Analysis Plan

Project Name: Encouraging flu vaccine uptake among pregnant patients at Duke

Project Code: 1735

Date Finalized: 4/23/18

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.

Outcome Variables to Be Analyzed:

The primary outcome is receipt of a flu vaccine (as recorded in Duke's electronic health record (EHR) system).

We will also analyze secondary outcomes including date of the flu vaccine (time elapsed since the initiation of the flu season), receipt of the TDAP vaccine, and the number of antenatal care visits.

Transformations of Variables:

The only transformation required will be the creation of the variable "time elapsed since the initiation of the flu season" using the date that the vaccine was provided. For this analysis, October 1 will be coded as the beginning of the flu season.

Imported Variables:

The treatment was administered in two waves and with clinic-level blocking in each wave. Accordingly, indicator variables for block (equal to the interaction of clinic and wave) will be constructed. The data is generated by Duke in a series of spreadsheets corresponding to each block; accordingly, an additional merge is not necessary, but we can code the blocks accordingly.

Treatment of Missing Data:

If the variable for flu vaccine receipt is missing, the observation will be dropped from the sample given that this is the primary variable of interest. We do not anticipate any missing values here, however. We are aware that the variable corresponding to the date of the vaccine is manually imported from patients' charts, and may be missing in some cases. Accordingly, we will estimate the specification for "time to flu shot" using the restricted sample of observations for which dates are reported. In the event that the number of missing observations exceeds 10%, or is statistically different between treatment and control, we will caveat this analysis appropriately and also explore constructing bounds on any effect.

In addition, we have requested that Duke provide information about flu vaccine receipt in the previous year. The availability of this data is somewhat uncertain and there may be a large number of missing values. This variable is intended as a covariate to be used in a secondary specification. We will proceed with this specification if the number of missing observations does not exceed 20%. However, this analysis should be considered exploratory and will be reported as such. We will also explore including dummy variables for missing values as covariates.

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:

Our primary specification will be ordinary least squares (OLS). We will regress the four outcomes of interest on a dummy variable for treatment assignment, conditional on clinic-wave strata fixed effects.

We will also explore re-estimating this specification controlling for past vaccination history (a dummy variable indicating whether the respondent received the flu vaccine in the previous year) if this variable is reported for at least 80% of the sample. This analysis should be considered exploratory and will be labeled as such.

Finally, we will also re-estimate the primary specifications for receipt of the flu and TDAP vaccine using logit models, given that these are binary variables. In the event the results differ between the two models, the OLS results will be considered primary.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:

We will be conducting two-tailed tests and will report results that are significant at the 10, 5 and 1 percent level. Given that the intervention is extremely cheap, no minimum effect size of interest has been specified by the partner.

Limitations:

- Some patients are not eligible to get the vaccine in Duke. This primarily applies to the undocumented and/or uninsured population. Uninsured women can get public coverage for specific periods in their pregnancy, but this may not coincide with the time when they get this message. Ineligible women are referred to a public health clinic (a few miles from the Duke clinic; accessible but not particularly convenient on public transit, limited open hours). This limitation also means that the "default" aspect of the circulated message is relatively weak: we cannot make a statement that the woman is scheduled to receive a flu vaccine (this may not be true), and can only say that they are "due" the vaccine at their next visit.
- Some patients may get the flu vaccine outside Duke but not report it to their doctor or their doctor may not enter it in the EHR system correctly. This would primarily reduce the available pool of women who could switch into receiving the flu vaccine. In addition it may have several potential effects on what we measure/observe, which we will not be able to tell apart:

  Women may get the flu vaccine elsewhere as a result of our message, who otherwise may not have gotten the vaccine (and who don't tell their doctor). This would downward bias our observed effect size.

  Women who would get the flu vaccine elsewhere anyway may become more likely to tell their doctor about this and get it properly recorded (increasing observed effect size)

- We do not know the open rates of MyChart messages. There will almost certainly be some messages that are never read.