given mask-wearer are uncertain; the population-wide consequences of infections prevented by a given mask-wearer may be larger given a higher reproduction number.

In summary, we found that mask distribution, role modeling, and promotion in a LMIC setting increased mask-wearing and physical distancing, leading to lower illness, particularly in older adults. We find especially robust evidence that surgical masks prevent COVID-19. Whether people with respiratory symptoms should generally wear masks to prevent respiratory virus transmission—including for viruses other than SARS-CoV-2—is an important area for future research. Our findings suggest that such behavior may benefit public health.

# 4 Methods and Materials

## 4.1 Sampling frame and timeline

We discuss our sample size calculations in Appendix B and discuss the selection and pairwise randomization in Appendix C. In brief, we stratified villages based on geographic location and available case data, and then selected one treatment and control village from each pair.

Village-level cluster randomization was important for three 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 (*45*). Third, this design allows us to assess the full impact of masks on symptomatic infections, including via source control. Individual-level randomization would identify only whether masks protect wearers.

Our intervention was designed to last 8 weeks in each village. The intervention started in different villages at different times, rolling out over a 6-week period in 7 waves. There were between 16 and 61 village-pairs grouped in each wave based on geographic proximity and 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.

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.

The intervention protocol, pre-specified analysis plan, and CONSORT checklist are available at https://osf.io/vzdh6/.

## 4.2 Outcomes

Our primary outcome was symptomatic seroprevalence of SARS-CoV-2. Our secondary outcomes were prevalence of proper mask-wearing, physical distancing, and symptoms consistent with COVID-19. For COVID-19 symptoms, we used the symptoms that correspond to the WHO case definition of probable COVID-19 given epidemiological risk factors: (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); or (c) loss of taste or smell. Seropositivity was defined by having detectable IgG antibodies against SARS-CoV-2.

## 4.3 Intervention Materials and Activities

Our entire intervention was designed to be easily adopted by other NGOs or government agencies and required minimal monitoring. We have made the materials public in multiple languages to ease widespread adoption and replication by other implementers (http://tinyurl.com/maskprotocol).

We provide design specifications for our masks in Appendix F. We used high-quality surgical masks that had had a filtration efficiency of 95% (standard deviation [SD] = 1%); this is substantially higher than the filtration efficiency of the cloth masks we designed, which had a filtration efficiency of 37% (SD = 6%). These cloth masks had substantially higher filtration than common commercial 3-ply cotton masks, but lower than hybrid masks that use materials not commonly available for community members in low-resource settings ( (*54*)). While cloth masks have less leakage because they fit the face more closely ( (*55*)) and can be sewn without specialized equip-

ment, they are an order of magnitude more expensive than surgical masks. The filtration efficiency of the high-quality surgical masks used in this study was 76% after washing them with bar soap and water 10 times (manuscript forthcoming). While surgical masks can break down into microplastics that can enter the environment if disposed of improperly, analysis of waste generated in Bangladesh's first lockdown finds that the mass of surgical mask waste was one-third that of polyethylene bags, which also break down into macro- and micro-plastics (*56–58*).

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 (*59*). 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 ($0.06 of which was the cost of a sticker reminding people they could wash and reuse the surgical mask).

Adult household members were asked to wear masks whenever they were outside their house and around other people. 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-randomized additional intervention elements in randomly chosen subsets of treatment villages to investigate whether those have any additional impact on mask-wearing. The basic intervention package consists of five main elements:

1. One-time mask distribution and information provision (about masks) at households

2. Mask distribution in markets on 3-6 days per week during all eight weeks of the intervention.

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.

Participants and mask surveillance staff were not told which villages were in which intervention arm, but the intervention materials were clearly visible. The pre-specified analyses and sample exclusions were made by analysts blinded to the treatment assignment.

## 4.4 Cross-randomization of behavior change communication and incentives

**Village-level Cross-randomizations** Within the intervention arm, we cross-randomized villages to four village-level and four household-level treatments to test the impact of a range of social and behavior change communication strategies on mask-wearing. 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 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.

3. Randomization of treated villages to no incentive, non-monetary incentive, or monetary incentive of 190 USD given to the village leader for a project benefitting the public. 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.

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 single village, only one of these household randomizations was operative. As our data collection protocols relied on passive observation at the village-level, we could not record the mask-wearing behavior of individual households. To infer the effect of the household-level treatments we therefore varied the color of the masks distributed to the household based on its cross-randomization status and had surveillance staff record the mask color of observed individuals. In surgical mask villages, a household received blue or green and promoters distributed and equal number of blue and green masks in public settings. In cloth mask villages, households received violet or red masks and promoters distributed blue masks in public settings. To avoid conflating the effect of the household-specific treatment with the effect of the mask color, we randomized which color corresponded to which treatment status across villages (this way a specific color was not fully coincident with a specific treatment). The household-level randomizations, described in further detail in Appendix D and visualized in S2, were:

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

2. 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.

3. 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 households receiving texts, and in the 50% villages, the specific households that received the texts was also random.

**Conceptual Basis for Tested Social and Behavior Change Communication** We selected intervention elements that had a reasonable chance of persuading rural Bangladeshis to wear masks by consulting 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 key deterrents to the adoption of welfare-improving products, and especially of technologies that produce positive health externalities, such as face-masks (*21, 60*). 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 (*25, 61*).

The additional village- and household-level treatment we experimented with were also motivated 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 (*62–65*). We experimented with incentives because it is unclear whether extrinsic rewards crowd out intrinsic motivation (*66–68*). We test whether soft commitment devices encourage targets to follow through with actual behaviour change (*69, 70*), whether public displays can promote social norms (*27*), whether an altruistic framing inspires people more or less than self-interest (*71*), whether social image concerns and signaling can lead to higher compliance (*22, 72*), and whether regular reminders are a useful tool to ensure adoption (*23*).

## 4.5 Piloting Interventions

IPA implemented two pilots: Pilot 1 from July 22-31 and Pilot 2 from August 13-26, 2020. The objective of the pilots was to mimic some of the major aspects of the main experiment to identify implementation challenges. Each pilot was conducted in 10 unions that were not part of the main study area. We used the difference between the pilots to better understand which elements of our full intervention were essential. 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.

## 4.6 Surveillance Strategies

Mask-wearing and physical distancing were measured through direct observation. Surveillance was conducted using a standard protocol that instructed staff to spend one hour at each of the following high-traffic locations in the village: market, restaurant entrances, main road, tea stalls, and mosque, changing the location and timing to record the mask-wearing and physical distancing practices of as many individuals as possible. While SARS-CoV-2 transmission is more likely in indoor locations with limited ventilation than outside, rural Bangladeshi villages have few non-residential spaces where people gather, so observations were conducted outside except at the mosque, where surveillance was conducted inside.

Surveillance staff were distinct from intervention implementation staff and conducted surveillance in paired intervention and control villages. 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 while passively observing mask-wearing and physical distancing practices in the communities. They recorded the mask-wearing behavior of all of the adults they were able to observe during surveillance periods; observations were not limited to adults from enrolled households. Surveillance staff noted whether adults were wearing any mask or face covering, whether the mask was one distributed by our project (and if so, the color), and how the mask was worn. We defined proper mask-wearing as wearing either a project mask or an alternative face-covering over the mouth and nose and improper mask-wearing as wearing a mask in any way that did not fully cover the mouth and nose. Surveillance staff observed a single individual and recorded that person as practicing physical distancing if s/he was at least one arm's length away from all other people. Additional details are in Appendix G

## 4.7 Symptomatic SARS-CoV-2 Testing

**Symptom reporting** The owner of the household's primary phone completed surveys by phone or in-person at weeks 5 and 9 after the start of the intervention. They were asked to report symptoms experienced by any household member that occurred in the previous week and over the pre-

vious month. COVID-like symptoms were defined by whether they were consistent with the WHO COVID-19 case definition for suspected or probable cases with an epidemiological link (*73*).

**Blood sample collection** We collected endline capillary blood samples from participants who reported COVID-like symptoms during the study period and consented to blood collection. We additionally collected samples on a subset of randomly-selected participants at baseline, independent of symptoms, to assess overall seropositivity. For the purposes of blood collection, endline was defined as 10-12 weeks from the start of the intervention. 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). 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). 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. Additional details are presented in Appendix H.

## 4.8 Symptomatic Seropositivity

Our primary outcome is symptomatic seropositivity. As noted above, individuals are symptomatic if they meet the WHO surveillance definition of probable COVID-19 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 tested all individuals who were symptomatic in either our 5-week or 9-week household survey.

Our goal is to estimate the impact of the intervention on symptomatic 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 influenza-like illness and COVID-19 based on reported

symptoms, as well as indicators for each pair of villages from our pairwise stratification method.

In our pre-registered specification, we estimate this parameter by ordinary least squares, clustering at the village-level using the approach in (*74–76*). The dependent variable is $Y_{ij}$, the independent variable of interest is $T_j$, and controls are included for the $x_j$ covariates, including baseline mask-use and baseline respiratory symptom rates in each village. We also report results from a generalized linear model with a Poisson family and log-link function to compute relative risk (*77*). More details of our statistical analyses are reported in Appendix K.

# 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. C. R. MacIntyre, A. A. Chughtai, *BMJ* **350**, h694 (2015).

4. H. Bundgaard, *et al.*, *Annals of Internal Medicine* **174**, 335 (2021).

5. C. N. Ngonghala, *et al.*, *Mathematical Biosciences* **325**, 108364 (2020).

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

7. W. Lyu, G. L. Wehby, *Health Affairs* **39**, 1419 (2020).

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

9. J. Abaluck, *et al.*, *SSRN* (2020).

10. Y. Cheng, *et al.*, *Science* (2021).

11. A. Mullard, *Nature* (2020).

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

13. L. M. Brosseau, M. Sietsema, *CIDRAP* (2020).

14. M. A. Johansson, *et al.*, *JAMA Network Open* **4**, e2035057 (2021).

15. Science brief: Community use of cloth masks to control the spread of sars-cov-2 (2021).

16. J. M. Brophy, Covid-19: Controversial trial may actually show that masks protect the wearer; (2020).

17. J. Pan, C. Harb, W. Leng, L. C. Marr, *Aerosol Science and Technology* **55**, 718–733 (2021).

18. D. Kahneman, D. T. Miller, *Psychological Review* **93**, 136 (1986).

19. J. Jordan, E. Yoeli, D. Rand, *PsyArXiv* (2020).

20. R. B. Cialdini, N. J. Goldstein, *Annual Review of Psychology* **55**, 591 (2004).

21. M. Bates, R. Glennerster, K. Gumede, E. Duflo, *Field Actions Science Report* **4**, 30 (2012).

22. A. Karing, *University of California, Berkeley* (2018).

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

24. N. J. Goldstein, R. B. Cialdini, V. Griskevicius, *Journal of Consumer Research* **35**, 472 (2008).

25. G. Miller, A. M. Mobarak, *Marketing Science* **34** (2014).

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

27. C. Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (Oxford University Press, 2016).

28. T. R. Bhuiyan, *et al.*, *medRxiv* (2021).

29. icddr,b, Higher covid-19 seropositivity observed among residents in Dhaka and Chattogram (2021). Accessed on 16 Aug 2021.

30. Coronavirus COVID-19 Dashboard, 2020 (2021). Accessed on 16 Aug 2021.

31. M. V. Murhekar, *et al.*, *International Journal of Infectious Diseases* **108**, 145–155 (2021).

32. A. Anand, J. Sandefur, A. Subramanian, *Center for Global Development* (2021). Working Paper No. 589.

33. J. Abaluck, A. M. Mobarak, Getting all Bangladeshis to wear masks (2020). Publisher: White-Board Magazine.

34. A. Jakubowski, *et al.*, *medRxiv* (2021).

35. K. K. Tithila, BRAC's efforts to mask up Bangladesh could be game-changer (2021). Publisher: Dhaka Tribune, Bangladesh.

36. S. Riaz, Punjab authorities kick off 'NORM' campaign to increase mask-wearing (2021). Publisher: Arab News, Pakistan.

37. S. Bhattacharjee, Covid-19 Crisis: India Draws Lessons from Bangladesh's Mask Study (2021). Publisher: The Business Standard, India.

38. Republica, Nepal Mask Campaign launches with the slogan 'Let's wear masks, let's save each other's lives' (2021). Publisher: Nagarik Network, Nepal.

39. G. J. Hollands, *et al.*, *Cochrane Database of Systematic Reviews* (2019).

40. S. Naikoba, A. Hayward, *The Journal of Hospital Infection* **47**, 173 (2001).

41. C. Houghton, *et al.*, *Cochrane Database of Systematic Reviews* (2020).

42. H. Seo, *et al.*, *The Journal of Hospital Infection* **102**, 394 (2019).

43. D. Biswas, *et al.*, *The American Journal of Tropical Medicine and Hygiene* **101**, 1446 (2019).

44. S. L. McGuinness, *et al.*, *Tropical Medicine & International Health* **23**, 816 (2018).

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

46. S. R. Patil, *et al.*, *PLoS Medicine* **11**, e1001709 (2014).

47. J. S. Solís Arce, S. S. Warren, N. F. Meriggi, *et al.*, *Nature Medicine* **27**, 1385 (2021).

48. Y. Yan, J. Bayham, A. Richter, E. P. Fenichel, *Scientific reports* **11**, 1 (2021).

49. A. Cohen, L. Einav, *American economic review* **97**, 745 (2007).

50. M. L. Kasting, G. K. Shapiro, Z. Rosberger, J. A. Kahn, G. D. Zimet, *Human vaccines & immunotherapeutics* **12**, 1435 (2016).

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

52. GiveWell, 2021 GiveWell cost-effectiveness analysis-version 1. Accessed on 4 Jan 2021.

53. J. Hadfield, *et al.*, *Bioinformatics* **34**, 4121 (2018).

54. L. H. Kwong, *et al.*, *ACS Nano* **15**, 5904 (2021).

55. S. Duncan, P. Bodurtha, S. Naqvi, *PloS one* **16**, e0258191 (2021).

56. O. O. Fadare, E. D. Okoffo, *The Science of the total environment* **737**, 140279 (2020).

57. Environment and Social Development Organization, COVID-19 Pandemic Pushes Single Use Plastic Waste Outbreak: No Management, No Protection: High Health and Environmental Risk Unveil (2020).

58. I. M. Steensgaard, *et al.*, *Environmental Pollution* **224**, 289 (2017).

59. E.-S. Jang, C.-W. Kang, *Infection & Chemotherapy* **52**, 583 (2020).

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

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

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

63. H. Allcott, *Journal of Public Economics* **95**, 1082 (2011).

64. R. Guiteras, J. Levinsohn, A. M. Mobarak, *Centre for Economic Policy Research* (2019). Discussion Paper No. DP13498.

65. L. Beaman, A. BenYishay, J. Magruder, A. M. Mobarak, *American Economic Review* **111**, 1918 (2021).

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

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

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

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

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

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

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

73. World Health Organization, WHO COVID-19 Case Definition. Accessed on 15 Oct 2020.

74. S. Correia, *Unpublished manuscript, http://scorreia. com/research/hdfe. pdf (last accessed 25 October 2019)* (2017).

75. S. Gaure, Ols with multiple high dimensional category dummies, *Tech. rep.*, Memorandum (2010).

76. P. Guimaraes, P. Portugal, *The Stata Journal* **10**, 628 (2010).

77. G. Zou, *American journal of epidemiology* **159**, 702 (2004).

78. C. Rutterford, A. Copas, S. Eldridge, *International Journal of Epidemiology* **44**, 1051 (2015).

79. IHME, COVID-19 Projections-Bangladesh (2021). Accessed on 17 Aug 2021.

80. WHO Bangladesh, *Morbidity and Mortality Weekly Update (COVID-19)* **76** (2021).

# 5 Acknowledgements

Thanks to 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 Dr. Michael Friedman and the CDC Bangladesh country office for assistance with antibody tests. Thanks to Asif Saleh, Executive Director of BRAC, Dr. Morseda Chowdhury, Tanjila Mazumder Drishti, Imran Ahmed Chowdhury, and hundreds of BRAC staff for help with implementation, especially in the scale-up phase of this project in Bangladesh. Similarly, Reema Nanavaty, Sahil Hebbar, and other staff of SEWA India, and Captain Usman and his team at the Lahore Commissioners office, the Nepal Rapid Action Taskforce (C19 RAT), Senator Carmen Sanguinetti, Mario Sanchez, and Florencia Lopez Boo (IADB) played crucial roles in replication and scale-ups in India, Pakistan, Nepal and Uruguay.

Thanks also to Judy Chevalier, Bhavani Prathap Kasina and Stacey Daves-Ohlin for their advice and assistance, Tom Schmidt for help in shipping, Peter Hull for econometrics consulting, 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, Adam Gsellman for help with visualizations, and many employees at IPA Bangladesh for assistance throughout this project.

Most importantly, our continued engagement in and scaling up of mask promotion would not be possible without the tireless efforts of the Mask-NORM team, including Neela Saldanha, Heidi McAnnally-Linz, Maha Rehman, Gautam Patel, Jose Pinilla, Janani Rajashekar, Preeti Adhikary, Sharon Barnhardt, Mehrab Ali, Urvashi Wattal, Islamul Haque, Ana Tamayo, Laura Burke, Jeffrey

Mosenkis, and many other staff of Yale University and of Innovations for Poverty Action.



## 5.1 Funding

This research was financially supported by a grant from GiveWell.org to Innovations for Poverty Action (grant GR-000000272).

J. Benjamin-Chung was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant K01AI141616).

## 5.2 Author Contributions

**Conceptualization:** JA, LHK, AS, SPL, AMM; **Methodology:** JA LHK, AS, JBC, PJW, SPL, AMM; **Software:** EC; **Validation:** EC; **Formal Analysis:** JA, EC; **Investigation:** LHK, AS, MH, HMR, AAJ, SGM, AR, FLB, TSH; **Resources:** EBJ, SB; **Data Curation:** SRahman, EC, NZ; **Writing:** JA, LHK, AS, EC, SPL, AMM; **Visualization:** EC, NZ; **Supervision:** JA, SPL, AMM; **Project Administration:** JA, LHK, AH, MAK, SRaihan, SRahman; **Funding Acquisition:** JA, AMM, LHK, AS, SPL.

## 5.3 Competing Interests

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

### 5.4 Data & Materials Availability

All data and code are provided in our online repository.

### 5.5 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. The Bangladesh Directorate General of Health Services under the Ministry of Health, Aspire to Innovate (a2i), an information and data-focused organization within the Bangladesh government, North-South University in Dhaka, and the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) partnered in the study design and discussions and reviewed protocols. We provide ethical justification for our decisions in our online ethics appendix.

## 6 Supplementary Materials

Figs. S1 to S7

Tables S1 to S37

Appendices A-R

References (78-80)

# Supplementary Materials for

## The Impact of Community Masking on COVID-19:

## A Cluster-Randomized Trial in Bangladesh

Jason Abaluck[1,*,†], Laura H Kwong[2,3,†], Ashley Styczynski[4,†]

Ashraful Haque[5], Md. Alamgir Kabir[5], Ellen Bates-Jefferys[6]

Emily Crawford[1], Jade Benjamin-Chung[7], Shabib Raihan[5]

Shadman Rahman[5], Salim Benhachmi[8], Neeti Zaman[5]

Peter J. Winch[9], Maqsud Hossain[10], Hasan Mahmud Reza[11],

Abdullah All Jaber[10], Shawkee Gulshan Momen[10],

Aura Rahman[10], Faika Laz Banti[10], Tahrima Saiha Huq[10],

Stephen P. Luby[2,4,‡], Ahmed Mushfiq Mobarak[1,12,*,‡]

**This PDF includes:**

Figs. S1 to S7

Tables S1 to S37

Appendices A-R

1. Yale School of Management, Yale University; New Haven, CT, USA. 2. Woods Institute for the Environment, Stanford University; Stanford, CA, USA. 3. Division of Environmental Health Sciences, University of California Berkeley; Berkeley, CA, USA. 4. Division of Infectious Diseases and Geographic Medicine, Stanford University; Stanford, CA, USA. 5. Innovations for Poverty Action Bangladesh; Dhaka, Bangladesh. 6. Innovations for Poverty Action; Evanston, IL, USA. 7. Department of Epidemiology and Population Health, School of Medicine, Stanford University. 8. Yale Research Initiative on Innovation and Scale, Yale University; New Haven, CT, USA. 9. Social and Behavioral Interventions Program, Johns Hopkins Bloomberg School of Public Health; Baltimore, MD, USA. 10. NGRI, North South University; Dhaka, Bangladesh. 11. Department of Pharmaceutical Sciences, North South University; Dhaka, Bangladesh. 12. Deakin University; Melbourne, Australia.

† denotes co-first author, ‡ denotes co-last author.

***Corresponding authors. Email: jason.abaluck@yale.edu and ahmed.mobarak@yale.edu**

# A List of Supplementary Materials

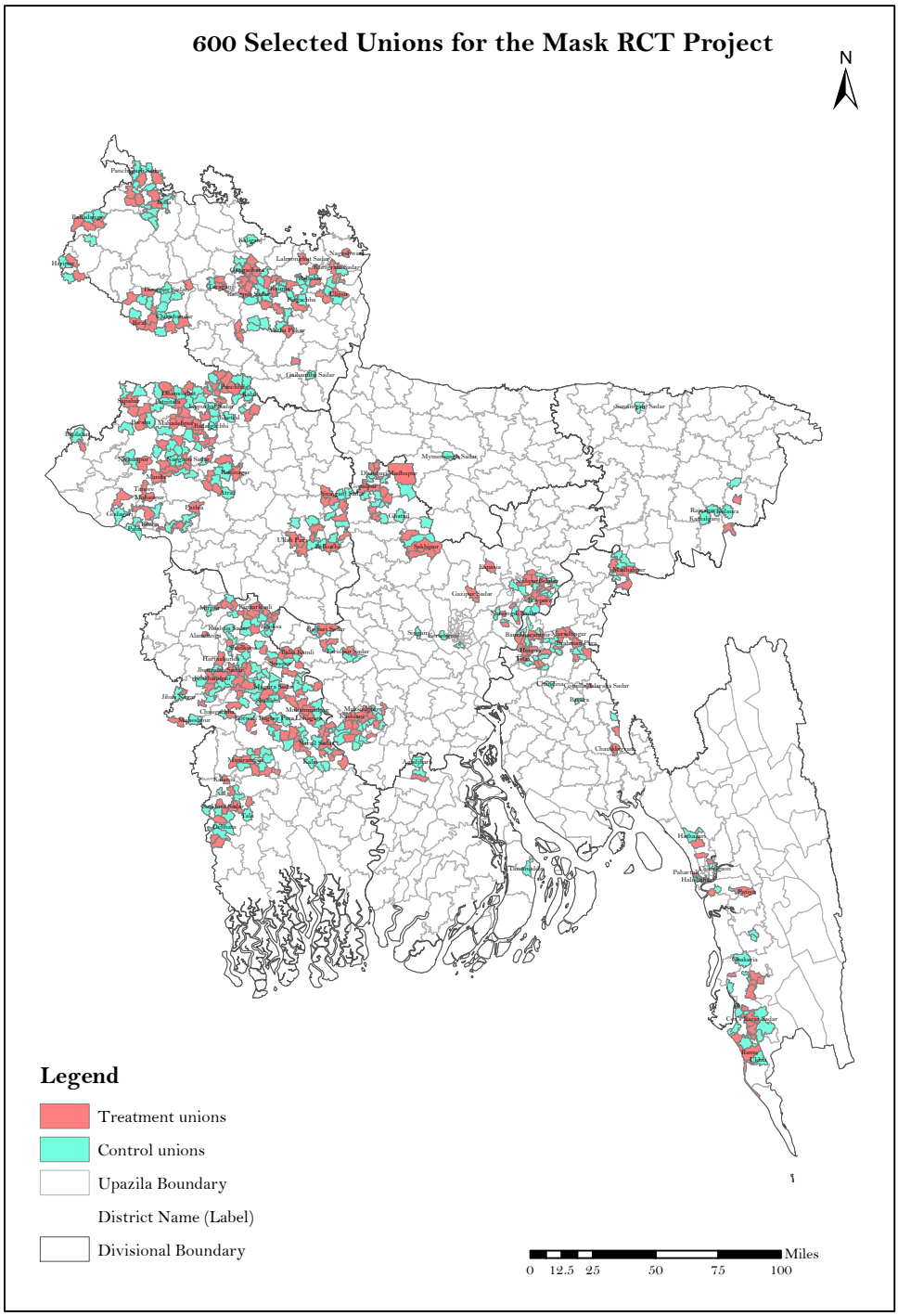
## Appendix Figures and Tables

## Appendices



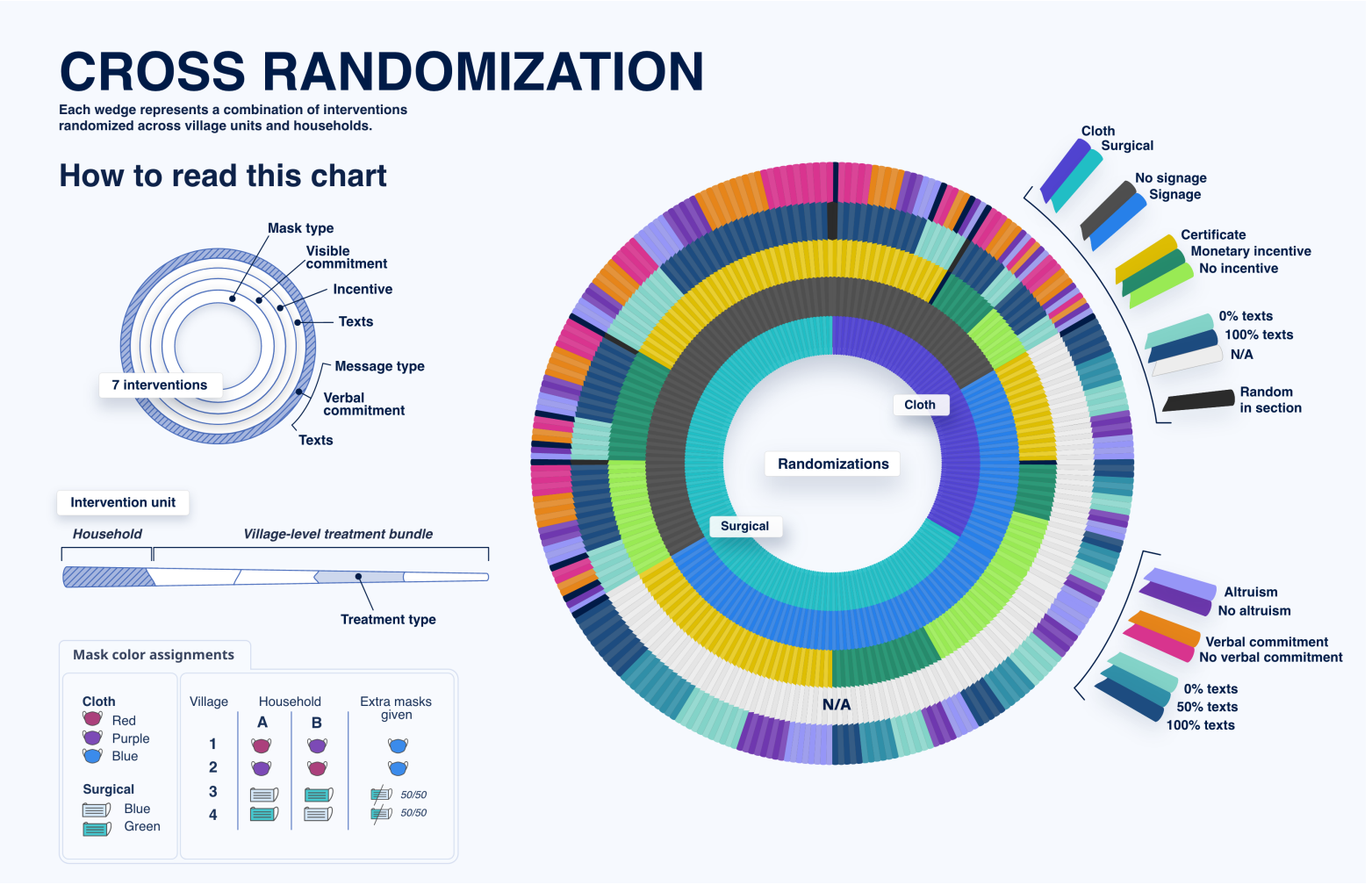
## Tables and Figures Referenced in Appendices

Figure S1: Map of 600 Treatment and Control Unions




The figure shows the location of the 600 treatment and control unions in the study.

Figure S2: Schematic of Cross-Randomizations




*Notes:* Each color represents a village-level or household-level randomization, with inner circles representing village-level randomizations and outer circles representing household level randomization. Different tones of the same hue represent the control or treatment status for each randomization. The fraction of villages receiving each randomization is proportional to the area of each concentric circle. The "Colors" box in the bottom left shows the color of masks used to denote households that were assigned to the control or treatment status of the household-level randomization.

Table S1: Enrollment and Consent (Individual-Level)

| | Treatment Villages | Control Villages | Total |
|---|---|---|---|
| Number of People Consented to Baseline Household Visits | 178,322 | 163,861 | 342,183 |
| Number of People Reached for Symptom Collection in the Midline and/or Endline Visits | 174,776 | 161,234 | 336,010 |
| Number of People with WHO-defined COVID-19 Symptoms | 13,307 | 13,853 | 27,160 |
| Number of Symptomatic Endline Blood Samples Collected | 5,345 | 5,445 | 10,790 |
| Number of Symptomatic Endline Blood Samples Tested | 4,714 | 4,798 | 9,512 |

All counts provided are at the individual level.
Table S2 shows household approached separately from households that consented to our baseline survey.

Table S2: Enrollment and Consent (Household-Level)

| | Treatment Villages | Control Villages | Total |
|---|---|---|---|
| Number of HHs Approached in Baseline Household Visit | 68,514 | 65,536 | 134,050 |
| Number of HHs that Consented to Baseline Household Visits | 64,851 | 60,202 | 125,053 |
| Number of HHs Reached for Symptom Collection in the Midline and/or Endline Visits | 63,489 | 59,163 | 122,652 |

All counts provided are at the household level.

Table S3: Endline Blood Collection Consent Rates by Demographic Characteristics

| | Treatment | Control | Total |
|---|---|---|---|
| Total | 40.2% | 39.3% | 39.7% |
| | | *By Sex* | |
| Female | 40.2% | 39.3% | 39.7% |
| Male | 40.2% | 39.3% | 39.7% |
| | | *By Age Group* | |
| $<$ 40 Y.O. | 40.4% | 38.7% | 39.5% |
| 40-49 Y.O. | 39.9% | 40.0% | 40.0% |
| 50-59 Y.O. | 41.0% | 40.2% | 40.6% |
| $\geq$ 60 Y.O. | 39.2% | 39.3% | 39.3% |

Consent rates are defined as the ratio of the number of individuals we successfully drew blood from to the number of eligible symptomatic individuals; those who met the WHO definition of a probable COVID-19 case.

Table S4: Balance Tests (Village-Level)

| | Baseline Symptomatic Seroprevalence Rate | Baseline WHO-Defined COVID-19 Symptoms | Baseline Mask-Wearing Rate |
|---|---|---|---|
| | | *Summary Statistics* | |
| Intervention Rate | 0.00022 | 0.02687 | 0.12297 |
| Control Rate | 0.00028 | 0.02540 | 0.12430 |
| | | *Balance Tests* | |
| Intervention Coefficient | -0.00005 | 0.00086 | 0.00069 |
| | (0.00007) | (0.00173) | (0.00549) |
| N villages | 572 | 572 | 572 |
| *F* | | 0.18 | |
| *Joint-Test Prob* $> F$ | | 0.8355 | |

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 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.

Table S5: Mask-Wearing and Physical Distancing, without Controlling for Baseline Variables

| | Full | No Active Promotion | Mosques | Markets | Other Locations | Surgical Mask Villages | Cloth Mask Villages |
|---|---|---|---|---|---|---|---|
| | | | *Proper Mask-Wearing* | | | | |
| Intervention Coefficient | 0.288*** (0.012) | 0.279*** (0.012) | 0.371*** (0.016) | 0.288*** (0.012) | 0.252*** (0.012) | 0.302*** (0.014) | 0.258*** (0.020) |
| Average Mask-Wearing Rate in Paired Control Villages§ | 0.133 | 0.134 | 0.123 | 0.120 | 0.146 | 0.129 | 0.143 |
| | | | *Physical Distancing* | | | | |
| Intervention Coefficient | 0.050*** (0.005) | 0.056*** (0.005) | 0.000 (0.000) | 0.073*** (0.007) | 0.067*** (0.007) | 0.053*** (0.006) | 0.044*** (0.011) |
| Average Distancing Rate in Paired Control Villages§ | 0.241 | 0.253 | 0.000 | 0.291 | 0.311 | 0.229 | 0.268 |
| N villages | 572 | 572 | 570 | 570 | 568 | 380 | 192 |

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 control-intervention pair.
§We report the mean rate of proper mask-wearing among the control villages during the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
'No Active Promotion" refers to any time that surveillance was conducted while promotion was not actively occurring (regardless of the week of the intervention). This excludes surveillance during the Friday Jumma Prayers in the mosque, when promoters were present and actively encouraged mask wearing.
"Other Locations" include tea stalls, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
"Surgical Villages" refer to all treatment villages which received surgical masks as part of the intervention, and their control pairs. "Cloth Villages" refer to all treatment villages which received cloth masks as part of the intervention, and their control pairs. These samples include surveillance from all available locations, equivalent to the to the column labeled "Full", but run separately for each subgroup.
Of the 572 villages included in the analyses sample, we exclude an additional village and its pair in the mosque and market sub-samples, and two villages and their pairs in the other location sub-sample because we did not observe them in the baseline period prior to the intervention. There are 190 treatment villages which received surgical masks as part of the intervention and 96 treatment villages which received cloth masks.

Table S6: Number of People Observed

| | Full | No Active Promotion | Mosques | Markets | Other Locations | Surgical Mask Villages | Cloth Mask Villages |
|---|---|---|---|---|---|---|---|
| | | | *No Baseline Control* | | | | |
| Intervention Coefficient | -31 | -53 | 35 | -20 | -46** | -9 | -75 |
| | (51) | (45) | (24) | (17) | (23) | (63) | (85) |
| Avg. Number People Observed in Paired Control Villages§ | 2820 | 2682 | 580 | 882 | 1358 | 2914 | 2635 |
| | | | *With Baseline Control* | | | | |
| Intervention Coefficient | -43 | -64 | 23 | -18 | -53** | -37 | -45 |
| | (45) | (40) | (20) | (15) | (21) | (58) | (76) |
| N villages | 572 | 572 | 570 | 570 | 568 | 380 | 192 |

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 control-intervention pair. The regressions "with baseline control" include controls for the number of people observed in the baseline visit.
§We report the average number of people observed among the control villages during the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
"No Active Promotion" refers to any time that surveillance was conducted while promotion was not actively occurring (regardless of the week of the intervention). This excludes surveillance during the Friday Jumma Prayers in the mosque, when promoters were present and actively encouraged mask wearing.
"Other Locations" include tea stalls, at the entrance of the restaurant as patrons enter, and the main road to enter the village.
"Surgical Villages" refer to all treatment villages which received surgical masks as part of the intervention, and their control pairs. "Cloth Villages" refer to all treatment villages which received cloth masks as part of the intervention, and their control pairs. These samples include surveillance from all available locations, equivalent to the to the column labeled "Full", but run separately for each subgroup.
Of the 572 villages included in the analyses sample, we exclude an additional village and its pair in the mosque and market sub-samples, and two villages and their pairs in the other location sub-sample because we did not observe them in the baseline period prior to the intervention. There are 190 treatment villages which received surgical masks as part of the intervention and 96 treatment villages which received cloth masks.

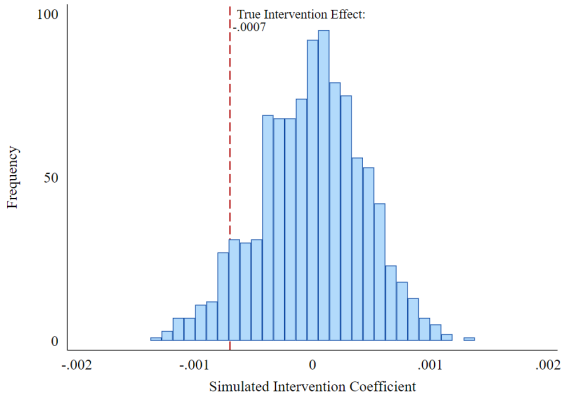
Table S7: Symptomatic Seroprevalence

| | Intervention Effect | Intervention Effect by Mask Type |
|---|---|---|
| *No Baseline Controls* | | |
| Intervention Coefficient | -0.0007** | |
| | (0.0003) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0008* |
| | | (0.0004) |
| Intervention Coefficient for Cloth Mask Villages | | -0.0005 |
| | | (0.0005) |
| Average Symptomatic Seroprevalence Rate in Paired Control Villages§ | 0.0076 | 0.0076 |
| *With Baseline Controls* | | |
| Intervention Coefficient | -0.0007** | |
| | (0.0003) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0009** |
| | | (0.0004) |
| Intervention Coefficient for Cloth Mask Villages | | -0.0003 |
| | | (0.0005) |
| N individuals | 304,726 | 304,726 |
| N villages | 572 | 572 |

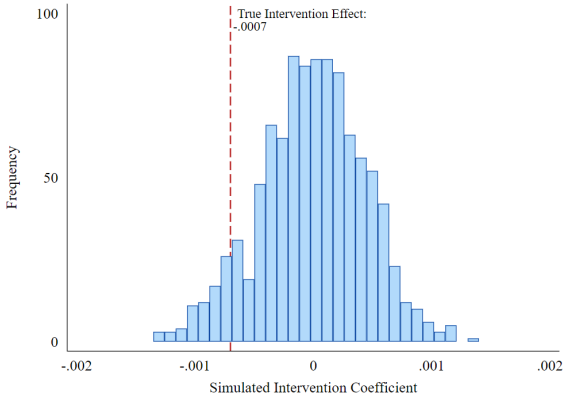
Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
All regressions include an indicator for each control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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. newline §We report the mean rate of symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.

Figure S3: Randomization Inference on Symptomatic Seroprevalence and Symptoms

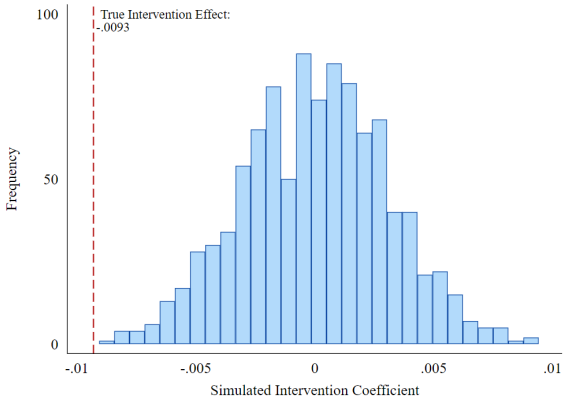
(a) Symptomatic Seroprevalence



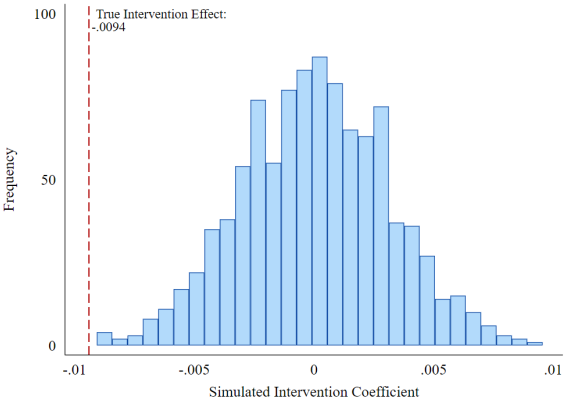

(b) Symptomatic Seroprevalence, Without Baseline Controls




(c) Symptoms




(d) Symptoms, Without Baseline Controls




The histograms are generated by plotting the frequency the coefficient on the intervention under 1,000 imputations of randomly assigning the treatment/control status within each village-pair. The regressions used to generate the intervention coefficient in panel (a) and (b) are equivalent to those in Table S7, top and bottom panel, respectively. The regressions used in panel (c) and (d) are equivalent to those in Table S8, top and bottom panel, respectively. The one-sided *p*-values for each panel is as follows:
(a) 0.070
(b) 0.074
(c) 0.000
(d) 0.000

Table S8: WHO-defined COVID-19 Symptoms

| | Intervention Effect | Intervention Effect by Mask Type |
|---|---|---|
| *No Baseline Controls* | | |
| Intervention Coefficient | -0.0094*** | |
| | (0.0022) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0103*** |
| | | (0.0028) |
| Intervention Coefficient for Cloth Mask Villages | | -0.0074** |
| | | (0.0034) |
| Average Symptomatic Rate in Paired Control Villages§ | 0.0860 | 0.0860 |
| *With Baseline Controls* | | |
| Intervention Coefficient | -0.0093*** | |
| | (0.0021) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0111*** |
| | | (0.0028) |
| Intervention Coefficient for Cloth Mask Villages | | -0.0058* |
| | | (0.0034) |
| N individuals | 321,948 | 321,948 |
| N villages | 572 | 572 |

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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

Table S9: WHO-defined COVID-19 Symptoms (Robustness Check)

| | Intervention Effect | Intervention Effect by Mask Type |
|---|---|---|
| | *No Baseline Controls* | |
| Intervention Coefficient | -0.0031*** | |
| | (0.0011) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0048*** |
| | | (0.0015) |
| Intervention Coefficient for Cloth Mask Villages | | 0.0002 |
| | | (0.0017) |
| Average Symptomatic Rate in Paired Control Villages§ | 0.0329 | 0.0329 |
| | *With Baseline Controls* | |
| Intervention Coefficient | -0.0030*** | |
| | (0.0011) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0051*** |
| | | (0.0015) |
| Intervention Coefficient for Cloth Mask Villages | | 0.0010 |
| | | (0.0017) |
| N individuals | 304,726 | 304,726 |
| N villages | 572 | 572 |

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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.

Table S10: Pilot Analyses of Mask Wearing

| | Main Intervention | Pilot 1 | Pilot 2 | Pilot 1 | Pilot 2 |
|---|---|---|---|---|---|
| | | *No Baseline Controls* | | | |
| Intervention Coefficient | 0.288*** (0.012) | 0.109 [-0.165, 0.306] | 0.284 [0.105, 0.410] | | |
| Difference from Main Intervention | | | | -0.181** (0.092) | -0.005 (0.058) |
| Average Control Mask Wearing Rate§ | 0.133 | 0.129 | 0.095 | | |
| | | *With Baseline Controls* | | | |
| Intervention Effect | 0.288*** (0.012) | 0.096 [-0.120, 0.302] | 0.341 [0.099, 0.489] | | |
| Difference from Main Intervention | | | | -0.189** (0.073) | 0.022 (0.053) |
| N villages | 572 | 10 | 10 | 592 | 592 |

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 regressions "with baseline controls" include controls for baseline rates of mask-wearing.
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 during the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.

Table S11: Subgroup Analyses of Mask-Wearing

| | Female Only | Male Only | Above Median | Below Median |
|---|---|---|---|---|
| | | *No Baseline Controls* | | |
| Intervention Coefficient | 0.225*** | 0.271*** | 0.247*** | 0.346*** |
| | (0.013) | (0.013) | (0.018) | (0.022) |
| Average Control Mask-Wearing Rate[§] | 0.313 | 0.116 | 0.175 | 0.087 |
| | | *With Baseline Controls* | | |
| Intervention Coefficient | 0.225*** | 0.271*** | 0.247*** | 0.350*** |
| | (0.013) | (0.013) | (0.019) | (0.022) |
| N villages | 566 | 566 | 200 | 200 |

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 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. 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 during the baseline observation. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The sex-specific subgroup is run on all locations except mosques because no females were observed at mosques. The sex-specific samples excludes 6 villages because of lack of data. The above-median and below-median samples includes 85 singleton observations which were dropped.

Table S12: Symptomatic Seroprevalence by 10-Year Age Groups

| | All | 18-29 Y.O. | 30-39 Y.O. | 40-49 Y.O. | 50-59 Y.O. | 60-69 Y.O. | $\geq$ 70 Y.O. |
|---|---|---|---|---|---|---|---|
| | | | *No Baseline Controls* | | | | |
| Intervention Coefficient | -0.0007** | -0.0004 | 0.0007 | -0.0008 | -0.0021*** | -0.0020** | -0.0018 |
| | (0.0003) | (0.0003) | (0.0005) | (0.0007) | (0.0008) | (0.0010) | (0.0012) |
| Avg. Symptomatic Seroprevalence in Paired Control Vill.§ | 0.0076 | 0.0066 | 0.0100 | 0.0136 | 0.0175 | 0.0203 | 0.0270 |
| | | | *With Baseline Controls* | | | | |
| Intervention Coefficient | -0.0007** | -0.0004 | 0.0007 | -0.0008 | -0.0021*** | -0.0019** | -0.0019 |
| | (0.0003) | (0.0003) | (0.0005) | (0.0007) | (0.0008) | (0.0010) | (0.0012) |
| N Individuals | 304,726 | 101,137 | 69,717 | 51,727 | 38,996 | 27,625 | 15,524 |
| N Villages | 572 | 572 | 572 | 572 | 572 | 572 | 572 |

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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.

Table S13: WHO-Defined COVID-19 Symptoms by Age Groups

| | All | < 40 Y.O. | 40-49 Y.O. | 50-69 Y.O. | ≥ 60 Y.O. |
|---|---|---|---|---|---|
| | | | *No Baseline Controls* | | |
| Intervention Coefficient for Surgical Mask Villages | -0.0103*** (0.0028) | -0.0085*** (0.0028) | -0.0091*** (0.0035) | -0.0122*** (0.0039) | -0.0172*** (0.0043) |
| Intervention Coefficient for Cloth Mask Villages | -0.0074** (0.0034) | -0.0057** (0.0027) | -0.0010 (0.0051) | -0.0161*** (0.0059) | -0.0112* (0.0058) |
| Average Symptomatic Rate in Paired Control Villages§ | 0.0860 | 0.0717 | 0.0983 | 0.1060 | 0.1080 |
| | | | *With Baseline Controls* | | |
| Intervention Coefficient for Surgical Mask Villages | -0.0111*** (0.0028) | -0.0093*** (0.0028) | -0.0097*** (0.0035) | -0.0127*** (0.0038) | -0.0180*** (0.0042) |
| Intervention Coefficient for Cloth Mask Villages | -0.0058* (0.0034) | -0.0045 (0.0027) | 0.0010 (0.0050) | -0.0139** (0.0058) | -0.0087 (0.0058) |
| N Individuals | 321,948 | 178,881 | 55,182 | 41,683 | 46,202 |
| N Villages | 572 | 572 | 572 | 572 | 572 |

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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

Table S14: WHO-Defined COVID-19 Symptoms by 10-Year Age Groups

| | All | 18-29 Y.O. | 30-39 Y.O. | 40-49 Y.O. | 50-59 Y.O. | 60-69 Y.O. | $\geq$ 70 Y.O. |
|---|---|---|---|---|---|---|---|
| | | | *No Baseline Controls* | | | | |
| Intervention Coefficient | -0.0094*** | -0.0079*** | -0.0068** | -0.0064** | -0.0135*** | -0.0122*** | -0.0192*** |
| | (0.0022) | (0.0020) | (0.0027) | (0.0029) | (0.0033) | (0.0037) | (0.0045) |
| Avg Symptomatic Rate in Paired Control Vill.§ | 0.0860 | 0.0607 | 0.0872 | 0.0983 | 0.1064 | 0.1083 | 0.1123 |
| | | | *With Baseline Controls* | | | | |
| Intervention Coefficient | -0.0093*** | -0.0080*** | -0.0067** | -0.0061** | -0.0131*** | -0.0119*** | -0.0188*** |
| | (0.0021) | (0.0020) | (0.0026) | (0.0029) | (0.0032) | (0.0036) | (0.0045) |
| N Individuals | 321,948 | 105,163 | 73,718 | 55,182 | 41,683 | 29,616 | 16,586 |
| N Villages | 572 | 572 | 572 | 572 | 572 | 572 | 572 |

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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

Table S15: WHO-Defined COVID-19 Symptoms by Age Groups, Expressed in Prevalence Ratios

| | All | < 40 Y.O. | 40-49 Y.O. | 50-59 Y.O. | ≥ 60 Y.O. |
|---|---|---|---|---|---|
| | | *No Baseline Controls* | | | |
| Intervention Coefficient for Surgical Mask Villages | 0.874*** [0.809, 0.939] | 0.876*** [0.800, 0.953] | 0.903*** [0.831, 0.975] | 0.879*** [0.805, 0.953] | 0.829*** [0.750, 0.908] |
| Intervention Coefficient for Cloth Mask Villages | 0.907** [0.823, 0.991] | 0.913** [0.835, 0.991] | 0.989 [0.886, 1.093] | 0.832*** [0.714, 0.950] | 0.888* [0.775, 1.002] |
| Average Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0860 | 0.0717 | 0.0983 | 0.1060 | 0.1080 |
| | | *With Baseline Controls* | | | |
| Intervention Coefficient for Surgical Mask Villages | 0.865*** [0.803, 0.928] | 0.866*** [0.792, 0.941] | 0.897*** [0.827, 0.967] | 0.870*** [0.797, 0.943] | 0.823*** [0.748, 0.899] |
| Intervention Coefficient for Cloth Mask Villages | 0.922* [0.838, 1.005] | 0.921* [0.842, 1.001] | 1.007 [0.905, 1.108] | 0.853** [0.737, 0.970] | 0.907 [0.795, 1.020] |
| N Individuals | 321,948 | 178,881 | 55,182 | 41,569 | 46,071 |
| N Villages | 572 | 572 | 572 | 570 | 570 |

Confidence Intervals are in brackets.
*** 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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

# B Sample Size

To determine the necessary sample size for a cluster randomized trial, we used equation 5 for binary outcomes from Rutterford, et al (*78*). We rearranged the equation to solve for delta, the clinically relevant difference in reduction of symptomatic seropositivity. This allowed us to determine the optimum power we could achieve within budget and logistical constraints. By enrolling 600 villages with an anticipated number of 250 households per village and two eligible persons per household, we estimated our per-arm sample size would be 150,000 adults. We assumed P1=P2 and conservatively estimated the proportion of seropositivity at endline to be 9% with 4% attributed to the study period. We estimated an intercluster correlation coefficient of 0.02. This gave us a delta of 8.29E-03. Dividing by P gave us a minimum detectable effect of 9.2%. To determine the number of blood tests needed, we estimated that 12% of people enrolled would develop COVID-like symptoms over the study period and that one-third these individuals would have a SARS-CoV-2 infection. This gave us a target of 36,000 blood tests.

# C Pairwise Randomization Procedure

To develop the sample frame, Innovations for Poverty Action (IPA) Bangladesh selected 1,000 rural and peri-urban unions out of 4,500 unions in Bangladesh. We excluded Dhaka district, because of high initial seroprevalence, and three hill districts, because of the logistical difficulties 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-districts (*upazilas*) (out of 485).

We used a pairwise randomization to select 300 intervention and 300 control unions within the same sub-districts. This randomization procedure was designed to pair unions that were similar in terms of (limited) COVID-19 case data, population size, and population density. Each union consists of roughly 80,000 people, or around 80 villages. Surveyors blind to treatment assignment followed a scoping protocol (Appendix E) to identify the union's largest market and co-located

village. Field staff sought consent for a baseline survey in every household in every selected village; in intervention villages every adult in consenting households was given a mask. Some unions are very small so to avoid spillover effects, so we selected only on village per union and we ensured that selected villages were at least 2 km apart. Treatment and control unions were scattered throughout the country (Figure S1).

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 4,562 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 sorted unions by "cases per person" based on data from July 11-24, 2020 and created pairs using adjacent unions in this sort order. 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 t-stat of the balance tests with respect to case trajectory and cases per person.

# D Cross-Randomization Procedure

Villages 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 was assigned to one and 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 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.

3. In the villages without the signage randomization and no household-level altruism randomization, we asked some households to make a verbal commitment to be a mask-wearing household while the remaining were not asked to make a commitment.

4. 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.

# E Scoping and Recruitment

All households in selected villages were eligible for participation in the study. At each household, field staff sought consent to participate in respiratory symptom surveys from the adult who answered the door. The scoping staff that mapped enrolled villages were blind to study arm assignment. However, the implementation staff that consented households was not blind to study arm assignment. 93.3% of households consented to participate in the study and completed a baseline symptom survey. Of the households that were surveyed in the baseline household visit, 83.2% of households provided a response to the week 5 symptom survey. 94.4% of households provided a response to the week 9 symptom survey. 98.1% of households provided a response to the week 5 or week 9 symptom survey. There were no statistically significant differences between response rates in the treatment and control groups.

Individuals who reported symptoms any time during the 8-week study period were sought out for collection of a blood sample; blood sample collection was conducted only after additional

informed, written consent was provided. 39.7% of symptomatic participants agreed to blood collection. Blood consent rates are not significantly different in the treatment and control group and are comparable across all demographic groups, we cannot rule out that the composition of consenters differed between the treatment and control groups. If we assume that consenters and non-consenters have similar seroprevalence rates, then we would expect true symptomatic seroprevalence to be perhaps 2.5 times than the rates we report.

Surveillance staff were instructed to record details about the mask-wearing behavior of every person they saw while stationed at public places throughout the community: in mosques, at (predominantly open-air) markets, at outdoor tea stalls, on the main road, and outside restaurants. In other words, they conducted a census of all individuals within their field of view during surveillance activities. Surveillance staff were provided an example schedule for surveillance that suggested visiting 9 locations over the course of the day, spending one hour at each location. This included 2 hour outside of a restaurant or at a tea stall, 2 hours on the main road near the entrance to the village and transportation stations, 3 hours at a mosque, and 1 hour at a market. However, staff were free to vary the timing and location of their surveillance activities to maximum surveillance in crowded locations or locations with relatively higher numbers of people.

This observed sample is representative of the rural Bangladeshi population that is present in crowded public places during the day; this population is largely men, who have more social contacts outside the home than women. This is reflected in our surveillance in at mosques, markets, tea stall, restaurants, and on the main road, in which men constituted 88.2% of all observed adults in these areas. (Men constituted 100% of all observed adults at mosques and 87-89% of all observed adults in each of the other locations.)

There was no difference in the number of people observed in public areas between treatment and control groups. The distinct appearance of project-associated masks and elevated mask-wearing in treatment villages made it impossible to blind surveillance staff to study arm assignment. However, study staff were not informed about the exact purpose of the study.

# F Details on Mask Materials and Design

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 masks were produced by Bangladeshi garment factories within 6 weeks after ordering.

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, 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. 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 an external layer 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%.

# G Details on Surveillance

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). The same staff member conducted surveillance at paired intervention and control villages at baseline and then once per week on weeks 1, 2, 4, 6, 8, and 10 after the intervention. The 10-week observation

was conducted two weeks after all intervention activities had ceased. We also collected longer-term data on mask-wearing behavior 20-27 weeks after the launch of interventions. 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 9 am to 7 pm. In 10 unions we conducted audits to assess the validity of surveillance data by pairing one monitoring officer with surveillance staff; in all cases the difference in their results was <10%, our pre-determined threshold.

Surveillance staff observed a single individual and recorded that person as practicing physical distancing if s/he was at least one arm's length away from all other people. This is consistent with the WHO guideline that defines physical distancing as one meter of separation https://www.who.int/westernpacific/emergencies/covid-19/information/physical-distancing. Accessed January, 30 2021. Note that compliance with WHO guidelines does not require physical distancing; for example, members of the same household need not remain physically distant (and presumably would not change their distancing behavior as a result of our intervention).

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.

# H Antibody Testing

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 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). 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. Individuals were coded as symptomatic seropositive if they reported symptoms consistent with the WHO COVID-19 case definition, their blood was collected, and the antibody test was positive.

# I Impact of Masks on Symptoms, Seroprevalence, and Seroconversions

Our primary outcome measures symptomatic seroprevalence: this is the fraction of individuals who are symptomatic during our intervention period and seropositive at endline. Some of these individuals may have antibodies from infections occurring prior to our intervention. If so, the impact of our intervention on symptomatic seroprevalence may understate the impact on symptomatic seroconversions occurring during our intervention (i.e. the fraction of symptomatic infections prevented by masks). In this section, we discuss the relationship between these two quantities.

Let $SC$, the symptomatic seroconversion rate, denote the probability that an individual is SARS-CoV-2 antibody-positive during our intervention and symptomatic. Then the symptomatic seroprevalence is $SS = SC + P_{prior}$, where $P_{prior}$ denotes the probability that an individual was infected prior to our intervention *and* is symptomatic during our intervention for some non-COVID reason.

The change in seroconversions between the treatment and control group is given by $\Delta SC = SC(1) - SC(0)$ where the notation $SC(T_i)$ denotes the potential outcome of seroconversions as a function of treatment status. Our goal is to estimate $\Delta SC/SC(0)$, the percentage change in sero-

conversions as a result of our intervention.

We observe $\Delta SS = \Delta SC + \Delta P_{prior}$. Additionally, we observe $SS(0) = SC(0) + P_{prior}(0)$. Suppose that masks prevent a fraction $\alpha$ of non-COVID symptoms. Then, $P_{prior}(1) = (1-\alpha)P_{prior}(0)$ and $\Delta P_{prior} = -\alpha P_{prior}(0)$. Then we have:

$$\frac{\Delta SS}{SS(0)} = \frac{\Delta SC - \alpha P_{prior}(0)}{SC(0) + P_{prior}(0)} \tag{1}$$

Rearranging (and substituting $SC(0) = SS(0) - P_{prior}(0)$), we obtain:

$$\frac{\Delta SC}{SC(0)} = \frac{\Delta SS}{SS(0)} + \frac{P_{prior}(0)(\alpha + \frac{\Delta SS}{SS(0)})}{SS(0) - P_{prior}(0)} \tag{2}$$

Note that if we assume that symptomatic seroconversions fall by exactly the same fraction as other symptomatic conditions, then we also have $SC(1) = (1-\alpha)SC(0)$, and solving equation 2 gives $\frac{\Delta SS}{SS(0)} = -\alpha = \frac{\Delta SC}{SC(0)}$. In other words, the percentage change in seroconversions equals the percentage change in seroprevalence provided either that $P_{prior} = 0$ *or* if the intervention works only by alleviating symptoms (and does so equally for COVID-19 and non-COVID diseases).

More generally, if the intervention both alleviates symptoms and reduces infections, then the relative impact on symptomatic seroconversions and symptomatic seroprevalence will depend on whether masks are more effective at preventing COVID-19 or other respiratory diseases (with a larger proportional reduction in symptomatic seroconversions in the former case). The magnitude of the difference between symptomatic seroconversions and symptomatic seropositives will depend on the fraction of symptomatic seropositives which are pre-existing at baseline.

# J Behavioral Mechanisms

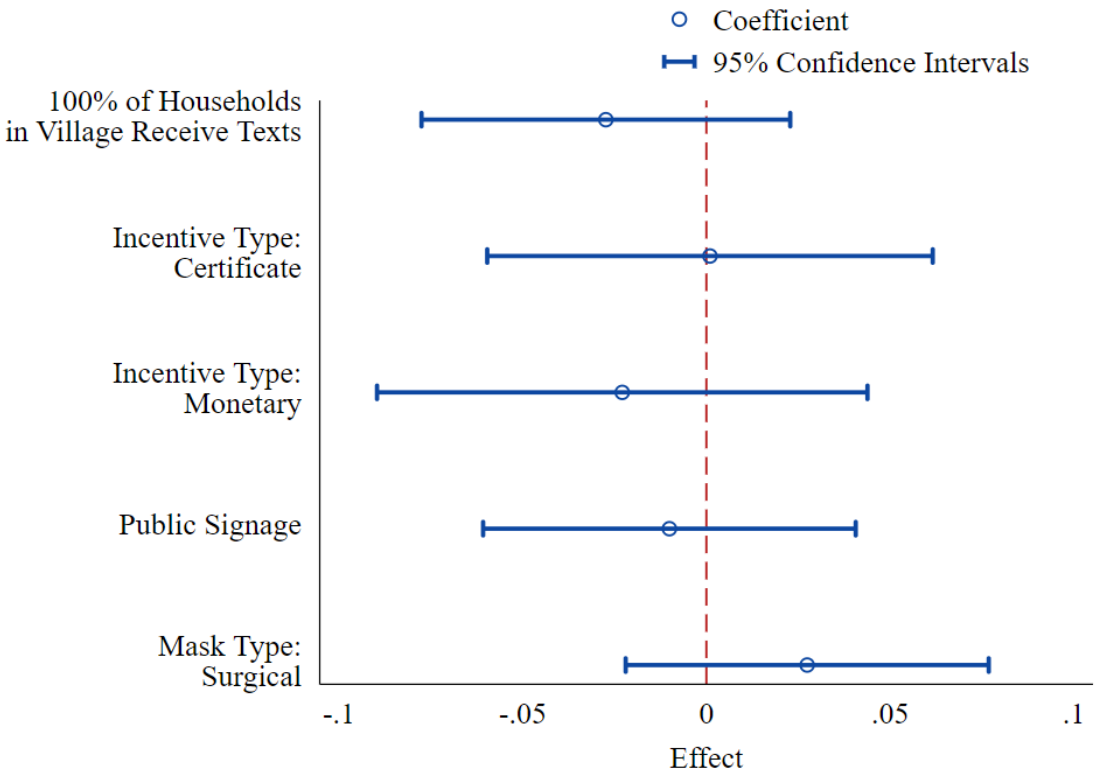
Our intervention combines multiple distinct elements: we provide people with free masks; we provide information about why mask-wearing is important; we conduct mask promotion in the form of monitors encouraging people to wear masks and stopping non-mask-wearing individuals

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.

## J.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 S4 and Table S16. *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 2.5-3.9 percentage points). Text message reminders, incentives for village-leaders, or explicit commitment signals explain little of the mask increase we document.

Figure S4: Village-Level Cross Randomizations




The figures corresponds to the regressions in S16, 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 4.4.

## J.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 S5 and Table S17. The coefficients indicate the impact of each cross-randomization relative to the core intervention (identified since some villages had no household randomization other than mask color). 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.

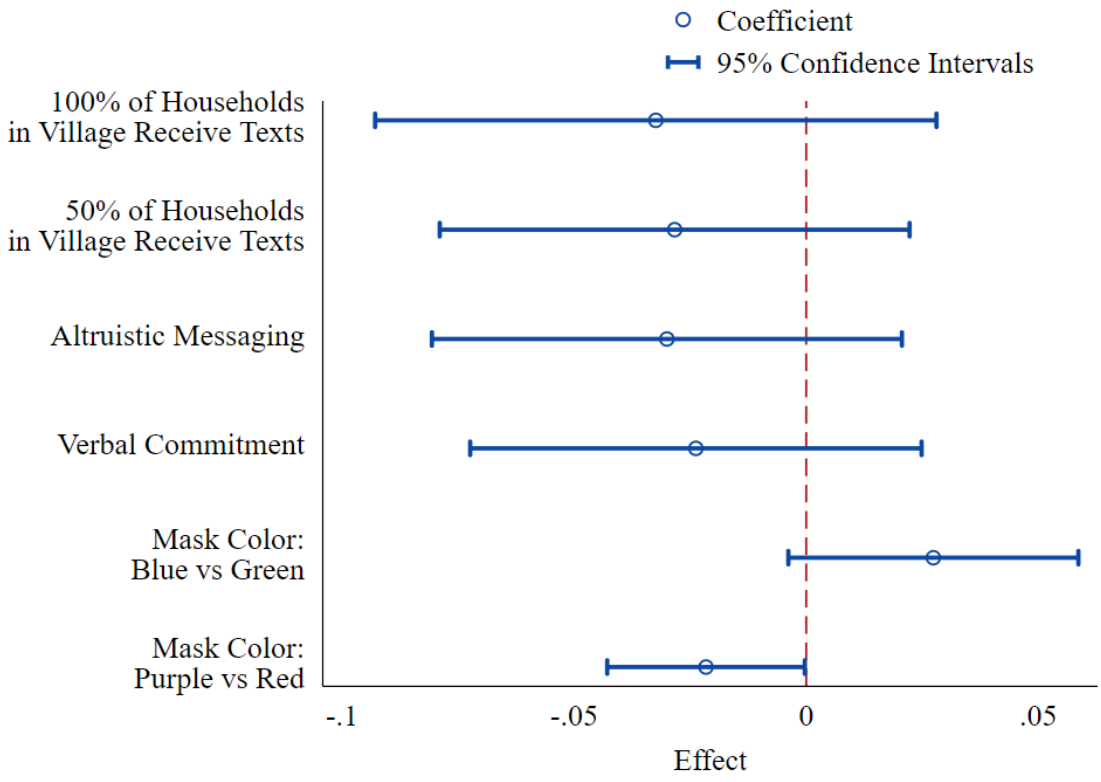
We did see an impact of mask color on mask adoption. In villages where surgical masks were distributed, blue surgical masks were 2.7 percentage points more likely than green surgical masks to be observed. In villages where cloth masks were distributed, purple masks were 2.2 percentage points less likely than red masks to be observed.

Table S16: Village-Level Cross Randomizations

| Coefficient | Full | No Active Promotion | Mosques | Markets | Other Locations |
|---|---|---|---|---|---|
| *No Baseline Controls* | | | | | |
| Mask Type (Surgical) | 0.027 | 0.027 | 0.062* | 0.017 | 0.018 |
| | (0.025) | (0.025) | (0.035) | (0.026) | (0.025) |
| Commitment w/ Signage | -0.010 | -0.007 | -0.019 | -0.008 | -0.009 |
| | (0.026) | (0.026) | (0.034) | (0.027) | (0.026) |
| Incentive Type | | | | | |
| Monetary | -0.023 | -0.026 | 0.013 | -0.034 | -0.026 |
| | (0.034) | (0.034) | (0.045) | (0.034) | (0.035) |
| Certificate | 0.001 | -0.002 | 0.021 | 0.003 | -0.009 |
| | (0.031) | (0.031) | (0.038) | (0.031) | (0.031) |
| 100% Text | -0.027 | -0.023 | -0.041 | -0.024 | -0.016 |
| | (0.026) | (0.025) | (0.033) | (0.026) | (0.026) |
| *With Baseline Controls* | | | | | |
| Mask Type (Surgical) | 0.029 | 0.029 | 0.063* | 0.018 | 0.022 |
| | (0.025) | (0.025) | (0.034) | (0.026) | (0.025) |
| Commitment w/ Signage | -0.007 | -0.003 | -0.021 | -0.004 | -0.005 |
| | (0.026) | (0.025) | (0.033) | (0.026) | (0.026) |
| Incentive Type | | | | | |
| Monetary | -0.021 | -0.024 | 0.013 | -0.031 | -0.026 |
| | (0.033) | (0.033) | (0.043) | (0.033) | (0.035) |
| Certificate | 0.006 | 0.003 | 0.026 | 0.008 | -0.006 |
| | (0.031) | (0.030) | (0.039) | (0.030) | (0.031) |
| 100% Text | -0.026 | -0.022 | -0.038 | -0.024 | -0.017 |
| | (0.025) | (0.025) | (0.033) | (0.026) | (0.026) |
| N villages | 286 | 286 | 286 | 286 | 286 |

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 control-intervention pair. The regressions "with baseline control" include controls for the number of people observed in the baseline visit.
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. 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.
'No Active Promotion" refers to any time that surveillance was conducted while promotion was not actively occurring (regardless of the week of the intervention). This excludes surveillance during the Friday Jumma Prayers in the mosque, when promoters were present and actively encouraged mask wearing.
"Other Locations" include tea stalls, at the entrance of the restaurant as patrons enter, and the main road to enter the village.

Figure S5: Household-Level Cross Randomizations




The figure corresponds to the regression presented in Table S17.
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 receive 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 purple or red.
For a more detailed description of the household-level cross-randomizations, see Section 4.4.

Table S17: Household-Level Cross-Randomizations

| Coefficient | Full |
|---|---|
| Household-Level Text Randomization | |
| 50% of Households in Village | -0.032 |
| | (0.031) |
| 100% of Households in Village | -0.028 |
| | (0.026) |
| Altruistic Messages | -0.030 |
| | (0.026) |
| Verbal Commitment | -0.024 |
| | (0.025) |
| Mask Color | |
| Blue vs Green | 0.027* |
| | (0.016) |
| Purple vs Red | -0.022** |
| | (0.011) |
| 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.
The regression includes a control for the mask type to separate the effect of mask colors.
Surgical masks distributed to households were blue or green. Cloth masks distributed to households were purple or red.

### J.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 4.3: 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 S10. 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.

## K 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).

## K.1 Estimating Effects of Village-level Cross-randomizations

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

$$E(p_j|T_j, x_j, D_k) = \beta T_j + \sum_k D_k \delta_k + x_j \gamma \tag{3}$$

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.

### K.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 $\Delta_j$, 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 $\alpha_{0m}$, given by:

$$E(\Delta_j | d_{jc}) = \alpha_{0m} + \sum_c \alpha_c d_{jc} + surgical_j \tag{4}$$

$\alpha_{0m}$ 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 $\Delta_j$) and heteroskedasticity-robust standard errors.

## L Additional Balance Tests

While our stratification procedure should have achieved balance with respect to variables observed at the time of randomization, given the many possible opportunities for errors in implementation, we nonetheless confirm in this Appendix L that our control and treatment villages resemble each other at baseline with respect to key variables of interest. This assessment was not preregistered. We find that the control and treatment groups are balanced with respect to our primary outcomes of interest: mask-wearing, symptoms and symptomatic seropositivity. In this Appendix we investigate several other covariates and find a few small imbalances, and conduct robustness checks.

For example, we find more 18-30 year olds in the treatment group, perhaps because households reported teenagers as 18 in order to receive more masks; our results are robust to dropping this age range.

In Table S4 we present balance test results for our mask-wearing specification (at the village level). In our main specification, this is a regression of mask-wearing on a constant, an intervention indicator, and indicators for each control-intervention pair with analytic weights proportional to the number of adults recorded in the baseline household survey as well as heteroskedasticity robust standard errors. For the balance tests, we replace the dependent variable with several variables measured at baseline: symptomatic seroprevalence, WHO-Defined COVID-19 symptoms, and baseline mask-wearing rate. We find that all of these variables appear balanced.

In Table S18, we report results from analogous balance tests based on the specification used for our primary biological outcome. We replace the dependent variable (symptomatic seroprevalence) with baseline covariates of interest to assess balance. We also report a bottom-line F-test which again fails to reject balance.

Table S18: Balance Tests (Individual-Level)

| | Baseline Symptomatic Seroprevalence | Baseline WHO-Defined COVID-19 Symptoms | Baseline Mask-Wearing Rate |
|---|---|---|---|
| | | *Summary Statistics* | |
| Intervention Rate | 0.00020 | 0.02468 | 0.11829 |
| Control Rate | 0.00022 | 0.02342 | 0.11990 |
| | | *Balance Tests* | |
| Intervention Coefficient | -0.00001 | 0.00081 | 0.00093 |
| | (0.00004) | (0.00113) | (0.00391) |
| N individuals | 304,726 | 304,726 | 304,726 |
| N villages | 572 | 572 | 572 |
| $F$ | | 0.76 | |
| *Joint-Test Prob* $> F$ | | 0.8596 | |

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 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.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.

We also ran balance tests with respect to several other covariates and detected a few balance failures. While small in magnitude, we investigate these further in order to understand the severity of the underlying problem.

Table S19 highlights these balance failures. Specifically, we find imbalances with respect to household count, age and household size. On average, treatment villages have 16 more households, the treatment villages have 0.4 percentage points more people younger than 30, and treatment households have 0.02 more members. While small in magnitude, these imbalances are unlikely to have arisen by chance given the size of our sample.

We believe the imbalances with respect to age and household size likely arose because households in the treatment group were more likely to report teenagers as being over 18 in order to receive additional masks. We believe the imbalance with respect to the number of households likely occurred for a similar reason, with implementers in the treatment group including more "borderline" households as part of the village in order to distribute masks to those households.

To check for these mechanisms, we drop from the sample individuals under 30 and villages with over 350 households – the latter only very coarsely targets "extra" households that lie on the border of villages. After imposing these restrictions, we find in Table S20 that the imbalances with respect to age and household size disappear entirely (this also occurs with the age restriction alone), and the imbalance with respect to household count shrinks by 25% but remains significant. In Table S21, we repeat our primary specification in this restricted sample with better balance and find that our results are qualitatively unchanged.

Table S19: Additional Balance Tests (Individual-Level)

| | Household Count | Proportion Female | Proportion Below 30 | Average Household Size |
|---|---|---|---|---|
| | | *Summary Statistics* | | |
| Intervention Group | 230 | 0.5127 | 0.3348 | 2.6506 |
| Control Group | 213 | 0.5114 | 0.3288 | 2.6158 |
| | | *Balance Tests* | | |
| Intervention Coefficient | 16*** | 0.0017 | 0.0044** | 0.0327*** |
| | (3) | (0.0011) | (0.0018) | (0.0094) |
| N individuals | 304,726 | 304,726 | 304,726 | 304,726 |
| N villages | 572 | 572 | 572 | 572 |
| *F* | | 55.62 | | |
| *Joint-Test Prob* $>$ *F* | | 0.0000 | | |

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table S20: Additional Balance Tests (Individual-Level, After Sample Selection)

| | Household Count | Proportion Female | Proportion Below 40 | Average Household Size |
|---|---|---|---|---|
| *Removing All People Below 30 & All Villages With More than 350 Households* | | | | |
| Intervention Coefficient | 12*** | 0.0034** | -0.0008 | 0.0086 |
| | (2) | (0.0013) | (0.0020) | (0.0057) |
| N individuals | 197,615 | 197,615 | 197,615 | 197,615 |
| N villages | 563 | 563 | 563 | 563 |
| $F$ | | 33.63 | | |
| *Joint-Test Prob* $> F$ | | 0.0000 | | |

Standard errors are in parentheses.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.
The sample excludes an additional 107,111 individuals up to the age of 30, and 9 villages that have more than 350 households.

Table S21: Symptomatic Seroprevalence (With Controls and Additional Sample Selection)

| | Intervention Effect | Intervention Effect by Mask Type |
|---|---|---|
| *Controlling for Number of Households and Sex* | | |
| Intervention Coefficient | -0.0006* | |
| | (0.0003) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0008* |
| | | (0.0004) |
| Intervention Coefficient for Cloth Mask Villages | | -0.0002 |
| | | (0.0005) |
| Average Symptomatic Seroprevalence Rate in Paired Control Villages§ | 0.0076 | 0.0076 |
| N individuals | 304,726 | 304,726 |
| N villages | 572 | 572 |
| *After Additional Sample Selection* | | |
| Intervention Coefficient | -0.0008* | |
| | (0.0004) | |
| Intervention Coefficient for Surgical Mask Villages | | -0.0011* |
| | | (0.0006) |
| Intervention Coefficient for Cloth Mask Villages | | -0.0001 |
| | | (0.0007) |
| Average Symptomatic-Seroprevalence Rate in Paired Control Villages§ | 0.0091 | 0.0091 |
| N individuals | 197,615 | 197,615 |
| N villages | 563 | 563 |

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 control-intervention pair.
The regression in the top panel includes controls for baseline rates of mask wearing, baseline symptom rates, number of households in a village, and sex.
The regression in the bottom panel controls for baseline rates of mask wearing and baseline symptom rates.
§We report the mean rate of symptomatic-seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.
The bottom panel runs sample excludes an additional 107,111 individuals up to the age of 30 and 9 villages that have more than 350 households.

# M Persistence of Mask-Wearing Behavior

In Table S22, we report estimates of our primary specification separately by week of surveillance. Week 10 is especially interesting, as it was two weeks after intervention activities ceased. This analysis was not preregistered.

We find no evidence that the impact of the intervention attenuates over the 10 weeks. In the 414 villages for which we have 10 weeks of surveillance, the point estimates are slightly smaller in week 10 (a 23.3 percentage point increase) than week 1 (30.4 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 that encourages subsequent adoption by others. If mask-wearing was driven by a "novelty factor" associated with 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 414 villages for which we have data in all weeks are plotted in Figure S6.

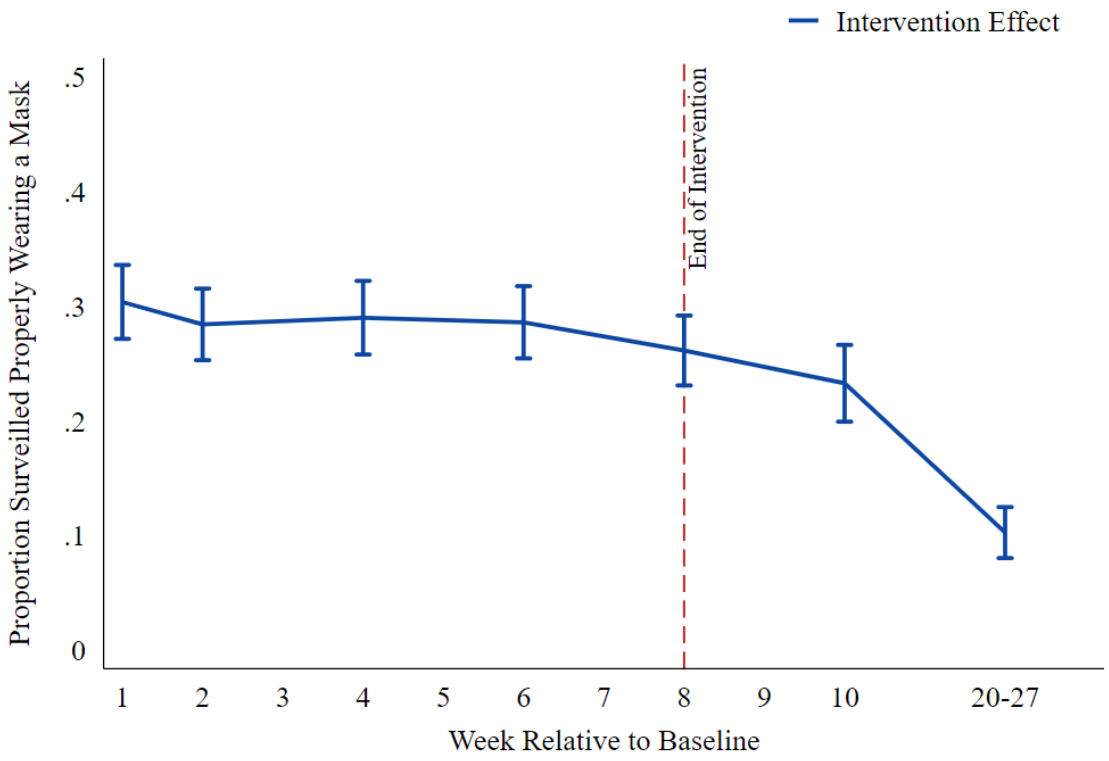
We additionally conducted a follow-up surveillance 5 months after the start of the intervention (20-27 weeks, depending on the wave). Mask-wearing had declined to 14.1% in the control group and 22.4% in the intervention group (a regression adjusted difference of 0.10 [0.08,0.13]).

Table S22: Persistence of Mask-Wearing

| | *Week from Baseline Observation* | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 4 | 6 | 8 | 10 | Followup |
| | *Consistent Panel* | | | | | | |
| Intervention Coefficient | 0.304*** (0.016) | 0.284*** (0.016) | 0.290*** (0.016) | 0.286*** (0.016) | 0.261*** (0.016) | 0.233*** (0.017) | 0.102*** (0.011) |
| N villages | 414 | 414 | 414 | 414 | 414 | 414 | 414 |
| | *All Villages* | | | | | | |
| Intervention Coefficient | 0.300*** (0.014) | 0.285*** (0.014) | 0.291*** (0.014) | 0.298*** (0.015) | 0.261*** (0.014) | 0.230*** (0.015) | 0.094*** (0.010) |
| N villages | 542 | 558 | 548 | 550 | 528 | 508 | 546 |

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 control-intervention pair, 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. 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.
"Followup" surveillance occurred between June 4th and June 8th 2021, which is anywhere from 20 to 27 weeks after baseline for each village.
This analysis estimates separate intervention effects 1, 2, 4, 6, 8, 10 weeks, and 20-27 weeks after baseline observation. The top panel runs the regressions only among a consistent panel of 414 villages that have all 10 weeks and the subsequent followup observation. The results of the analysis are displayed graphically in Figure S6.
The bottom panel is run among all villages which have surveillance data for that period of observation, as well as the baseline period.
The 10th week of observation and the followup observation occur after all active promotion of mask-wearing has ceased.

Figure S6: Persistence of Mask-Wearing




The figure corresponds to the regressions presented in Table S22, top panel. We present the effect of the intervention separately across weeks 1, 2, 4, 6, 8, 10, and 20-27 weeks after the baseline observation with 95% confidence intervals. The 20-27 week observation was collected during our "Followup" surveillance between June 4th and June 8th 2021, which is anywhere from 20 to 27 weeks after baseline for each village.
The analysis is run across a panel of 414 villages with observation through the entirety of the study. The 10th week of observation and the followup observation occur after all active promotion of mask-wearing has ceased.

# N Imputed Symptomatic Seroprevalence

In this section, we analyze the results of our intervention by wave, as well as assessing the sensitivity of our analysis to alternative methods of imputing missing values. These analyses turn out to be related, as there is one wave with imbalanced consent rates, meaning that dropping non-consenters distorts the treatment and control comparison.

In Table S23, we analyze the impact of our intervention on mask-wearing and physical distancing by waves. In all waves, mask-wearing increased by between 24.2 and 35.7 percentage points, and physical distancing increased by between 4.2 and 7.7 percentage points.

When we analyze our second stage results by wave, reported in the top panel of Table S24, we find that most waves have comparable effect sizes with two exceptions: in wave 7 we find an especially large impact of masks on symptomatic seropositivity and in wave 2 we find an opposite signed impact on symptomatic seropositivity.

Probing further, we believe the opposite signed result in wave 2 is due to imbalanced consent rates in that wave. In Table S25, we show consent rates for control and treatment groups by waves. Since we drop individuals who do not consent in our primary specification, lower consent rates appear as lower rates of symptomatic seropositivity. In Table S1, we found that consent rates were balanced across treatment and control groups, as well as by age and gender. In Table S25, we find that consent rates are generally comparable across waves, although they appear notably lower in the control group of wave 2 relative to the treatment group. Consent rates are somewhat lower in the control group of wave 5 and the treatment group of wave 7, although differences are not as stark.

To check whether our results are driven by differential consent, we consider an alternative method of dealing with missing data. Instead of dropping symptomatic individuals who did not consent to blood collection, we impute for those individuals the mean (conditional) seropositivity observed among all individuals in the data.

These results are shown in the bottom panel of Table S24. Several points are worth noting. First, the point estimate for the main effect of masks on seropositivity becomes substantially larger.

There is a mechanical effect due to the fact that rates of symptomatic seropositivity in the data now increase by a factor of 2.5 since we previously dropped the 60% of symptomatic individuals who did not consent to blood samples. Scaling our original estimate by this factor would give an effect size of -0.0018. The effect size with seropositivity imputed is slightly larger, at -0.0022, and much more precisely estimated than the main effect in our data. Additionally, with this imputation method, the anomalous Wave 2 result disappears.

Table S26 shows our main results broken out by age, with the top panel using our pre-registered sample and the bottom panel showing the same results with the imputation method. The results are quite similar in both cases, with larger effects at older ages.

In Table 4 in the main text, we further disaggregate the results by cloth and surgical masks. Two points are notable in these results. First, when we drop non-consenters (in the original specification), cloth masks appear to impact symptomatic seropositivity only in the 40-50 age group (and have an insignificant effect pooling all ages). However, if we instead impute seropositivity at the average value for non-consenters, cloth masks appear most effective at older ages, and about as effective as surgical masks.

Tables S28 and S7 report analogous results with respect to gender for symptoms and symptomatic seropositivity. We see a similar pattern to the age results: we see similar effects for both genders for symptoms and symptomatic seroposivity when we impute seropositivity at the average value for non-consenters. If we instead drop non-consenters, the symptomatic seropositivity estimates for men become less precise and are no longer significantly different from zero.

Table S23: Mask-Wearing and Physical Distancing by Wave, Controlling for Baseline Variables

| | Full | Wave 1 | Wave 2 | Wave 3 | Wave 4 | Wave 5 | Wave 6 | Wave 7 |
|---|---|---|---|---|---|---|---|---|
| | | | | *Proper Mask-Wearing* | | | | |
| Intervention Coefficient | 0.288*** | 0.282*** | 0.350*** | 0.258*** | 0.242*** | 0.310*** | 0.257*** | 0.357*** |
| | (0.012) | (0.053) | (0.044) | (0.021) | (0.028) | (0.025) | (0.025) | (0.033) |
| | | | | *Physical Distancing* | | | | |
| Intervention Coefficient | 0.051*** | 0.077*** | 0.056*** | 0.060*** | 0.045*** | 0.042*** | 0.048*** | 0.068*** |
| | (0.005) | (0.024) | (0.015) | (0.017) | (0.010) | (0.014) | (0.013) | (0.016) |
| N villages | 572 | 28 | 52 | 78 | 102 | 118 | 110 | 84 |

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 control-intervention pair. The regressions include controls for baseline rates of physical 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. 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.

Table S24: Symptomatic-Seroprevalence by Wave, With Baseline Controls

| | All | Wave 1 | Wave 2 | Wave 3 | Wave 4 | Wave 5 | Wave 6 | Wave 7 |
|---|---|---|---|---|---|---|---|---|
| *Pre-Registered Sample: Drop Individuals Without Blood Draws* | | | | | | | | |
| Intervention Coefficient | -0.0007** | -0.0008 | 0.0031*** | -0.0011 | -0.0009 | -0.0004 | -0.0006 | -0.0031*** |
| | (0.0003) | (0.0011) | (0.0009) | (0.0009) | (0.0008) | (0.0006) | (0.0008) | (0.0008) |
| Avg. Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0076 | 0.0061 | 0.0036 | 0.0067 | 0.0091 | 0.0067 | 0.0085 | 0.0095 |
| N Individuals | 304,726 | 14,606 | 27,132 | 43,602 | 56,607 | 62,419 | 57,397 | 42,963 |
| N Villages | 572 | 572 | 572 | 572 | 572 | 572 | 572 | 572 |
| *Imputing Symptomatic-Seroprevalence for Missing Blood Draws* | | | | | | | | |
| Intervention Coefficient | -0.0022*** | 0.0001 | 0.0007 | -0.0010 | -0.0027 | -0.0013 | -0.0028** | -0.0054*** |
| | (0.0005) | (0.0014) | (0.0010) | (0.0014) | (0.0017) | (0.0012) | (0.0011) | (0.0011) |
| Avg. Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0189 | 0.0104 | 0.0113 | 0.0143 | 0.0221 | 0.0192 | 0.0221 | 0.0217 |
| N Individuals | 321,948 | 14,995 | 28,168 | 45,358 | 60,237 | 66,303 | 61,185 | 45,702 |
| N Villages | 572 | 572 | 572 | 572 | 572 | 572 | 572 | 572 |

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 control-intervention pair. The regressions include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis in the top panel utilizes the pre-registered sample, equivalent to Table S7; it includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.
The analysis in the bottom panel replicates the regression in the top panel, but imputes the seropositivity of individuals for who we did not draw blood. For symptomatic individuals we did not draw blood from, we simulate their symptomatic-seroprevalence status by using the average rate of conditional seropositivity among all symptomatic individuals. This analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

Table S25: Endline Blood Collection Consent Rates by Wave

| | Number of Individuals | | Rate of COVID Symptoms | | Blood Draw Consent Rate | |
|---|---|---|---|---|---|---|
| | Treatment | Control | Treatment | Control | Treatment | Control |
| Wave 1 | 10,412 | 8,810 | 4.0% | 4.1% | 51.0% | 49.7% |
| Wave 2 | 18,378 | 14,862 | 4.8% | 5.2% | 44.2% | 33.9% |
| Wave 3 | 25,042 | 23,700 | 6.4% | 6.3% | 46.3% | 44.7% |
| Wave 4 | 31,999 | 28,598 | 8.8% | 10.4% | 40.9% | 42.0% |
| Wave 5 | 35,216 | 33,777 | 8.3% | 8.4% | 35.9% | 31.8% |
| Wave 6 | 33,424 | 31,954 | 8.8% | 10.1% | 38.5% | 37.8% |
| Wave 7 | 23,851 | 22,160 | 7.2% | 9.9% | 38.8% | 44.0% |

"Rate of COVID symptoms" reports the proportion of individuals that report WHO-defined COVID symptoms in the midline or endline surveys.
"Blood Draw Consent Rate" reports the proportion of individuals that consented to a blood draw in the endline, conditional on being symptomatic.

Table S26: Symptomatic Seroprevalence by Age Groups, With Baseline Controls, Expressed in Prevalence Ratios

| | All | < 40 Y.O. | 40-49 Y.O. | 50-59 Y.O. | ≥ 60 Y.O. |
|---|---|---|---|---|---|
| *Pre-Registered Sample: Drop Individuals Without Blood Draws* | | | | | |
| Intervention Prevalence Ratio | 0.905** [0.815, 0.995] | 0.995 [0.886, 1.104] | 0.917 [0.763, 1.070] | 0.791*** [0.634, 0.948] | 0.779** [0.610, 0.947] |
| Avg. Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0076 | 0.0055 | 0.0095 | 0.0108 | 0.0104 |
| N Individuals | 287,349 | 146,306 | 35,839 | 24,086 | 27,943 |
| N Villages | 538 | 480 | 384 | 348 | 360 |
| *Imputing Symptomatic-Seroprevalence for Missing Blood Draws* | | | | | |
| Intervention Prevalence Ratio | 0.879*** [0.820, 0.938] | 0.899*** [0.831, 0.967] | 0.929 [0.836, 1.022] | 0.924 [0.820, 1.028] | 0.731*** [0.627, 0.834] |
| Avg. Symptomatic-Seroprevalence in Paired Control Villages§ | 0.0189 | 0.0152 | 0.0226 | 0.0229 | 0.0251 |
| N Individuals | 321,383 | 177,708 | 51,676 | 37,340 | 43,431 |
| N Villages | 570 | 566 | 528 | 504 | 534 |

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 control-intervention pair. The regressions also include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis in the top panel utilizes the pre-registered sample, equivalent to Table 2; it includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.
The analysis in the bottom panel replicates the regression in the top panel, but imputes the seropositivity of individuals for who we did not draw blood. For symptomatic individuals we did not draw blood from, we simulate their symptomatic-seroprevalence status by using the average rate of conditional seropositivity among all symptomatic individuals. This analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

Table S27: Symptomatic Seroprevalence by Sex with Baseline Controls, Expressed in Prevalence Ratios

| | All | Male | Female |
|---|---|---|---|
| *Pre-Registered Sample: Drop Individuals Without Blood Draws* | | | |
| Intervention Prevalence Ratio | 0.905** | 0.947 | 0.869** |
| | [0.815, 0.995] | [0.837, 1.058] | [0.755, 0.984] |
| Intervention Prevalence Ratio for Surgical Villages | 0.889** | 0.905 | 0.870* |
| | [0.780, 0.997] | [0.779, 1.031] | [0.724, 1.016] |
| Intervention Prevalence Ratio for Cloth Villages | 0.942 | 1.049 | 0.869 |
| | [0.781, 1.103] | [0.829, 1.268] | [0.694, 1.044] |
| Average Symptomatic Seroprevalence in Paired Control Villages§ | 0.0076 | 0.0068 | 0.0083 |
| N Individuals | 287,349 | 129,308 | 133,898 |
| N Villages | 538 | 496 | 486 |
| *Imputing Symptomatic-Seroprevalence for Missing Blood Draws* | | | |
| Intervention Coefficient | 0.879*** | 0.881*** | 0.876*** |
| | [0.820, 0.938] | [0.811, 0.952] | [0.801, 0.951] |
| Intervention Coefficient for Surgical Villages | 0.873*** | 0.847*** | 0.894** |
| | [0.801, 0.945] | [0.766, 0.929] | [0.797, 0.990] |
| Intervention Coefficient for Cloth Villages | 0.890** | 0.953 | 0.842*** |
| | [0.787, 0.993] | [0.819, 1.087] | [0.725, 0.959] |
| Average Symptomatic Seroprevalence in Paired Control Villages§ | 0.0189 | 0.0178 | 0.0200 |
| N Individuals | 321,383 | 156,302 | 164,004 |
| N Villages | 570 | 568 | 566 |

Confidence Intervals are in brackets.
*** 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 control-intervention pair. The regressions include controls for baseline rates of proper 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls. The analysis in the top panel utilizes the pre-registered sample, equivalent to Table S7; it includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood. The analysis in the bottom panel replicates the regression in the top panel, but imputes the seropositivity of individuals for who we did not draw blood. For symptomatic individuals we did not draw blood from, we simulate their symptomatic-seroprevalence status by using the average rate of conditional seropositivity among all symptomatic individuals. This analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

Table S28: WHO-Defined COVID Symptoms by Sex, Expressed in Prevalence Ratios

| | All | Male | Female |
|---|---|---|---|
| *No Baseline Controls* | | | |
| Intervention Coefficient | 0.885*** | 0.884*** | 0.886*** |
| | [0.834, 0.934] | [0.834, 0.933] | [0.823, 0.950] |
| Intervention Coefficient for Surgical Villages | 0.874*** | 0.864*** | 0.882*** |
| | [0.809, 0.939] | [0.803, 0.926] | [0.800, 0.964] |
| Intervention Coefficient for Cloth Villages | 0.907** | 0.922* | 0.895** |
| | [0.823, 0.991] | [0.838, 1.005] | [0.798, 0.991] |
| Avg. Symptomatic Seroprevalence in Paired Control Villages§ | 0.0860 | 0.0824 | 0.0894 |
| *With Baseline Controls* | | | |
| Intervention Coefficient | 0.884*** | 0.884*** | 0.885*** |
| | [0.834, 0.934] | [0.837, 0.932] | [0.822, 0.947] |
| Intervention Coefficient for Surgical Villages | 0.865*** | 0.857*** | 0.873*** |
| | [0.803, 0.928] | [0.800, 0.915] | [0.792, 0.953] |
| Intervention Coefficient for Cloth Villages | 0.922* | 0.938 | 0.908* |
| | [0.838, 1.005] | [0.854, 1.022] | [0.812, 1.004] |
| N Individuals | 321,948 | 156,846 | 165,102 |
| N Villages | 572 | 572 | 572 |

Confidence Intervals are in brackets.
*** 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 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. 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 symptomatic seroprevalence at endline. This is not equivalent to the coefficient on the constant due to the inclusion of the pair indicators as controls.
The analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.

# O Variation of Effects

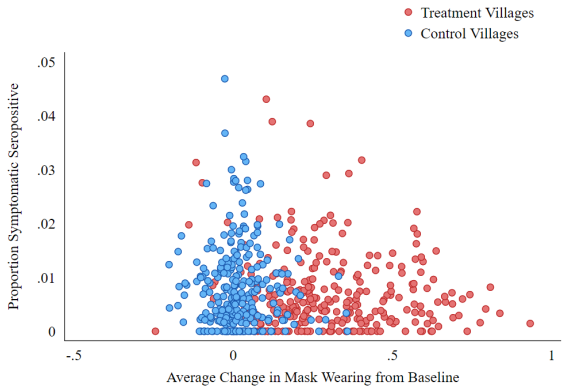
In this Appendix, we investigate how WHO-Defined COVID symptoms and symptomatic seropositivity relate cross-sectionally to changes in mask-wearing and changes in physical distancing relative to baseline. This comparison should be interpreted with caution, since the observational variation across villages in mask-wearing and measured physical distancing is not random. For example, within the treatment or control group, some villages might have more mask-wearing precisely because people were observed with COVID-19 symptoms. Were this the case, even if masks reduced COVID-19, we might see a positive relationship between mask-wearing and biological outcomes; a similar bias could be present for physical distancing.

With these caveats in mind, Figure S7 shows the relationship between each biological outcome variable and the changes in mask-wearing and physical distancing graphically. Table S29 shows coefficients from a regression of each outcome on the respective change, controlling for the same covariates as our baseline regression, except for pair fixed effects (omitting these effects is necessary if we want to study cross-sectional variation across villages, rather than only pairwise comparisons). We report these results for each covariate separately, as well as both together (note that the latter specification makes sense as a causal model only if mask-wearing does not directly cause physical distancing).

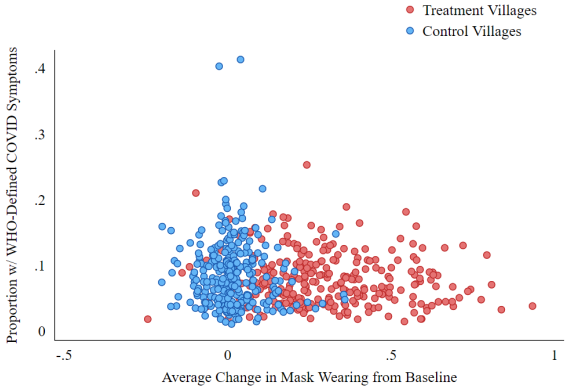
We find clear evidence of a negative relationship between mask-wearing and both symptoms and symptomatic seropositivity. Once we control for mask-wearing, we see no significant relationship between physical distancing and symptomatic seropositivity. The standard deviation of the change in mask-wearing across villages is also considerably larger than the change in physical distancing, at 0.21 vs. 0.13 respectively, so even were the coefficients the same, the change in mask-wearing in the causal interpretation would account for more of the variation outcomes.

Figure S7: Variation of Effect within Treatment Arms

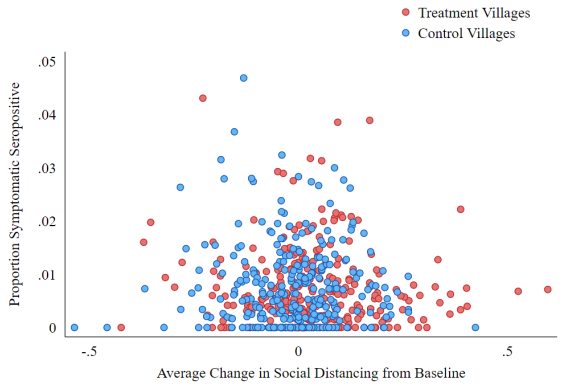
(a) Symptomatic Seroprevalence by Change in Mask Wearing



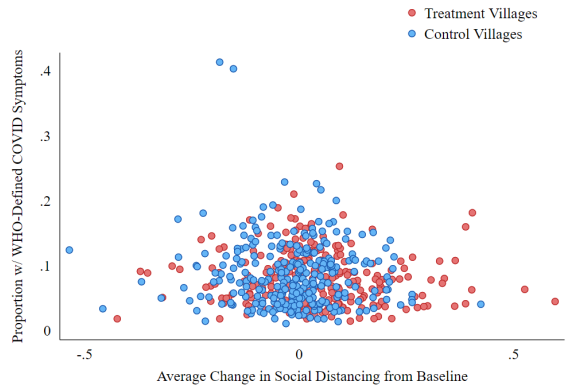

(b) Symptoms by Change in Mask Wearing




(c) Symptomatic Seroprevalence by Change in Physical Distancing




(d) Symptoms by Change in Physical Distancing




Panels S7a and S7b plot the average change in mask wearing for each village against the proportion of people in their village that are symptomatic-seropositive (panel S7a) or have WHO-defined COVID symptoms (panel S7b). Panels S7c and S7d plot for each village, the average change in physical distancing against the proportion of individuals that are symptomatic-seropositive (panel S7c) or have WHO-defined COVID symptoms (panel S7d). Change in mask wearing [physical distancing] is the difference in the average rate of mask wearing [physical distancing] during the intervention between the average rate of mask wearing [physical distancing] in the baseline. Each point represents a village, with color indicating treatment status.

Table S29: Symptomatic Seroprevalence & COVID Symptoms by Mask Wearing & Physical Distancing

| | Mask Wearing | Physical Distancing | Mask Wearing & Physical Distancing |
|---|---|---|---|
| *WHO-Defined COVID Symptoms* | | | |
| Coefficient on Change in Mask-Wearing from Baseline | -0.0279*** | | -0.0241*** |
| | (0.0081) | | (0.0081) |
| Coefficient on Change in Social-Distancing from Baseline | | -0.0301* | -0.0174 |
| | | (0.0157) | (0.0159) |
| N Individuals | 321,948 | 321,948 | 321,948 |
| N Villages | 572 | 572 | 572 |
| *Symptomatic-Seropositivity* | | | |
| Coefficient on Change in Mask-Wearing from Baseline | -0.0032** | | -0.0028** |
| | (0.0013) | | (0.0014) |
| Coefficient on Change in Social-Distancing from Baseline | | -0.0031 | -0.0016 |
| | | (0.0023) | (0.0025) |
| N Individuals | 304,726 | 304,726 | 304,726 |
| N Villages | 572 | 572 | 572 |

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 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. 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 analysis in the top panel includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.
The analysis in the bottom panel includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for, symptomatic individuals that we did not collect blood from, and individuals that we drew blood from but did not test their blood.

# P Additional Preregistered Specifications

In this section, we discuss additional preregistered specifications not reported in the text. For reference, our pre-analysis plan is available at: https://osf.io/vzdh6/. Our initial intention had been to collect blood from a single high-risk individual within each household at endline. When we failed to collect as many baseline bloodspots as hoped, we decided to test all symptomatic individuals at endline rather than a single high-risk individual in each household. The only data observed at the time of this decision was the count of total baseline bloodspots collected.

We had initially planned to do only telephone surveys at weeks 5 and 9. Near the start of our Week 9 activities, we switched to in-person household surveys in Week 9 in order to increase the survey response rate. At the time this decision was made, analysts were still blinded to treatment and control villages. In total, we surveyed 104,063 households in week 5 (all phone surveys), and 118,018 households in week 9. Of these, 102,871 were household surveys.

Our pre-registration document suggests that we can compute the impact of our intervention on seroconversions by comparing our effect size to the difference between endline and baseline seropositives among individuals symptomatic during our intervention. As the analysis in Appendix I makes clear, this is not quite correct. If $P_{prior}$, the fraction of symptomatic seropositives due to infections prior to baseline, is zero, then the estimated impact on symptomatic seropositives equals the impact on symptomatic seroconversions and no further adjustment is needed. More generally, the impact on symptomatic seropositives incorporates both seroconversions, as well as reductions in symptomatic seroconversions due to non-COVID respiratory diseases. We cannot determine the impact on seroconversions without knowing both $P_{prior}(0)$ and the relative impact of masks on COVID-19 and non-COVID respiratory diseases. If the latter two quantities are equal in proportion, the impact on symptomatic seropositives again equals the impact on symptomatic seroconversions with no further adjustment needed.

Given that we find no evidence of an impact of any of the cross-randomizations, we did not estimate the specification flexibly interacting them.

We did not proceed with the "individual intervention" described in the pre-registration docu-

ment which was designed to test the protective benefits of masks to the wearer, because we were unable to entice a sufficient number of markets and vendors to participate in that trial and switch mask-wearing behavior.

We did not collect the intended pharmacy data to use as an auxiliary outcome, and we did not collect follow-up hospitalization and mortality data due to the expense of revisiting households. We also do not yet have data on distance to nearby city or estimated average village-wealth.

In Table S30, we report our pre-specified instrumental variable regressions. If we assume that the entire impact of our intervention is via proper mask-wearing, then we estimate that going from zero percent to one hundred percent of villagers wearing masks would reduce symptomatic seroprevalence by -0.0024, a 32% reduction. Essentially, this specification scales our "intent-to-treat" estimates by a factor of 3.33, the reciprocal of the first stage.

Table S30: IV Regressions

| | Symptomatic Seroprevalence | WHO-Defined COVID-19 Symptoms |
|---|---|---|
| *No Baseline Controls* | | |
| Proper Mask-Wearing Coefficient | -0.0024** | -0.0327*** |
| | (0.0012) | (0.0077) |
| *With Baseline Controls* | | |
| Proper Mask-Wearing Coefficient | -0.0024** | -0.0325*** |
| | (0.0012) | (0.0075) |
| N Individuals | 304,726 | 321,948 |
| N Villages | 572 | 572 |

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 control-intervention pair. The regressions "with baseline controls" include controls for baseline rates of proper 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. 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 analysis includes all people surveyed in the baseline household visits, excluding individuals that we did not collect midline or endline symptoms for.
Proper Mask-Wearing is defined as the village-level rate of individuals observed properly wearing mask during the intervention period. The instrument is the treatment status of the village.

# Q 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 51,660 people to regularly wear masks, or 173 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), 68,775 cloth masks, and 136,770 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, $347.35 worth of cloth masks per village, and $89.35 of surgical masks per village. We estimate that we induced 598 x 29% = 173 people per village to wear masks, which amounts to fixed costs of $3.75 per adult induced to wear a mask in cloth mask villages, and $2.26 per adult in surgical mask villages.

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

- 1,089,947 masks distributed through promotion over an average of 29 days per village. Of these, there were 301,868 cloth masks distributed (105 cloth masks per day per village) and 788,079 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 $52.57 per village-day for cloth masks and $17.75 per village-day for surgical masks. The promotion costs were $24 per village-day. Dividing by the number of people induced to wear masks per village (173), we obtain costs of $0.44 per person-day in cloth mask villages and $0.24 per person-day in surgical mask villages. Using these figures, we calculated that after subtracting surveillance costs, our intervention cost $17.00 for each person induced to regularly wear a cloth mask and $9.49 for each person to regularly wear a surgical mask.

**Cost-effectiveness** To determine the impact of the intervention using surgical masks in reducing mortality from COVID-19 in Bangladesh, we used estimates of current and projected deaths from COVID-19, including excess deaths that occurred over the same time period (May 1, 2021-September 1, 2021) (*79*). The lower bound includes only COVID-19 reported deaths. The mid-range estimates include 50% of excess deaths as being directly attributable to COVID-19. The upper bound includes all excess deaths that occurred over the same time period as being directly attributable to COVID-19. We projected the impact of the intervention using surgical masks on deaths over four months following one month of intervention. We calculated the absolute risk reduction as the difference in death rate over the intervening period with and without the surgical mask intervention. We applied a 35% reduction of deaths among those 60 and older and a 23% reduction of deaths among those aged 50-60 based on the study findings and age-adjusted COVID-19 mortality rates for Bangladesh (*80*). We assumed no change in deaths for those under age 50. We determined the number needed to treat by taking the inverse of the absolute risk reduction.

As shown in Table S31, for one month of the intervention, the number needed to treat to prevent one death ranges from 6,682 to 35,001. Our estimates above suggest that the total cost of our intervention per person induced to wear a mask for a month was: $3.75 + $0.44 x 30 = $17.00

in cloth mask villages and $2.26 + $0.24 x 30 = $9.49 in surgical mask villages. By multiplying the number needed to treat times the cost per person induced to wear a mask, we estimate that after four months, the intervention as we conducted it (with cloth and surgical masks) cost between $63,408 and $332,161 per life saved, depending on mortality estimates. 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. Furthermore, this does not account for reductions of morbidity associated with hospitalization or other complications of COVID-19.

Table S31: Calculation of Number Needed to Treat and Cost per Life Saved

| | COVID-19-related Deaths (May 1 - Sept 1, 2021)* | Estimated Deaths with Intervention† | ARR | NNT | Cost per Life Saved - Intervention (USD) | Cost per Life Saved - at Scale (USD) |
|---|---|---|---|---|---|---|
| Lower bound | 17,984 | 13,233 | 2.86E-05 | 35,001 | $332,161 | $52,502 |
| Mid-range | 56,097 | 41,276 | 8.91E-05 | 11,221 | $106,487 | $16,831 |
| Upper bound | 94,209 | 69,319 | 1.50E-04 | 6,682 | $63,408 | $10,022 |

ARR = Absolute Risk Reduction; NNT = Number Needed to Treat
*https://covid19.healthdata.org/bangladesh
†Applying 35% reduction to deaths in the 60+ age group and 23% reduction to deaths in the 50-59 age group

Many cost elements can be brought down further through "at-scale implementation". 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. Additionally, surgical masks are about 8 times cheaper than cloth masks, and factory production costs can be brought down at scale. We calculate based on our current at scale activities that conducting the intervention for one month for the entire country of Bangladesh would cost $1.50 USD/person. Following out the effects for four months after one month of intervention, this translates to sub-

stantially lower costs per life saved: $10,022-$52,502 (Table S31).

For context, (*51*) estimate that the value of a statistical life is $205,000 in Bangladesh, implying that our intervention at scale is 4-20 times more cost-effective than what the typical Bangladeshi would be willing to pay to reduce mortality risk, and therefore a "very good buy" for policymakers. This cost-effectiveness analysis was not pre-specified.

# R Polling Policy Makers

## R.1 Polling and Policy-Maker Priors

To assess how our findings compared to the priors of relevant policy makers, we polled participants during presentations to the World Health Organization, the World Bank, and the National Council of Applied Economic Research in Delhi, India. In total, more than 100 audience members with expertise and specific interest in public health and mask-wearing were surveyed and asked to make predictions about the impact of our various interventions on mask-wearing and physical distancing, just before we showed them our empirical results (at the time, our biological outcomes were unavailable).

There are three main takeaways from this polling exercise: first, only a tiny fraction of policymakers correctly predicted the impact of our core intervention on mask-wearing and physical distancing. Second, policy maker predictions varied widely, both for effects of the intervention on mask-wearing and physical distancing. Third, policy-makers systematically underestimated the overall impact of our intervention and especially the impact of in-person reinforcement on mask-wearing.

When asked if they thought the intervention would increase mask-wearing by 5, 10, 20, 30, or 40 percentage points, only 21% of respondents correctly predicted that the intervention increased mask-wearing by 30 percentage points (about what we would expect if they guessed randomly). The expected value of the predicted increase in mask-wearing was 22 percentage points whether we described the intervention with or without mask promotion included. The difference in mask-

wearing observed in our two pilot studies suggests that in-person reinforcement increased mask-wearing by 18 percentage points. In other words, policy-makers makers believed that in-person reinforcement would have no additional impact, despite our piloting suggesting it is the single most important element of our intervention. With regard to behavioral adjustments, 64% of respondents predicted that physical distancing would either decrease or remain unchanged as a result of the mask-promotion interventions, when in fact, it increased.

Policy-makers consistently believed that our cross-randomizations would increase mask-wearing, when in fact, we find that none of them had a significant effect (often with fairly precise zeros). 68% of respondents believed that text messages would help (they didn't), 62% of respondents believed that incentives for village-leaders would help (they didn't), and 77% of respondents believed that verbal commitments or commitments made using signs on one's door would increase mask-wearing (they didn't). More results from this polling exercise are presented in the tables below.

Table S32: What do you think was the increase in mask-wearing as a result of household mask distribution and mask promotion in the community?

| | WHO | NCAER | World Bank | Frequency | Percent |
|---|---|---|---|---|---|
| No change | 0 | 1 | 3 | 4 | 3% |
| Increased by 5 percentage points | 5 | 10 | 8 | 23 | 20% |
| Increased by 10 percentage points | 4 | 12 | 8 | 24 | 21% |
| Increased by 20 percentage points | 4 | 19 | 9 | 32 | 28% |
| Increased by 30 percentage points | 4 | 7 | 11 | 22 | 19% |
| Increased by 40 percentage points | 2 | 6 | 3 | 11 | 9% |
| Total | 19 | 55 | 42 | 116 | 100% |

These are polls taken in response to the prompt: "We provided free masks to all households and promoted mask-wearing in mosques and markets with community leaders and imams. What do you think happened to mask-wearing relative to the 13% proper mask usage rate in the control villages without any interventions?" The results were collected from audience participants during live presentations to the World Health Organization (WHO), the National Council of Applied Economic Research (NCAER) in Delhi, and the World Bank.

Table S33: What do you think was the additional effect of mask promoters reminding people to wear masks?

| | WHO | NCAER | World Bank | Frequency | Percent |
|---|---|---|---|---|---|
| No change | 0 | 1 | 4 | 5 | 4% |
| Increased by 5 percentage points | 2 | 4 | 5 | 11 | 9% |
| Increased by 10 percentage points | 6 | 20 | 5 | 31 | 26% |
| Increased by 20 percentage points | 2 | 10 | 14 | 26 | 22% |
| Increased by 30 percentage points | 4 | 10 | 11 | 25 | 21% |
| Increased by 40 percentage points | 5 | 10 | 7 | 22 | 18% |
| Total | 19 | 55 | 46 | 120 | 100% |

These are polls taken in response to the prompt: " In addition to the mask distribution and promotion activities described previously, we had mask promoters periodically monitor passers-by and remind them to wear masks. What do you think happened to mask-wearing relative to the 13% proper mask usage rate in the control villages without any interventions?"
The results were collected from audience participants during live presentations to the World Health Organization (WHO), the National Council of Applied Economic Research (NCAER) in Delhi, and the World Bank.

Table S34: Do you think text message reminders to wear masks further increased mask-wearing?

| | WHO | NCAER | World Bank | Frequency | Percent |
|---|---|---|---|---|---|
| Yes | 0 | 33 | 32 | 65 | 68% |
| No | 0 | 19 | 11 | 30 | 32% |
| Total | 0 | 52 | 43 | 95 | 100% |

These are polls taken in response to the prompt: "We sent text reminders to wear masks. Do you think this increased mask-wearing further?"
The results were collected from audience participants during live presentations to the World Health Organization (WHO), the National Council of Applied Economic Research (NCAER) in Delhi, and the World Bank.

Table S35: How do you think mask distribution and promotion affected physical distancing?

| | WHO | NCAER | World Bank | Frequency | Percent |
|---|---|---|---|---|---|
| Physical distancing decreased | 5 | 0 | 8 | 13 | 22% |
| Physical distancing was unchanged | 9 | 0 | 16 | 25 | 42% |
| Physical distancing increased | 5 | 0 | 17 | 22 | 37% |
| Total | 19 | 0 | 41 | 60 | 100% |

These are polls taken in response to the prompt: "How did mask distribution and promotion affect individuals' physical distancing?"
The results were collected from audience participants during live presentations to the World Health Organization (WHO), the National Council of Applied Economic Research (NCAER) in Delhi, and the World Bank.

Table S36: Do you think incentive payments to village leaders further increased mask-wearing?

| | WHO | NCAER | World Bank | Frequency | Percent |
|---|---|---|---|---|---|
| Yes | 0 | 32 | 0 | 32 | 62% |
| No | 0 | 20 | 0 | 20 | 38% |
| Total | 0 | 52 | 0 | 52 | 100% |

These are polls taken in response to the prompt: "We promised the village and leaders an incentive payment if we saw increases in mask-wearing. Do you think this increased mask-wearing further?"
The results were collected from audience participants during live presentations to the World Health Organization (WHO), the National Council of Applied Economic Research (NCAER) in Delhi, and the World Bank.

Table S37: Do you think verbal commitments and signage to wearing masks further increased mask-wearing?

| | WHO | NCAER | World Bank | Frequency | Percent |
|---|---|---|---|---|---|
| Yes | 0 | 40 | 0 | 40 | 77% |
| No | 0 | 12 | 0 | 12 | 23% |
| Total | 0 | 52 | 0 | 52 | 100% |

These are polls taken in response to the prompt: "We had households verbally committing to wear masks and putting up signs to display to others that they were a mask-wearing household. Do you think this increased mask-wearing further?"
The results were collected from audience participants during live presentations to the World Health Organization (WHO), the National Council of Applied Economic Research (NCAER) in Delhi, and the World Bank.