Encouraging Flu Vaccination Among High-Risk Patients Identified by a Machine-Learning Model of Flu Complication Risk

NCT04323137

January 8, 2021

Statistical Analysis Plan

Project Status

All intervention messages have been disseminated. Data collection is ongoing, but an interim data analysis step will be conducted to inform COVID-19 vaccination efforts.

Planned Analyses

Primary Analyses:

Question 1: Does informing patients that they are at high risk for flu complications (a) increase the likelihood that they will receive flu vaccine; and (b) decrease the likelihood that they receive diagnoses of flu and/or flu-like symptoms in the ensuing flu season?
Analysis 1: We will evaluate if patients who are informed about their risk status (patients who were randomized to “high risk” message groups 2, 3, or 4 in the Study Design) exhibit improved flu vaccination and diagnosis rates compared with patients who were not informed (randomized to the control condition, group 1). To assess this, we will employ a generalized linear model with a binary distribution and log-link function (i.e. a logistic regression).

Question 2: Does informing patients that their high-risk status was determined (a) by analyzing their medical records (group 3) vs. by no specified method (group 2); and (b) by an AI/ML algorithm analyzing their medical records (group 4) as opposed to via unspecified methods or human medical records analysis affect the likelihood that they receive the flu vaccine and/or diagnoses of flu and/or flu-like symptoms in the ensuing flu season?
Analysis 2: We will evaluate this using 3 logistic regression models, comparing patients in groups 2 and 3, 2 and 4, and 3 and 4. Reported p-values will be adjusted for family-wise error rate across the three comparisons using Holm’s procedure.

For both analyses described above, we will first assess the intraclass correlation (ICC) for patient-visited PCPs and/or clinics; if the ICC is greater than 0.1, a mixed model will be used with this variable included as a random effect.

Question 3: Patients in each of the conditions were stratified based on whether they were in the top 3rd percentile of risk, or in the subsequent 3rd-10th percentile. Patients in the former group were always told they were in the “top 3%” of risk. Patients in the latter group were either told they were in the “top 10%” of risk or that they were at “high” risk. These different risk conditions will be used to ask whether a) patients with the same risk level are affected by a vague verbal ("high") vs. specific numeric ("top x%") risk framing and b) patients with the same numeric risk phrasing are differentially affected due to different specific risk levels (3% vs. 10%).
Analysis 3: We will evaluate this using a single logistic regression model with "top 10%" condition dummy coded as the reference level. This will allow the contrasts of "top 10%" vs. "high" (for question 3a) and of "top 10%" vs. "top 3%" (for question 3b).

Secondary Analyses:

We will use the approaches in Analyses 1 and 2 to evaluate the impact of the intervention on the secondary outcome measures listed in the pre-registration.

Exploratory Analyses

Impact on timing of shot
This analysis will focus on the subgroup of patients who were vaccinated. We will employ linear regression models to compare the time it took to receive a flu shot for patients who were informed about their risk status with those in the control group (as in Analysis 1) and among the 3 risk status communication conditions (as in Analysis 2) . Time will be specified as the number of days since the beginning of the intervention (September 21, 2020). If the distribution of the time variable is skewed so as to violate the assumptions of a linear regression, data will be log transformed or analyzed using a gamma model.

Impact of age
While older patients tend to be aware of their increased vulnerability, younger patients may be more surprised to learn of their high-risk status. We will therefore examine the differential response of patients above vs. below age 65 by introducing this age variable as moderator and assessing if it interacts with experimental group to affect vaccination behavior. We will also conduct this analysis using age as a continuous variable.

Impact of modality
Patients who received an intervention message (groups 2, 3, and 4) may have received communications via up to 3 modalities. We will employ logistic regression to evaluate the impact of modality on outcomes of interest, assessing pairwise comparisons of the following modality combinations: letter only, letter + myGeisinger, letter + SMS, or letter + myGeisinger + SMS. The small number of patients who did not experience one of these modality combinations will be excluded for this analysis.

Impact of risk level and framing
We will evaluate the impact of risk level (top 3rd percentile vs. remaining 3-10th percentile) as a moderator for Analysis 1 and risk framing ("top 3%", "top 10%", "high") as a moderator for Analysis 2, assessing if the moderators interact with experimental group to affect study outcomes. As described in Analysis 3, risk framing will be dummy coded with "top 10%" as the reference level.

Analysis Exclusion Criteria
A small number of deceased patients and patients who received the flu vaccination before the start of the intervention were unintentionally included in this study. These patients will be

excluded from all analyses. Another small population of patients' print letters were returned to sender; these patients will remain included in analyses but will be documented as having not been contacted through this modality. Finally, due to delays in intervention rollout, there were patients who were vaccinated prior to the start of the study (9/21/20). These patients will be excluded from all above analyses, as they were not intended to receive the intervention.