

Evaluating the Impact of Behavioural Messages on Honest Reporting of COVID-19 Symptoms via the Mobile COVID-19 Connect Platform in South Africa

Pre-analysis plan | May 2022

INTRODUCTION

IDinsight and Praekelt.org are conducting multiple randomized control trials (RCTs) to assess the impact of mobile-based behavioral messages on the COVID-related health decisions of COVID-19 Connect users in South Africa.

This pre-analysis plan documents the key research questions the evaluation seeks to answer and specifies the analysis that will be performed for each question. Since the authors completed the plan before any data was analyzed, the plan will provide a useful reference in evaluating the final results of the study.

Project Context

South Africa reported its first COVID case at the beginning of March 2020, and by June-July, it reported almost 14,000 new cases and 200-300 deaths a day. Following a reduction in new reported cases between August and November, COVID cases and deaths began to rise again in December, possibly due to a new and more transmissible strain of the virus (NDTV, 2020). As of February 2, 2021, there have been a total of 1,456,309 reported COVID cases and 44,399 deaths due to the disease (World Health Organization, 2021) in the country.

The South African National Department Of Health (NDOH), in partnership with Praekelt.org, launched the digital COVID-19 Connect platform to disseminate accurate and timely information on COVID transmission, prevention, symptoms, and up-to-date statistics to the public. With 6.2 million users joining the platform in the first 7 weeks after its launch and approximately 100,000 daily users,¹ the platform has massive potential to deliver accurate COVID-related information and influence health-seeking behaviors at scale.

¹ <https://www.praekelt.org/covid-19-response-in-sa>

Intervention Overview

The COVID-19 Connect platform consists of two sets of digital tools: HealthAlert and HealthCheck. HealthAlert disseminates accurate and up-to-date information on COVID transmission, prevention, symptoms, testing, and treatment news updates as well as up-to-date statistics to the public. HealthCheck helps users assess their COVID risk through a COVID symptom checker and receive appropriate health behavior recommendations in return.

In collaboration with Praekelt and the NDOH, IDinsight will evaluate the effectiveness of **light-touch behavioral messaging nudges on new and existing HealthCheck users**. The intervention tests whether messaging appealing to a new or existing user's commitment to honesty can lead to more truthful responses. We have identified several promising messaging nudges. These were selected based on the COVID-19 Connect Theory of Change (see Appendix), behavioral decision frameworks, analysis of existing HealthCheck data, as well as a comprehensive evidence review of similar interventions, and discussions with the Praekelt team.

Evaluation Overview

The impact of the intervention will be evaluated using randomized controlled trials (RCTs). Randomization will be at the individual level, where an individual refers to the "unique user" (henceforth user), which is defined by a combination of their phone number and the channel they are using to access the COVID-19 Connect platform (USSD, WhatsApp).

Randomizing control and treatment allocation can ensure that the study groups will be statistically identical on observable and unobservable variables, on average. Therefore, any difference in indicators at the end of an intervention can be directly attributed to the intervention. The majority of the outcomes in this research will be collected through existing Praekelt backend user data, though several of the proposed outcomes will be collected through user surveys.

The proposed research will provide evidence on the value of light-touch behavioral nudges as a policy tool for improving the practice of COVID related health behaviors. Since these experiments will be conducted in close collaboration with Praekelt and the NDOH, the lessons learned will be applied to improve the COVID-19 Connect platform and amplify its impact to improve health outcomes for the South African population.

In what follows, we describe the intervention details and analytical framework for the study..

Increasing the honest reporting of COVID symptoms

Background/Literature

We test whether messaging appealing to a user's commitment to honesty can lead to more truthful responses to HealthCheck symptom tracker questions.

This set of interventions targets college students, faculty, and other employees of colleges in South Africa who regularly use HealthCheck. The "Higher Health" initiative led by the Higher Education, Science and Technology Ministry requires students to complete HealthCheck to be granted entry to public spaces on college campuses to reduce the spread of COVID.² The HealthCheck tool was adapted for this purpose: after university students and staff complete the symptom tracker questions, they are shown a digital receipt which declares their risk level (low, medium, high). To enter public spaces, users must show receipt of a "low-risk" result within the last 24 hours.

Since HigherHealth users must repeatedly complete the COVID-risk self-assessment on a daily or near-daily basis and produce a low risk result to gain entry to college campuses, we hypothesize that over time responses are likely to become less truthful on average, with a higher proportion of "low risk" users relative to the true prevalence of low-risk symptoms in the population. One reason is that HealthCheck represents a barrier to entry and inconvenience: incentivized by freedom of movement within public spaces on campus, HigherHealth users may be likely to downplay symptoms to achieve a "low risk" result, particularly those who are on the margins of low and moderate risk. Another reason could simply be that over time, the task becomes more mechanical and less deliberate, so users are likely to fill out the same set of responses automatically each time, even when their symptoms may have changed, also leading to a greater proportion of "low risk" categorizations than is true.

² <https://www.praekelt.org/higher-health-launch>

Figure 1: HealthCheck risk distribution, by recruitment platform

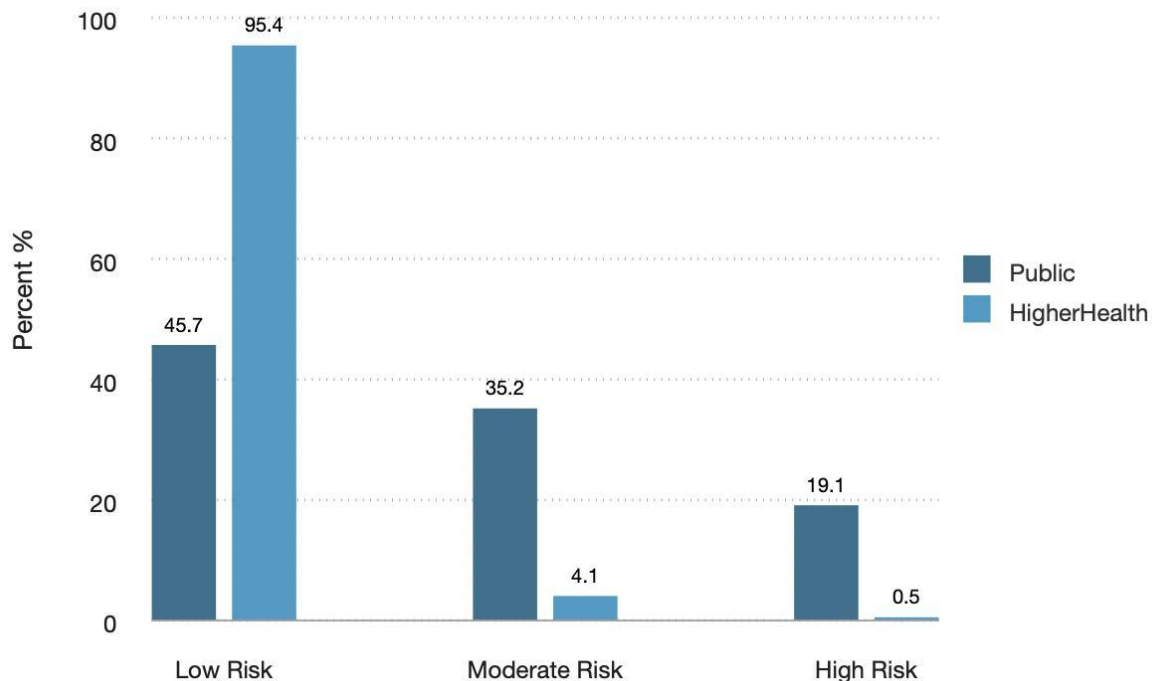


Figure 1 shows the distribution of HealthCheck risk categorizations by recruitment platform (Public vs HigherHealth). As seen from the graph above, HealthChecks completed on the HigherHealth platform fall overwhelmingly in the low risk category, compared to HealthChecks completed on the Public platform. Though it is possible that HigherHealth users may actually be lower risk due to their younger demographic or better COVID management policies on college campuses, it is difficult to rule out the hypothesis that users on this platform downplay their symptoms.

Mazar, Amir, and Ariely (2008) established the notion of increasing honesty through influencing one's self-concept (the way people view and perceive themselves), through targeting internal reward levers (e.g. intrinsic motivation) instead of external levers (e.g. punishment). People typically view honesty as part of their internal reward system, have strong beliefs in their own morality, and want to maintain this aspect of their self-concept (Greenwald 1980; Griffin and Ross 1991; Josephson Institute of Ethics 2006; Sanitioso, Kunda, and Fong 1990). People are often torn between two competing motivations: gaining from cheating versus maintaining a positive self-concept through being honest. They balance these two competing motivations through either rationalizing their actions in more compatible terms with their morality or by paying more or less attention to their own moral standards.

Several lab experiments have found that messaging can increase honesty by targeting these internal levers and increasing people's attention to their own moral standards. These experiments include

appealing to people's sense of morality via religious reminders (Mazar, Amir, and Ariely 2008) and making it difficult for participants to break a promise by leveraging commitments and reminding subjects of an existing honesty norm (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006). Previous research on commitments and honesty has found that signing a no-cheating declaration at the beginning rather than at the end of a self-reported survey can increase honesty (Shu et al, 2012). However, a replication study by Kristal et al (2020) with a larger sample size failed to find any differences between no-cheating declarations placed at the beginning or end of the survey.

Recent work by Cagal et al (2019) finds that message framing affects the impact of honesty appeals: signing a no-cheating declaration increases truth-telling if the declaration is morally charged, does not affect behavior if it is morally neutral, and reduces truth-telling if it is morally neutral and threatens to punish. Hence there is some scope to test whether placing a morally charged appeal to honesty at the beginning of the HealthCheck survey could increase honest reporting of symptoms from HigherHealth users.

A related strand of literature on improving honesty has focused on highlighting the consequences of dishonest behavior and reinforcing those consequences through verification threats. A field experiment run by Prima et al (2020) in Indonesia on the reporting of physical assets finds that verification threats instead of honesty pledges are more effective at reducing the likelihood of asset misreporting. This is corroborated by Bing et al (2011), who conduct an experiment with business school students and find that explicit reminders of an honor code combined with a realistic penalty for cheating reduces academic dishonesty amongst business school students. Though there is insufficient scope to enforce penalties for dishonesty from HigherHealth users, the available evidence suggests that making the consequences of honesty salient could be effective at increasing honest reporting of symptoms. Since the costs of dishonesty in the context of the COVID pandemic are not just private (such as financial or reputational losses) but also societal (health), we propose testing whether messages emphasizing social benefits of honest behavior could have an added effect on improving honesty.

Research questions

For this study, we propose the following research questions:

- **Research question 1:** "What is the effect of asking individuals to pre-commit to truthful reporting of COVID symptoms?"
- **Research question 2:** "Does framing the pre-commitment message as a pro-social appeal, or increasing the salience of consequences have a differential effect on truthful reporting of COVID symptoms compared to neutral (value-free) framing?"

Intervention

This study will consist of 1 control arm and 3 treatment arms to investigate the impact of using different behavioral science insights on the honest reporting of symptoms. Each of the arms are discussed below.

Control (*Status-quo*) - The HealthCheck survey currently includes an honesty declaration at the end of the survey where the user is asked to confirm that the information they have provided is accurate and whether they consent to being contacted by the NDOH if necessary. It is possible that in its current state where both questions are asked as one, the user may muddle their response. The control arm will use this existing HealthCheck feature but separate out the two components (honesty declaration, consent to be contacted) into two distinct questions.

Treatment 1 (*Status-quo + beginning*): This treatment arm will use the same honesty declaration used in the control arm. However, instead of placing this declaration at the end of the survey, it will place the declaration at the beginning of the survey.

Treatment 2 (*Pro-social appeal*): This treatment arm will be similar to treatment 1 as it will also situate the honesty declaration at the beginning of the survey. However, instead of using a neutral framing, the declaration will have a “pro-social” framing, by appealing to the user’s motivation to protect the health of those around them (i.e. their campus community) by honestly reporting their symptoms.

Treatment 3 (*Salience of consequences*): Similar to treatment arms 1 and 2, this treatment arm will also situate the honesty declaration at the beginning of the HealthCheck survey. However, the honesty declaration will include additional information emphasizing that the user may regret passing COVID to a vulnerable family member.

Target Population and Recruitment

We will target new and existing users on the “HigherHealth” channel. Analysis of HealthCheck data shows that a majority of the HealthChecks completed on this channel are from users in the age 18-40 category (88%), followed by the 40-65 category (9.1%), with very little usage from the under 18 and above 65 age categories. The users in the 18-40 category likely correspond to university students whereas the 40-65 category are likely academic and administrative staff. Similar to the public channel, there is an uneven distribution of completed HealthChecks across provinces on the “HigherHealth” channel, as 64.9% of the completed HealthChecks are from the ZA-GT, ZA-LP, and ZA-NL provinces while only 0.3% of the completed HealthChecks are from the ZA-NC province. Lastly, the majority of HealthChecks on the “HigherHealth” channel are completed via USSD (70%) while WhatsApp and Web account for the remaining 30%. Only users aged 18 or higher from the “HigherHealth” WhatsApp and USSD channels will be recruited for this study.

Sample Size Calculations

The available sample size is bounded by the estimated number of COVID-19 Connect users over the intervention period. Based on existing usage data, in the month of January 2021, 125,984 HigherHealth users accessed COVID-19 Connect on USSD or WhatsApp. Assuming a 2-4-week intervention period, this represents an estimated upper bound of the available sample size.

For this study, given the relatively large available sample size, low cost of enrolling the additional user, and low/no cost of data collection, we aim to deploy the interventions to the maximum available pool of users over the intervention period. This will enable this experiment to be well-powered, or sensitive enough, to detect very small minimum detectable effect sizes (MDES) of about 1.5 percentage points (pp) in binary outcomes between individual treatment arms versus control, with 80 percent chance of statistical significance. Though these effect sizes are relatively small in relative terms, when scaled to the population of HealthCheck users, they represent meaningful changes of policy relevance or public health significance. For example, if treatment arms successfully increases truthful reporting of moderate risk symptoms, the effect size of 1.5 pp translates to 225 students in the sample abstaining from campus due to being symptomatic; when scaled to the monthly population of HigherHealth users, that constitutes 3,000 students.

Therefore this study will aim to recruit a minimum of 5,000 users per intervention arm, and with 4 treatment arms and 1 control arm, will have a total minimum sample size of 25,000 users. Note that these numbers represent minimum sample sizes, not final. If this minimum is reached before the planned end of the study (2-4 weeks), we will continue to recruit additional users until the “end” date of the intervention.

Randomization Procedure

Randomization is performed at the level of the individual user, where a “unique user” is defined by a combination of their phone number and the channel they are using to access the COVID-19 Connect platform (USSD, WhatsApp). Randomization is done automatically whenever a user opts into HealthCheck - they will be assigned to one of the treatment arms and see the corresponding message. We stop the intervention once we have randomized enough individuals based on desired sample size indicated above. We will stratify treatment by source (USSD, WhatsApp) and province (9 provinces), with a total of 18 strata.³

³ Province has relatively stable proportions of users over time, enabling us to use past HealthCheck data (from January and February of 2021) to get strata proportions. Comparatively, the distribution of other socio-demographic variables vary month to month.

Data Collection Activities

No primary data will be collected for this study. All outcomes of interest will be measured through back-end user-data collected on the COVID-19 Connect platform.

Analytical Framework

Outcomes

Table 1: Outcomes for Study

| Outcome | # | Metric | Unit of analysis | Type | Primary/Secondary |
|-------------------|---|--|-----------------------|------------|-------------------|
| User behavior | 1 | User completed full set of HealthCheck symptom questions | HealthCheck | Binary | Primary |
| | 2 | Number of days user avoided visiting campus in a week, where proxy for avoidance =1 if healthcheck is either moderate risk, high-risk, or incomplete | User-week | Continuous | Primary |
| Symptoms reported | 3 | HealthCheck resulted in a “medium” or “high” risk categorization compared to “low risk” | Completed HealthCheck | Binary | Secondary |
| | 4 | Number of symptoms reported | Completed HealthCheck | Continuous | Secondary |
| | 5 | Type of symptom reported | Completed HealthCheck | Binary | Secondary |
| | 6 | Time taken to complete HealthCheck (in minutes) | Completed HealthCheck | Continuous | Secondary |
| Usage pattern | 7 | User completed HealthCheck more than once per calendar day | User-day | Binary | Secondary |
| | 8 | User completed HealthCheck more than once per calendar day and moved from a higher risk category to a lower one | User-day | Binary | Secondary |

| | | | | | |
|--|--|---|--|--|--|
| | | (Moderate > Low, High > Moderate, High > Low) | | | |
|--|--|---|--|--|--|

Analytical model

Outcomes of interest will be measured through back-end user-data collected on the COVID-19-Connect platform. Since users may complete multiple HealthChecks over the intervention period, the analysis of primary outcomes will be done at the HealthCheck level. However, 2 of the secondary outcomes (5,6) will be analysed at the user-day level. The following linear regression models will be estimated by Ordinary Least Squares (OLS). Note for binary outcomes, we will estimate treatment effects using a Linear Probability Model (LPM).

Specification 1. Estimating treatment effect - pooled and separate arms (HealthCheck level)

Pooled specification

$$Y_{ih} = \alpha + \beta_1 T_i + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{ih}$$

Where:

- Y_{ih} is the outcome for individual i completing healthcheck h
- T_i is the treatment variable (binary indicator for whether individual i received any of the treatment messaging arms). β_1 is the coefficient representing the causal effect of interest.
- X_i is a vector of individual-level covariates including age, designation (student, staff member, visitor), type of educational institution (campus, office, other), and educational institution
- γ_s is a vector of strata fixed effects (source, province)
- π_t are calendar week fixed effects to control for time trends
- λ_d are day of week fixed effects
- δ_c are institution/campus fixed effects
- ϵ_{ih} is the health-check level error term, clustered at the individual level.

Separate arms:

To estimate the effect of each treatment arm separately, the same equation is used above but the single treatment effect will be replaced with a series of separate effects:

$$Y_{ih} = \alpha + \sum_j \beta_j T_{ij} + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{ih}$$

Where:

- T_{ij} is an indicator for assignment to treatment arm j . β_j is the impact of treatment arm j relative to control.
- Note, p-values estimated in this model correspond to the hypothesis that each of the β_j s, considered individually, equal zero.

We will use the coefficients of the treatment indicators to measure the effect of each treatment arm relative to the control, and to measure the effect of treatment arms relative to each other. We will use F-statistics to test two omnibus null hypotheses: that all treatment effects (relative to control) are zero and that all treatment effects are the same (that is, that no treatment arm has an effect relative to any other treatment arm, though all may differ from the control).

Note these regressions will be run on the relevant sample specified for outcomes 1-5 in Table 1.

Specification 2. Estimating treatment effect - pooled and separate arms (user-day level)

Pooled specification

$$Y_{id} = \alpha + \beta_1 T_i + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{id}$$

Where:

- Y_{id} is the outcome for individual i on intervention-day d
- T_i is the treatment variable (binary indicator for whether individual i received any of the treatment messaging arms). β_1 is the coefficient representing the causal effect of interest.
- X_i is a vector of individual-level covariates including age, designation (student, staff member, visitor), type of educational institution (campus, office, other), and educational institution
- γ_s is a vector of strata fixed effects (source, province)
- π_t are calendar week fixed effects to control for time trends
- λ_d are day of week fixed effects
- δ_c are institution/campus fixed effects
- ϵ_{id} is the individual level error term, clustered at the day level.

Separate arms:

To estimate the effect of each treatment arm separately, the same equation is used above but the single treatment effect will be replaced with a series of separate effects:

$$Y_{id} = \alpha + \sum_j \beta_j T_{ij} + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{id}$$

Where:

- T_{ij} is an indicator for assignment to treatment arm j . β_j is the impact of treatment arm j relative to control.
- Note, p-values estimated in this model correspond to the hypothesis that each of the β_j s, considered individually, equal zero.

We will use the coefficients of the treatment indicators to measure the effect of each treatment arm relative to the control, and to measure the effect of treatment arms relative to each other. We will use F-statistics to test two omnibus null hypotheses: that all treatment effects (relative to control) are zero and that all treatment effects are the same (that is, that no treatment arm has an effect relative to any other treatment arm, though all may differ from the control).

Specification 3. Estimating treatment effect - pooled and separate arms (user-week level)

Pooled specification

$$Y_{iw} = \alpha + \beta_1 T_i + \theta X_i + \gamma_s + \pi_t + \delta_c + \epsilon_{iw}$$

Where:

- Y_{iw} is the outcome for individual i on intervention-week w
- T_i is the treatment variable (binary indicator for whether individual i received any of the treatment messaging arms). β_1 is the coefficient representing the causal effect of interest.
- X_i is a vector of individual-level covariates including age, designation (student, staff member, visitor), type of educational institution (campus, office, other), and educational institution
- γ_s is a vector of strata fixed effects (source, province)
- π_t are calendar week fixed effects to control for time trends
- δ_c are institution/campus fixed effects

- ϵ_{iw} is the individual level error term, clustered at the week level.

Separate arms:

To estimate the effect of each treatment arm separately, the same equation is used above but the single treatment effect will be replaced with a series of separate effects:

$$Y_{id} = \alpha + \sum_j \beta_j T_{ij} + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{id}$$

Where:

- T_{ij} is an indicator for assignment to treatment arm j. β_j is the impact of treatment arm j relative to control.
- Note, p-values estimated in this model correspond to the hypothesis that each of the β_j s, considered individually, equal zero.

We will use the coefficients of the treatment indicators to measure the effect of each treatment arm relative to the control, and to measure the effect of treatment arms relative to each other. We will use F-statistics to test two omnibus null hypotheses: that all treatment effects (relative to control) are zero and that all treatment effects are the same (that is, that no treatment arm has an effect relative to any other treatment arm, though all may differ from the control).

Specification 4. Effects over time (HealthCheck level)

We will assess treatment effects over time to see whether the impact of honesty messaging changes by length of exposure. The following pooled specification will be estimated for all primary outcomes:

$$Y_{ih} = \alpha + \sum_{t=0} \beta_w Week_w * T_i + \beta_5 T_i + \sum_{t=0} Week_w + \theta X_i + \gamma_s + \pi_t + \delta_d + \epsilon_{ih}$$

(User-week level)

$$Y_{iw} = \alpha + \sum_{t=0} \beta_w Week_w * T_i + \beta_5 T_i + \sum_{t=0} Week_w + \theta X_i + \gamma_s + \pi_t + \delta_d + \epsilon_{iw}$$

Where:

- $Week_t$ is an indicator referring to the person-week individual i was part of the study (not calendar week); starting from Week 0 to Week T where T is the last person-week for individual i in the study. $\beta_w + \beta_5$ is the impact of the pooled treatment in Week w .

We will also estimate effects over time for each separate treatment group relative to control.

Subgroup Analysis

We will consider heterogeneous effects of treatments on primary outcomes on key groups:

- Source (WhatsApp vs. USSD)
- Age categories (18 - 39 vs. 40 and older)
- Designation (student vs. staff member/visitor/other)

WhatsApp vs. USSD

Pooled specification

$$Y_{ih} = \alpha + \beta_1 T_i + \beta_2 WhatsApp_i + \beta_3 T_i * WhatsApp_i + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{ih}$$

Where:

- Y_{ih} is the primary outcome for individual i on HealthCheck h
- T_i is the treatment variable (binary indicator for whether individual i received any of the treatment messaging arms)
- $WhatsApp_i$ is a binary variable for whether individual i accesses HealthCheck on Whatsapp (=1) or USSD (=0)
- β_n denotes the coefficients determined by the regression model (β_3 is the coefficient of interest)
- X_i is a vector of individual-level covariates including age, designation (student, staff member, visitor), type of educational institution (campus, office, other), and educational institution
- γ_s is a vector of strata fixed effects (source, province)
- π_t are calendar week fixed effects to control for time trends
- λ_d are day of week fixed effects
- δ_c are institution/campus fixed effects
- ϵ_{ih} is the individual level error term, clustered at the day level.

Age Categories

Pooled specification

$$Y_{ih} = \alpha + \beta_1 T_i + \beta_2 Age_i + \beta_3 T_i * Age_i + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{ih}$$

Where:

- Y_{ih} is the primary outcome for individual i on HealthCheck h
- T_i is the treatment variable (binary indicator for whether individual i received any of the treatment messaging arms)
- Age_i is a binary variable for whether individual i is over the age of 39
- β_n denotes the coefficients determined by the regression model (β_3 is the coefficient of interest)
- X_i is a vector of individual-level covariates including age, designation (student, staff member, visitor), type of educational institution (campus, office, other), and educational institution
- γ_s is a vector of strata fixed effects (source, province)
- π_t are calendar week fixed effects to control for time trends
- λ_d are day of week fixed effects
- δ_c are institution/campus fixed effects
- ϵ_{ih} is the individual level error term, clustered at the day level.

Student vs Staff/Visitor/Other

Pooled specification

$$Y_{ih} = \alpha + \beta_1 T_i + \beta_2 Student_i + \beta_3 T_i * Student_i + \theta X_i + \gamma_s + \pi_t + \lambda_d + \delta_c + \epsilon_{ih}$$

Where:

- Y_{ih} is the primary outcome for individual i on HealthCheck h

- T_i is the treatment variable (binary indicator for whether individual i received any of the treatment messaging arms)
- $Student_i$ is a binary variable for whether individual i is a student (=1) or a staff members/visitor/other (=0)
- β_n denotes the coefficients determined by the regression model (β_3 is the coefficient of interest)
- X_i is a vector of individual-level covariates including age, designation (student, staff member, visitor), type of educational institution (campus, office, other), and educational institution
- γ_s is a vector of strata fixed effects (source, province)
- π_t are calendar week fixed effects to control for time trends
- λ_d are day of week fixed effects
- δ_c are institution/campus fixed effects
- ϵ_{ih} is the individual level error term, clustered at the day level.

Subgroup analysis will also be done for the *separate arms* specification in addition to pooled.

Missing Values

Missing values take the form of incomplete HealthChecks. As specified above, we will first assess the effect of treatment on a binary outcome of whether users complete the full set of HealthCheck questions. Then, among the completed HealthChecks for which there is a complete set of symptom and risk categorization data (thereby allowing us to construct the specified outcomes), we will examine the effects of treatment on the distribution of symptoms and probability of getting a medium or high risk categorization. Therefore there is minimal risk of having missing data for any of our outcomes or covariates.

Multiple Hypothesis Adjustments

Given that subgroup analyses and multiple outcomes per outcome category increase the number of hypotheses being tested, standard statistical significance levels would likely result in finding significant outcomes by chance alone (i.e. false positives). To correct for this, analysis will be adjusted for multiple inferences. The correction will impose a more conservative threshold for statistical significance. Normal, uncorrected p-values will also be reported. Multiple hypothesis adjustments will focus on primary outcomes specified, as these are the outcomes that potential scale-up decisions will be based on. Secondary outcomes are exploratory. We adjust for multiple hypotheses by applying a false discovery

rate (FDR) adjustment following Benjamini et al. (2006)⁴. Under this approach, in expectation, less than 5% of null hypotheses will be incorrectly rejected. Table 3 below summarizes the outcomes described and indicates the number of hypotheses being tested within each “family” of primary outcomes for each treatment group vs. control comparison.

Table 2 - Number of Hypothesis Tested for Each Family of Primary Outcomes

| Outcome category | # of outcome indicators | # of group comparisons in sub-group analysis | Total hypotheses tested within a "family" of outcomes |
|-----------------------------|-------------------------|--|---|
| Symptoms Reported (primary) | 1 | 4 | 4 |

⁴ Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika* 93, no. 3 (2006): 491-507.

References

- Abroms L.C. & Boal A.L. & Simmens S.J. & Mendel J.A. & Windsor R.A. (2014) *"A randomized trial of Text2Quit: a text messaging program for smoking cessation"*. Am. J. Prev. Med. 2014;47(3):242–250.
- Akhu-Zaheya L.M. & Shiyab W.Y. (2017) *"The effect of short message system (SMS) reminder on adherence to a healthy diet, medication, and cessation of smoking among adult patients with cardiovascular diseases"*. Int J Med Inform. Feb;98:65-75. doi: 10.1016/j.ijmedinf.2016.12.003. Epub 2016 Dec 8. PMID: 28034414.
- Armitage, C.J. (2004). *"Evidence that implementation intentions reduce dietary fat intake: A randomized trial."* Health Psychology, 23, 319-323.
- Bing, M. N., Davison, H. K., Vitell, S. J., Ammeter, A. P., Garner, B. L., & Novicevic, M. M. (2012). *"An experimental investigation of an interactive model of academic cheating among business school students. Academy of Management Learning & Education"*, 11(1), 28–48. <https://doi.org/10.5465/amle.2010.0057>
- Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. "Adaptive linear step-up procedures that control the false discovery rate." Biometrika 93, no. 3 (2006): 491-507.
- Bokemper S, Huber G, James E, Gerber A, Omer S. (2020) *"Persuasive Messaging to Encourage COVID-19 Risk Reduction,"* Research Square; 2020. DOI: 10.21203/rs.3.rs-96261/v1.
- Boruchowicz, C., & Bóo, F. & Finamor F., & Pfeifer, R. & Pacheco, T. (2020). *"Are Behaviorally Informed Text Messages Effective in Promoting Compliance with COVID-19 Preventive Measures?: Evidence from an RCT in the City of São Paulo."*
- Brewer, N.T., & Chapman, G.B., & Gibbons, F.S., & Gerrard, M., & McCaul, K., & Weinstein, N.D. (2007). *"Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination,"* Health Psychology, 26(2), 136–145.
- Bruhn, M., & McKenzie, D. (2009). *"In pursuit of balance: Randomization in practice in development field experiments"* American economic journal: applied economics, 1(4), 200-232.
- Burton, J. & Brook, G. & McSorley, J. & Murphy, S. (2013) *"The utility of short message service (SMS) texts to remind patients at higher risk of STIs and HIV to reattend for testing: a controlled before and after study"*
- Cagala, T. & Glogowsky, U. & Rincke, J. (2019). *"Content Matters: The Effects of Commitment Requests on Truth-Telling."*

Capraro, V. & Barcelo, H. (2020). *"The effect of messaging and gender on intentions to wear a face covering to slow down COVID-19 transmission"*

https://www.researchgate.net/publication/341299724_The_effect_of_messaging_and_gender_on_intentions_to_wear_a_face_covering_to_slow_down_COVID-19_transmission

Charness, G. & Dufwenberg, M. (2006). *"Promises and Partnership"*. *Econometrica*, 74 (6), 1579–1601.

Cocosila, M. & Archer, N. & Haynes, R.B. & Y. Yuan Can (2009) *"Wireless text messaging improve adherence to preventive activities? Results of a randomised controlled trial"* *Int. J. Med. Inform.*, 78 (4) , pp. 230-238, 10.1016/j.ijmedinf.2008.07.011

Dal, S. & MacDonald T.K. & Fong G.T. & Zanna M.P. & Elton-Marshall T.E. (2006) *"Remembering the message: the use of a reminder cue to increase condom use following a safer sex intervention"* *Health Psychol.* 2006 May;25(3):438-43. doi: 10.1037/0278-6133.25.3.438. PMID: 16719617.

Day, G. (1971), *"Attitude change, media, and word of mouth"*, *Journal of Advertising Research*, 11(6), 31-40.

Deke, J., 2014. *"Using the Linear Probability Model to Estimate Impacts on Binary Outcomes in Randomized Controlled Trials"*, Evaluation Technical Assistance Brief 6. Washington, DC: Office of Adolescent Health, United States Department of Health and Human Services.

Dupas, P., & Huillery E., & Seban, J. (2018). *"Risk Information, Risk Salience, and Adolescent Sexual Behavior: Experimental Evidence from Cameroon."* *Journal of Economic Behavior and Organization* 145: 151-175

Ellingsen, T., & Johannesson, M. (2004). *"Promises, Threats and Fairness,"* *Economic Journal*, 114 (495), 397–420.

Everett, J. A. C., & Colombatto, C., & Chituc, V., & Brady, W. J., & Crockett, M. (2020). *"The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic."* <https://doi.org/10.31234/osf.io/9yqs8>

Greenwald, G. (1980), *"The Totalitarian Ego: Fabrica- tion and Revision of Personal History,"* *American Psychologist*, 35 (7), 603–618.

Godes, D. & Mayzlin, D. & Chen, Y. & Das, S. & Dellarocas, C. & Pfeiffer, B. & Liba, B. & Sen, S. & Shi, M. & Verlegh, P. (2005). *"The Firm's Management of Social Interactions,"* *Marketing Letters*, Springer, vol. 16(3), pages 415-428, December.

Gollwitzer, P.M. (1993). *"Goal achievement: The role of intentions,"* In W. Stroebe and M. Hewstone (Eds.), *European Review of Social Psychology*, 4, 141-185. Chichester, England: Wiley.

Gollwitzer, P.M. (1999). *"Implementation intentions: Strong effects of simple plans,"* American Psychologist, 54, 493-503.

Gomila, R., 2020. *"Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis."* Journal of Experimental Psychology: General, pp.1-27.

Griffin, W. & Ross, L. (1991), *"Subjective Construal, Social Inference, and Human Misunderstanding,"* in Advances in Experimental Social Psychology, Vol. 24, Mark P. Zanna, ed. New York: Academic Press, 319–59.

Josephson Institute of Ethics (2006), *"Report Card on the Ethics of American Youth,"* (October 15), (accessed August 1, 2008), [available at <http://www.josephsoninstitute.org/reportcard/>].

Kristal, A., & Whillans, A., & Bazerman, M., & Gino, F., & Shu, L. & Mazar, N. & Ariely, D. (2020) *"Signing at the beginning versus at the end does not decrease dishonesty"* Proceedings of the National Academy of Sciences Mar 2020, 117 (13) 7103-7107; DOI: 10.1073/pnas.1911695117

Leeuwen, L. & Onrust, S. & Van Den P. & Bas & Kleijnjan & Marloes & Lemmers & Lex & Engels, Rutger & Hermans, Roel. (2019). *"Cue-Reminders to Prevent Health-Risk Behaviors: A Systematic Review."* Frontiers in Public Health. 7. 10.3389/fpubh.2019.00097.

Lisa, S., & Mazar, N., & Gino, F., & Ariely, D., & Bazerman, M. (2012) *"Signing at the beginning decreases dishonesty"*, Proceedings of the National Academy of Sciences Sep 2012, 109 (38) 15197-15200; DOI: 10.1073/pnas.1209746109

Mazar, N., & Amir, O., & Ariely, D. (2008). *"The Dishonesty of Honest People: A Theory of Self-Concept Maintenance,"* Journal of Marketing Research.;45(6):633-644. doi:10.1509/jmkr.45.6.633

National Academies of Sciences, Engineering, and Medicine. (2020). *"Encouraging Adoption of Protective Behaviors to Mitigate the Spread of COVID-19: Strategies for Behavior Change"*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25881>.

Nicole M. & Freiling I. & Beets B., & Brossard D. (2020) *"Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19"*, Journal of Risk Research, 23:7-8, 1052-1059, DOI: 10.1080/13669877.2020.1756385

O'Dell, B.L., Rosser, B.R.S., Miner, M.H., & Jacob, S.M. (2008). *"HIV prevention altruism and sexual risk behavior in HIV-positive men who have sex with men,"* AIDS Behavior 12, 713–720.

Prestwich, A. & Perugini, M. & Hurling, R. (2009). *"Can the effects of implementation intentions on exercise be enhanced using text messages?"* Psychology & health. 24. 677-87. 10.1080/08870440802040715.

Prestwich, A.J. & Conner, M.T. & Lawton, R.J. & Bailey, W. & Litman, J. & Molyneaux, V. (2005). *“Individual and collaborative implementation intentions and the promotion of breast self-examination,”* Psychology & Health, 20, 743-760.

Price, L., & Feick, L. (1984), *“The Role of Interpersonal Sources in External Search: An Informational Perspective”*, in Advances in Consumer Research, Vol. 11, Kinnear, Thomas C. , ed. Ann Arbor, MI: Association for Consumer Research. 250–53.

Prima, R. & Feeny, S. & Hoffmann, R. & Satriawan, E. (2020) *“Threats, Pledges, and Asset Misreporting: A Framed Field Experiment In Indonesia”*.

Sanitioso, R., & Kunda, Z., & Fong, G. (1990), *“Motivated Recruitment of Autobiographical Memories,”* Journal of Personality and Social Psychology, 59 (2), 229–41.

Sheeran P, & Harris P.R., & Epton T. (2014) *“Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies.”* Psychol Bull. Mar;140(2):511-43. doi: 10.1037/a0033065. Epub 2013 Jun 3. PMID: 23731175.

Stephenson M.T., & Witte K. (1998) *“Fear, threat, and perceptions of efficacy from frightening skin cancer messages,”* Public Health Rev. ;26(2):147-74. PMID: 10327830.

Williams, S.N., & Armitage, C.J., & Tampe, T. & Dienes, K., (2020). *“Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: A UK-based focus group study.”*

Wilson, R.F. & Sharma A.J. & Schluechtermann S. (2020) *“Factors Influencing Risk for COVID-19 Exposure Among Young Adults Aged 18–23 Years — Winnebago County, Wisconsin, March–July 2020.”* MMWR Morb Mortal Wkly Rep 2020;69:1497–1502. DOI: <http://dx.doi.org/10.15585/mmwr.mm6941e2>external icon.

Appendix

Figure 2: COVID-19 Connect Theory of Change

