



## Analysis Plan

Project Name: Increasing flu vaccine uptake among veterans at the New York Harbor VA

Project Code: 1743

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: treatment (1); control (0)
- Patient SID: a unique identifier
- 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 10/1/2017
- Vaccine\_1yr: indicator of whether patient received a flu shot between 10/1/2016 and 9/30/2017
- Vaccine\_GT1: indicator of whether patient received a flu shot between 10/1/2014 and 9/30/2016
- 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

See the [Project Design Document](#) for more detail on the intervention.

### **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 10/1/2017). 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 10/1/2017) 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 10/1/2017 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 10/1/2017) 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.

### **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:**

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 primary questions of interest are: Were the emails influential in increasing flu shot take-up? Did people who got a flu shot after being sent an email receive their flu shot earlier in the flu season, compared to people in the control condition who got a flu shot?

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 an indicator variable for assignment to treatment, 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.

Thus, our main regression specification is:

$$\begin{aligned} S_i &= \alpha + \beta E_i + X'_i \theta + \varepsilon_i \\ T_i &= \alpha + \beta E_i + X'_i \theta + \varepsilon_i \end{aligned}$$

where  $E_i$  is an indicator for whether the individual was assigned to receive the email reminder treatment;  $X'$  is a vector of individual covariates; and  $\varepsilon_i$  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.<sup>1</sup>

### *Standard Error Adjustments*

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

### **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 adhering to antiretroviral therapy, any effect we can detect statistically is likely policy relevant.

We will not perform any adjustment for multiple comparisons.

### **Limitations:**

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.

### **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 -- by adding interaction terms to our primary specifications to represent interaction between treatment and the covariates listed above.

#### *Timing effects*

We will explore whether emails affected the timing of flu shots (e.g., Did more people get their shots in the weeks immediately following the email sends?)

#### *Survival Analysis*

Additionally, the structure of the data and the multiple treatment reminders sent at semi-regular intervals (three times during the sample period), will allow us to also explore the application of a survival analysis technique. We anticipate using a cox proportional hazard rate to model the impact of treatment reminder for various subgroups (i.e. older population, prior flu shot recipients, etc.)

---

<sup>1</sup> Freedman DA (2008) Randomization does not justify logistic regression. *Statistical Science* 23: 237–249..

### *Prior Flu and Flu Shot Experiences*

After receiving the data, we will explore the possibility of extending our follow-up analyses to include an interaction term that estimates the differential effect of the treatment on those prior flu shot recipients who were diagnosed with the flu. This population may have a negative perception of the effectiveness of flu shots and those may be less likely to engage in future flu shot behaviors.

### *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').