

# STUDY PROTOCOL AND STATISTICAL ANALYSIS PLAN

## **The Doctors for Coronavirus Prevention Project Thanksgiving / Christmas Messaging Campaign**

NCT Number: 04644328  
Unique Protocol ID: 2003000118A010

Note: This document contains the pre-specified written description of the clinical study and statistical considerations for analyzing the data in the study. Please note that the study was titled differently in this document (“COVID-19 Thanksgiving Messaging at Scale”) as it was unclear for a period whether resources for the Christmas experiment would be available.

# Pre-Analysis Plan

## COVID-19 Thanksgiving Messaging at Scale

This Draft: November 25, 2020

### 1. Introduction

In prior work, we have found that video messages delivered by doctors increase knowledge of COVID-19, and in some cases, lead to viewers taking actions consistent with the messaging. In a scale-up phase, we plan to use Facebook ads to show a 15 second video clip recorded by 6 MGH, Harvard and Lynn Community health center doctors to approximately 40,000,000 Facebook users. The ads will be shown between November 14-29 and focus on staying safe – limiting travel, social distancing and mask-wearing. We are randomizing exposure to the ad campaign at the ZIP code and county level to study whether the videos change mobility and Thanksgiving holiday travel and whether there are any detectable impacts on the spread of COVID-19. Using information about Facebook social network connections and mobility patterns between counties, we also plan to measure whether exposure to the ad campaign spilled over along the network.

### 2. Sample Frame and Treatment

#### 2.1 Treatment

Facebook users will be exposed to an ad campaign featuring a set of videos about Thanksgiving travel. 6 doctors from MGH, Harvard and Lynn Community health center have each recorded one short message (15 seconds) using an identical script: “This Thanksgiving, the best way to show your love is to stay home. If you do visit, wear a mask at all times. I’m Dr. XX from XX, and I’m urging you: don’t risk spreading COVID. Stay safe, stay home.”

The video messages will be posted to a project Facebook page titled "The Doctors for Coronavirus Prevention Project". Users will see the ads and the name of the Facebook group associated with the ads on their feeds. Watching the videos in the ad posts is completely optional for Facebook users.

Within treated geographies, Facebook’s algorithm will be used to place the ads to users 18 and older. We expect each treated individual to see the ads approximately 3 times over the course of the ad campaign.

## 2.2 Sample Frame

We rolled out the campaign in 2 phases:

- Phase 1:
  - 13 states were chosen based on the availability of zip-code level COVID-19 data. We used a 2-stage randomization with both county and zip-code level variation.
  - 820 counties
  - 6998 zip codes
  - Campaign duration: November 14-29
- Phase 2:
  - Facebook was interested in scaling up the campaign to 15 additional states with high COVID-19 case rates and relatively low levels of compliance with preventative behaviors.
  - 1069 counties
  - Campaign duration: November 23-29

We expect that approximately 40,000,000 Facebook users will be exposed to the videos across the two phases.

## 2.3 Treatment Assignment

The experimental design differed across phases.

- Phase 1: 2-stage randomization with both county and zip-code level variation
  - 50% of counties in the 13 states assigned to Low Intensity treatment
    - \* 25% of zip codes in the low intensity counties treated (exposed to the campaign)
  - 50% of counties in the 13 states assigned to High Intensity treatment
    - \* 75% of zip codes in the high intensity treated (exposed to the campaign)
- Phase 2: 1-stage randomization at the county-level.
  - 80% of counties the 15 target states treated (exposed to the campaign)

Treatment intensities were decided based on Facebook's ad budget constraints.

In the Zip-code level treatments, ad targeting may not be perfect. If individuals do not share their location data with Facebook, then IP address is used to proxy for geographic location.

### 3. Hypotheses

Main hypotheses:

1. Do the ads change the behaviors of those in the treated zip areas?
  - Do they travel less?
    - Within zip
    - To places outside of zip? Do they travel shorter distances?
  - Do they host fewer visitors?
  - Do they wear masks more?
2. Do the ads change the type of interactions that happen?
  - Super-spreading interaction (high degree)
  - Sneaky interaction (low-degree)
3. Do the ads affect COVID rates and symptoms in treated areas?
4. Do the ads spill over onto other areas that are connected to the treatment ZIPs?
  - Behavioral spillovers: travel/mobility
  - COVID impacts
  - Are spillovers more pronounced on the mobility network or on the Facebook friendship network?

### 4. Empirical Analysis

#### 4.1 Outcomes

1. **Mobility and Travel:** different outcomes are observable from Facebook at the county and zip code level due to data privacy concerns:
  - **Zip code-level:**
    - (a) **Fraction of individuals staying in the same geographical location (1mi x 1mi area) (measured daily).** An individual is recorded as staying home if they are in the same (1mi x 1mi) area at the beginning and the end of an 8 hour interval. The measure misses people who leave their neighborhood and return within the interval.
    - (b) **Fraction of individuals leaving the geographical location (1mi x 1mi area) each day (measured daily)**
  - **County-level:**
    - (a) **“Stay put” metric: how much of the time are people really staying in their homes? (measured daily)** This measure is much more precise than what is available at the zip code level. It reflects the fraction of the day that is spent in a single 600m x 600m geographical area. [Source](#).

- (b) **“Change in movement” metric:** “We quantify how much people move around by counting the number of level-16 Bing tiles (600m x 600m) they are seen in within a day. People seen in more tiles are probably moving around more, while people seen in fewer are probably moving around less. Each day we take all the eligible people in a given region and compute the number of distinct tiles they were seen in.” (measured daily). [Source](#).
- (c) Colocation data: Data marks counts of people from county pairs (A,B) who are observed 5 minutes in the same place. It counts this as a share relative to all possible meetings. (Measured weekly). This gives a county-by-county transition matrix that is symmetric. (We do not observe where the meeting takes place).
- Timing: We are interested in mobility outcomes for Thanksgiving week. Baseline observations are based on Facebook data availability. When possible, we will include mobility measures from other holidays in 2020.

## 2. Covid 19:

### (a) Case rates/hospitalization/deaths:

- **Zip: available from local public health departments for the 13 states in the Phase I experiment.**
- County: available from local public health departments, nation-wide. Available for download from the New York Times.

### (b) Symptoms:

- Facebook COVID-19 symptoms survey. [Source](#).
- Timing: Daily data during the 1 month following Thanksgiving. We began collecting baseline COVID data at the zip code level on November 11.

## 3. Intended behaviors and beliefs:

- The data all come from Facebook’s “Brand Lift” tool. In each zip code of the Phase 1 study, a hold-out sample is randomly selected in target zips to not receive the ad. We can compare treated and control responses within zip.
- Questions:
  - (a) Ad recall
  - (b) Mask intentions
  - (c) Thanksgiving travel intentions
  - (d) How much should people try to stay home?

Our primary outcomes are in bold. We are most likely to be able to detect changes on mobility given the study design. Therefore, mobility is our primary outcome. We are also extremely interested in impacts on Covid-19 rates and symptoms. However, detecting impacts on Covid-19 is significantly more demanding from a statistical power perspective given that it is relatively rare at the weekly time horizon. We view the intended behaviors and beliefs outcomes as secondary.

Additional outcomes, data pending:

1. Mobility data from Safegraph (zip code level).
2. MGH COVID-19 symptom tracking app data.
3. Stay put, Change in movement, co-location at the zip code level.
4. Dynata survey data on plans to travel for Thanksgiving, sampled from Nov. 13-24

## 4.2 Regression Analysis

Direct Impacts (Phase 1):

1. Analysis using Zip-level randomization (Covid cases and mobility outcomes):

(a) Direct impacts (pooled):  $Y_{zt} = \alpha_t + \gamma_c + \beta Treat_{zt} + \delta Y_{z0} + \varepsilon_{zt}$

- $Y_{z0}$  is a baseline measure of the outcome.
- $\gamma_c$  denotes either:
  - Indicator for High Intensity county
  - County fixed effects. For counties with homogenous treatment assignment across zips, pool them into a larger “county unit” to be able to add FEs without losing observations.

(b) Direct impacts (disaggregated):  $Y_{zt} = \alpha_t + \beta_1 Treat_{zt} + \beta_2 1(HighIntensity)_{ct} + \beta_3 Treat_{zt} 1(HighIntensity)_{ct} + \delta Y_{z0} + \varepsilon_{zt}$

- $\beta_2$  is a measure of spillovers, related to the geographic network.

(c) Distributional direct impacts: Quantile regression.

- Repeat the exercises, above, with quantile regressions. Motivation, right tail (super-spreader) and left tail (sneaky) interactions.

2. Analysis using County-level randomization (Covid symptoms and mobility outcomes):

(a) Direct impacts:  $Y_{ct} = \alpha_t + \beta 1(HighIntensity)_{ct} + \delta Y_{c0} + \varepsilon_{zt}$

(b) Distributional direct impacts: Quantile regression.

3. Analysis using Individual-level randomization (“Brand Lift” outcomes)

- We will obtain from FB: Control mean, Treatment mean, number of observations. With this we will need to construct standard errors.
- 2 methods for inference, using binary outcomes:
  - Construct t-tests on equality of treatment and control
  - Conduct permutation tests
- We will also be able to test if the individual response is different in Urban versus Rural. FB will give us means in T and C and number of observations separately for Urban vs. rural.

Direct Impacts (Phase 2):

4. Analysis using County-level randomization (Covid symptoms and cases, mobility outcomes):

(a) Direct impacts:  $Y_{ct} = \alpha_t + \alpha_s + \beta 1(Treated)_{ct} + \delta Y_{c0} + \varepsilon_{zt}$

- $\alpha_s$  are state fixed effects due to stratification.

(b) Distributional direct impacts: Quantile regression.

Indirect impacts (Phase 1 and Phase 2 variation):

5. Indirect zip code-level impacts

- Sample-frame = Phase 1 states, but spillovers can come from both phase 1 and phase 2 treatments.
- We only have detailed facebook network information (SCI) at the zip code level (no colocation data).

(a)  $Y_{zt} = \alpha + \beta_{SCI,T} ExpPhase1Treat_z^{SCI} + \beta_{SCI,HI} ExpPhase1HighInt_z^{SCI} + \delta_{SCI} ExpPhase2Treat_z^{SCI} + \gamma_{SCI,1} ExpPhase1Sample_z^{SCI} + \gamma_{SCI,2} ExpPhase2Sample_z^{SCI} + \lambda Y_0 \varepsilon_{zt}$

• Notation:

- SCI is the Facebook friend network, Mobility is the Facebook colocation mobility network.
- *Phase1* denotes the first treatment phase. *Phase2* denotes the second treatment phase.
- *Treat* denotes whether a connected zip is treated (for Phase 1 and Phase 2).
- *HighInt* denotes whether a connected zip is in a High Intensity county.
- All non-phase 1 and phase 2 zips are marked non-treated
- We will measure exposure (*Exp*) in 2 different ways:
  - (a) Share of network connections exposed
  - (b) Number of network connections exposed

6. Indirect county-level impacts:

- Sample-frame = All states (Phase 1 and Phase 2 states, and non-intervened states), but need to control for relevant exposure

(a)  $Y_{ct} = \alpha + \beta_{SCI} ExpPhase1Treat_c^{SCI} + \beta_{Mob} ExpPhase1Treat_c^{Mobility} + \delta_{SCI} ExpPhase2Treat_c^{SCI} + \delta_{Mob} ExpPhase2Treat_c^{Mobility} + \gamma_{SCI,1} ExpPhase1Sample_c^{SCI} + \gamma_{Mob,1} ExpPhase1Sample_c^{Mobility} + \gamma_{SCI,2} ExpPhase2Sample_c^{SCI} + \gamma_{Mob,2} ExpPhase2Sample_c^{Mobility} + \phi_{SCI} Degree_c^{SCI} + \phi_{Mob} Degree_c^{Mob} + \lambda Y_0 + \varepsilon_t$

(b)  $Y_{ct} = \alpha + \beta_{SCI} ExpPhase1Treat_c^{SCI} + \delta_{SCI} ExpPhase2Treat_c^{SCI} + \gamma_{SCI,1} ExpPhase1Sample_c^{SCI} + \gamma_{SCI,2} ExpPhase2Sample_c^{SCI} + \phi_{SCI} Degree_c^{SCI} + \lambda Y_0 + \varepsilon_t$

(c)  $Y_{ct} = \alpha + \beta_{Mob} ExpPhase1Treat_c^{Mobility} + \delta_{Mob} ExpPhase2Treat_c^{Mobility} + \gamma_{Mob,1} ExpPhase1Sample_c^{Mobility} + \gamma_{Mob,2} ExpPhase2Sample_c^{Mobility} + \phi_{Mob} Degree_c^{Mob} + \lambda Y_0 + \varepsilon_t$

- Notation:
  - SCI is the Facebook friend network, Mobility is the Facebook colocation mobility network.
  - Phase1 denotes the first treatment phase. Phase2 denotes the second treatment phase.
  - Treat denotes whether a connected county is either High Intensity (for Phase 1) or treated (for phase 2)
  - We will measure exposure in 2 different ways:
    - \* Share of network connections exposed
    - \* Number of network connections exposed

#### Supplemental Analysis:

7. Indirect impacts: (“Link-level” Analysis), using nation-wide county co-location data: Construct county-pair level analog to the Indirect county-level regressions, above.
8. Test for presence of any spillovers on the network: [reference](#).
9. Analysis using County-level non-random variation
  - Goal, use untreated states to generate a “pure control”
    - Can be used to measure impacts of the whole campaign.
  - Differences-in-differences regression
    - Not in sample frame
    - Low intensity treatment
    - High intensity treatment
  - Use standard diff in diff setup (log(cases) on county FE + time FE + treatment \* FE.
    - Additionally use Synthetic Diff-in-Diff approach: [reference](#).
    - Also Consider non-linear Diff-in-Diff to capture extreme covid growth rates: [reference](#).

### 4.3 Controls

For all regressions, we will use Machine Learning techniques to select the controls. Potential controls include: Baseline covid rates, Baseline mobility rates, State fixed effects, zip and county-level demographics from the census, distance to major city, distance to state capitol,...

### 4.4 Inference

For our core analyses at both the county and zip level, we will use heteroskedasticity-robust standard errors, as our randomization is at those levels. We will check for additional statistical robustness following the approaches outlined in Section 5 of the Supplementary Appendix of [this paper](#).



## 4.5 Heterogeneous Effects

We will explore heterogeneity on the following dimensions:

- Urban vs. rural ZIPs: (this is what we stratified on for the brand lift study.) Caveat: ad targeting is running differently in urban vs. rural, so the “first stage” is likely also different.
- Travel restrictions and local government messaging: are PSAs and policy substitutes or complements? (If substitutes, we get power on the other stuff by dummifying them out.)
- COVID rates at origin and destination. Does the campaign discourage travel to/from high covid places?
- Prior infection rates (e.g. infections in April?)
- Red/Blue counties and zips
- Weather: rain or cold weather on Thanksgiving (or weekend more generally)

## 5. Funding and Human Subjects Review

Facebook is supplying sufficient ad credits to fund the campaign. The IRB at MIT is serving as the primary institution of record and has entered into a reliance agreement with Harvard, Massachusetts General Hospital, and Yale.

	<b>Massachusetts Institute of Technology</b> Committee on the Use of Humans as Experimental Subjects COUHES	<b>COUHES Protocol #</b> 2003000118  Updated December 11, 2020
--	--	--

**APPLICATION FOR COMPREHENSIVE REVIEW**

*Please complete all questions and provide sufficient detail. Indicate 'N/A' if a question does not pertain to your research. An incomplete application will be rejected and returned for completion.*

**I. BASIC INFORMATION**

<b>1. Title of Study</b>	
Using Social Media to Spread Public Health Messages for COVID19	
<b>2. Principal Investigator</b>	
Name: Benjamin Olken	Building and Room #: E52-542
Title: Professor of Economics	Email: bolken@mit.edu
Department: Economics	Phone: 617-253-6833
<b>3. Funding</b>	
<i>If the research is funded by an outside sponsor, please enclose one copy of the research proposal with your application. A draft of the research proposal is acceptable.</i>	
<i>Do not leave this section blank. If your project is not funded, check No Funding in section C.</i>	
<b>A. Sponsored Project Funding:</b>	
<input type="checkbox"/> Current Proposal Grant/Proposal # _____ Sponsor _____ Title _____	
<input type="checkbox"/> Current Award Grant/Account # _____ Sponsor <u>JPAL Innovation in Governance Initiative</u>  Title <u>Using Social Media to Spread Public Health Messages for COVID19</u>	
<input checked="" type="checkbox"/> Current Award Grant/Account # <u>030890-00001</u> Sponsor <u>NSF</u> Title: <u>RAPID: Covid-19 Information Campaigns for Vulnerable populations</u>	
<b>B. Institutional Funding:</b>	
<input type="checkbox"/> Gift <input type="checkbox"/> Departmental Resources <input checked="" type="checkbox"/> Other (explain) <u>Esther Duflo MIT discretionary account, J-PAL director pilot fund, Facebook (in ind)</u>	

(For the Facebook US Doctors Messaging Scale-up, Facebook is donating ad credits to show the study video to approximately 10,000,000 users, worth about \$350,000). As of November 21, 2020, Facebook has also planned to fund a scale up beyond the initial study zip codes and will also cover all costs. As of December 10, 2020, Facebook has agreed to fund a second round of messaging around the December holidays which will support showing ads to up to 70,000,000 users (most likely fewer but this is the maximum).

C. No Funding

This protocol will not be funded

**4. Statement of Financial Interest**

A. Does the investigator, study personnel, or their Family have a financial interest in a company or other organization involved in this study?

Yes X No

B. Could the work contemplated in this project reasonably appear to affect a company or other organization in which the investigator, study personnel, or their Family have a financial interest?

Yes X No

C. Does this study contemplate:

i. Receiving or using any data (e.g., proprietary data sets, data sets, confidential information) from a company or other entity organization in which the investigator, study personnel, or their Family have a financial interest

Yes X No

ii. Receiving or using any materials (e.g., drugs, devices, biological agents, investigational medical devices) from a company or other entity organization in which the investigator, study personnel, or their Family have a financial interest

Yes X No

iii. Granting subawards to a company or other entity organization in which the investigator, study personnel, or their Family have a financial interest

Yes X No

iv. Making purchases from a company or other entity organization in which the investigator, study personnel, or their Family have a financial interest

Yes X No

If 'yes' was checked for any of the questions above, then attach a **Supplement for Disclosure of Financial Interest** for each individual with an interest. This supplement and detailed guidance are available on the COUHES website under Policies & Procedures in the [Financial Conflicts of Interest](#) section.

**5. Anticipated Dates of Research**

Start Date: 3/16/2020	Completion Date: 12/31/2024
<b>6. Collaborating Institutions</b> <i>If you are collaborating with another institution(s), then you must obtain approval from that institution's institutional review board (IRB) and forward the approval to COUHES.</i>	
Harvard University Yale University Stanford University Massachusetts General Hospital (MIT will be prime IRB, and we will request these other institutions cede to MIT)	
<b>7. Location of Research</b> <i>If on the MIT campus, indicate where on campus. If you plan to use the facilities of the Clinical Research Center you will need to obtain approval of the MIT Clinical Research Center.</i>	
Location of the research will depend on the country site. All US research will take place online. Moreover, any time we use a social media platform such as Twitter, YouTube, or Instagram, the research will all be online.  For the India component of the research targeting rural areas or low income urban neighborhoods, the research will be done all by phone. This is because in much of India, the main social media platform is WhatsApp, which does not have a centralized platform. Instead individuals form size-limited groups and forward information to group members. None of this information is public. So we will use phone surveys to measure the outcomes of the research.	
<b>8. International</b> <i>Research conducted outside the United States may be subject to additional requirements.</i>	
A. Are you collecting or receiving identifiable data from subjects within the European Union (EU) and/or European Economic Area (EEA)? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No B. Is the project in, related to, or funded by a person or entity from China (including Hong Kong), Russia or Saudi Arabia? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No <i>If yes, additional review and approval is required. Please see <a href="#">Additional Review</a> for additional information.</i>	

## II. STUDY INFORMATION

<b>1. Purpose of Study</b> <i>Provide a concise statement of the background, nature and reasons for the proposed study. Use non-technical language that can be understood by non-scientists.</i>
Given the developing global coronavirus pandemic, we propose to expand our previous research in Indonesia (COUHES Protocol # 1406006433), which shows that celebrities have an important role in public health messaging, to encouraging COVID19 preventative health measures. 1) We plan to recruit a new set of social media influencers in the U.S. and other countries in a similar fashion to what we did in Indonesia, but this time with all of the messaging being about coronavirus (e.g., social distancing, hygiene). 2) We also plan to conduct surveys with a random sample of followers of the social media influencers participating in the Covid-19 awareness campaign. The purpose of this is to allow us to

understand which messages work in real time and continually improve them. We are including new recruitment scripts for both the influencers and the followers with the consent scripts.

We are also expanding the initial research idea to focus on how the medical community can best reach vulnerable populations of color through short online videos (US Doctors Videos).

2) We have teamed up with physicians from Massachusetts General Hospital (MGH) to design video messages to reach communities of color and to speak to their specific concerns.

3) Because we know that video messages cannot be too long and we do not know which types of message content reaches the target population in the most impactful, sincere and trustworthy way, we are testing many different variants of the basic message. Experimental manipulations include a) varying the racial and ethnic identity of the doctor in the video; b) varying whether the medical professional addresses the so-called “elephant in the room” pertaining to historical inequitable treatment and lack of trust in the medical community or pertaining to deportation fears when seeking healthcare; c) specifically for masks – whether we provide information from a nationally representative group of respondents on perceptions of people of color wearing masks.

Drawing on the initial round of US Doctors Videos research, we intend to launch a phase 2 project (US Doctors Round 2) to be able to answer an additional set of related questions:

4) Do the impacts of concordance look different during Summer 2020, post-George Floyd? The country has moved very quickly on some issues involving race in America. In order to be able to advise governments and public health organizations on messaging for minority communities, it is important that we re-run a component of our study in the current moment.

5) What are the impacts of racial concordance on white subjects?

6) Following the widespread protests of June 2020, many organizations have issued statements about racial injustice. The American Medical Association is one organization that has done this. We would like to understand how these types of statements affect the impact of whatever messaging follows (in our case, information about Covid-19). Does the impact of these statements vary by the race of the messenger, the race of the viewer, or the political affiliation of the viewer?

7) An article in the July 6 2020 New York Times documents that African Americans are 3 times more likely to contract Covid-19 than white Americans. How does acknowledging this fact affect the impact of messaging on Covid-19. Again, does the impact of hearing this information vary by the race of the messenger, the race of the viewer, the political affiliation of the viewer, and prior beliefs about Covid-19 incidence?

We found that the messages in both rounds of the US Doctors Videos project increased knowledge of COVID-19, and some treatment conditions also led to viewers taking more actions in response to the videos. In a scale-up phase, we plan to use Facebook ads to show a 15 second video clip recorded by MGH, Harvard and Lynn Community health center doctors (6 people in all who all recorded in previous versions) to approximately 40,000,000 Facebook users (maximum 70,000,000). The ads will be shown before the Thanksgiving holiday and will focus on staying safe – limiting travel, social distancing and mask-wearing. We will randomize exposure to the ad campaign at the ZIP code, county, or DMA level to ask:

8) Do the videos change mobility and Thanksgiving holiday travel? Do they reduce the spread of COVID-19?

9) Are there spillover impacts from the video messages? For example, if individuals decide to stay home, then do the geographical regions that tend to be visited by people from treated areas experience any effects, either through information spillovers, or through a reduction in travel?

In preliminary analysis, we are finding that the Thanksgiving round of Facebook messages decreased travel at the county level. Because of these impacts, we are planning a similar campaign in coordination with Facebook about December holiday travel. Facebook will again show 15 second video clips recorded by the partner doctors. We have also expanded the set of health-care providers recording videos to include nurses. We have also broadened the geographic footprint of the health care professionals in the videos. Again, the videos will be shown to Facebook users with the same protocols. We again are attempting to reach approximately 40,000,000 users (70,000,000 maximum) with the campaign. The research hypotheses are the same, but with December holiday travel and covid as our target outcome.

We are also expanding our research to India where the difference in social media usage requires a different research strategy:

10) We plan to recruit celebrities to record video messages about Covid19.

11) We will then distribute the messages to individuals that we have identified as information hubs in local communities. We will send the videos to these information hubs and ask them to share the messages on WhatsApp with others in the community. We will tell them not to share the messages via face-to-face interactions.

12) We plan to conduct surveys with others in the study communities to understand a) if they have seen the messages and b) measure their beliefs about Covid19 their practice of social distancing and good hygiene, and their financial situations.

13) Because of the importance of frontline health workers (FHWs) in the official, rural response, we are also planning to conduct complementary surveys in the study areas. This will help us understand the impacts of the information intervention.

14) We plan to also launch a “light-touch” intervention with the FHWs themselves. This will include setting up WhatsApp groups comprised of FHWs from other villages in the same state. We would encourage them to support one another and to share information about their own experiences. We will also study how well the groups work if we additionally: i) give information about how confusing and fast-paced the landscape is (giving a license to ask to FHWs) or ii) give information about common rumors that our household surveys uncover or that the media reports in the local areas. (Health Worker Script)

## 2. Study Plan

*This section determines if the study plan meets the Federal definition of a clinical trial. COUHES will assist with any additional requirements based on the responses below. For more information available on COUHES website for Clinical Trials: <http://couhes.mit.edu/clinical-trials-mit>*

A. Are the participants prospectively assigned to an intervention?

X Yes     No

B. Is the study designed to evaluate the effect of the intervention on the participants?

X Yes     No

C. Is the effect being evaluated a health-related biomedical or behavioral outcome?

X Yes     No

### 3. Experimental Procedures

*Provide an outline of your experimental procedures with a detailed description of your proposed study. When applicable, include copies of any questionnaires or standardized tests.*

*Do not attach or copy sections of a grant application.*

*When applicable, include a detailed description of the experimental devices or procedures, detailed information on the exact dosages of drugs or chemicals to be used, total quantity of blood samples to be used, and descriptions of any special diets.*

*Provide sufficient information for effective review by non-scientists. Define all abbreviations and use simple words. This section should not exceed 5 pages unless justification is provided for additional length.*

The following procedures are for sites where Twitter, Instagram, or YouTube are the relevant social media platforms. This pertains to the US, for example.

We will recruit influencers -- i.e. people with large social media followings -- and encourage them to share public health guidance about COVID19. The influencers have full discretion about what they post on social media, but we plan to give randomized recommendations about the content of each message to understand which messages are most effective.

All of the recommended message content will be vetted by the MD on our study team. All of the messages will be consistent with the recommendations of the CDC or equivalent public health authorities in other countries. We are attaching a set of sample instructions for the influencers consistent with this design.

The different treatments will vary the emphasis of the message in four different dimensions to understand which types of messages are most effective, so that we can encourage more of the most effective message types:

1. Target behavior: Messages will be requested to emphasize either social distancing or hygiene.
2. Rationale for the behavior: Messages will either emphasize the internality (the benefits accrue to the person doing that behavior) or the externality (the benefits accrue to others in the community)
3. Timing of the messages: We will vary the order and the variety of the content that we request people to tweet about.
4. Numbers versus narrative: We will emphasize the way the message is delivered in terms of its emphasis on the numbers versus the human narrative behind that feature.

We will also include a fifth randomized nudge:

5. Video vs. written message. We will give the influencer full discretion over this choice.

We are also planning to conduct *online* surveys with a random set of followers of each of the participating influencers. We would like to survey the followers repeatedly. This information will be the key outcomes that allow us to understand the impacts of the messages, so that we can dynamically adjust the messages to focus on those messages which appear to be most effective. The survey questions are attached.

### **US Doctors Study (Round 1)**

After receiving approval in our last IRB amendment, we quickly ran a nationally representative pilot survey to understand current knowledge gaps about covid and also to measure people's perceptions of others, especially those belonging to communities of color, when wearing masks.

After running this pilot, we would like to launch a video messaging intervention aimed at communities of color. In the past few weeks, data has shown that covid19 is disproportionately infecting and killing African Americans and Latinx people. We have teamed up with physicians from MGH to develop video messages targeted at these groups. The doctors bring their expertise care providers for a diverse patient population.

The following procedures will be followed. The research subjects will participate online.

1. Doctors from MGH who express interest in participating will record messages according to our scripts (**US Doctors Script African American, US Doctors Script Hispanic**). They will be instructed to film several different versions each.
2. We will recruit a sample of Hispanic and African American study participants from across the country through Lucid, an online survey firm that has access to a large subject pool.
3. Participants will first read a consent script and give us their informed consent (**US Doctors Messaging Consent**)
4. Participants will then navigate through the following steps
  - a. Brief demographics survey questions (US Doctors Messaging Survey)
  - b. Videos 1 and 2: Introduction, Elephant in the Room, Social Distancing (**US Doctors Script African American, US Doctors Script Hispanic**). Individuals in the control group will only see the introduction portion of the videos.
  - c. Beliefs survey questions (US Doctors Messaging Survey)
  - d. Video 3: Masks (**US Doctors Script African American, US Doctors Script Hispanic**). Again, individuals in the control group will not see a video here.
  - e. Main outcomes survey (**US Doctors Messaging Survey**)
  - f. Control group is shown a version of all 3 videos (**US Doctors Script African American, US Doctors Script Hispanic**).
5. The video messages each participant in one of the treatment groups sees will be randomized on the following dimensions. The different versions are clearly demarcated in the US Doctors Scripts document.
  - a. Racial or ethnic identity of the doctor delivering the messages: concordant vs. discordant identity to the subject.
  - b. Whether the message includes an acknowledgment of "elephant in the room" issues for each target group: trust in the medical system or fear of deportation.



- c. Whether the social distancing component of the message is delivered by Dr. Birx of the CDC or recorded by the MGH physicians.
- d. Whether individuals are given information about how representative individuals perceive mask wearers of color. This information comes from results of our nationally representative pilot survey.
- e. Some individuals in a control group will only see messages after all surveying has been completed.

## US Doctors Study (Round 2)

After completing the initial US Doctors Study, we learned that racial concordance is extremely important for African American viewers. We would like to run a similar video messaging study with very similar protocols. Our goal is to better understand the impacts in the context of July 2020. The past 6 weeks have seen large scale protests and rapid social change. Do the strategies that worked in early May still work now? We would also like to study new messaging variants – namely 1) what is the role of statements addressing racial justice? 2) what is the role of making racial disparities in the incidence of Covid-19 salient in the videos?

The following procedures will be used:

- 6. Doctors from MGH who express interest in participating will record several new messages according to our scripts (**US Doctors Script Round 2**).
- 7. We will recruit a sample of white and African American study participants from across the country through Lucid, an online survey firm that has access to a large subject pool. We will stratify treatments on age, education, geographical location and political leaning.
- 8. Main study session
  - a. Participants will first read a consent script and give us their informed consent (**US Doctors Messaging Consent Round 2**). Consent will pertain to the main study session and to the follow-up survey activities. The consent text will also inform individuals that if they provide their consent, they might be recontacted in the future.
  - b. Participants will then answer a short set of baseline questions
  - c. Treatment videos are shown (see below)
  - d. After each video, respondents give evaluations (see **US Doctors Survey Round 2**)
  - e. Participants answer a set of Endline questions (**US Doctors Survey Round 2**)
- 9. Several days after the main study session, Lucid will attempt to recontact all of the initial participants for follow-up activities.
- 10. Follow-up session
  - a. Follow-up survey (**US Doctors Survey Round 2**)

**Treatment variants:** All respondents (including the control) will initially view one of two messages with content developed by the AMA

- 11. Racial injustice (treatment of interest) or drug pricing (placebo)
  - a. The race of the doctor in the video message will be randomized
  - b. Some respondents will see a video of the physician speaking. Others will only hear that physician's voice.

All respondents will then be shown the main video messages

12. Control: One meaningful difference in round 2 compared to round 1 is the protocol for the control group. The control group will be shown video messages with content developed by the AMA, the Mayo Clinic or Mount Sinai Hospital on topics about health that do not directly relate to Covid-19.
  - a. The race of the doctor in the video will be randomized across participants
13. Treatment: Subjects will watch three videos about Covid-19 in short succession.
  - a. Content
    - i. Video 1: introduction to Covid-19 and its symptoms. We will use the same videos here that were recorded for the prior iteration of the study.
    - ii. Video 2: social distancing and hygiene. We are recording new videos here because now many states have reopened and the formal guidance is different.
      1. We will randomize from two different versions of this video. There will be a baseline version and a version that additionally emphasizes that Black Americans and other minority groups are three times as likely to get Covid-19 as white Americans.
    - iii. Video 3: masks. Again, we will use videos that were recorded for the prior iteration of the study.

The racial concordance between the doctor and the viewer will be randomized. The race of the doctor will be the same across all three videos.

### **US Doctors Facebook Scale-Up - November 2020**

After completing the prior 2 US doctors studies, we learned that messages improve knowledge overall, regardless of specific sub-treatment. In the Round 2 study, seeing any message, on average also improved behaviors. This evidence suggests that scaling simple messaging could be a useful tool in fighting the pandemic. We would now like to scale up a very simple message to a segment of the US in advance of the Thanksgiving holiday. We will do this through a Facebook advertising campaign. Instead of ads, individuals will see a short video message about staying safe during the Thanksgiving holiday.

The following procedures will be used:

14. Doctors from MGH, Harvard and Lynn Community health center who express interest in participating will record one short message (15 seconds). Script: “This Thanksgiving, the best way to show your love is to stay home. If you do visit, wear a mask at all times. I'm Dr. XX from XX, and I'm urging you: don't risk spreading COVID. Stay safe, stay home.”
15. The video messages will be posted to a project Facebook page (*see Facebook page\_ Doctors for Coronavirus prevention.2020.11*)
16. Within our sample frame, described below, we will randomize ZIP codes and counties into treatment and control
17. Control: no intervention. Treatment: these ZIP codes and counties will participate in the Facebook ad campaign.
18. Facebook ad campaign: Facebook will allocate ad credits across users in treatment areas. Users selected to receive the ads from this campaign may see a video up to 10 times in a 2-week period. Like for any Facebook ad, individuals can choose to whether or not to watch the video and can close the ad at any time.

19. Treatment and control status at the ZIP code or county level will be merged with aggregated and completely de-identified datasets (see below) to measure COVID symptoms, cases and mobility. The COVID data is publicly available. The mobility data is prepared by Facebook which already has in place a sharing agreement for researchers access with MIT, Harvard, Stanford and Yale.
20. We will also obtain aggregated data from four questions Facebook will pose through its Brand Lift tool. A subset of individuals seeing the ad and a small holdout sample will see one of the following four questions on their feeds. Answering the question is completely optional.
  - a. Do you recall seeing an ad from doctors about Thanksgiving?
  - b. How likely are you to travel this holiday season? [Very likely, somewhat likely, somewhat unlikely, very unlikely, I don't know]
  - c. How likely are you to wear a mask when you visit with family and friends this holiday season? [Very likely, somewhat likely, somewhat unlikely, very unlikely, I don't know]
  - d. This holiday season, do you think people should stay home if they can? [yes definitely, it depends, no they should do what they want]
21. We will not have access to any individual information: we won't know who was targeted by the ad, whether they watched it, or what they answered to the brandlift question.

### **Facebook page**

The document *Facebook page\_Doctors for Coronavirus prevention.2020.11* screenshot is a mockup of the Facebook page that will host the videos for the campaign. If users click on the ads, they will be taken to this page. The final name of the campaign will be "The Doctors for Coronavirus Protection Project." The project will also link to a project page on the JPAL North America, so that users can verify that the project is linked to a reputable research organization. The "about" section has a list of all the doctors involved in the project. We will host the December ad campaign on the same Facebook account.

### **US Doctors Facebook Scale-Up - December 2020**

Preliminary findings from the November 2020 Scale-up show that the videos did decrease travel around the Thanksgiving holiday and did have beneficial spillovers on average to Facebook connections. Because of these encouraging findings, we are planning to run a similar campaign for December holiday travel in coordination with Facebook. Again, we will use a Facebook advertising campaign. Instead of ads, individuals will see a short video message about staying safe during the holidays.

The following procedures will be used:

22. A range of doctor and nurse volunteers from across the country will record three script segments:
  - a. Everybody records: "We are nurses and doctors, and as hard as it is, we are staying home for the holidays."
  - b. Everybody picks one (or more) of the following:
    - i. I am doing this because I love you mom and dad
    - ii. I am doing this because it is the safest way to celebrate
    - iii. I am doing this because I have seen enough suffering in my hospital
    - iv. I am doing this because there is finally light at the end of the tunnel and we need to hang on just a little while longer.
  - c. Everybody records: "I am [XX] from [XXX] and I am urging you, don't spread COVID. Stay safe, stay home."

23. The video messages will be posted to the same project Facebook page as the Thanksgiving campaign.
24. Within our sample frame, described below, we will randomize ZIP codes and counties into treatment and control
25. Control: no intervention. Treatment: these ZIP codes and counties will participate in the Facebook ad campaign.
26. Facebook ad campaign: Facebook will allocate ad credits across users in treatment areas. Users selected to receive the ads from this campaign may see a video up to 10 times in a 2-week period. Like for any Facebook ad, individuals can choose to whether or not to watch the video and can close the ad at any time.
27. Treatment and control status at the ZIP code or county level will be merged with the same aggregated and completely de-identified datasets as the Thanksgiving campaign.
28. We will again obtain aggregated data from four questions Facebook will pose through its Brand Lift tool. A subset of individuals seeing the ad and a small holdout sample will see one of the following four questions on their feeds. Answering the question is completely optional.
  - a. Do you recall seeing an ad from doctors or nurses about holiday travel?
  - b. How likely are you to travel this holiday season? [Very likely, somewhat likely, somewhat unlikely, very unlikely, I don't know]
  - c. How likely are you to wear a mask when you visit with family and friends this holiday season? [Very likely, somewhat likely, somewhat unlikely, very unlikely, I don't know]
  - d. This holiday season, do you think people should stay home if they can? [yes definitely, it depends, no they should do what they want]
29. We will not have access to any individual information: we won't know who was targeted by the ad, whether they watched it, or what they answered to the brandlift question.

### **WhatsApp:**

The following procedures are for sites where WhatsApp is the relevant social media platforms (e.g., rural and low income urban India):

The messages have been approved by Marcella Alsan MD from Harvard University and by Abhijit Chowdhury, MD from Birbhum Health and Demographic Survey System, India.

30. Recruit major celebrities: Given the context, we do not need many celebrities. In fact, we only require a minimum of one celebrity, though we certainly welcome more than one.
31. Record messages: Ask each celebrity to record a "bank" of 16 messages.
  - a. See attachment (India Celebrity Guidance)
32. Treatments: We will have 3 dimensions of treatment differences:
  - a. Internality vs Externality
  - b. Hygiene vs Social Distance
  - c. Type 1 vs. Type 2 Error (Type 1 error: individuals with symptoms need not have covid, so it's okay to report. Type 2 error: individuals without any symptoms may still have covid, so it's even more important to practice hygiene and social distancing.)
33. Selection of information hub: contact individuals in each study village and ask them to name the person that they would recommend for spreading trusted information, who also has a smartphone (Initial Contact Script and Survey).
34. Intervention: in every period we

- a. Choose a set of villages to receive a(nother) message (villages will receive many messages over time.)
  - b. Seed the gossip with one of the videos in the message bank. See attachment (Seeding Script)
  - c. Make the info drop "common knowledge" among our outcome sample individuals.
35. Surveys: In each survey round, we contact a randomly chosen subset of all sampled individuals in each village and conduct the survey over the phone. See attachment (IRB India Phone Survey). Surveyors will use tablets to record answers. However, if we run out of tablets, or the tablets malfunction, they will use paper surveys.

**4. Drugs and Devices**

*If the research involves the administration of a novel drug not approved by the Food and Drug Administration (FDA) for the use outlined in the protocol, then the principal investigator (or sponsor) must obtain an Investigational New Drug (IND) approval from the FDA. If the study involves the use of an approved drug in an unapproved way, the investigator (or sponsor) must submit an application for an IND approval. If applicable, include a copy of the IND approval (new drug) or application (new use).*

*If the study involves the use of a novel medical device and the device poses significant risk to human subjects, the investigator (or sponsor) must obtain an Investigational Device Exemption (IDE) approval from the FDA.*

*COUHES may determine an IDE or IND is appropriate during review.*

A. Will drugs or biological agents requiring an IND be used?  Yes  No

*If yes, please provide details:*

B. Will an investigational medical device be used?  Yes  No

*If yes, please provide details:*

**5. Radiation**

*Research involving the use of radiation or radioactive materials may require review from MIT's Environment, Health, and Safety (EHS) Office. COUHES may determine EHS review is appropriate.*

A. Will radiation or radioactive materials be used?  Yes  No

*If yes, please provide details:*

B. Will any type of lasers be used?  Yes  No

*If yes, please provide details:*

**6. Diets**

A. Will special diets be used?  Yes  No *,If yes, please provide details:*

**III. PERSONNEL**

**Fill out the personnel list at the end of this form.**

**IV. HUMAN SUBJECTS**

**1. Subjects**

*The number of subjects must corresponded with the maximum number of subjects investigators will consent for the study.*

<p>A. Maximum number of subjects:</p> <p>Adults: 132,500 Minors:</p> <p>India: We are planning to recruit subjects in 5000 villages. This will mean 50,000 total participants in the India site, including initial contacts, seeds, and outcome survey participants.</p> <p>50,000 frontline health workers</p> <p>We would like to run a pilot online survey in the US with 1,500 individuals.</p> <p>We would like to conduct the two phases of the US Doctors messaging intervention with 31,000 individuals in total.</p> <p>(Round 1 recruited approximately 11,000 respondents. We hope to target an additional 20,000 respondents in Round 2)</p> <p><b>November 2020: US Doctors Facebook Scale-up:</b> approximately 20,000,000 people (at most 40,000,000) will be targeted by the message campaign in the first round (many fewer will pay attention to it). On November 22, Facebook asked to scale up the ads to more states with an additional 20,000,000 – 30,000,000 users. The total maximum number of people viewing the ads will be 70,000,000 (initial randomization + scale-up).</p> <p><b>December 2020: US Doctors Facebook Scale-up:</b> We are again going to target 40,000,000 users with 70,000,000 maximum seeing our videos.</p>	<p>B. Specify age range(s):</p> <p>Adults: 18+                      Minors:</p>
<p>C. Inclusion and exclusion criteria:</p> <p>i. What are the criteria for inclusion or exclusion? Twitter, YouTube, Instagram sites: Adults who are followers of influencers who join our campaign.</p>	

For our pilot survey for the MGH doctor's project, we plan to survey 1,500 individuals using a sample provided to us by Lucid. We would like to oversample Hispanics and African Americans with High School education or below. We believe that these groups face the largest information gaps while also shouldering a disproportionate burden from covid19.

For the main US Doctors messaging study (Round 1) we will recruit a sample from Lucid of African Americans and Latinos with High School education below. This is the group that we believe are hardest to reach with current public health messaging campaigns and who bear the most disproportionate impact from covid19.

For the main US Doctors messaging study (Round 2) we will try to recruit a sample from Lucid of African Americans and white Americans with below college education. We plan to achieve balance in the white sample on political ideology. Basic demographics about the sample will be shared with us by Lucid.

For the US Doctors Facebook Scale-up (November 2020 and December 2020), the sample frame is all users of Facebook. We will focus the campaign on ZIP codes and counties depending on the availability of public-release COVID-19 case information, and randomized at the zipcode/county level. We may focus on a subset of states based on public health guidance. Within regions receiving the ad campaign, we will instruct Facebook to target the ads to a representative group of Facebook users in the region. They will only target adults over the age of 18.

WhatsApp sites: WhatsApp does not allow us to view any information about users, publicly. Also, in the current climate it is not appropriate to make face-to-face contact with participants. Therefore, we need to recruit individuals in rural villages or urban neighborhoods whom we can contact through by phone through any of the following methods:

1. Contact individuals who participated in prior studies involving this study's investigators and who shared their phone numbers with us.
2. Contact beneficiaries of partner NGOs such as Pratham India who are willing to share their information with us.
3. Contact local elected politicians (ward members). The phone numbers of ward members are publicly available.
4. Ask individuals we contact through all of the above channels to give us contact information of other people they know, building out a so-called "snowball" sample.

FHW activities: In each study village, we will seek to enroll the government front-line health worker in the study.

ii. Are any inclusion or exclusion criteria based on age, gender, or race/ethnic origin?  
(Investigator must explain why and provide justification.) No.

iii. Explain the inclusion of any vulnerable population(s) (e.g. children, cognitively impaired persons, educationally disadvantaged persons, non-English speakers, MIT students) and why. COVID-19 is a global problem. We are interested in social media campaigns in the US, but we also think that the methodology and perhaps lessons from the study can immediately be deployed in other places where the spread is in an earlier stage, including Indonesia and India. We will conduct surveys in the local languages.

## 2. Subject Recruitment

*Identification and recruitment of subjects must be ethically, legally acceptable, and free of coercion. Describe below what methods will be used to identify and recruit subjects. Include copies of recruitment documents (i.e. flyers, e-mails, advertisements, etc.).*

Twitter, YouTube, Instagram sites:

We are planning to conduct *online* surveys with a random set of followers of each of the participating influencers. Lists of followers on the social media platforms we plan to study are public, and we will send direct messages on the social media platform to recruit participants. We are including a sample recruitment email as an attachment.

**MGH Doctors project (baseline survey):** We will use a sample provided to us by Lucid.

**MGH Doctors messages experiment (Rounds 1 and 2):** We will use a sample provided to us by Lucid, specified to our target demographics, above.

**US Doctors Facebook Scale-up (November and December 2020):** Facebook will show ads up to 10 times over the two weeks prior to Thanksgiving (or Christmas) directly on users' Facebook feeds. As with any ad, users will be able to turn the ad associated with this messaging campaign off at any time. Facebook will also ask "Brand lift: questions to a subset of those receiving the ads and a small hold-out sample. Again, responding is completely optional.

WhatsApp sites:

WhatsApp is not a centralized platform. Instead, individuals join groups (which cap the number of members) and then frequently share video content across these groups. Our unit of recruitment will be the village or the urban neighborhood (we often use the term village, but in urban settings this refers to neighborhood). We will identify one person of high network centrality within each village through phone surveys. Then this individual will be "seeded" with a video created under our guidance by a celebrity.

Given the WhatsApp context, we need to recruit villages, "seed" individuals, and outcome survey respondents ourselves, as there is no clear set of social media followers to study. Moreover, we are not able to do door-to-door recruitment during the current public health environment. Therefore all recruitment must be done over the phone. We propose to recruit participants in four ways:

1. Contact individuals who participated in prior studies involving this study's investigators and who shared their phone numbers with us.



2. Contact beneficiaries of partner NGOs such as Pratham India who are willing to share their information with us.
3. Contact local elected politicians (ward members). The phone numbers of ward members are publicly available.
4. Ask individuals we contact through all of the above channels to give us contact information of other people they know, building out a so-called “snowball” sample.

Once we have received any contact information for a given village we plant to recruit:

1. At least one “seed” individual who will be given the video message by text directly from a surveyor. He/she will be asked to spread the message using WhatsApp. (We have several messages instructing individuals to only share the information via Social media, not through face-to-face interactions or phone sharing. This individual will be selected by asking the phone contacts we have access to the following question: Can you please think of all of the people in this village who you know with smartphones. Which of these people would be best at informing the largest number of people about a new job opportunity, a fair or a festival, or important health information?
2. Several (between 3 and 30) “outcome survey respondents.” We will sample from all of the phone numbers we can manage to collect in the village though all of the strategies listed above.

For the FHWs, we will use publicly available phone numbers to contact them initially. We will go through a consent process before conducting any surveys or interventions.

### 3. Informed Consent

*Informed consent is required from all human subject research studies involving participants. Templates are available on the COUHES website under Forms & Templates (<https://couhes.mit.edu/forms-templates>). Under very limited circumstances, COUHES may waive the elements or requirement for informed consent. If you are requesting a waiver or alteration of consent, include the Waiver or Alteration of Informed Consent Request form.*

**Attach informed consent form(s) with this application.**

### 4. Subject Compensation

*Payment must be reasonable in relation to the time and trouble associated with participating in the study. It cannot constitute an undue inducement to participate.*

A. Describe all plans to pay subjects in cash or other form of payment:

Twitter, YouTube, Instagram sites:

None. Subjects will not be compensated in general, but we will have a lottery for one subject respondent to win an iPad. The online surveys will be short and costs of participation are minimal.

MGH doctors project (Round 1 and Round 2): Lucid will compensate participants their standard, agreed-upon rates. In the Round 2 study, we will also assess each individual’s willingness to pay for masks. Individuals will be asked their valuation for 2 reusable masks. Some individuals will randomly be chosen to receive a prize worth at most \$30, either paid as an Amazon gift card or a coupon for 2 masks from an online mask store. We will not collect any PII to execute this reward system. Individuals will be given single use coupon codes. We will not know the identity of the users.

US Doctors Facebook Scale-up (November and December 2020): There will be no compensation.

WhatsApp sites:

None. There will be no compensation to participants.

B. Will subjects be reimbursed for travel and expenses?

N/A

### **5. Potential Risks**

*A risk is a potential harm that a reasonable person would consider important in deciding whether to participate in research. Risks can be categorized as physical, psychological, sociological, economic and legal, and include pain, stress, invasion of privacy, embarrassment or exposure of sensitive or confidential data. All potential risks and discomforts must be minimized to the greatest extent possible by using e.g., appropriate monitoring, safety devices and withdrawal of a subject if there is evidence of a specific adverse event.*

A. What are the risks/discomforts associated with each intervention or procedure in the study

We believe risks to be minimal. All suggested messages will be vetted by the MD(s) on our team and in accordance with CDC guidelines.

For the US Doctors Facebook Scale-up (November and December 2020), the participation of the subjects is minimal and there are no risk to the subjects. They will have the option to watch a 15 second video, posted as a Facebook ad. Some will receive 1 question to answer on the Facebook feed. There will be no PII collected at any time. The researchers will never know which specific Facebook users were shown the video messages or who answered the Brand lift questions.

There is a risk of discomfort of watching a 15 second advertisement you disagree with; in the November 2020 round, the Facebook page received comments, some of which were negative, from about 0.04 percent of all users shown ads. Facebook reports that this is consistent with their experience on other similar campaigns (e.g. promoting flu vaccination.)

For evaluation we are primarily using aggregate data which is either publicly available or already available to MIT and other universities. The only exception is the Facebook symptoms survey, which is completely de-identified and already available to researchers at MIT.

For the India site, we are also running all of our materials by a MD based in Kolkata, India to ensure that our guidance is consistent with local practice.

In this current climate, we do not understand which types of messages are more successful at informing rural households or encouraging behaviors like social distancing and hygiene. After consulting with local experts, we believe that each message could be beneficial ex ante. For the social media and India projects, we plan to use an adaptive research design to dynamically learn about which treatment is the best. Once we learn which one is the best, we will only distribute

the best message(s). If we find that some treatment is doing particularly badly (or even might be backfiring) we will discontinue the use of that treatment. Moreover, we will give the more effective videos to all study villages once we learn what that treatment is.

B. What procedures will be in place to prevent/minimize potential risks or discomfort?

All suggested messages will be vetted by the MD on our team and in accordance with CDC or other appropriate local guidelines. Survey participants can skip any question they choose and withdraw participation at any time.

## 6. Potential Benefits

A. What potential benefits may subjects receive from participating in the study?

We aim to encourage compliance with CDC guidelines about hygiene, mask-wearing and social distancing, reducing the spread of COVID19. We also believe that our treatments might help to make the FHWs more effective.

B. What potential benefits can society expect from the study?

We aim to encourage compliance with CDC and other local guidelines about hygiene, mask-wearing, and social distancing, reducing the spread of COVID19. Also we aim to understand how to encourage information sharing among FHWs, who are stretched very thin during a health crisis.

## 7. Data Collection, Storage, and Confidentiality

A. How will data be collected?

Twitter, YouTube, Instagram sites and US Doctors messaging project:  
Online survey platform (Qualtrics or equivalent, Lucid)

For the November and December 2020 Facebook studies, the only data directly collected from the project will include: number of users in treatment areas receiving the messages, aggregated information about engagement with the videos (length of time spent watching, likes, shares).

We will use 4 supplemental sources of information to measure outcomes:

1. Public release data on COVID-19 cases at the ZIP code or county level.
2. Facebook Data for Good mobility dataset in the period between January 2019 and March 2021. This data set includes information aggregated geographical units comparable to ZIP codes about travel patterns of users who agree to share location data with Facebook. We will also observe flows of individuals between geographical locations, again aggregated to protect privacy. The data is already licensed to MIT and other universities.
3. Facebook's Social Connectedness Index. This dataset shows measures of ZIP code to ZIP code connections, measured using Facebook friendships. Again, this information is aggregated a level sufficient to preserve privacy of Facebook's users.
4. Facebook COVID symptom survey. Facebook is collaborating with researchers at University of Maryland and Carnegie Mellon University to collect a panel survey of COVID symptoms. The de-identified dataset is available to researchers, via university agreement. We will obtain geographically aggregated data and match the geographic locations of respondents to whether or not our ad campaign was run in that geography.

- 
5. Facebook brandlift questions: shortly after the ad campaign, Facebook has the capacity to ask 4 questions to a subset of people who were targeted by the ad and a small hold-out sample (one question is asked to each person). They will provide us with aggregated responses at the demographic (age bins, gender) and geographical (Zip code or bigger) level.
6. Mobility data from SafeGraph publicly available to all researchers aggregated to the census block group level.

WhatsApp sites:

Phone surveys. The enumerators will record answers using SurveyCTO on tablets. If we run out of tablets, or the tablets malfunction, the enumerators will record responses on paper surveys.

- 
- B. Is there audio or videotaping?  Yes  No

*Explain the procedures you plan to follow:*

- 
- C. Will data be associated with personal identifiers or will it be coded?

Personal Identifiers  Coded

*Explain the procedures you plan to follow.*

*Twitter, YouTube, Instagram sites:*

We will collect email addresses and social media user names to follow up with participants since we will survey the same respondents multiple times. We will separate out identifiers from the coded dataset and analyze data on the de-identified dataset.

*MGH doctors project pilot:*

We will collect phone numbers to be able to contact respondents in the future. We will separate out identifiers from the coded dataset and analyze data on the de-identified dataset.

*MGH Doctors Messaging Project:*

We are currently not planning to collect any contact information or personal identifiers from participants. If we decide to collect identifiers such as phone numbers or email addresses, we will apply for an amendment. This is also true in Round 2. Unlike the previous round, we plan for Lucid to do the re-contacting of respondents for the follow-up survey. This means that we still will not need to collect PII from the respondents. We will only share coded identifiers with Lucid.

*US Doctors Facebook Scale Up:* We are not collecting any PII. We will not have any individual-level data from the project, and we will never know who was selected by Facebook for the ad campaign or for the Brand Lift questions.

*Whatsapp sites:*

We will collect names and phone numbers to follow up with participants since we will survey the same respondents multiple times. We will separate out identifiers from the coded dataset and analyze data on the de-identified dataset.

- 
- D. Where will the data be stored and how will it be secured?
-

All identifiers will be kept in Veracrypt-encrypted dropboxes.

E. What will happen to the data when the study is completed?

All PII will be destroyed once the study is complete.

F. Can data acquired in the study affect a subject's relationship with other individuals (e.g. employee-supervisor, patient — physician, student-teacher, family relationships)?

N/A

### 8. Deception

*Investigators must not exclude information from a subject that a reasonable person would want to know in deciding whether to participate in a study.*

A. Will information about the research purpose and design be withheld from subjects?

Yes     No

*If yes, explain and justify:*

### 9. Adverse Effects

*Serious or unexpected adverse reactions or injuries, and/or unanticipated problems involving risks to subjects or others must be reported to COUHES within 48 hours. Other adverse events should be reported within 10 working days.*

A. What follow-up efforts will be made to detect any harm to subjects, and how will COUHES be kept informed?

We do not expect harm from the study.

### 10. Health Insurance Portability and Accountability Act ("HIPAA")

*If your study involves individually identifiable health information and is sponsored by MIT Medical, an MIT Health Plan or another healthcare provider, then you must complete the questions below because HIPAA likely applies to your study. For more information regarding the applicability of HIPAA to human subjects research, please [click here](#).*

A. Do you plan to obtain, use or disclose identifiable health information in connection with your research study?

Yes     No

While the study is about COVID-19, we are not asking participants any information about their own health status in the celebrities studies.

In the MGH doctors project pilot and the main messaging project, we are asking an individual about his/her covid19 status and about pre-existing conditions that are co-morbidities for covid19.

However, we are not obtaining health information from any third party.

*If YES, then all participants must complete an Authorization for Release of Protected Health Information Form. Please attach a copy of this draft form. You must use the [template](#) available on the COUHES website.*

*Alternatively, COUHES may grant a Waiver of Authorization in certain very limited circumstances when use of individually identifiable health information would pose only minimal risk to study*

participants (among other requirements). For additional information regarding whether your study might qualify for a waiver, please [click here](#).

B. Are you requesting a Waiver of Authorization?

Yes     No     N/A

*If yes, explain your rationale for concluding that:*

- (i) *use of participant health information poses no more than minimal risk;*
- (ii) *the research could not be conducted without the waiver and*
- (iii) *the research could not be conducted without the information.*

*In addition, please explain your plan for (i) ensuring the participant health information is not improperly used or disclosed either within MIT or to any outside third parties and (ii) destroying identifiers at the earliest possible opportunity.*

C. Will the health information you will receive for use in this study be de-identified?

Yes     No     N/A

The health information will all come directly from the respondent.

*If yes, you do not need to obtain a signed Authorization for Release of Protected Health Information Form from study participants. Note, however, that if you receive identifiable participant health information that you plan to convert into de-identified information for use by other researchers, then you must obtain a signed Authorization for Release of Protected Health Information Form from each participant before receiving their identifiable health information for use in your study.*

D. Will you be using or disclosing a limited data set?

Yes     No

*If yes and you will only receive participant health information in limited data set form, then you do not need to obtain a signed Authorization for Release of Protected Health Information Form from study participants. You must complete a formal data use agreement with the party from whom you will receive the limited data set information in order for your application to be approved.*

*If yes and you will receive identifiable participant health information that you plan to convert into limited data set form for use by other researchers, then you must obtain a signed Authorization for Release of Protected Health Information Form from each participant before receiving their identifiable health information for use in your study. You must complete a formal data use agreement in order for your application to be approved.*

## V. INVESTIGATOR'S ASSURANCE

**I certify the information provided in this application is complete and correct.**

**I understand that I have ultimate responsibility for the conduct of the study, the ethical performance of the project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by COUHES**

**I agree to comply with all MIT policies, as well all federal, state and local laws on the protection of human subjects in research, including:**

- **ensuring all study personnel satisfactorily complete human subjects training;**
- **performing the study according to the approved protocol;**
- **implementing no changes in the approved study without COUHES approval;**
- **obtaining informed consent from subjects using only the currently approved consent form;**
- **protecting identifiable health information, to the extent required by law, in accordance with HIPAA requirements; and**
- **promptly reporting significant or untoward adverse effects.**

**Signature of Principal Investigator \_\_\_\_\_ Date \_\_\_\_\_**

**Print Full Name and Title \_\_\_\_\_**

**Signature of Department Head \_\_\_\_\_ Date \_\_\_\_\_**

**Print Full Name and Title \_\_\_\_\_**

**By signing this form, you confirm a scientific review of the proposed research has been conducted and that the proposed research is of scientific and scholarly validity.**

*Signed copies of the Comprehensive Review Application and supporting documents should be e-mailed to [couhes@mit.edu](mailto:couhes@mit.edu). In addition, two single sided hardcopies must be submitted to the COUHES office: Building E25-Room 143b.*

	<b>Massachusetts Institute of Technology</b> Committee on the Use of Humans as Experimental Subjects COUHES	<b>COUHES Protocol #</b> 2003000118  August 2020
--	--	---

### PERSONNEL LIST

*This form must be attached with the Application for Comprehensive Review. Any application submitted without a completed personnel list will be returned to you.*

*Personnel is defined as anyone that plays a role in research involving human subjects, including direct contact, indirect involvement, analysis of data, blood or tissue samples. This extends to principal investigators, associate investigators, student investigators, study coordinators, visiting scientists, consultants, laboratory technicians and assistants.*

*All study personnel must be listed below. This listing must include contact information, a brief statement of qualifications and their study role.*

*Important note: all study personnel are required to complete [Human Subject Training](#) before work begins on the project.*

#### I. MIT AFFILIATES

<i>Personnel name and e-mail address</i>	<i>Briefly describe qualifications</i>	<i>Study role(s)</i>	<i>Obtaining consent</i>
<b>Contact*</b> Name: Benjamin Olken Email: bolken@mit.edu	Professor of Economics Prior work on social messaging for public health	PI	X
Name: Esther Duflo Email: eduflo@mit.edu	Professor of Economics	PI	<input type="checkbox"/>
Name: Abhijit Banerjee Email: banerjee@mit.edu	Professor of Economics	PI	<input type="checkbox"/>
Name: Robert Dulin Email: rdulin@povertyactionlab.org	Research Analyst		<input type="checkbox"/>
Name: Pierre-Luc Vautrey Email: vautrey@mit.edu	PhD student in Economics	Co-Investigator	X
Name: Anirudh Sankar Email: asankar@povertyactionlab.org	Research Analyst		X
Name: Harsh Goyal Email: hdgotal@mit.edu	MIT DEDP student Research assistant		
Name: Ritesh Das Email: dritesh@mit.edu	MIT DEDP student Research assistant		



Name: Sirena Yu Email: sirenayu@mit.edu	MIT Undergraduate Research Assistant (UROP)		X
Name: Mohit Karnani Email: mohitkarnani@gmail.com	PhD student in Economics		X
Name: Advik Shreekumar Email: adviks@mit.edu	PhD student in Economics		X

*\*NOTE: Please designate a person with whom COUHES should communicate regarding issues or questions about the protocol.*

## **B. NON-MIT AFFILIATES**

*Proof of training must be attached for all non-MIT affiliates. Documentation from collaborating institutions may be submitted in lieu of training certificates.*

<i>Personnel name, affiliation, and e-mail address</i>	<i>Briefly describe qualifications</i>	<i>Study role(s)</i>	<i>Obtaining consent</i>
Name: Marcella Alsan Affiliation: Harvard Email: marcella_alsan@hks.harvard.edu	MD, MPH, PhD Professor of Public Policy	CO-PI Will also vet all messaging to ensure consistency with CDC/public health authority guidelines	X
Name: Arun Chandrasekhar Affiliation: Stanford Email: arungc@stanford.edu	Assistant Professor of Economics Prior work on social messaging for public health	Co-PI	X
Name: Emily Breza Affiliation: Harvard Email: ebreza@fas.harvard.edu.	Assistant Professor of Economics Prior work on social networks	Co-PI	X
Name: Paul Goldsmith-Pinkham Affiliation: Yale Email: paul.goldsmith-pinkham@yale.edu	Assistant Professor of Finance Prior work on social networks	Co-PI	X
Name: Fatima Cody Stanford Affiliation: Massachusetts General Hospital/ Harvard Medical School Email: FSTANFORD@mgh.harvard.edu	MD, MPH, MPA, FAAP, FACP, FAHA, FTOS	The MGH doctors are involved in the following activities: - advising on message content - recording video messaging	

		- advising on data analysis with deidentified data	
Name: Lucy Ogbu-Nwobodo Affiliation: MGH/ Harvard Medical School Email: logbu-nwobodo@mgh.harvard.edu	MD, MS Resident, Clinical Fellow	The MGH doctors are involved in the following activities: - advising on message content - recording video messaging advising on data analysis with deidentified data	
Name: Erica Warner Affiliation: MGH/ Harvard Medical School Email: ewarner@mgh.harvard.edu	Assistant Professor, Medicine, Harvard Medical School Assistant Investigator, Medicine, Massachusetts General Hospital	The MGH doctors are involved in the following activities: - advising on message content - recording video messaging advising on data analysis with deidentified data	
Name: Carlos Torres Affiliation: MGH Chelsea HealthCare Center and MassGeneral Hospital for Children Email: <a href="mailto:CTORRES4@PARTNERS.ORG">CTORRES4@PARTNERS.ORG</a>	MD	The MGH doctors are involved in the following activities: - advising on message content - recording video messaging advising on data analysis with deidentified data	
Name: Shobitha Cherian Affiliation: J-PAL South Asia at IFMR Email: shobitha.cherian@ifmr.ac.in	Prior involvement in education, diseases	Research Associate	X
Name: Tithee Mukhopadhyay	Prior involvement in social networks	Associate Director-Rsearch	X

Affiliation: J-PAL South Asia at IFMR Email: tithee.m@ifmr.ac.in			
Name: Vasu Chaudhary Affiliation: J-PAL South Asia at IFMR Email: vasu.chaudhary@ifmr.ac.	Prior involvement in education, diseases	Research Associate	X
Name: Sarah Eichmeyer Affiliation: Stanford University Email: saraeich@stanford.edu	PhD student in Economics		