Analysis Plan

Project Name: Encouraging the Use of USCIS' Online Services
Evaluation B: Email Notifications to Increase Online Account Creation
Project Code: 2208-B
Date Finalized: 02/16/2023

Project Description

The purpose of the project is to support the goals of US Citizenship and Immigration Services (USCIS) to reduce barriers to accessing immigration benefits and efficient adjudications. A priority for USCIS is increasing the number of customers who utilize online services. Overall, this project aims to (1) build evidence on effective methods of encouraging the use of online tools, and (2) build evidence on who is most responsive to this type of intervention in order to inform future efforts and policies.

In this evaluation, we aim to test methods of encouraging USCIS customers, who have recently filed a form by mail that was available to file online, to create an online account and link their paper-filed case to their account. Linking submitted paper forms to an online account allows filers to track the status of their form online and communicate with USCIS agents via secure message. The Digital Services Division (DSD) currently sends regular email outreach to customers who submit via paper a form that is available online to inform them about the availability of online accounts. The current rate of online account usage among this population is unknown, and a secondary benefit of this project for the agency includes the creation of a dataset that can quantify that rate.

Customers may underutilize online tools for a variety of reasons, including a lack of awareness that they exist, a (mis)belief that such tools are not secure, or a perception that online tools are complicated or cumbersome to use. This evaluation will test variations in the email message language that target potential barriers to creating online accounts and linking their open cases. We will then measure the impact of receiving modified communications (relative to the status quo communication) on customers' creation of online accounts and linking their paper-filed forms to their online accounts.

This study will take advantage of the outreach that is already being conducted. Rather than sending all customers the same email message, we will randomly assign customers to one of three experimental conditions:

1. Status quo email (the current standard language - a longer email describing benefits of online accounts and the steps to link an online account to a paper-filed case)

2. Operational transparency email (simplified language emphasizing the benefits of increased transparency available through online accounts)
3. Social norms email (simplified language emphasizing that many other customers have already signed up for online accounts)

Preregistration Details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before randomization.

Hypotheses

Primary

The experiment is designed to address the following primary hypotheses (H):

RQ1: Online Account Creation[1]

Does an email emphasizing operational transparency or social norms increase online account creation relative to the status quo email, among USCIS customers who recently filed a paper form?

- **H1:** Customers who receive the *operational transparency* email (group 2) will be more likely to create online accounts than customers who receive the status quo email (group 1).
- **H2**: Customers who receive the *social norms* email (group 3) will be more likely to create online accounts than customers who receive the status quo email (group 1).

RQ2: Form Linkage

Does an email emphasizing operational transparency or social norms increase the likelihood that customers link their paper-filed cases to online accounts relative to the status quo email, among USCIS customers who recently filed a paper form?

- **H3:** Customers who receive the *operational transparency* email (group 2) will be more likely to link their paper-filed cases to their online accounts than customers who receive the status quo email (group 1).
- **H4**: Customers who receive the *social norms* email (group 3) will be more likely to link their paper-filed cases to their online accounts than customers who receive the status quo email (group 1).

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Data Source(s):

[1] Evaluation of RQ 1 depends on the availability of online account creation data. In the event that the data is not accessible, we will make note of this during analysis and we will not evaluate RQ 1.

The raw data for this project will come from the following USCIS systems: Electronic Immigration System (ELIS), Central Index System 2 (CIS2), Electronic Correspondence Handling Online (ECHO), and GovDelivery.

Outcomes to Be Analyzed:
The primary outcomes for this evaluation are:

- A binary indicator reflecting the customer's creation of an online account after the outreach date (ONLINE_ACCT)
- A binary indicator reflecting the customer's linkage of their paper-filed case to their online account after the outreach date (ACCT_LINK_PAPER)

We expect to receive an outcome data snapshot that reflects outcomes approximately 3 to 4 months after the last outreach in our sample. We may receive a second data snapshot, but the timing and capacity for the second data transfer has not yet been determined. If we do receive a second snapshot, we will analyze the data using the same specifications as outlined here.

Imported Variables:
The agency project team will randomize customers using SAS. They will maintain the randomization datasets created during this process on their agency's internal server until they are ready to transfer the datasets to OES. The treatment condition indicators will be merged with the raw dataset that includes covariates and outcome variables exported from USCIS databases.

Transformations of Variables:
<u>Independent variables</u>
We will create the following new independent variables using a transformation on the raw data:

- TREAT_OPTRANS: A binary indicator for assignment to the treatment condition that received the email emphasizing operational transparency
- TREAT_SOC: A binary indicator for assignment to the treatment condition that received the email emphasizing social norms
- TREAT_CONTROL: A binary indicator for assignment to the control condition that received the email with the status quo language

<u>Covariates</u>
We will create the following covariate using transformations or calculations on the raw data:

- AGE: The length of time in years between the date the data was transferred to OES and customer's date of birth (DOB). Since DOB will be provided as month and year to protect privacy, the first day of the customer's birth month will be used to calculate age.

Other covariates may be transformed depending on the format of the raw data. For instance, if there are too few observations within categorical variable labels, labels may be combined to form groups of larger sizes.

Expected Sample Sizes:

The agency currently conducts this outreach on a monthly basis, so customers will be randomized into treatment conditions on an ongoing basis until the necessary sample to detect the expected minimum treatment effect has been collected or until it is no longer feasible to continue implementing the evaluation. According to power analyses based on a range of possible base rates, to detect a minimum effect of 1 percentage point between the Status quo group and either group 2 or group 3, we must accrue a total sample size of 76,750 (power = .8, alpha = 0.05, equal probability of group assignment, assumed Status quo mean = 20%). We estimate that the upper range of this sample can be accrued within 2 to 3 months, based on the agency's historical outreach data.

Transformations of Data Structure:
At this time, no transformation of data structure is expected.

Data Exclusion:
We do not expect to exclude any observations from the main analyses.

Treatment of Missing Data:
The analyses of these research questions are not expected to have missing outcome data.

If any covariate required for the analyses is missing for over 5% of the sample, we will run two models: (1) one model that excludes that covariate; and (2) one model that includes the covariate, with missing data imputed with multiple imputation. We will report results from both, noting any meaningful differences.

Descriptive Statistics, Tables, & Graphs

- A table containing descriptive statistics and balance tests for baseline covariates among the treatment groups, including customer demographics. The balance test will be an omnibus F-test of differences in characteristics between each treatment arm pair, for a total of three F statistics. We will make note of these imbalances, but these tests will not be used to select random assignments, as they will be conducted ex-post.
- Graphs of treatment effects to illustrate the results
- Tables summarizing results from the statistical models

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Statistical Models:

We will conduct an intent-to-treat analysis using the following linear model using the Lin (2013) estimator:

$$Y_i^j = \beta_0 + \beta_1 TREAT\_OPTRANS_i + \beta_2 TREAT\_SOC_i + \delta X_i + \varepsilon_i$$

where Y_i represents the outcome of interest for $j \in \{ONLINE\_ACCT,\ ACCT\_LINK\_PAPER\}$.

The first coefficient of interest, β_1, will be interpreted as the average effect of assignment to the operational transparency condition on the outcome of interest, compared to the status quo condition. The second coefficient of interest, β_2, will be interpreted as the average effect of assignment to the social norms condition on the outcome of interest compared to the status quo condition. X_i represents a vector of the following covariates, if available: age, gender, race, ethnicity, and disability status. The vector of covariates is included to increase the precision of the estimate for the coefficient of interest.

Confirmatory Analyses:
Our confirmatory analyses are the hypothesis tests described for the coefficients of interest in the model evaluating the hypotheses associated with RQs 1 & 2.

Exploratory Analysis:
We will also evaluate the following research questions, for which we do not have any *a priori* hypotheses:

- **RQ3:** Which language framing, emphasizing operational transparency (group 2) or social norms (group 3), is most effective at encouraging customers to create online accounts and link their paper-filed cases?
    - We will answer this question using a linear model similar to the main model, except we will use the social norms (group 3) treatment condition as the reference group so that β_1 will be interpreted as the average effect of assignment to the operational transparency condition on the outcome of interest, compared to the social norms condition.
- **RQ 4:** Are there any heterogeneous treatment effects by demographics (if available)?
    - Groups of interest may include any or all of the following:
        - Hispanic country of origin vs. non-Hispanic country of origin and/or Hispanic/Latino vs. non-Hispanic/Latino ethnicity
        - Older adult vs. not an older adult (exact age cut-off to be determined based on sample age characteristics)
        - Has a disability vs. no disability indicated
        - White vs. non-White
        - Black vs. non-Black
        - Used an attorney or preparer vs. did not

        - Depending on the final data received, we may expand or modify the groups of interest
    - We will estimate a model for each group of interest, similar to the main model, with additional independent variables: a binary indicator for the group of interest and interaction terms between the binary indicator for the group of interest and each treatment indicator.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
For all OLS models, we will use HC2 standard errors for statistical inference. In all models, we will reject the null hypothesis if $p < 0.05$ for a two-tailed test on the coefficient of interest. To adjust for multiple comparisons, we will use the Holm-Bonferroni procedure to control the family-wise error rate (FWER) across our confirmatory hypotheses.

Limitations:
It is possible that our outcome measurements will underestimate engagement with online tools due to matching errors across datasets. For example, customers may create online accounts with a different email address than the one they have on file. This sample also only includes customers who have an email address on file with USCIS, which may limit the generalizability of findings to the full population of customers that USCIS serves.