Analysis Plan

Project Name: Increasing SSI Uptake Among Newly-Eligible Population

Project Code: 1723

Date Finalized: August 23, 2018

This document serves as a basis for distinguishing between planned (confirmatory) analysis and any unplanned (exploratory) analysis that might be conducted on project data. This is crucial to ensuring that results of statistical tests will be properly interpreted and reported. In order that the Analysis Plan fulfill this purpose, it is essential that it be finalized and date-stamped before we begin looking at the data — ideally, before we take possession of the data. Once this plan is finalized, a date is entered above, and the document is posted publicly on our team website.

Data and Data Structure

Data come from the administrative records of the SSA. These are individual-level monthly data. They include a small number of covariates (e.g., age, sex, state of residence).

Outcome Variables to Be Analyzed:

The primary outcomes of interest are:

- SSI application filed
- SSI application allowed

As secondary outcomes, we will examine:

- SSI application started
- SSI application denied
- Average amount of SSI payments received for months with SSI payments
- Total amount of SSI payments received
- Received SSI for at least one month
- Number of months received SSI

We will measure these outcomes at 3, 6, and 9 months.

In addition to these outcomes of interest for our regression analysis, we will examine the reasons SSI applications are denied by calculating the frequency of each reason SSI application is denied. Reason SSI application is denied will not be an outcome of interest in the regression analysis.

Statistical Models & Hypothesis Tests

Balancing checks:

We will conduct balance checks on study randomization as follows. For each of the following variables, we will calculate sample means and standard deviations for the control group, for each of the treatment arms, and for "any letter" (i.e. pooled treatments): age, sex, potential SSI payment, previously applied for SSI, and state. We will use t-tests to test for equality of means (or proportions) for all variables except state of residence and Chi-squared tests to test for equivalence of distribution for state of residence. We will interpret the randomization as having resulted in balanced assignment if the fraction of tests that reject the null hypothesis of no difference in means (or proportions or distributions) is less than 10%. If more than 10% of tests reject the null hypothesis of no difference, then we will measure these differences relative to the full sample means and interpret the randomization as having resulted in balanced assignment if these differences are less than 1% relative to the full sample means.

Treatment effects:

Our primary analysis will be an intent-to-treat (ITT) specification. We will estimate the causal effect of the intent to treat using linear ordinary least squares (OLS) regression.

We will augment this analysis with a treatment-on-the-treated calculation. For the treatment-on-the-treated calculation, we will assume a 10% return rate for the study letters (i.e. a 90% delivery rate for the study letters). We base the 10% return rate assumption on previous SSA administrative experience with return rates for similar mailings using administrative records.

Part 1 our analysis of treatment effects focuses on measuring the effects of each of the letters. Part 2 of our analysis of treatment effects focuses on measuring the effects of each of the letter elements. That is, the statements on the letters. For example, the eligibility statement or the maximum payment statement.

Part 1: Effects of letters

The first part of our analysis of treatment effects focuses on measuring the effects of each of the letters.

We will pool the full study sample (i.e. control and all treatment observations) to form our main regression sample.

We will regress the outcome of interest (e.g., SSI application completed) for individual *i* on the full set of indicator variables for each of the letters (e.g. the Basic Letter). That is, we will estimate the parameters of the following regression equation:

$$outcome_i = \propto + \beta_1 BasicLetter_i + \beta_2 MaximumPaymentLetter_i + \beta_3 SimplifyingApplicationLetter_i$$

$$+ \beta_4 CombinedLetter_i + \varepsilon_i \qquad (1)$$

The coefficient on the Basic Letter indicator variable, β_1, is the estimate of the causal effect of the intent to treat of the Basic Letter. The coefficient on the Maximum Payment Letter indicator variable, β_2, is the estimate of the causal effect of the intent to treat of the Maximum Payment Letter. The coefficient on the Simplifying Application Letter indicator variable, β_3, is the estimate of the causal effect of the intent to treat of the Simplifying Application Letter. The coefficient on the Combined Letter indicator variable, β_4, is the estimate of the causal effect of the intent to treat of the Combined Letter.

In additional specifications, we will add control variables such as study participant age, sex, potential SSI payment amount, previously applied for SSI, and state of residence.

Part 2: Effects of letter elements (i.e. statements on letters)
The second part of our analysis of treatment effects focuses on measuring the effects of each of the letter elements. That is, the statements on the letters. For example, the eligibility statement or the maximum payment statement.

We will pool the full study sample (i.e. control and all treatment observations) to form our main regression sample.

We will regress the outcome of interest (e.g., SSI application completed) for individual *i* on the full set of indicator variables for each of the letter elements (i.e. statements) and the interaction between the maximum payment statement and the simplifying application process statement. That is, we will estimate the parameters of the following regression equation:

$$outcome_i = \propto + \beta_1 eligibilitystatement_i + \beta_2 maximumpaymentstatement_i + \beta_3 simplifyingstatement_i$$

$$+ \beta_4 maximumpaymentstatement * simplifyingstatment_i + \varepsilon_i \qquad (2)$$

The coefficient on the eligibility statement indicator variable, β_1, is the estimate of the causal effect of the intent to treat of the eligibility statement. The coefficient on the maximum payment statement indicator variable, β_2, is the estimate of the causal effect of the intent to treat of the maximum payment statement. The coefficient on the simplifying statement indicator variable, β_3, is the estimate of the causal effect of the intent to treat of the simplifying application process statement. The coefficient on the interaction term (i.e. maximum payment statement indicator variable interacted with the simplifying application process statement indicator variable), β_4, is the estimate of the causal effect of the intent to treat of the interaction of the maximum payment statement and simplifying application process statement.

In additional specifications, we will add control variables such as study participant age, sex, potential SSI payment amount, previously applied for SSI, and state of residence.

Heterogenous treatment effects:
After conducting our main analysis, we will test for several sets of heterogeneous effects in our ITT regressions. First, we will test for heterogeneous effects by age group, using following age categories: age 65, age 66-70, age 71-75, and age 76-80. Second, we will test for heterogeneous effects by potential SSI payment amount, where potential SSI payment amount is the 2017 SSI Federal Benefit Rate of $735/month plus $20 minus SSA Old Age Security monthly payment amount. Third, we will test for heterogeneous effects by whether the individual resides in a state that bundles Medicaid with SSI.

For all of these heterogeneity analyses, we will follow specifications similar to those in our main analysis. For the age analysis, we will include controls for these age categories and interact these age categories with the treatment indicators specified in the main analysis.

For the potential SSI payment amount analysis, we will use two approaches. First, we will include a control for potential SSI payment amount in US dollars and interact this variable with the treatment indicators specified in the main analysis. Second, we will include controls for quintile of potential SSI payment amount and interact these quintile indicator variables with the treatment indicators specified in the main analysis.

For the Medicaid bundle analysis, we will include a control for whether the individual resides in a state that bundles Medicaid with SSI and interact this indicator variable with the treatment indicators specified in the main analysis.

Standard error adjustments:
We will estimate heteroscedasticity-robust standard errors and cluster our standard errors at the state level.

Software:
SSA will use SAS and Stata to conduct the randomization assignment and the regression analyses.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
We do not perform corrections for multiple hypothesis testing. The number of hypotheses we are testing is small, existing literature strongly supports investigating these hypotheses, and we are examining a relatively small set of outcomes that conform with a clear theory of change (e.g. schedule an application, start an application, complete an application, SSA approves/denies, etc).

Limitations:
One limitation of the study design is that we will not be able to verify whether recipients opened and read the letter. Another limitation of the study design is that SSA does not know SSI eligibility with certainty until after an individual applies and SSA reviews their application. SSA cannot observe assets and hence can only make a preliminary assessment of SSI eligibility based on amount of SSDI and SS retirement payments received.