Analysis Plan

Project Name: Increasing flu vaccine uptake among veterans at the St. Cloud VA
Project Code: 1740
Date Finalized: May 1, 2018

This document serves as a basis for distinguishing between planned (confirmatory) analysis and any unplanned (exploratory) analysis that might be conducted on project data. This is crucial to ensuring that results of statistical tests will be properly interpreted and reported. In order that the Analysis Plan fulfill this purpose, it is essential that it be finalized and date-stamped before we begin looking at the data — ideally, before we take possession of the data. Once this plan is finalized, a date is entered above, and the document is posted publicly on our team website.

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

We expect the data to include the following variables:

- Treatment condition: postcard version = 1, 2, or 3
- Patient SID: a unique identifier used in random assignment
- Age: <50, 50–65, >65
- Rurality: U = urban, R = rural, H = highly rural
- ImmunizationDate: date of flu shot if one was received on or after 9/14/2017
- Vaccine_1yr: indicator of whether patient received a flu shot between 9/14/2016 and 9/14/2017
- Vaccine_GT1: indicator of whether patient received a flu shot between 9/14/2014 and 9/14/2016
- Previous_Flu_DX: indicator of whether patient has ever received a flu diagnosis, meaning any ICD10 code that contains "Influenza" in the name. This includes:
    - J09.X1   Influenza due to identified novel influenza A virus with pneumonia
    - J09.X2   Influenza due to identified novel influenza A virus with other respiratory manifestations
    - J09.X3   Influenza due to identified novel influenza A virus with gastrointestinal manifestations
    - J09.X9   Influenza due to identified novel influenza A virus with other manifestations
    - J10.00   Influenza due to other identified influenza virus with unspecified type of pneumonia
    - J10.01   Influenza due to other identified influenza virus with the same other identified influenza virus pneumonia
    - J10.08   Influenza due to other identified influenza virus with other specified pneumonia
    - J10.1   Influenza due to other identified influenza virus with other respiratory manifestations
    - J10.2   Influenza due to other identified influenza virus with gastrointestinal manifestations
    - J10.89   Influenza due to other identified influenza virus with other manifestations
    - J11.00   Influenza due to unidentified influenza virus with unspecified type of pneumonia
    - J11.08   Influenza due to unidentified influenza virus with specified pneumonia

- ○ J11.1 Influenza due to unidentified influenza virus with other respiratory manifestations
- ○ J11.2 Influenza due to unidentified influenza virus with gastrointestinal manifestations
- ○ J11.83 Influenza due to unidentified influenza virus with otitis media
- ○ J11.89 Influenza due to unidentified influenza virus with other manifestations

For reference, postcard versions are numbered as follows:

- Postcard version 1 - information only
- Postcard version 2 - social norms (motivation)
- Postcard version 3 - planning prompt (implementation)

See the Project Design Document for more detail on the postcard versions.

Outcome Variables to Be Analyzed:
We have two primary outcomes. The first is flu shot take-up, a binary indicator of whether the patient received a flu shot in the current flu season (defined as beginning on 9/14/2017, since postcards were mailed on or around this date). The second primary outcome, conditional on the first, is flu shot take-up time, or days elapsed between the beginning of the flu season (again, defined for this analysis as 9/14/2017) and the date when a flu shot was received. Both of these will be derived from variables in the raw data as described in the following section.

Transformations of Variables:
There will be three transformations. First, we will create an indicator of flu shot take-up based on whether 'ImmunizationDate' has a valid value (a date between 9/14/2017 and the end of the observation period). Second, conditional on flu shot take-up, we will create a variable representing flu shot take-up time, or days elapsed between the beginning of the flu season (defined for this analysis as 9/14/2017) and the date when a flu shot was received. Third, we will combine 'Vaccine_1yr' and 'Vaccine_GT1' to yield 'Vaccine_3yr,' a single indicator of whether the patient received a flu shot in the 3 years before the current flu season.

Imported Variables:
Based on preliminary data received from VA St. Cloud, we expect that the final outcome data will include treatment condition (postcard version = 1, 2, or 3). If not, we will import it from the crosswalk used during random assignment using 'Patient SID' as the merge variable.[1]

Data Exclusion:
We will exclude data for any individuals on the treatment assignment list that were not matched to VA health records, identified as records with missing age. (We assume here that any usable health record should include date of birth and thus age.) We do not anticipate any such exclusions.

Treatment of Missing Data:

---

[1]Even if treatment condition (postcard version) is included in the final data received from VA St. Cloud, we will still perform this merge to check that the assignment of 'Patient SID' to treatment conditions is identical to what was specified earlier in the crosswalk.

We expect no missing values on treatment condition, age, or rurality. On the flu shot and flu diagnosis variables, missing values will be interpreted as absence of a shot or diagnosis.

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Statistical Models:

Randomization Test

Before continuing with analysis, we will check the initial randomization by conducting d^2 omnibus balance tests using observable characteristics — in particular, age, rurality, and prior flu shot behaviors and flu diagnosis.

Treatment Effects

We will estimate the causal effect of the treatment (more specifically, intent to treat) using OLS regression. The two questions of interest are: Were either or both of the behaviorally informed postcards (postcards 2 and 3) more effective than the information-only postcard (postcard 1) in promoting flu shot take-up? Did people who got a flu shot after receiving either of the behaviorally informed postcards receive their flu shot earlier in the flu season, compared to people who got their flu shot after receiving the information-only card?

In our basic OLS specification, we will regress each of our primary outcomes — flu shot take-up (S_i) and flu shot take-up time (T_i) — on two indicator variables for assignment to treatment (postcard version 2 and postcard version 3, with postcard version 1 as a reference group), both with and without individual covariates representing age, rurality (potentially a proxy for proximity to a VA medical facility), and specific features of the patient's health history that might reflect propensity to get a flu shot. Coefficients on the treatment indicator variables will estimate the causal effects of sending the behaviorally informed postcards, as compared to the information-only postcard.

Thus, our main regression specification is:

$$S_i = \propto + \beta_2 P_{2i} + \beta_3 P_{3i} + X'_i \theta + \varepsilon_i$$
$$T_i = \propto + \beta_2 P_{2i} + \beta_3 P_{3i} + X'_i \theta + \varepsilon_i$$

where P_2 and P_3 are indicators for whether the individual was assigned to receive postcard 2 and postcard 3, respectively; X' is a vector of individual covariates; and ε_i is an idiosyncratic error term.

The individual covariates X' will represent

- age
- rurality
- whether the patient received a flu shot in the last 3 years ('Vaccine_3yr')
- whether the patient has a prior flu diagnosis on record ('Previous_Flu_DX')

We include these covariates to adjust for factors that might reflect pre-treatment differences in propensity to get a flu shot.

Note, to analyze effect on flu shot take-up time (T_i), we will restrict our analytical sample to those who received a flu shot in the current flu season. We will only conduct this analysis if there is sufficient power (80%) to detect a difference of 1 month in take-up time.

To analyze flu shot take-up (the binary outcome S_i), in addition to the OLS specification, we will run an analogous logistic regression. We do this in anticipation of publishing in outlets for which logistic regression is the norm. If the OLS and logistic regressions yield substantively different results, the OLS results will take precedence. We will use Freedman's (2008) plug-in estimator in order to report differences in log-odds among treatment conditions.[2]

Standard Error Adjustments

In our primary analysis, we will estimate heteroskedasticity robust (HC2) standard errors.

Follow-Up Analyses:

Difference Between Treatment Arms

If we find an effect of either or both of the behaviorally informed postcards (as compared to the information-only postcard), then we will test the difference between the two ($\beta_2 = \beta_3$).

Inference Criteria, Including Any Adjustments for Multiple Comparisons:

We will use standard inference criteria. We will use two-tailed tests and three threshold *p*-values: 1%, 5%, and 10%. Given the very small cost of the intervention and the large health benefits of getting a flu shot, any effect we can detect statistically is likely policy relevant.

To account for multiple inferences, we will use the Holm-Bonferroni method and adjust for two inferences in a first step ($\beta_2 = 0$ and $\beta_3 = 0$) followed by one inference in a second step ($\beta_2 = \beta_3$), where the second step is performed only if at least one of the inferences in the first step yielded a significant result.

Limitations:

---

[2] Freedman DA (2008) Randomization does not justify logistic regression. *Statistical Science 23*: 237–249..

There are at least two main limitations for this study. First, while we have a few covariates to test balance and heterogeneous effects, we do not have access to covariates related to income or other potentially relevant factors. Second, our measurement of flu shot take-up is limited to veterans who receive their flu shot from a VA facility. Flu shots received from any non-VA source (e.g. a local pharmacy) may not be captured within our data. In other words, strictly speaking, we may be measuring the effect of treatments on getting flu shots in VA facilities, not on getting flu shots in general.

Exploratory Analysis:

Heterogeneous treatment effects

Power calculations for our main analysis suggest that our study is not powered to identify heterogeneous effects. Nonetheless, we will test for several sets of heterogeneous effects in case the study has a larger than expected sample size.

If we find an effect of either or both of the behaviorally informed postcards (as compared to the information-only postcard), then we will test for heterogeneous treatment effects by adding in two-way interactions between treatment assignment and each of covariates listed above.

Flu shot take-up time

In our main analysis (above), we analyze flu shot take-up time conditional on receipt of a flu shot. To address the fact that this analysis will be conditional on a post-treatment outcome, we will also conduct an exploratory analysis using survival analysis techniques and treating patients who never received a flu shot during the observation period as right-censored. We might also conduct an additional exploratory analysis using an instrumental variables approach, in which treatment assignment is the instrument and take-up time is regarded as a measure of compliance with treatment.

Recency of prior flu shot

If we find a significant interaction between either treatment assignment variable and prior flu shots ('Vaccine_3yr'), then we will attempt to assess whether recency of the prior flu shot is relevant by substituting two separate indicators for whether the patient received a flu shot last year ('Vaccine_1yr') and whether the patient received a flu shot two or three years ago ('Vaccine_GT1').

Downstream impact on flu diagnosis

If the data include an indicator for flu diagnosis in the current flu season, then we will conduct an exploratory analysis of whether rates of flu diagnosis might also have differed across treatment conditions as a consequence of vaccination.