

I-SCREEN: Increasing Screening for Cancer using a Randomized Evaluation of EHR-based Nudges

University of Pennsylvania

Dr. Amol Navathe

Dr. Shivan Mehta

Dr. Joshua Liao

NCT# 06177795

Statistical Analysis Plan

07/02/2025

Objectives

Our objective is to test the effectiveness of multicomponent, personalized EHR-based nudges to both clinicians and patients for increasing screening mammogram completion among eligible primary care patients. Secondly, we are interested in determining if EHR-based nudges increase screening mammogram completion among high-risk subgroups and whether response to the multicomponent intervention varies across different types of clinics, clinicians, or patients.

Relevant Study Design Details

Interventions

30 primary care clinics will be randomized at Penn. Clinicians within the clinics assigned to intervention will receive a default pended mammogram order if their primary role is physician or advanced practice provider. Residents and fellow physicians within the intervention clinics will also receive the automated, default pended order nudge. Patients seen at intervention clinics at Penn who are identified to be at high risk of non-completion based on a model built from historical data will be randomized to receive standard text messaging or more intensive bi-directional text messaging.

At UH, 82 primary care providers will be randomized to intervention or usual care. Providers assigned to the intervention will receive a default pended mammogram order if their primary role is physician or advanced practice provider. Residents and fellow physicians within the intervention clinics will also receive the automated, default pended order nudge. There is no high-risk intensification and therefore no individual patient-level randomization at UH. Data collected from Penn and UH will be analyzed separately.

Randomization

At Penn, primary care clinics will be randomized 2:1 to either a nudge intervention or usual care arm using covariate-constrained randomization. The covariate-constrained randomization will aim to balance the distribution of clinic panel baseline mammogram completion rate, panel size, and panel risk

across the intervention and usual care arms. Baseline completion rate, panel size, and panel risk will be calculated using historical data extracted from the EHR from May 1, 2023 to October 31, 2023. Baseline completion rate is defined as the proportion of eligible patients at a clinic who received a screening mammogram within 3 months of the eligible visit during this time. Panel size is defined as the number of unique eligible patients seen by the clinic during this time. Panel risk is defined as the proportion of a clinician's patients with Medicare and/or Medicaid insurance, no active EHR patient portal account, or no log-in to the patient portal in the prior year. Patients who have an eligible appointment at an intervention clinic and are identified as high-risk for not completing a screening mammogram will be further randomized using a 1:1 ratio at the patient level to receive either an additional, intensification nudge (bi-directional texting) or the standard nudge intervention (standard text messaging) (**Figure 1A**).

At UH, there will be only one level of randomization. Clinicians will be randomized using a 1:1 ratio to either a nudge intervention or usual care arm using covariate-constrained randomization (**Figure 1B**). Here, the covariate-constrained randomization will aim to balance the distribution of clinician panel baseline mammogram completion rate, panel size, and panel risk, where these variables are computed using data from March 15, 2023 to September 15, 2023 extracted from the EHR. At UH, panel size is defined as the number of unique patients seen by the provider during this time. Panel risk is defined as the proportion of a clinician's patients with Medicare and/or Medicaid insurance during this time.

Blinding

All investigators, statisticians, and analysts will be blinded during the randomization, intervention, and analysis phases. They will be unblinded after the analysis of all primary and secondary outcomes has been completed. The trials at Penn and UH will be analyzed separately.

Outcomes

For both trials, the primary outcome is patient mammogram completion within 3 months of the eligible primary care visit. The secondary outcome is mammogram completion within 6 months of the

eligible primary care visit. These patient-level outcomes will be coded as a 1 to indicate completion and 0 to indicate non-completion.

Statistical Analysis

The primary analyses for all primary and secondary outcomes will follow the intent-to-treat principle. Accordingly, patients who were identified as eligible (1 day prior at Penn and 1-3 days prior at UH) are included in the analysis regardless of whether the appointment was kept, cancelled, rescheduled, or the patient did not show up. Sensitivity analyses (described later) will include per-protocol and other approaches to evaluate robustness to protocol deviations for some patients.

Analytical Plan: Penn Trial

The primary analysis will fit generalized estimating equations (GEE) with a patient-level binary outcome indicating mammogram completion within 3 months (primary outcome) or 6 months (secondary outcome) of the eligible primary care visit. We will fit a single GEE model to estimate the following co-primary endpoints: 1) effect of the clinic-level intervention (pending orders and post-visit text messaging for patients with a signed order) 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. We will also compare the effect of the most intensive multicomponent intervention (pending orders and intensive texting for high-risk patients) 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. In the Penn trial, 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 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 $\frac{1}{2}$) while each row of a patient whose record was replicated will receive a weight of 1 (since their individual-level intervention “assignments” were deterministic).

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

$$\text{logit} (\Pr(Y = 1 | A_1, A_2, X)) = \beta_0 + \beta_1 A_1 + \beta_2 A_1 A_2 + \beta_3^T X, \quad (1)$$

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 month fixed effects and mean-centered versions of the three variables used in the constrained randomization procedure: 1) clinic size (number of visits); 2) baseline clinic risk (proportion of panel deemed high risk); and 3) baseline clinic completion rate (proportion of panel with completed mammography). We will also include a patient-level

indicator of whether the patient received a mammogram order within the 6 months prior to their eligible visit.

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 need to 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 that when using an independent working correlation. 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 for two co-primary endpoints (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 and marginal probability scales, including 97.5% confidence intervals for effects 1 and 2 and 95% confidence intervals for contrasts 3 and 4.

To examine the effect of the intervention on mammogram ordering, a GEE with logit link and a binary outcome indicating whether the pended order was signed or cancelled will be fit. The rates of

mammogram ordering will be compared according to contrasts (1)-(4) above. We will again use an independence working correlation.

Per-protocol Analysis Plan

Over 500 patients at Penn did not receive pended orders per the study protocol, but they did receive text messaging. We will leverage this protocol deviation to estimate the isolated effect of text messaging alone. We will re-fit model (1) using the weighted and replicated approach after excluding all patients in active clinics who received pended orders. Inference will focus on the main effect of A_1 which will capture the effect of text messaging versus control, averaging over the randomized texting options for high-risk patients in the active clinics. We will also test contrasts 3) and 4). Contrast 3) will provide an estimate of the effect of standard text messaging for all patients relative to control. Contrast 4) will compare the effect of standard texting for low-risk and intensive texting for high-risk relative to control. We do not *a priori* expect a significant effect of time on completion rates, however, if we find evidence of significant variation in completion rates over the study period in the primary models, we will restrict the control arm to eligible visits during times when pended order were not received in the intervention arm.

Analytical Plan for Exploratory Analyses

Exploratory subgroup analyses will examine the effect of the multicomponent nudge versus usual care by risk group. For the low-risk group, 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 independent working correlation since we are interested in the effect of the nudges on individual level outcomes. The weighted and replicated approach is not necessary for the analysis of low-risk patients. For the high-risk group, we will subset the data to patients deemed high-risk in both active and control arms. Then 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 also test contrasts 3) and 4) in the high-risk group.

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:

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

where Z is a single high-risk criterion. We will fit model (2) separately for each high-risk criteria (i.e., insurance type and patient portal status) 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:

$$\begin{aligned} & \beta_0 + \beta_1 + \beta_2 A_2 + \beta_3 + \beta_4 + \beta_5 A_2 + \beta_6^T X - (\beta_0 - \beta_1 - \beta_2 A_2 + \beta_3 - \beta_4 - \beta_5 A_2 + \beta_6^T X) \\ &= 2\beta_1 + 2\beta_2 A_2 + 2\beta_4 + 2\beta_5 A_2 \end{aligned}$$

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

$$\begin{aligned} & \beta_0 + \beta_1 + \beta_2 A_2 + \beta_6^T X - (\beta_0 - \beta_1 - \beta_2 A_2 + \beta_6^T X) \\ &= 2\beta_1 + 2\beta_2 A_2 \end{aligned}$$

Thus, $2\beta_4 + 2\beta_5 A_2$ captures the additional effect of A_1 amongst high-risk individuals with $Z = 1$ relative to those with $Z = 0$. Averaging over levels of A_2 results in $2\beta_4$. 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:

$$\begin{aligned} & \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6^T X - (\beta_0 + \beta_1 - \beta_2 + \beta_3 + \beta_4 - \beta_5 + \beta_6^T X) \\ & = 2\beta_2 + 2\beta_5 \end{aligned}$$

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

$$\begin{aligned} & \beta_0 + \beta_1 + \beta_2 + \beta_6^T X - (\beta_0 + \beta_1 - \beta_2 + \beta_6^T X) \\ & = 2\beta_2 \end{aligned}$$

Thus, $2\beta_5$ is the additional effect of A_2 amongst high-risk individuals with $Z = 1$ relative to those with $Z = 0$. 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 exploratory effect modification analyses.

Additionally, we will test for effect modification by clinic-level baseline screening rates, hypothesizing that clinics with low baseline screening rates will experience the greatest benefit of the nudges. The model specification will be:

$$\text{logit}(\Pr(Y = 1 | A_1, X, Z)) = \beta_0 + \beta_1 A_1 + \beta_2 Z + \beta_3 A_1 Z + \beta_4^T X, \quad (4)$$

where Z is measured at the clinic level and will be the percentage (100*proportion) of the clinic's eligible patient panel that did **not** complete mammogram screening during the period of time from May 1, 2023, to October 31, 2023. Statistical inference will target the interaction parameter, β_3 , which will capture the additional effect of A_1 (on the log odds scale) at the individual level for every 1 percentage point increase in a clinic's baseline unscreened rate. Our hypothesis corresponds to a positive β_3 coefficient, i.e., the clinics with lower baseline screening rates experience a greater effect of the nudges.

Analytical Plan: UH Trial

The primary analysis will fit generalized estimating equations (GEE) with a patient-level binary outcome indicating mammogram completion within 3 months (primary outcome) and 6 months (secondary outcome) of the eligible primary care visit. The model will be specified as follows:

$$\text{logit}(\Pr(Y = 1 | A_1, X)) = \beta_0 + \beta_1 A_1 + \beta_2^T X, \quad (5)$$

where A_1 is an indicator of whether the patient was seen by an active (1) or control (-1) clinician, 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 clinician level constrained randomization procedure: baseline rate of mammogram completion in the clinician's panel, total number of patients in the clinician's panel from March 15, 2023 to September 15, 2023, and proportion of the panel deemed high risk. For the primary analysis, X will also include a patient level indicator of whether the patient had a prior mammogram order within the 6 months prior to the eligible visit. We will use a logit link, and an independent working correlation will be specified at the clinician level so that standard errors will reflect clustering at the clinician level. We will use a Type I error of 0.05 and report 95% confidence intervals for the effects of interest in the UW analysis.

Like the Penn trial, the planned analysis for the UH trial also includes examining the effect of the intervention on mammogram ordering rates. The mammogram ordering rate will be compared between the clinicians in the intervention arm versus clinicians in the usual care arm using a GEE with a binary outcome, logit link, and robust standard errors clustered at the clinician level using an independent working correlation.

Per-protocol Analysis Plan

The primary analyses will follow the intent to treat principle. At UH, there was a protocol deviation caused by an error in the UH enrollment query code that resulted in eligible patients who were not enrolled in the study. We will also perform per-protocol analyses that exclude data from the time period where the incorrect query code was used to enroll patients. Furthermore, we will compare characteristics and outcomes of patients who were enrolled using the incorrect query (Phase 1) to those enrolled using the corrected query (Phase 2) as a sensitivity analysis. All other details of the per-protocol analyses will follow the analysis plan for the intent-to-treat analyses.

Analytical Plan for Exploratory Analyses

Lastly, there are several exploratory analyses planned. First, we will evaluate whether the effect of the clinician nudge is modified by provider baseline screening rates. The model specification will be the same as model (4), but Z will be measured at the provider level and represent the percentage (100*proportion) of the provider's eligible patient panel that did **not** complete mammogram screening within 3 months of the eligible visit during the time period from March 15, 2023, to September 15, 2023. Statistical inference will target the interaction parameter, β_3 , which will capture the additional effect of A_1 (on the log odds scale) at the individual level for every 1 percentage point increase in a provider's baseline unscreened rate.

Second, we will evaluate whether the effect of the clinician nudge is modified by patient-level risk status. The model specification will be:

$$\text{logit}(\Pr(Y = 1 | A_1, X, Z)) = \beta_0 + \beta_1 A_1 + \beta_2 Z + \beta_3 A_1 Z + \beta_4^T X, \quad (6)$$

where Z is an individual-level indicator of whether the patient is deemed high-risk for non-completion. Statistical inference will target the interaction parameter, β_3 , which will capture the additional effect of A_1 (on the log odds scale) for high-risk patients. Third, we will determine which high-risk intensification subgroups experienced the greatest benefits of the provider-level nudge. Using the data set restricted to high-risk individuals, we will fit model (6) where Z will be a single high-risk criterion. We will fit model (6) separately for each high-risk criteria which will allow us to determine which high-risk intensification subgroups experienced the greatest benefits of the provider-level nudges

Finally, we will conduct a post-hoc qualitative assessment of the overall effect of the full multicomponent intervention at Penn versus UH to understand how individual nudges may have impacted mammogram completion and provider ordering. This assessment will be descriptive, and any observed effect differences between the two trials could be at least partially attributable to differences in patient population or study implementation by the different health systems rather than the individual nudges.

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.

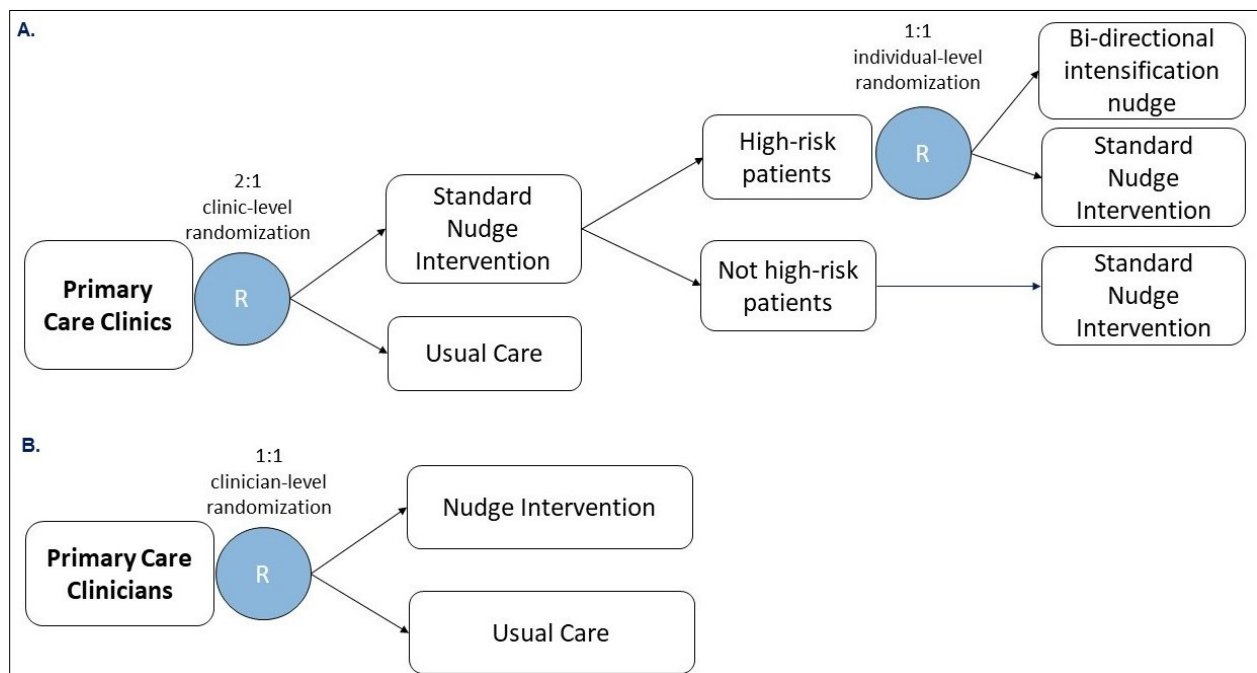


Figure 1. Trial design, including randomization, for Penn (1A) and UH (1B)