Analysis Plan

Project Name: Encouraging uptake of the recommended vaccine sequence for seniors in Louisiana

Project Code: 1738

Date Finalized: February 12, 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.

Outcome Variables to Be Analyzed

Main outcome of interest:

- Proportion of vaccinations received out of overdue vaccinations (ie. if individual is missing flu and Zoster, and receives Zoster only, the outcome is 0.5); interval [0,1]
    - Difference in means

> **Clarification in February 28, 2018:**
> Equation for outcome = # of received vaccinations/3.
>
> (We use 3 vaccinations as the full set of vaccinations in this analysis plan. This outcome gives us the proportion of received vaccinations out of all vaccinations.)

Other main outcomes of interest:

- Vaccinated for flu; dichotomous [0,1]
    - Difference in proportions
- Vaccinated for Zoster; dichotomous [0,1]
    - Difference in proportions
- Vaccinated for TD; dichotomous [0,1]
    - Difference in proportions
- Vaccinated for pneumonia; dichotomous [0,1]
    - Difference in proportions

Indices:

- Proportion of vaccinations received out of overdue vaccinations (ie. if individual is missing flu and Zoster, and receives Zoster only, the outcome is 0.5).
    - This outcome is created by dividing the number of vaccinations received by the vaccinations missing.

- Interval [0,1]

Clarification in February 28, 2018:
Equation for outcome = # of received vaccinations/3.

(We use 3 vaccinations as the full set of vaccinations in this analysis plan. This outcome gives us the proportion of received vaccinations out of all vaccinations.)

Statistical Models:
To estimate the ATE, we compare the number of vaccines that are received when an individual is in his/her last month in the control group, vs. the number of vaccines that are received when the individual is in his/her last month in the treatment group in the study.[1]

For more precision, we account for the month when the individual receives treatment and the length of time in which the individual is in the treatment group. We make treatment month (`Z_month`) a blocking variable. In this study individuals receive the treatment in October, November, December, or January, and therefore, are in the treatment condition for 1, 2, 3, or 4 months.

For more precision, we also account for the vaccination history of the individual. We make vaccination history into a blocking variable called (`vac_block`). We make a `month_block` variable that is `Z_month` and `vac_block` as a final blocking variable. (See "design variables" section in Transformations.) This blocking variable allows us to adjust for both length of time in treatment and vaccination effects to estimate the ATE.

Change made in February 28, 2018. (After analysis started.)

Issue: Time trends in the study may inflate results from the study. For example, as more time goes by, more individuals receive vaccinations, regardless of treatment or control. *Z_month* as blocking variable cannot remove time trends and therefore does not give an estimate that adheres to the objectives of the study.

Change to blocking variable: The new blocking variable is no longer *z_month*, but *month*, which is an indicator for the month of the observation. We increase precision and remove the time trend by measuring the effect of treatment and control individuals within month.

The month_block blocking variable is the combination of the month of observation and the block of the individual.

Change to dataset: We use the entire dataset and no longer consider only the last month that individuals are in control vs. the last month that individuals are in treatment.

[1] See Gerber and Green (2012) for estimating the combined immediate and lagged effect (p.280).

Specification for main outcome:

$$Y_{it} = \beta_0 + \beta_1 Z_{it} + \beta_2 X + \varepsilon_{it}$$

Where

- Y = vaccinations received out of overdue vaccinations
- *X* = blocking variable `month_block`
- *i* = individual
- *t* = last month when the individual is in control and the last month of the experiment.
- β_1 = average treatment effect

Change in February 28, 2018.

- Y = vaccinations received out of all vaccinations
- *X* = blocking variable `month_block`
- *i* = individual
- *t* = month of observation
- β_1 = average treatment effect

We will run the same specification above for each of the vaccines as well, where *Y* is whether an individual received a specific vaccine. We will use the Holm-Bonferroni method to deal with the multiple comparisons.

Transformations:

- Panel data format: we will be transforming the data received into panel data format. In the current format, each observation is by individual denoted by an unique identifier. Variables for each individual include the dates for when they received each of the 4 vaccines, and a month indicator recording the month that the individual received the postcard reminder. The dataset will be transformed so that each observation is by individual/month.
    - Note: For the main analysis, we will not use the entire dataset, but only the last control month and the last treatment month of the study per individual. This is because we want the average treatment effect, which is the combined immediate and lagged effects. Therefore, we want to know if an individual is up-to-date right before he/she moves into treatment, and then again when he/she has been in treatment for at least one month. With this set up, we keep all of the information about whether an individual receives in vaccine in treatment or control.

      Change in February 28, 2018.
      We use the entire dataset with the change in blocking variable.

    - For follow up analysis, we will use the entire dataset because we want to compare between months.

- Additional indicator variables for vaccine:

    - If vaccine is up-to-date, then 1; otherwise 0.

- Additional indicator variable for last month in control and last month in experiment:
    - If month is last month that the individual is in control, 0. If month is last month that individual is in treatment, 1. Otherwise, NA.

- Design variables for randomization: Prior to fielding the experiment, we put individuals into 4 blocks according to their vaccination history. Individuals were then randomized by block to a treatment-month. `Z_month` is the treatment-month variable; `block` is the blocking variable. Blocks are determined by
    - Vaccinator: up-to-date on flu (last year), shingles, tetanus (n=7,657)
    - Non-vaccinator: overdue on flu, shingles, tetanus (n=99,669)
    - Partial vaccinator (all but flu): up-to-date on shingles, tetanus, overdue on flu (n=5,457)
    - Partial vaccinator (mix): mix of up-to-date/overdue on flu, shingles, tetanus (n=96,084)

- Other design variables: For the purposes of clarity in analysis, we will call `block`, `vac_block`. We make a `month_block` variable that is `Z_month` and `vac_block` as a final blocking variable for the main analysis.

Follow-Up Analyses:
The experimental design was set up as a block-randomized, stepped-wedge design. In addition to the main analysis, we will do the following analyses with the entire dataset.

Comparison between blocks: We will compare ATE between blocks because we expect heterogeneous effects between blocks. We expect individuals who are partial vaccinators (all but flu) to be most likely to vaccinate. We expect non-vaccinators to be the least likely to vaccinate.

Specifications:

$$Y_{it} = \beta_0 + \beta_1 Z_{it} + \beta_2\, block + \beta_3 (Z\ x\ block) + \varepsilon_{it}$$

Where

- Y = vaccinations received out of overdue vaccinations
- *block* = blocking variable `block` based on vaccination history
- *i* = individual
- *t* = month
- $\beta_{\{3,\ldots,n\}}$ = differences in ATE by blocks

Comparison between months: We will compare ATE between months to determine if the postcard has greater effects in different months. We expect that the postcard will have a greater effect in the early flu season (October) than the late flu season (December).

Specifications:

$$Y_{it} = \beta_0 + \beta_1 Z_{it} + \beta_2 Z_{month} + \beta_3 (Z \; x \; Z_{month}) + \varepsilon_{it}$$

Where

- Y = vaccinations received out of overdue vaccinations
- *Z_month* = blocking variable `Z_month`
- *i* = individual
- *t* = month
- $\beta_{\{3,\ldots,n\}}$ = differences in ATE by months

Change in February 28, 2018

Specifications:

$$Y_{it} = \beta_0 + \beta_1 Z_{it} + \beta_2 \, month + \beta_{\{3,\ldots,n\}} (Z \; x \; month) + \varepsilon_{it}$$

- Y = vaccinations received out of overdue vaccinations
- *month* = blocking variable `month`
- *i* = individual
- *t* = month
- $\beta_{\{3,\ldots,n\}}$ = differences in ATE by months

Immediate vs. lagged effects: We will compare ATE immediately after the postcards are sent out and a month after the postcards are sent out. This comparison is similar to our first specification, but we are first comparing individuals in the month before and after they receive a postcard (immediate effect). Then we compare individuals who have received a postcard for 2 months vs. 1 month. and in the months (lagged effect). We expect that the immediate effect will be larger than the lagged effect.

Inference Criteria:

HC2 standard errors. Clusters at the individual level.

For all analyses, we need to do a FWER adjustment (Holm-Bonferroni method).

Data Exclusion:

All data will be used in the study.

Treatment of Missing Data:

A large percentage of our data include NAs. The "no data" records in the baseline data report below indicate the number of NAs per vaccine. After asking Louisiana, we learned that the NAs means that the registry has no data about that person for that vaccine.

Some reasons we think that the Louisiana Immunization Registry does not have data on these individuals include: 1) they do not receive their vaccines in Louisiana, or 2) their doctors/pharmacists do not report the vaccines to the Registry, or 3) they have never had vaccines.

We plan to keep NAs in the study, and they will be treated as 0s in the study. Treating an NA as a 0 means that we assume individuals with NAs are not up-to-date on certain vaccines. We think this is valid because:

1) The Louisiana Immunization Registry treats individuals with NAs as if they were not up-to-date on a vaccine, and
2) Assuming that they did not receive the vaccine will give us a conservative estimate of the effect of the postcard.

Note that by considering NAs as 0s, we are taking the average treatment effect of people who take up the vaccine in a recorded way vs individuals who do not take up the vaccine or take it up in a non-visible way to us.

- Up-to-date on flu vaccine: 68,259 individuals (33% of sample)
  - Last flu shot between 9/1/2016 - 9/30/2017
  - No data: 57,051
- Up-to-date on TD/Tdap vaccine: 44,413 individuals (21% of sample)
  - Last TD/Tdap booster between 9/1/2007 - 9/30/2017
  - No data: 145,498
- Up-to-date on shingles vaccine: 39,551 individuals (19% of sample)
  - Received the shingles vaccine
  - No data: 169,316
- Up-to-date on PCV13: 20 individuals (.009% of sample)
  - Received the PCV13
  - No data: 208,846
- Up-to-date on PPSV23: 52,451 individuals (25% of sample)
  - Received the PPSV23
  - No data: 156,416

Limitations:
We do not know anything about our experimental population except that they are between the ages of 65-70 and are not up-to-date for at least 1 vaccine. We have no demographic information and cannot learn more about how demographics, locations, etc. might affect the take up of the treatment or the likelihood of vaccination.

Exploratory Analysis (Optional):
None for now. See follow-up analysis.

Link to an Analysis Code/Script (Optional):
Randomization code sent to Louisiana