BE IMMUNE: Behavioral Economics to IMprove and Motivate vaccination Using Nudges through the EHR

University of Pennsylvania

Dr. Shivan Mehta
Dr. Amol Navathe
Dr. Joshua Liao

NCT#06057727

Statistical Analysis Plan

01/07/2025

Objectives

Our primary objective is to test the effectiveness of multicomponent, personalized EHR-based nudges to both clinicians and patients for increasing flu vaccination completion among eligible primary care patients.

Relevant Study Design Details

Randomization

Primary care clinics will be randomized 2:1 to either a nudge intervention or usual care arm using covariate-constrained randomization, stratified by health system (Penn, UW). For the clinic level randomization, the covariate-constrained randomization will aim to balance the distribution of clinician panel baseline flu vaccination completion rate, panel size, and panel risk. Panel size is defined as the number of patient visits from September 1, 2022 through February 28, 2023. Panel risk is defined as the proportion of a clinician's patients who meet at least one of the following criteria: age ≥ 70 years, Non-Hispanic Black race, residence in lower income communities (lowest quartile according to zip code), or no documented history of flu vaccine in the year prior. Second, patients who have an eligible appointment at an intervention clinic and are identified as high-risk for not completing a flu vaccination will be further randomized using a 1:1 ratio at the patient level—stratified by clinic—to receive either an additional, intensification nudge (bi-directional texting) or the standard nudge intervention (standard text messaging).

Blinding

All investigators, statisticians, and analysts will be blinded during the randomization, intervention, and analysis phases. They will be unblinded at the end of the study analyses, after all outcomes have been obtained.

Outcomes

The primary outcome is flu vaccination completion during the first eligible primary care visit. Secondary outcomes are flu vaccination completion within 3 months of the first eligible primary care visit and flu vaccination ordering. These patient-level outcomes will be coded as a 1 to indicate completion (or signed order) and 0 to indicate non-completion (or unsigned/cancelled order).

Statistical Analysis

The primary analysis for all primary and secondary outcomes will follow the intent-to-treat principle. Sensitivity analyses (described later) will include per-protocol and other approaches to evaluate robustness to protocol deviations for some patients.

Analytical Plan for Primary and Secondary Outcomes

The primary analysis will fit generalized estimating equations (GEE) with a patient-level binary outcome indicating flu vaccination completion at the eligible primary care visit (primary outcome) and within 3 months (secondary outcome) of the eligible visit. We will fit a single GEE model to estimate the following: 1) effect of the clinic-level intervention (pended orders) only versus usual practice (main effect averaging over individually randomized bi-directional versus standard texting for high-risk patients at intervention clinics), and 2) effect of bi-directional vs standard text messaging amongst high-risk patients seen at intervention clinics, and 3) effect of the multicomponent intervention (pended orders and intensive texting for high-risk patients) compared to usual practice. These effects of interest are analogous to comparisons of intervention sequences embedded within Sequential Multiple Assignment Randomized Trial (SMART) designs. Therefore, we will adapt statistical methods commonly used for the analysis of data from SMART designs and use them to analyze data from the trial. More specifically, we will adapt the weighted and replicated approach that is used to compare embedded treatment sequences in SMART designs.[1]

The weighted and replicated method is an analytic trick that allows for standard statistical software to be used to fit a single model for statistical estimation and inference regarding multiple effects of interest. Only high-risk patients seen at active intervention clinics are randomized to standard or intensive text messaging. Low-risk patients and all patients seen at control clinics are only exposed to the clinic-level randomized assignment. As a result, they do not have an individual-level randomized assignment in the analytic data set. To fit a single model, we must assign these individuals a value for the individual-level randomization. To do so, we first replicate their row of data and assign one of their records to standard messaging and the other to intensive messaging. However, after replicating these individuals, they will be over-represented in the analytic data set compared to high-risk, intervention clinic patients who received an individual-level randomized assignment and are therefore not replicated. To re-balance the representation in the pseudo-population created by replicating some patient observations, we weight each row by the inverse of the probability of receiving the observed value of the individual-level intervention. That is, patients who were truly randomized to standard or intensive texting will receive a weight of 2 (inverse of ½) while each row of a patient whose record was replicated will receive a weight of 1 (since their intervention "assignments" were deterministic).

Once the replicated data set is created, we will fit the following GEE model to the patient-level binary outcomes using the observation weights mentioned above:

$$logit\,(\Pr(Y = 1\,|A1,\,A2,\,X)) = \beta_0 + \beta_1 A_1 + \beta_2 A_1 A_2 + \beta_3^T X, \qquad (1)$$

where A_1 is an indicator of whether the patient was seen at an active (1) or control (-1) clinic; A_2 is an indicator of intensive (1) or standard (-1) text messaging (patients who were actually randomized will have only one value for A_2, while replicated patients will have one 1 and one -1); $A_1 A_2$ is the interaction between the patient's cluster assignment and their individual-level assignment(s); and X is a vector of all other covariates we want to control for which will at minimum include mean-centered versions of the three variables used in the constrained randomization procedure: 1) clinic size (number of patient visits); 2) baseline clinic risk (proportion of panel deemed high risk); and 3) baseline clinic completion rate (proportion of panel with completed flu vaccination).

We will specify an independence working correlation when fitting the primary GEE model. Although we expect some amount of correlation amongst observations within clinic, the replicated subject observations should be treated as independent. Fortunately, estimates from GEE models are robust to misspecification of the working correlation structure, meaning they are consistent for the true values even when the working correlation is specified incorrectly. Furthermore, when correlation amongst observations within clinic is low, the efficiency of using an exchangeable working correlation is similar to or worse than when using an independent working correlation. Since we will randomize 45 clinics, we will not use a small cluster correction.[2] We will use this approach for both the primary and secondary outcome analyses. Given model (1) with effect coding (i.e., -1 and 1 for levels of A_1 and A_2), we have the following:

1. The main effect of A_1 is $2\beta_1$ (i.e., the main effect of clinic-level nudges averaging over the differential effects of standard and intensive texting).
2. The main effect of A_2 given $A_1 = 1$ is $2\beta_2$. This is the comparison of text messaging intensity (standard versus intensive) given clinic-level nudges are active.
3. The comparison of clinic-level nudges with intensive texting versus control is $2\beta_1 + \beta_2$.
4. The comparison of clinic-level nudges with standard texting versus control is $2\beta_1 - \beta_2$.

We will test each of the four linear contrasts of the parameters estimated from the GEE model. We will use a Bonferroni correction for the first two contrasts, aligning with how the study was powered (i.e., two-sided significance level of 0.025 for the first two contrasts). We will consider contrasts 3 and 4 as exploratory, and thus we will not correct the corresponding p-values for multiple testing. Results will be reported on the odds ratio scale, including 95% confidence intervals. Confidence intervals for contrasts 1 and 2 will incorporate the Bonferroni correction used for the two tests.

Sensitivity Analyses to Evaluate Model Specifications

As a sensitivity analysis, we will restrict the data to high-risk patients at active clinics and estimate the main effect of intensive vs standard text messaging, specifying an exchangeable working

correlation within clinics. As another sensitivity analysis, we will use an exchangeable working correlation at the clinic level after duplicating each clinic's data, assigning one 1 and one -1 for patients without an individual-level assignment, and removing the record for each high-risk patient an intervention clinic corresponding to the opposite of the individual-level randomized assignment that was received. Duplicated clinics will be treated as independent while observations within clinic will be treated as exchangeable. In this way, duplicated records will be treated as independent. Using this approach, we will re-test the main effect of A_1, i.e., $2\beta_1$, to evaluate sensitivity to specification of the working correlation. We will use a t-distribution with degrees of freedom equal to the number of clinics (47) minus the number of mean model parameters to construct a confidence interval for $2\beta_1$. If the interval does not include 0 (log odds ratio scale) we will reject the null and conclude there is a significant effect of clinic nudges compared to usual care.

Sensitivity Analyses to Evaluate Protocol Deviations

Patients observed in the first six weeks at Penn did not receive the full intervention due to a switch from using API (Application Programming Interface) to RPA (Robotic Process Automation). Specifically, patients at intervention clinics at Penn did not receive pended orders. In our primary analyses, we will include these patients as having received pended orders, following the intent-to-treat principle. In sensitivity analyses, we will repeat our analyses of primary and secondary outcomes after excluding this group who did not receive the full intervention. To mitigate any confounding effects of time (i.e., patients more likely to complete vaccinations earlier in the season) we will also exclude all Penn control arm patients during the first six weeks of the trial. Since this issue was not encountered at UW, all UW patients during this six week time period will be included in this per-protocol analysis.

Next, we will leverage this subset of intervention clinic patients who did not receive pended orders to estimate the effect of texting alone (standard texting versus control and intensive texting versus control) on completion at the day of eligible visit. We will also leverage this subset of patients to estimate the effect of ordering amongst low-risk patients who received text messaging as well as amongst high-risk

subjects who received either standard or intensive text messaging. There was a different subset of patients in intervention clinics who did not receive text messaging. We will utilize this group to estimate the effect of ordering alone versus control. These analyses will be exploratory, and their feasibility will depend on the number of patients in each group who did not receive the full planned intervention components. Analyses will mimic the approach specified for the primary and secondary outcomes (e.g., GEE, weighting and replicating when appropriate).

Analytical Plan for Exploratory Analyses

To examine the effect of the intervention on flu vaccination ordering, a GEE with a binary outcome indicating whether the pended order was signed or cancelled will be fit. The rates of flu vaccination completion will be compared between (1) usual care, (2) clinic level nudge only, and (3) clinic-level nudges with high-risk intensification using the same weighted and replicated approach described for the primary outcome and the mean model given in equation (1). We will again use an independence working correlation for these comparisons and perform sensitivity analyses to assess sensitivity to specification of the working correlation.

Exploratory subgroup analyses will examine the effect of the clinic-level nudge versus usual care among patients who are 1) identified as low-risk for completion and 2) identified as high-risk for completion, and 3) seen at Penn versus UW. For (1), we will restrict the data set to low-risk patients in the active and control clinics. Then, we will fit a GEE model with an indicator for clinic randomized assignment and variables used in the covariate constrained randomization. We will fit the model with a logit link and exchangeable working correlation. For (2), we will subset the data to patients deemed high-risk in both active and control arms, we will fit a GEE model with an indicator for clinic randomized assignment, indicator for individual-level randomized assignment, interaction between the clinic- and individual-level randomized assignments, and variables used in the covariate constrained randomization. We will use the weighted and replicated approach described previously, replicating each high-risk patient in the control arm and assigning them each a weight of 1 while assigning each high-risk patient in the

intervention clinics a weight of 2. We will fit the model with a logit link and independent working correlation. We will test the main effect of the clinic-level nudge and the main effect of standard versus intensive texting. We will test contrasts 1), 3), and 4). For investigating possible effect heterogeneity by health system, we will re-fit models for our primary and secondary outcomes separately on data from Penn and UW.

To investigate potential effect heterogeneity over time, we will fit GEE models for our primary and secondary outcomes with the addition of month and month-by-treatment interactions. The model will take the following form:

$$logit\ (\Pr(Y = 1\ |A_1, A_2, T, X)) = \beta_0 + \beta_1 A_1 + \beta_2 A_1 A_2 + \ \beta_3 T + \beta_4 A_1 T + \beta_5 A_1 A_2 T + \beta_6^T X, \qquad (2)$$

where T is a categorical variable indicating study month. We will use the weighted and replicated approach and an independent working correlation structure. We will first assess whether the effect of text messaging style differs over time. If we do not find evidence of effect heterogeneity for text messaging style, we will drop A_2 from the model and focus only on the effect of pended orders over time using the following model:

$$logit\ (\Pr(Y = 1\ |A_1, A_2, T, X)) = \beta_0 + \beta_1 A_1 + \beta_2 T + \beta_4 A_1 T + \beta_5^T X, \qquad (3)$$

Inference will focus on the interaction effect, β_4, and we will not use any weighting or replicating to fit this model. We will fit this model using an independent working correlation.

Lastly, we will determine which high-risk intensification subgroups experienced the greatest benefits of the clinic- and individual-level nudges. Using the data set restricted to high-risk individuals, we will fit the following model:

$$logit\ (\Pr(Y = 1\ |A_1, A_2, Z, X)) = \beta_0 + \beta_1 A_1 + \beta_2 A_1 A_2 + \ \beta_3 Z + \beta_4 A_1 Z + \beta_5 A_1 A_2 Z + \beta_6^T X, \qquad (4)$$

where Z is a single, high-risk criteria. We will fit model (2) separately for each high-risk criteria which will allow us to determine which high-risk intensification subgroups experienced the greatest benefits of the clinic- and individual-level nudges. We will again use the weighted and replicated approach,

duplicating each control arm patient and assigning them a weight of 1 while assigning each intervention clinic patient a weight of 2. We will use an independent working correlation structure.

We now provide details about the interaction effects of interest. First, note that the main effect of A_1 amongst $Z = 1$ is:

$$\beta_0 + \beta_1 + \beta_2 A_2 + \beta_3 + \beta_4 + \beta_5 A_2 + \beta_6^T X - (\beta_0 - \beta_1 - \beta_2 A_2 + \beta_3 - \beta_4 - \beta_5 A_2 + \beta_6^T X)$$

$$= 2\beta_1 + 2\beta_4$$

Similarly, the main effect of A_1 amongst $Z = 0$ is:

$$\beta_0 + \beta_1 + \beta_2 A_2 + \beta_6^T X - (\beta_0 - \beta_1 - \beta_2 A_2 + \beta_6^T X)$$

$$= 2\beta_1$$

Thus, $2\beta_4$ captures the additional effect of A_1 amongst high-risk individuals with $Z = 1$ relative to those with $Z = 0$. We will test this quantity ($2\beta_4$) for statistical significance separately for each high-risk category. Next, note that the main effect of A_2 amongst $Z = 1$ setting $A_1 = 1$ (i.e., within intervention clinics) is:

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6^T X - (\beta_0 + \beta_1 - \beta_2 + \beta_3 + \beta_4 - \beta_5 + \beta_6^T X)$$

$$= 2\beta_2 + 2\beta_5$$

Similarly, the main effect of A_2 amongst $Z = 0$ with $A_1 = 1$ is:

$$\beta_0 + \beta_1 + \beta_2 + \beta_6^T X - (\beta_0 + \beta_1 - \beta_2 + \beta_6^T X)$$

$$= 2\beta_2$$

Thus, $2\beta_5$ is the additional effect of A_2 amongst high-risk individuals with $Z = 1$ relative to those with Z $Z = 0$. We will test this quantity ($2\beta_5$) for statistical significance separately for each high-risk category. We will not correct for multiple testing in any of the effect modification exploratory analyses.

References

1. Almirall, D., Nahum-Shani, I., Sherwood, N. E., & Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Translational behavioral medicine*, *4*(3), 260–274.
2. Liang, K.-Y., & Zeger, S. L. (2000). Longitudinal Data Analysis of Continuous and Discrete Responses for Pre-Post Designs. Sankhyā: The Indian Journal of Statistics, Series B (1960-2002), 62(1), 134–148.