

# Pre-Analysis Plan

## Increasing the Efficacy and Diffusion of COVID-19 Messaging for Vaccination

This Draft: January 24, 2022

### 1. Introduction

This study will distribute videos of health care professionals encouraging Covid-19 vaccination to a large sample of Facebook and Instagram users, and will test the most effective ways to maximize diffusion of this vaccine-related content to increase vaccination rates. The study sample will be U.S. states where vaccination rates remained low in fall 2021. The experimental design is an RCT with 4 groups, randomized at the county level: 1) a control group which receives no intervention, 2) a treatment group in which Facebook users receive ads which include videos of health care professionals telling them to get vaccinated, 3) a treatment group in which Facebook users receive ads which include videos of health care professionals encouraging them to help their friends to get vaccinated, and 4) a treatment group in which Facebook users receive ads which include videos of health care professionals encouraging them to get their most influential friends to help their friends get vaccinated. The study team has built a website to host the videos of health care professionals which answer common questions about Covid-19 vaccination. The primary outcome of interest is Covid-19 vaccination rates. The investigators will also measure engagement with vaccine-related content in each treatment group. Secondary outcomes which may be considered include Covid-19 case and symptoms prevalence, Covid-19 hospitalization or mortality, and outcomes related to movement, travel, and closures.

### 2. Sample and Treatment

#### 2.1 Sample Population

The experimental sample includes all states where less than 60% of the total population had received a first dose of Covid-19 vaccine by October 21, 2021. There are 1402 counties in the 19 states satisfying those criteria (Alabama, Alaska, Arkansas, Georgia, Idaho, Iowa, Indiana, Louisiana, Michigan, Mississippi, Missouri, Montana, North Dakota, Ohio, Oklahoma, South Carolina, Tennessee, West Virginia, Wyoming). Excluding the five counties with missing data, there are 1,397 counties in the experiment, of which there are 468 counties in the control group, 310 counties in T1 ("direct" messaging) treatment, 309 counties in T2 ("friends" messaging), and 310 counties in T3 ("gossips" messaging).

## 2.2 Treatments

The treatments are all initiated through Facebook platforms. Facebook users located in treatment counties will be exposed to one of three ad campaigns featuring a set of videos about Covid-19 vaccination. The videos feature doctors and nurses from Massachusetts General Hospital (MGH), Harvard Kennedy School (HKS), Johns Hopkins School of Nursing, Harvard Medical School, Lynn Community Health Center, St. Anthony North Family Medicine, and McGovern Medical School. In the videos, doctors and nurses are usually wearing scrubs or other medical uniforms. The ad videos were filmed by doctors and nurses using their phones, and edited by professionals at Code3. The videos are short (approximately 30 seconds in length).

All of the ads are associated with the project Facebook page titled “The Doctors for Coronavirus Prevention Project.” Users see the ads and name of the Facebook group associated with the ads on their Facebook feeds. Some ads are also served on Instagram. Watching the videos is completely optional for Facebook users. The ads are all targeted to Facebook users aged 18 and older. The ad campaign is programmed to optimize for the Facebook “reach” objective, with budget allocations roughly proportional to population.

**Treatment 1: “Direct” messaging:** Facebook users receive ads with a video of a doctor or nurse answering a common question or addressing a common myth about the Covid-19 vaccine.

**Treatment 2: “Friends” messaging:** Facebook users receive ads which encourage them to help their friends to get vaccinated.

**Treatment 3: “Gossips” messaging:** Facebook users receive ads which encourage them to ask their most influential friends to help their friends to get vaccinated.

In treatments 3 and 4, participants will have the option to sign up to be a “vaccine ambassador,” in which case they will get notifications when the study team posts new vaccine-related content, and will receive reminders about encouraging their friends to be vaccinated.

## 2.3 Sample Frame

We began to roll out the campaigns on December 22, 2021. We plan to run the Facebook campaigns for five weeks (until January 27, 2022).

## 2.4 Study website

The Facebook ads for all treatment groups link to a website built for the study:

<https://www.doctorsforcovidprevention.org>. Treatment 1 participants are directed to a page of video thumbnails with all of the videos recorded by doctors and nurses which answer questions and address myths about the Covid-19 vaccine. Treatments 2 and 3 participants are directed to a website which includes these same videos, and also encourages website visitors to share the vaccine-related content with their friends. The website for treatments 2 and 3 also allows visitors to sign up to be a “vaccine ambassador.” Interested website visitors can enroll formally as vaccine ambassadors through our site. They can share their contact information with the study to a) be entered into prize lotteries, b) receive reminders to share content with others, c) be contacted in the future for possible follow-up surveys. These individuals are asked to give their informed consent at this time and are screened based on whether they live in a treated county and whether they are 18 years or older.

The display of doctor/nurse videos on the study website is randomized in two ways. First, website visitors are randomized into two possible video framings: the “myth” site or the “question” site. In the “myth” version, doctors/nurses respond to common myths about the Covid-19 vaccine (e.g. the Covid-19 vaccine can give me Covid), while in the “question” version, the doctors/nurses respond to a question (e.g. can the Covid-19 vaccine give me Covid?). Second, the order that videos are shown is randomized. For the page which only shows video thumbnails, this means the order of the thumbnails is random for each unique website visitor. For the page which encourages people to share vaccine-related content, the random ordering of videos means that the “featured” video is randomized between unique website visitors.

## 2.5 Treatment Assignment

The counties in the study sample were randomized into three treatment groups and a control group: 1) a “direct” messaging group (T1), 2) a “friends” messaging group (T2), 3) a “gossips” messaging group (T3), and 4) a control group which did not receive any Facebook ads from us.

Randomization was conducted at the county level. County-level randomization was stratified by three characteristics: 1) state, 2) political leaning (according to 2020 election results), and 3) baseline vaccination rates. For political leaning, counties were divided into below- and above- median GOP vote in the 2020 presidential election. For baseline vaccination rates, counties were divided into above- and below- median percentage of the population that had received the 1st dose of Covid-19 vaccine.

After stratifying on these three variables, strata were adjusted so that no stratum was smaller than 9 counties. Strata with fewer than nine counties were dissolved by baseline percentage of population with first dose of vaccine. For the three states (South Carolina, Michigan, and Wyoming) where this does not result in strata that have at least nine counties, we dissolve instead by baseline GOP vote share. In total, this left us with 47 strata.

**Exceptions:** County-level GOP votes were not available for Alaska, and so this stratification variable was used for the other 18 states in the sample only.

Out of the 1,397 counties which fit the eligibility criteria for the experiment, we assigned 468 counties to the control group, 310 counties to T1 (“direct” messaging) treatment, 309 counties to T2 (“friends” messaging), and 310 counties to T3 (“gossips” messaging).

## 3. Hypotheses

Primary hypotheses:

1. Do the ads affect the new Covid-19 vaccinations in treated areas?
  - 1st dose of vaccine
  - 2nd dose/completed vaccination
  - Booster shots
2. Which type of messaging (“direct,” “friends,” or “gossips”) is most effective at driving engagement and shares of vaccine-related content, as well as new vaccinations?

3. Is myth- or question-based messaging more effective at eliciting engagement with vaccine-related content?

Secondary hypotheses:

4. Do the ads affect Covid-19 cases, hospitalizations, or symptoms in treated areas?
5. Do the ads spill over onto other areas that are connected to the treatment counties?
  - Behavioral spillovers: new vaccinations
  - Covid-19 case, hospitalization, symptom impacts
  - Are spillovers more pronounced on the mobility network or the Facebook friendship network?

## 4. Empirical Analysis

### 4.1 Primary Outcomes

1. **Covid-19 vaccinations:** this includes 1st doses, 2nd doses/completed vaccination, and booster shots when available. These data are available at the county level.

### 4.2 Secondary Outcomes

1. **Engagement with content related to Covid-19:** this includes website views, video views, and social media shares of vaccine-related videos and messaging, disaggregated at the county or treatment group level.
2. **Other Covid-19 outcomes:** depending on data availability, this will include Covid-19 symptoms, cases, hospitalizations, and deaths. The investigators do not expect the study to be well-powered on this outcome.
3. **Other survey outcomes:** if available, we may assess outcomes on vaccine hesitancy according to surveys. The investigators do not expect the study to be well-powered on this outcome.
4. **Other behavioral outcomes:** we may also study whether the treatments have an effect on movement and travel, as well as school or other institutional closures. The investigators do not expect the study to be well-powered on this outcome.

### 4.3 Website data

The study website collects detailed information on visitors, including the following:

1. County of origin (according to the Facebook ad which referred them to the website) – for part of the experimental timeframe
2. IP Address – for part of the experimental timeframe

3. Version of the website that the user saw
4. Any actions taken on the website along with timestamps (e.g. page view, video views, video shares, message shares, vaccine ambassador sign ups)
5. Sharing chains (e.g. if someone visits the website after it is shared with them from County A, we can follow whether they viewed the website due to a share from someone in County B)

## 4.4 Regression Analysis

1. Effects of treatment on county-level outcomes (Covid-19 vaccinations, content engagement, other behavioral or survey outcomes):

(a) Main analysis:  $Y_{ct} = \alpha_t + \gamma_c + \beta_1 Direct_{ct} + \beta_2 Friends_{ct} + \beta_3 Gossips_{ct} + \delta Y_{c0} + \epsilon_{ct}$

- $Y_{c0}$  is a baseline measure of the outcome
- *Direct*, *Friends* and *Gossips* are the three treatment groups. Treatment is assigned at the county level. We may also pool together friends and gossips, or all three treatment groups in some specifications.

(b) Heterogeneous effects of treatment:  $Y_{ct} = \alpha_t + \gamma_c + \beta_1 X_{ct} + \beta_2 Treat_{ct} + \beta_3 Treat_{ct} X_{ct} + \epsilon_{ct}$

- We plan to explore heterogeneity along the following dimensions:
  - Political leaning (red/blue) – this was a stratification variable
  - Baseline vaccination rates – we stratified on baseline population percentage with 1st dose of vaccine
  - Percentage white/non-white
  - Relevant vaccine mandates
  - Prior/baseline infection rates
  - Urban vs rural

We may look at heterogeneity in specifications which consider each treatment group separately, as well as specifications which pool together treatment groups.

(c) Distributional impacts using quantile regression

- We may repeat the exercises above with quantile regressions.

(d) Do people who visit the study website from different treatment groups click on different content (e.g. vaccine-related videos)? For example, do website visitors respond to myth vs fact videos differentially by treatment group?

2. Comparing the efficacy of myth- vs question-based messaging

(a)  $Y_{ict} = \alpha_t + \gamma_c + \beta \mathbb{1}(Myth_{ict}) + \epsilon_{ict}$

- $Y_{ict}$  stands for engagement-related outcomes such as video views, website visits, and video shares
- We will examine myth- vs question-based messaging at both the Facebook ad level and at the website level. (The videos vary in whether they are myth- or question-based both in the Facebook ads which are served to users, as well as in the videos which are visible to website visitors.)

## 4.5 Spillovers and Networks Analysis

We may use Facebook network information (SCI) at the zip code level to study the dynamics of information sharing about Covid-19 vaccination.

## 4.6 Controls

For all regressions, we will use machine learning techniques to select the controls. Potential controls include: baseline Covid-19 vaccination or case rates, state fixed effects, county-level demographic information from the census, county- or state-level vaccination policies, and many other attributes.

## 4.7 Inference

For our core analyses, we will use heteroskedasticity-robust standard errors. We will check for additional statistical robustness following the approaches outlined in Section 5 of the Supplementary Appendix of [this paper](#).

# 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, Stanford, Massachusetts General Hospital, and Yale.