Pre-analysis plan:
Behavioral interventions, economic preferences and vaccination uptake

Pol Campos-Mercade, Armando N. Meier, Stephan Meier,
Devin Pope, Florian H. Schneider, Erik Wengström

27 May 2021

1) Motivation

We study the impact of four interventions to increase COVID-19 vaccination uptake in Sweden. We look at actual vaccination behavior, vaccination intentions and the intention-behavior gap. We also collect data on peoples' preferences, vaccine beliefs and vaccine knowledge.

Our aim is to investigate 1) what interventions work to encourage people to get vaccinated quickly, 2) what interventions work best for different types of people (based on peoples' preferences, vaccine beliefs and vaccine knowledge) and what this means for treatment allocation, 3) whether and how self-reported vaccination intentions differ from actual vaccination uptake and 4) whether economic preferences can explain heterogeneities in vaccination uptake and the intention-behavior gap.

2) Design

To address our research questions, we conduct a survey with a general population sample of the Swedish population. In the online survey, we first measure participants' preferences. Then we randomly allocate participants to an intervention. Next, we measure participants' intentions to get vaccinated. Last, we examine whether people did or did not get vaccinated using data from administrative registers.

2.1) Survey

Enkätfabriken, a well-established Swedish survey company, sends the survey to a general population sample of the Swedish population.

First, we measure participants' preferences using the following survey questions:

- **Altruism:** How willing are you to give to good causes without expecting anything in return? (Response scale from 0 to 10)
- **Time preferences:** How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? (Response scale from 0 to 10)
- **Risk preferences:** In general, how willing are you to take risks? (Response scale from 0 to 10)

- **Reciprocity:** When someone does me a favor, I am willing to return it. (Response scale from 0 to 10)
- **Trust:** I assume that people have only the best intentions. (Response scale from 0 to 10)
- **Present focus:** I postpone starting on things I dislike to do. (Response scale from 0 to 10)
- **Norm following:** It is important for me to always behave properly and to avoid doing anything people would say is wrong. (Response scale from 0 to 10)

Second, we measure participants' beliefs, knowledge and worries related to COVID-19 vaccines:

- **Beliefs about vaccine risk:** In general, COVID-19 vaccines are safe.
- **Vaccine knowledge:** Diseases like autism, multiple sclerosis, and diabetes might be triggered through vaccination.
- **Worries vaccine:** I am worried about the side effects from COVID-19 vaccines.
- **Worries needles:** I am afraid of the needles used for vaccination.

We also collect variables on COVID-19 history, vaccine eligibility, risk group status and socio-demographics.

Then, we randomly allocate participants to one of six conditions. First, we have a control condition:

- **Control condition:** We encourage participants to take the COVID-19 vaccine.

Second, we study the impact of four behavioral interventions. All interventions include the same encouragement as in the control condition plus the treatment:

- **Benefit others condition:** We ask participants to make a list of 4 people that would benefit from the vaccine to make them aware of the vaccine's social impact.
- **Arguments condition:** We ask participants to write down arguments for why one should get vaccinated.
- **Information condition:** We inform people about the safety and effectiveness of the vaccination using a quiz.
- **Incentives condition:** We offer people SEK 200 if they get vaccinated within 1 month after they are eligible to get vaccinated, which we check using administrative data.

A final condition allows us to study the impact of vaccine appointment information and reminders on vaccination uptake (all of which are part of the Control condition):

- **Minimal condition:** Participants in this treatment condition do not receive information about making a vaccination appointment and do not receive reminder emails.

Next, we ask participants about their intentions to receive the COVID-19 vaccine:

- **Intention 1 (main outcome measure for intentions):** Do you think you will get a first shot of a COVID-19 vaccine within the first month after the vaccine becomes available to you? (No/Yes)
- **Intention 2:** We understand that there is always some uncertainty regarding all decisions. From 0% to 100%, what do you think are the chances that you will choose to get a first shot of a COVID-19 vaccine within the first month after the vaccine becomes available to you?
- **Intention 3:** When do you think you will get a COVID-19 vaccine after the vaccine becomes available to you? (Response scale: within 1 week, within 2 weeks, within 3 weeks, within 1 month, within 2 months, within 3 months, within 6 months, within 12 months, after 12 months, never)

Finally, we provide participants that are not in the Minimal condition with a link to a governmental website where they can receive information about how they can sign up for a vaccine appointment in their region. We record whether they click on the link:

- **Survey behavior:** Did the participant click the link to get information of how to make a vaccination appointment? (0/1)

We send participants that are not in the Minimal condition two treatment-specific reminders.

2.2) Administrative data

We will be able to match our survey responses with administrative data on vaccination uptake. For all participants we will see whether and when they received a vaccine. Using these data, we construct the following variables:

- **Behavior 1 (main outcome measure for behavior):** Did the participant get a first shot of a COVID-19 vaccine within the first month after the vaccine became available to him/her? (0/1)
- **Behavior 2:** How many days did the participant take to get a first shot of a COVID-19 vaccine?
- **Behavior 3:** Did the participant get a first shot of a COVID-19 vaccine at all within the time window we observe?

Using the actual behavior of the participant, we can also study the intention-behavior gap for those who intend to vaccinate:

- **Intention-behavior gap:** Intention 1 - Behavior 1

3) Analysis

3.1) Main analysis

We will study whether the Benefit others, Arguments, Information and Incentives conditions affect the intention to vaccinate (Intention 1) and actual vaccination uptake (Behavior 1). We compare vaccination uptake in each of these treatment conditions to the uptake in the Control condition using OLS. To do so, we regress our outcome variables y_i on a set of treatment condition dummies:

$$y_i = b_0 + b_1 1(\text{Benefit others})_i + b_2 1(\text{Arguments})_i + b_3 1(\text{Information})_i \\ + b_4 1(\text{Incentives})_i + b_5 X_i + e_i$$

where y_i is either Intention 1 or Behavior 1, $1(t)_i$ has a value of 1 if participant i is in treatment condition t and a value of 0 otherwise, X_i is a vector of control variables (consisting of gender dummies, age dummies, region dummies, interactions between age and region[1], being in an at-risk group for COVID-19, civil status dummies, a dummy for children in the household, dummies for employment status, dummies for education, dummies for parents' place of birth, and income, see section 3.5 for details on how we exactly define those variables), and e_i is an individual specific error robust to heteroscedasticity. We will use a two-sided test to examine whether b_1, b_2, b_3 and b_4 are statistically significantly different from zero.[2] Moreover, we will jointly test (F-test) whether we can reject the hypothesis that all four coefficients are zero.[3]

In the OLS regression above, we do not use the data from the Minimal condition. We will also study whether the information about booking a vaccination appointment and reminders to get vaccinated in the Control condition affect intention to vaccinate (Intention 1) and actual vaccination uptake (Behavior 1) relative to the Minimal condition. We will also compare the effect of the four behavioral interventions to the Minimal condition.

3.2) Heterogeneous treatment effects according to economic preferences, beliefs, and knowledge

We will then study whether there are heterogeneities in treatment effects for people with different preferences, vaccine beliefs and vaccine knowledge (we will call these different dimensions: "measure_i"). Our main focus lies on altruism, time preferences and risk preferences. We use a fully interacted OLS model where we prespecify all simple interactions of the four behavioral interventions with each of the different preference (and belief) measures. That is, we regress Behavior 1 on a set of treatment condition dummies and the interaction between treatment dummies and preference measures:

$$\text{Behavior } 1_i = b_0 + b_1 1(\text{Benefit others})_i + b_2 1(\text{Arguments})_i$$

[1] If there are heterogeneities in the rollout of the vaccination program, e.g., across age groups and regions, we will control for the exact level of the rollout dimension if we have the data available, see section 3.5 for details.

[2] The Benefit others, Arguments, Information and Incentives conditions are designed to increase vaccination uptake. Hence, one might expect b_1, b_2, b_3 and $b_4 > 0$. However, we want to allow for the possibility that some treatments could have a $b<0$ and hence pre-register two-sided tests.

[3] As secondary analysis, we will also look at the impact of the interventions on Intention 2 and 3, Survey behavior, the Intention-behavior gap, and, depending on the time window we finally observe, on Behavior 2 and 3. We will use similar specifications as described above. Last, we will explore whether some treatments are more effective than others.

$$+ b_3 1(\text{Information})_I + b_4 1(\text{Incentives})_i + b_5 1(\text{Benefit others})_i * \text{measure}_i + b_6 1(\text{Arguments})_i * \text{measure}_i + b_7 1(\text{Information})_i * \text{measure}_i + b_8 1(\text{Incentives})_i * \text{measure}_i + b_9 X_i + e_i$$

where $1(t)_i$ has a value of 1 if participant i is in the treatment condition t and a value of 0 otherwise, X_i is a vector of control variables (consisting of measure_i, gender dummies, age dummies, region dummies, interactions between age and region, being in an at-risk group for COVID-19, civil status dummies, a dummy for children in the household, dummies for employment status, dummies for education, dummies for parents' place of birth, and income, see section 3.5 for details), and e_i is an individual specific error robust to heteroscedasticity. We will use a two-sided test to examine whether b_5, b_6, b_7 and b_8 are statistically significantly different from zero. Moreover, we will jointly test (F-test) whether we can reject the hypothesis that all four coefficients are zero.[4]

We will look at the following measures of economic preferences, vaccine beliefs and vaccine knowledge:

- **Main analysis:** Altruism, Time preferences, Risk preferences.
- **Secondary analysis:** Reciprocity, Trust, Present focus, Norm following, Vaccine knowledge, Worries vaccine, Beliefs about vaccine risk.

We implement the above analysis for each of these variables as "measure_i" separately. Moreover, we will also report results when we add variables jointly.[5]

Note that in the OLS regressions above, we do not use data from the Minimal condition. However, we will also replicate the above analysis with the Minimal condition.

3.3) Targeting and individual treatment effect heterogeneity

We want to understand whether we can improve the effectiveness of the interventions by targeting interventions to participants.

We will first document the existence or absence of individual-level treatment effect heterogeneity using the survey measures, including socio-demographics and economic preferences.[6] (Our main analysis here will rely on the appropriate machine learning methods which are currently tree-based methods, see, e.g., Zhou, Wager, and Athey, 2018, auxiliary analyses may use the estimates of treatment effect heterogeneity using OLS.)

We will then use related methods to assess whether targeting can work:

---

[4] As secondary outcomes, we will also look at Intention 2 and 3, Survey behavior, Intention-behavior gap, and, depending on the time window we finally observe, also on Behavior 2 and 3.

[5] We will also explore whether there are relevant heterogeneities using the rest of the variables (e.g., gender, education, parental place of birth, income, COVID-19 history, fear of needles, etc.).

[6] We will use adjustments in case there are heterogeneities in age and region across time in vaccination rollout.

- The first algorithm we use is based on estimates of effect heterogeneity using a fully interacted OLS model to estimate the counterfactual treatment effects of each treatment for each individual.[7] We then use these estimates to assess whether one can increase vaccination take-up by reassigning treatments. Among others, we compare the overall effectiveness of a reassigned intervention to random assignment, to all people being in the control group, and to all people being in the group that has the largest average treatment effect across all participants.
- In addition, we will use machine learning methods to directly estimate the optimal targeting function without having an explicit parametric model (for instance, using tree-based methods such as in Wager and Athey, 2021).
- We will also look at how using i) only socio-demographics, and ii) socio-demographics, economic preferences, vaccine knowledge and vaccine beliefs affect policy targeting. The latter allows us to examine whether targeting can improve vaccination uptake, that is, what role economic preferences can play for optimal policy targeting.

3.4) Correlations between vaccine behaviors and preferences/beliefs

We will also study how preferences, vaccine beliefs and vaccine knowledge (“$measure_i$”) relate to vaccine behaviors and the behavior intention gap. To do so, we only focus on the data in the Control and Minimal condition and estimate OLS models of the form[8]:

$$y_i = b_0 + b_1 measure_i + b_2 X_i + e_i$$

where y_i is either the Intention-behavior gap or Behavior 1,[9] and X_i is a vector of control variables (consisting of gender dummies, age dummies, region dummies, interactions between age and region, being in an at-risk group for COVID-19, civil status dummies, a dummy for children in the household, dummies for employment status, dummies for education, dummies for parents’ place of birth, and income, a dummy for being in the Minimal condition, see section 3.5 for details), and e_i is an individual specific error robust to heteroscedasticity. We will use a two-sided test to examine whether b_1 is statistically significantly different from zero.

We will then look at the following measures of preferences, vaccine beliefs and vaccine knowledge (that is, we implement the above analysis for each of the following variables as “$measure_i$” separately): Altruism, Time preferences, Risk preferences, Beliefs about vaccine risk, Reciprocity, Trust, Present focus, Norm following, Vaccine knowledge, Worries vaccine and Worries needles.[10] We will also report results when we add measures jointly.

3.5) Definition of Variables

[7] We include all simple interactions between treatments and the variables.
[8] We will also report results using all data. We will then add five treatment dummies as controls.
[9] As secondary outcomes, we will also look at Intention 1 to 3 and Behavior 2 and 3.
[10] In addition, we will also explore whether there are relevant individual differences using the rest of the variables (e.g. gender, education, parental place of birth, income, COVID-19 history, etc.).

We will treat all measures, in particular the preference measures, as continuous for regression analyses (unless indicated otherwise or when measures are binary). For instance, when participants are asked whether a statement describes them, we will code "does not describe me at all" as 0 and "describes me perfectly" as 10.

We will code the variables below for all OLS analyses indicated before as follows:

- Age: We will select age controls that as best as possible capture the region- and age-specific vaccination strategy of Sweden and the timing of the survey waves. For instance, regions have offered vaccinations to ages 43+ and ages 45+. In this case, we then include separate age fixed effects for ages 45-49 and ages 43-44 and ages 40-42. We interact these fixed effects with regional fixed effects. As we will field the survey across different age groups, the fixed effects will also capture the wave-specific timing of the rollout. However, if this strategy does not capture the final survey or Sweden's vaccination rollout adequately, we will take this into account and adapt the coding.
- Gender: dummies for the categories indicating male/female/other.
- Region: dummies for each of the counties in Sweden.
- Being in a COVID-19 risk group: dummy for the category "yes."
- Civil status dummies for each status: single, sarbo, couple, married, others.
- A dummy for whether children live in the participant household: dummy for number of children in the household >0.
- Employment status dummies for each status: full-time, part-time, work, unemployed, student, pensioner, others.
- Educational attainment dummies for each group: elementary, high-school, professional training, ongoing university studies, university studies, research studies.
- Parental place of birth dummies: dummies for each place of origin of the mother and the father.
- Income dummies for each category of incomes used in the survey.

4) Data collection and sample size

4.1) Data collection

The surveys are collected through an online survey with the help of the Swedish survey company Enkätfabriken. We aim to collect data from about 10,000 participants. However, there is some uncertainty about the number of people that Enkätfabriken can recruit, and we could end up with a higher or small sample size. Importantly, we will not have access to the vaccination data until we have finished data collection. We will field the survey in age-group-specific waves based on the vaccination rollout across Sweden, such as to reach people when they consider signing up for vaccination.

4.2) Exclusion criteria

We exclude participants aged <18 and >=50, participants that already received the vaccine and participants who are at risk of side effects from the vaccine according to guidelines from the Swedish public health agency.

4.3) Power

We use simulations to estimate our power. With a control group of about 1/3 of the sample size and 2/15 participants in each condition. With 10’000 observations and using a baseline in which 70% of the participants in the control group vaccinate, we will have 80% power to detect an effect size of about 4 percentage points.

References

Zhou, Z., Athey, S., and Wager, S. (2018). Offline multi-action policy learning: Generalization and optimization. Working paper.

Athey, S., and Wager, S. (2021). Policy learning with observational data. Econometrica, 89(1), 133-161.