# A Prospective Randomized Trial of Personalized Nudges to Increase Influenza Vaccinations

Statistical Analysis Plan

NCT06566534

1/13/25

Statistical Analysis Plan

## Objective & Hypotheses

The purpose of the current study is to test the effectiveness of patient-facing, message-based nudges in promoting flu vaccination. Personalized nudges are selected for each patient as most likely to be effective for that individual based on demographic and health characteristics, as determined by a retrospective machine learning algorithm. The algorithm was trained and tested on Geisinger patients who were sent 19 different nudge messages in a 2020 "megastudy" (Milkman et al., 2021). 12 of those nudges are included in the present study.

As detailed below, participants are randomly assigned to one of four arms: the Personalized Nudge arm; the Reserved For You arm, in which they receive the best-performing (on average) nudge from the 2020 megastudy; the Active Control arm, or the Passive Control arm. The primary hypothesis is that Personalized Nudge will outperform each of the other three arms.

## Methods

Patients eligible for the study as of September 9, 2024 were randomized to one of the following arms:

1. Personalized Nudge: nudge predicted to be most effective for the patient on the basis of the machine learning-derived treatment assignment
2. Reserved For You Nudge: nudge found to be on average most effective in the megastudy, including language that a flu vaccine is "reserved" for patients at their upcoming appointment
3. Active Control: a simple message notifying patients they can get a flu shot at their upcoming appointment
4. Passive Control: no message beyond the ambient healthcare system and public health campaigns to increase vaccination

## Power analysis

Based on retrospective data from the 2023 flu season, at least 90,000 patients are estimated to be eligible for enrollment in the current study. The number of patients per arm will vary according to the expected effect sizes for planned analyses. A total of 15,000 patients will be targeted for the passive control arm, with 25,000 patients targeted for each of the other experimental arms. With a baseline primary outcome vaccination rate of 25% expected in the passive control arm, the experimental interventions are expected to increase vaccinations by at least 2 percentage points to 27% on average. By allocating 15,000 patients to passive control and 25,000 to each of the other arms, there will be 80% statistical power to detect, at minimum, a 1.3% increase in vaccination from 25% to 26.3%, with a two-tailed $p<.05$. Although personalization is expected to greatly improve nudge effectiveness, differences between active experimental arms tend to be smaller than differences between active and passive arms. Therefore, samples have been

allocated to detect, at minimum, a slightly smaller effect of a 1.1 percentage-point increase from 27.0% to 28.1% across active arms, with 80% power and two-tailed $p<.05$.

**Project status**

Study recruitment began on September 9, 2024. The anticipated recruitment end date is December 28, 2024. Outcome data have not yet been extracted from electronic health records by the study team.

**Planned Analyses**

***Primary Outcome*:** Number of patients vaccinated for influenza between enrollment date and target appointment date.

[ *Time Frame: Between the enrollment date and target appointment date (at least 4 days and up to 4 months)* ]

This field experiment will be conducted with Geisinger patients via SMS messages sent prior to their first flu shot-eligible appointment during the study period, referred to as the "target appointment." The key dependent variable is whether patients receive a flu shot at or before their target appointment (as recorded in their electronic health record).

If patients cancel or do not show up for their target appointment after they have been randomized to an arm and then schedule a new appointment during the study period, their new flu-shot eligible appointment becomes the target appointment and the outcome window extends from three days prior to the original appointment through the date of the new target appointment.

Patients who miss their target appointment and do not reschedule within the study period will still be included in the analysis. Their outcome window is from three days prior to the original appointment through the date of the original canceled appointment.

**Primary Question:** Are personalized flu shot messages more effective in increasing flu vaccination rates than the best-performing message in a prior megastudy (Reserved For You), an active control message, and/or no message?

**Secondary Question:** Among the Reserved For You message, the active control message, and no message, how do flu vaccination rates compare? Specifically, does the Reserved For You message lead to higher vaccination rates than both the active and passive control messages, and does the active control message result in higher rates than the passive control?

**Analysis (Confirmatory):** We will test the hypothesis that patients in different arms will exhibit different flu vaccination rates. Specifically, we will compare the following groups:

- Personalized Nudge
- Reserved For You Nudge
- Active Control
- Passive Control

For this analysis, one OLS regression model will be run in the following estimating equation:

$$Y_i = \beta_0 + \beta_1 ReservedForYou_i + \beta_2 ActiveControl_i + \beta_3 PassiveControl_i + \varepsilon_i$$

Where:

- $Y_i$: Binary outcome variable for patient *i* (1 = vaccinated, 0 = unvaccinated)
- $\beta_0$: Mean vaccination rate in the Personalized Nudge arm (reference category)
- $\beta_1$: The difference in vaccination rate between the Reserved For You nudge arm and the Personalized Nudge arm
- $\beta_2$: The difference in vaccination rate between the Active Control nudge arm and the Personalized Nudge arm
- $\beta_3$: The difference in vaccination rate between the Passive Control nudge arm and the Personalized Nudge arm
- $\varepsilon_i$: Error term for patient *i*

Based on this model, we will assess the following pairwise comparisons:

- Personalized Nudge vs. Reserved For You Nudge
- Personalized Nudge vs. Active Control
- Personalized Nudge vs. Passive Control

To estimate the secondary comparisons between study arms (Reserved For You, Active Control, and Passive Control), post hoc pairwise tests will be conducted using the estimated marginal means from the primary OLS model:

- Reserved For You Nudge vs. Active Control
- Reserved For You Nudge vs. Passive Control
- Active Control vs. Passive Control

Recent work suggests that OLS regressions are appropriate in randomized experiments such as this one, using binary outcome variables (Gomila, 2021). We will estimate heteroskedasticity-robust standard errors.

We will not correct for multiple comparisons in the analyses above, as adjustment is not appropriate for individual testing. Each result must be statistically significant to reject its associated individual null hypothesis (Rubin, 2021). Thus, each significant result is an indicator of a specific null hypothesis rejection and has no direct bearing on the success of other hypotheses or the overall study.

EHR flu vaccination data have been finalized and extracted as of this writing but have not yet been analyzed. Insurance claims for flu vaccinations can take up to 90 days to process. To ensure completeness, we will extract the final flu vaccination claims data on April 1, 2025—90 days after the last appointments in the study—plus an additional day to account for potential data refresh delays. In the interim, we will extract an earlier dataset with flu vaccination claims data in the coming weeks for use in preliminary analyses. If there are discrepancies between EHR and claims data regarding vaccine dates, we will use EHR data unless the EHR record is marked as "historic." In such cases, we will use the claims data.

**Heterogeneity Analyses (Exploratory):** We will conduct heterogeneity tests to determine if personalized nudges work better for certain subsets of the population. This will include, but not be limited to, examining factors such as:

- Alignment of the assigned personalized nudge with the received nudge: For participants in the Reserved For You or Active Control arms, we will evaluate whether the effectiveness of the intervention differs based on whether the nudge they received was predicted by the algorithm to be the most effective for them. For example:
    - Aligned: Participants randomized to the Reserved For You or Active Control arm, where the received nudge is the one predicted by the algorithm to be their optimal nudge
    - Misaligned: Participants randomized to the Reserved For You or Active Control arm, where the assigned nudge is not the one predicted by the algorithm to be their optimal nudge
- Feasibility Constraints: Receipt of the optimal nudge versus a different nudge due to feasibility constraints
- Pre-defined sub-groups: Patient groups based on demographic factors such as age and race
- Generalizability: We will assess how the effectiveness of the nudge varies across different clinical contexts, including:
    - Care Setting: The original flu shot megastudy was restricted to primary care target patient appointments. The current study also includes specialty care target appointments. We will compare the relative effectiveness of nudges among patients with target appointments in primary care vs. specialty care to determine whether the impact of nudges depends on the type of healthcare interaction.
    - Patient Population: We will compare the relative effectiveness of nudges among patients who were part of the original training set (i.e., patients in the megastudy) vs. those who are new to the current study to determine whether algorithm-driven predictions maintain validity in previously unseen patients.

**Sensitivity analyses and robustness checks**

Statistical analyses will use an intent-to-treat approach, including all patients randomized. However, the analyses described above may also be run with the subset of patients who, according to text message records from Twilio, actually received all their messages as intended and attended their target appointment. This robustness check may exclude patients who were randomized but never sent messages for a variety of reasons (e.g., their phone numbers were inactive, they were on a do-not-contact list the research team did not have access to prior to randomization). This analysis may also exclude patients who received repetitions of the same message, received their texts late, or experienced other message distortions due to technical events outside the study team's control.

**Other Prespecified Outcomes**

Similar exploratory analyses will also be run for the following outcomes, using the same OLS regression models described above.

1. Number of patients with flu shot receipt on or before December 31, 2024

   [Time Frame: Up to 4 months after randomization]

2. Number of patients with flu shot receipt between enrollment date and first eligible appointment date

   Patients received the flu shot at or before their first eligible appointment date (regardless of whether the patient rescheduled this appointment to a later date within the study period)

   [Time Frame: 4 Days]

3. Number of patients with flu diagnosis (encounter diagnosis or flu test)

   Patients received a flu diagnosis via encounter diagnosis or flu test between enrollment and April 30, 2025.

   [Time Frame: Up to 8 months after randomization]

4. Number of patients with flu-related complications

   Patients experienced flu-related complications before as defined by relevant diagnosis, hospitalization, or death, between enrollment and July 31, 2025 as recorded in the electronic health record

   [Time Frame: Up to 11 months after randomization]

References

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, *150*(4), 700–709. https://doi.org/10.1037/xge0000920

Milkman, K. L., Patel, M. S., Gandhi, L., Graci, H. N., Gromet, D. M., Ho, H., Kay, J. S., Lee, T. W., Akinola, M., Beshears, J., Bogard, J. E., Buttenheim, A., Chabris, C. F., Chapman, G. B., Choi, J. J., Dai, H., Fox, C. R., Goren, A., Hilchey, M. D., … Duckworth, A. L. (2021). A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment. *Proceedings of the National Academy of Sciences*, *118*(20), e2101165118. https://doi.org/10.1073/pnas.2101165118

Rubin, M. (2021). When to adjust alpha during multiple testing: a consideration of disjunction, conjunction, and individual testing. *Synthese*, *199*(3–4), 10969–11000. https://doi.org/10.1007/s11229-021-03276-4