Analysis Plan

Project Name: Increasing Flexible Spending Account (FSA) Participation Among Government Employees

Project Code: 1733

Date Finalized: 11/22/2017

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

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.

Outcome Variables to Be Analyzed:

The main outcome variables are an indicator variable for contributing to an FSA and FSA contributions as a continuous variable. Secondary outcome variables are an indicator variable for contributing $500, an indicator variable for contributing at least $500, and an indicator variable for "click-through" on the GSA Today messaging. No other data (e.g., claims) are available. (The indicator variables for the quantity of contributions will be created; the other variables are in the raw data.)

Transformations of Variables:

No transformations are planned other than the creation of the indicator variables above.

Imported Variables:

Additional employee-level covariates will be merged in from the GSA D2D employee dataset. We already have access to this data.

Treatment of Missing Data:

If data on whether or not an individual contributed is missing, that observation will be excluded from the analysis. However, the share of observations with missing data will be reported, and we will also report whether the share of observations with missing data is different comparing across treatment and control. If data on the amount contributed is missing, the observation will still be included in the analysis using the dummy variable for contribution.

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.

Analysis will be conducted in Stata. The dependent variables are an indicator variable for any contribution (D_i) and the absolute amount of contributions (C_i). First, we will estimate a specification regressing these variables on an indicator variable for assignment to any of the three treatment arms (T_i), both without individual covariates and conditional on individual covariates (X_i) (i.e. employee grade (e.g., GS level), GSA region indicator variables, year of hire indicator variables, an indicator variable for supervisory/non-supervisory employee, and (if available) previous year contributions).

$$D_i = \beta T_i + X'_i \Pi + \varepsilon_i$$
$$C_i = \beta T_i + X'_i \Pi + \varepsilon_i$$

where ε_i is an idiosyncratic error term.

The first specification can be estimated employing ordinary least squares (OLS) regression.

In addition, we will separately evaluate the effect of each treatment by estimating the following specifications.

$$D_i = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + X'_i \Pi + \varepsilon_i$$

$$C_i = \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + X'_i \Pi + \varepsilon_i$$

where T_1,T_2, and T_3 are indicator variables for assignment to each of the three treatment arms, X_i is the aforementioned vector of individual-level covariates, and ε_i is an idiosyncratic error term. Again, the first specification will be estimated using OLS regression.

In all regression analyses, we will estimate heteroskedasticity robust standard errors.

Follow-Up Analyses:

Using the second regression estimated above, we will conduct the post-estimation tests $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, and $\beta_2 = \beta_3$ to evaluate whether the treatments have significantly different effects.

We will also test for heterogeneous effects by contributions (i.e. any contribution and amount of contribution) in previous year (if available) by interacting our treatment indicator variables with previous year contributions (while including previous year contributions as controls). We will also test for heterogeneous effects by employee grade (e.g., GS) level, by GSA region, by year of hire, and by supervisory/non-supervisory position, by interacting our treatment indicator variables with indicator variables for the covariate of interest, while including the covariate of interest as a control. We will estimate a separate regression for each interaction variable. This analysis will be regarded as secondary to the main outcome variables.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
All tests will be two-tailed. We do not plan on any adjustments for multiple inference in our primary analysis (i.e., excluding the heterogeneous effects analysis).

For the analysis of heterogeneous effects, we will use the Benjamini-Hochberg method for multiple comparisons.

Limitations:
While the experimental design is straightforward, one key limitation is that we will not observe individuals' spending patterns. (This information is not reported by the vendor to GSA Human Resources.) Accordingly, it may be challenging to assess whether any particular level of contributions is optimal or welfare-increasing. We will try to look at second-year (i.e. the year after the study reference year) enrollment (and contributions) if we are able to acquire these data from GSA Human Resources.