Statistical Analysis Plan

COVID-19 Vaccination Take-Up
CT ID: 27165
November 16, 2021
Mireille Jacobson

Our analysis was guided by our-pre-registered pre-analysis plan that is available here: https://www.socialscienceregistry.org/trials/7405

Any analysis not pre-registered we describe below as exploratory.

To analyze the impact of our interventions on vaccination, we estimate the following linear regression model:

$$(1) \qquad Vaccinated_i = \alpha + \beta_1 Financial_i + \mathrm{M}_i\theta + \beta_3 Link_i + \delta X_i + \varepsilon_i$$

where $Vaccinated_i$ is an indicator (0/1) for whether a respondent received a SARS-CoV-2 vaccination within one month after completing the survey. $Financial_i$ is an indicator for whether the individual was randomized into the financial incentive arm, M is a vector of indicators for randomization into each of the three message types (CDPH, safety, or health consequences video) and $Link_i$ is an indicator for whether the individual was randomized to receive the highlighted link. The excluded group, the control condition, received no extra prompting to get vaccinated. To increase precision, we estimate versions of (1) that include X_i, a vector of predetermined characteristics including age and its square, race, gender, self-reported income, education, the language the respondent took the survey in (English/Spanish), whether the respondent was "impaneled," meaning primary health care is provided at Contra Costa Regional Medical Center, and indicators for calendar date. Our main hypotheses are that all of the interventions will increase vaccination rates, $\beta_1 > 0$, $\theta > 0$ and $\beta_3 > 0$.

To differentiate across the financial incentive amounts, we expand on (1) to estimate:

$$(2) \; Vaccinated_i = \alpha + \beta_1 1_i^{\$10} + \beta_2 1_i^{\$50} + \mathrm{M}_i\theta + \beta_3 Link_i + \delta X_i + \varepsilon_i$$

where $1_v^{\$10}$ and $1_v^{\$50}$ are indicators for being randomized into financial incentives of \$10 or \$50. Our hypothesis is that the magnitude of the effect is increasing in the incentive amount: $0 < \beta_1 < \beta_2$.

Our analysis of vaccinations intentions is based on a modification of equation (1):

$$(3) \qquad Intention_i = \alpha + \beta_1 Financial_i + \mathrm{M}_i\theta + \beta_3 Link_i + \delta X_i + \varepsilon_i$$

where $Intention_i$, a respondent's self-assessed probability of getting vaccinated in the next 30 days, takes the place of $(2)\; Vaccinated_i$. Note that since the financial incentive and scheduling link are presented to individuals after survey completion, they are included here only as indicators of treatment stratum and are not meant to generate causal estimates of their impact on vaccine intentions. Our main hypothesis is that messaging increases vaccine intentions, $\theta > 0$. We further hypothesize that the health consequences message will have the largest effect on intentions such that $\theta_3 > \theta_1$, and $\theta_3 > \theta_2$, where 3 denotes the health consequences message, 2 denotes the safety message and 1 denotes the CDPH message.

We analyze heterogeneity in the impacts of our interventions by respondent gender, race/ethnicity, age-group, and support for Trump or Biden during the 2020 presidential election.

To analyze whether race and gender concordance affects the impact of health messages, we rerun the models specified by equations 1 and 3, but include interactions between the relevant video messages and $Race\ Concord_i$ and $Gender\ Concord_i$, which are indicator variables equal to 1 if the physician messenger and the recipient share the same race/ethnicity or gender, respectively.

In exploratory analysis, we check the robustness of our results to model choice. First, we estimate probit regression models of vaccine uptake (equations (1) and (2)) and censored regression models (tobit models) of vaccination intentions (3). Unlike the linear regression model, the probit model bounds the predictions of the outcome to 0 or 1. Similarly, we estimate tobit regression models of vaccination intentions to account for the fact that intentions are censored at 0 and 100.