Attractive Flu Shot: A Behavioral Approach to Increasing Influenza Vaccination Uptake Rates

Hypothesis, Plan for Data Analysis and Stopping Rule for Data Collection

Background

In all treatments (control and four treatments) the first question asks whether the participant intends to take the shot or not. The second question is a free text explanation of their choice. Third, we ask how certain the participant is with respect to his/her intention to get the shot. The answer will be given on a scale of 1, not certain at all, to 5, very highly certain. In question 4 we ask whether or not he/she took the shot last year and finally in question 5 how many times (according to his/her memory) he/she took the shot in the past 5 years. In addition, the following individual characteristics (background demographics) will also be provided by the panel (the "sample provider"): age, gender, self-reported education, marital status and self-reported income.

Hypothesis

In order to discuss our hypothesis we first define two types of participants according to their answer to question number 3 (the certainty question):

- Certain – Participants who answer 4 or 5 on the certainty scale.
- Uncertain – Participants who answer 1 -3 on the certainty scale.

We hypothesize that our treatment groups will have a positive effect on the intentions to receive the flu shot among the uncertain participants. The larger effects are expected in the two treatments involving a cost or a monetary benefit. A weaker effect is expected for the two other treatments (the weaker of these two, is expected to be the one which includes a recommendation only. This treatment may be too weak to show any effect compared to the control as was found in a pilot study). As certain participants have strong views regarding the flu shot, we do not expect this subpopulation to be affected by any of the treatments.

Let Pr(Vaccinate) be the proportion of subjects willing to vaccinate. Then Pr(Vaccinate|Uncertain) is the proportion of subjects willing to vaccinate, given that they expressed low certainty levels (1 – 3). Then **for each treatment group** we test the following hypotheses:

H_0: $\Pr(\text{Vaccinate}|\text{Uncertain, Treatment}) - \Pr(\text{Vaccinate}|\text{Uncertain, Control}) = 0$
H_1: $\Pr(\text{Vaccinate}|\text{Uncertain, Treatment}) - \Pr(\text{Vaccinate}|\text{Uncertain, Control}) > 0$

Planned Analysis

Before we turn to the research question analysis we will examine individual characteristics (background demographics) in the treatments and control. We expect no significant differences since we will have a completely random assignment into treatments.

When coming to analyze the data we will first cluster the certainty level of participants into certain (4-5 on the certainty scale) or uncertain (1-3 on the certainty scale). While some of the analysis may take place with the actual certainty levels, based on our pilot study we expect very few participants marking very low levels of certainty (either 1 or 2). That is the main reason for the planned clustering. We will then look at the percentage of subjects who are uncertain (or the different distributions of certainty levels) in the different treatments. Once again, we expect no difference between the distributions across treatments (unless the treatment itself affects the answer to the certainty question. If the treatment does influence the answer to this question according to our test, we will take this into consideration in the analysis). Note that the exact terminology we use to ask the certainty question (see surveys in attached document) separates it from the specific treatment as it is a general question regarding the level of certainty of the participant with respect to his/her intention to get the shot. The exact wording has been chosen precisely in order to have the measure of certainty independent from the treatment.

Our data analysis will include at least the following tests: A two-way ANOVA examining how the percentage of those willing to accept the shot is affected by the interaction between certainty level and treatment. Following the result of this test we may run a one-way ANOVA for each of the certainty types. As mentioned above, we conjecture that the uncertain will be affected by the treatment but not the certain. Finally, and depending on the result of the one-way ANOVA tests, we will run Dunnett's test comparing each of the treatments to the control while accounting for multiple comparisons.

In addition to the above tests, we will run Logistic regression models where the dependent variable is the decision of whether to get the shot (1) or not (0) and the explanatory variables are the treatments, certainty type and interaction between certainty type and treatments. The controls will vary by the specification and will include age, gender, reported family income level and education. We will also add as a control the answer to the question regarding vaccination last year (0 or 1) and the answer to the question regarding number of shots received in the past 5 years. These last two variables may be viewed as a proxy for true preferences as expressed through recent years' actual behavior.

We will also look into participants' explanations of their intention to get the shot (early/late if in the treatment groups) or not. Specifically we will examine the percentage of explanations that refer to the decoy option or a comparison between the decoy and the attractive option, i.e., explanations that hint at the attraction effect

playing a role. We will look at these percentages by treatments as well as by certainty types for each treatment. This could hint at the psychological procedure behind participants' choices.

Other Analysis

In our last question we ask participants to mark how many times, to the best of their memory, they received the flu vaccine in the past 5 years (multiple choice question ranging from 0 to 5). The answer to this question will serve as a control in some of the regression specifications but it may also serve as an observable proxy for the level of certainty based on actual past behavior. Intuitively, a person who received 2 or 3 shots in the past 5 years seems not very certain about the shot while a person who received 0 or 5 shots acts as one who is very certain regarding the flu vaccine. Answering 1 or 4 may reflect medium levels of certainty.

We will examine the correlation between the answers to both questions (the direct certainty level and the number of shots in the past 5 year) to assess whether this is a reasonable proxy for certainty level. Given that this measure will show reasonably high correlation with our direct measure of certainty level we will run the same analysis described above when the certainty type is determined by this variable rather than the direct measure.

This question serving as a proxy for certainty level has two advantages compared to the direct question about the level of certainty: First, unlike the direct question, which may be affected by the treatment, this question is clearly independent of treatment. Second, the information conveyed through the answer to this question is observable by policy makers, unlike direct certainty levels, and may assist in tracking the population which is more likely to be affected by our design and save important resources. However, it also has a clear disadvantage in that it may be inaccurate. First, participants may not remember how many times they received the shot in the past 5 years. Second, it may be that uncertain participants opted not to get the shot more often than not due to the cost involved with getting the shot. Finally, a person may have changed his/her certainty level regarding his/her intention to receive the flu shot over the past 5 years and that may not be reflected well onto the certainty scale according to his/her answer to this question.

Pilot Study

The number of participants in each treatment is based on a pilot study and the effects found in that study. The pilot study was run during 2016 and 2017 and included 1,151 subjects. The pilot study included all 4 treatments plus control. It included all questions except the one about number of shots in the past 5 years. The question regarding direct certainty levels was also not the same as the one we will use in this study (it has been adjusted as we thought the previous version allowed the treatment to affect the answer since it referred to the decision made rather than

the general attitude towards influenza vaccination). Other minor wording changes have also been made.

The data in the pilot study was analyzed using the statistical tests described above: A two-way ANOVA, followed by a one-way ANOVA and Dunnett's test. We also conducted the regression analysis using the controls described above (except for the number of vaccination in the past 5 years which was not collected).

We found a 25% difference in intentions between the increased costs treatment and the control among the uncertain participants. A smaller effect of 16% for the uncertain subgroup was found between two of the other treatments (out of stock and benefit) and the control (but the number of uncertain participants in these treatments was very small, 12 and 24 participants only). The recommendation treatment, which is weakest in terms of the difference between the early and late shots, did not show any effect compared to the control. No significant effects were shown for the certain participants in any of the treatments.

Stopping Rule for Data Collection based on Pilot Study

We took the average expected effect (20.5%) and, accounting for multiple comparisons and the percentage of uncertain participants, we came up with 650 participants per treatment. We used the pwr package in R[1] and took alpha=0.05*/T* (divide by the number of tests T, i.e., Bonferroni correction. In our case *T=4* as we have 4 comparisons), and a power of 80%. The percentage of uncertain participants in the pilot was around 20%, and hence we multiplied the result of the required sample size by a factor of 1/0.2 to make sure we have enough "uncertain" subjects in each control/treatment group.

[1] Stephane Champely (2018). pwr: Basic Functions for Power Analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr