Analysis Plan

Project Name: Evaluating form modifications to make it easier to apply for benefits Project Code: 2307A Date Finalized: 2/23/2024



Project description

Objectives

This project aims to improve applicants' experiences with reporting their income on the Supplemental Nutrition Assistance Program (SNAP) application. The SNAP application requires applicants to provide information about their income in the past 30 days to assess eligibility and calculate benefits amounts. Additionally, applicants have to upload verification documents that confirm their responses on the application. This project uses an application digital assister, designed by Code For America (CfA), to help individuals in a large U.S. state complete the SNAP application. This evaluation examines which of three different sets of options for reporting income are more effective at encouraging applicants to complete the SNAP application.

In the federal context, multiple benefits applications ask applicants to report their income, and administrative burdens in form completion can impede access to federal programs among eligible beneficiaries. This project also will build generalizable evidence regarding how to ask people about income that can be applied across forms, benefits applications, and surveys administered by the federal government that include modules on income reporting.

Intervention and evaluation design

CfA designed and fielded this randomized evaluation in 2023. OES's role in this project is to provide analytic support and generate learnings that may be applicable in other federal contexts.

OES will not receive outcome data until the analysis plan is publicly posted.

In this evaluation, SNAP applicants were randomized to see one of three sets of options for reporting their income. Each of the **reporting choice sets** displays a different set of **income reporting options** as depicted in Figure 1 on page two.

Figure 1. Income reporting options



I can give an estimate

The choice sets vary in terms of the number and type of options available to report income. The primary objective of this evaluation is to learn which set of income reporting options results in the highest application completion rates. We also will leverage the evaluation design to descriptively learn which reporting options were preferred by applicants. We have a range of secondary outcomes that are relevant to federal forms that ask about income (described in the Data section below), including a set of outcomes that proxy for the customer experience.

The income reporting module occurs after applicants have answered basic questions about themselves, including the county they live in, disability status, student status, citizenship status, and if they receive income support in the form of SSI or disability payments. Applicants are then asked if they are self-employed, if there are multiple jobs in their household,¹ and if they have stable or variable income.² The evaluation sample consists of applicants who are self-employed, have a single job in their household, and have stable income.³

¹ An applicant can have multiple jobs if they themselves have more than one job or if there is more than one job across all the members of their household. The definition of a SNAP household is a SNAP "unit", which is defined as one person or a group of people who usually buy and prepare their food together (<u>U.S. Department of Agriculture (USDA) SNAP eligibility rules</u>). We use SNAP household, unit, and applicant interchangeably in this analysis plan.

² Applicants who were excluded from the evaluation include those who said that their income is expected to fluctuate in the next month *and* applicants who reported that they were not sure whether their income will fluctuate.

³ Based on baseline data received from the collaborator, we believe that roughly a third of applicants using the digital assister meet these inclusion criteria.

When applicants see one of the reporting choice sets, they choose an income reporting option then are directed to a page that allows them to enter their income. Each reporting option has a different format for entering their income:

- Applicants who select "hourly wage" are directed to a page where they report their hourly wage and the number of hours they worked in the last 30 days;
- Applicants who select "exact amount from my paystubs" are directed to a page where they enter an amount in a text box and report how often they are paid per month;
- Applicants who select "yearly salary" are directed to a page where they enter in their annual salary; and
- Applicants who select "I can give an estimate" are directed to an unstructured text box that asks them to enter in their pre-tax earned income for their job for the past 30 days.

We will observe which of the three choice sets (shown in Figure 1) applicants were randomized to see and their final selection of a single reporting option. Although applicants have the option of hitting the back button and changing their selections, we cannot observe who has done this, nor can we observe any answer selections prior to the final answer they submitted. We discuss how this affects inclusion in or exclusion from our analysis later in this document, but note that treatment assignment to the choice set was fixed and stable even if applicants used the back button and returned to the intervention screen several times in the session.

Preregistration details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before outcome data are analyzed. In addition, this project will be pre-registered in the AEA RCT Registry at https://www.socialscienceregistry.org/.

Hypotheses

We have two primary research questions, which are associated with the following hypotheses and which we plan to highlight in the project abstract. These research questions are:

- 1. **Impact of choice set on submission rate:** Does the reporting choice set that was offered to SNAP applicants in the digital assister impact application submission rates?
- 2. **Descriptive analysis of reporting option selection:** Which income reporting option was selected most frequently by applicants?

Primary hypotheses:

Impact evaluation of choice set on submission rate: Our primary hypothesis is that the different choice sets for income reporting will produce differences in submission rates.

H1: There will be a statistically significant difference in application submission rates between the three reporting choice sets.

Should we find evidence in support of H1, we will then test the following hypotheses:

H1a: There is a statistically significant difference in application submission rates between the control reporting choice set and the reporting choice set of Intervention A.

H1b: There is a statistically significant difference in application submission rates between the control reporting choice set and the reporting choice set of Intervention B.

H1c: There is a statistically significant difference in application submission rates between the reporting choice set of Intervention A and the reporting choice set of Intervention B.

1) Descriptive analysis of reporting option selection: We also plan to descriptively analyze the proportion of applicants who selected each reporting option in Intervention Arm 2 (the intervention arm that included all reporting options). Ex ante, we do not have strong expectations about whether applicants will opt for a particular reporting option. Therefore, our primary descriptive analysis will measure the proportion of applicants who selected each reporting option (including "give an estimate" or no selection). This may inform our understanding of how to improve the experience of completing the SNAP application, but note that this will not provide a causal estimate of which options perform best.

Secondary hypotheses:

- **Impact evaluation of choice set on secondary outcomes:** We will use the same approach outlined above to test if the following secondary outcomes differ across applicants exposed to the different income reporting choice sets:
 - Likelihood of completing the income module
 - Likelihood of uploading any SNAP verification documents
 - Rate of selecting the reporting option "I can give an estimate"
 - Rate of seeing an estimated (in)eligibility screen that reports an applicant is not eligible for SNAP
 - Total time spent on the application⁴
 - Time to submission
 - Total amount of income reported
- Application device subgroup effects:
 - Among applicants who use a mobile device for the digital assister application, there will be at least one non-zero pairwise difference between **application submission rates** across applicants exposed to the three income reporting choice sets.
 - Among applicants who use a desktop device for the digital assister application, there will be at least one non-zero pairwise difference between **application**

⁴ Note that we are only able to estimate time spent on the application as a whole. We cannot calculate time spent on particular application screens or modules because the use of the back button overwrites the timers for each screen.

submission rates across applicants exposed to the three income reporting choice sets.

- There will be at least one non-zero pairwise difference in the effects between mobile device and desktop users in **application submission rates** across applicants exposed to the three income reporting choice sets.

Exploratory hypotheses:

<u>Extension of reporting option selection</u>: We will analyze the proportion of people who selected each reporting option (including "give an estimate" or no selection) in each of the conditions separately.

<u>Instrumental variables analysis</u>: We are interested in understanding the impact of selecting a given reporting option on application behaviors. We plan to conduct an instrumental variables analysis where we use random assignment to the different choice sets to instrument for choosing a particular option. We will run this analysis if there is a strong first stage where assignment to the choices sets strongly predicts selection into one or more reporting options.

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):

Data for this randomized evaluation come from CfA's digital assister database. The sample for this evaluation is all SNAP applicants who used the digital assister application between May 8th and June 23rd, 2023, and who met the inclusion criteria described in the Intervention and Evaluation Design section.

There is a possibility that we could access data from the county SNAP administrators that share data with CfA. In this case, we would have more information on the outcomes of the applications submitted (i.e., whether they were approved for SNAP, their actual (as opposed to reported) income, etc.). However, we do not list those data in the outcomes below because it is likely that they will not be available.

Variables

The following table outlines the variables that we will receive from the digital assister application data. Note that the values of these variables represent the **final observed value** in the digital assister. In other words, if applicants answered a screen in one way, moved on, and then used the back button to return to that screen and provide a new answer, we will only observe the final answer in the data.

The table in the section on <u>Transformations of Variables</u> outlines how we will recode this data (where applicable) and translate these variables to outcomes or covariates.

Table 1. Raw unprocessed data received from the CfA digital assister application data base

Raw variable	Description
cfa_application_id	Unique application ID that CfA assigns each application
treatment_group⁵	 Randomly assigned intervention control = Choice Set A hourly_or_salaried = Choice Set B hourly_or_salaried_or_exact = Choice Set C
final_controller	 Income reporting option selected exact_earned_incomes = Use exact amount from my paystubs hourly_earned_incomes = Use my hourly wage salaried_earned_incomes = Use my yearly salary estimated_earned_incomes = I can give an estimate N/A = No response selected
created_at	Timestamp for when the applicant started the income module
finished_at	Timestamp for when the applicant completed the income module
application_start_time	Timestamp for when the applicant started the digital assister application
application_end_time	Timestamp for when the applicant submitted the digital assister application
final_time_stamp	Timestamp for the last screen the applicant completed
Reported monthly income	Total income for past thirty days, as calculated from applicant's responses by the CfA software
Uploaded verification documents	Binary variable equal to 1 if the applicant uploaded any income verification document through the CfA digital assister before exiting the application; 0 else
Language	Language in which the applicant elected to fill out the application:EnglishSpanish

⁵ Note that randomization is most precisely understood to be at the "application" level, because an applicant could initiate the application but return later and begin a new application. These applicants would not necessarily be assigned to the same treatment condition if they used a device without cookies saved, a new browser, or a new device (e.g. switching from desktop to mobile and beginning the application again).

	Chinese (Mandarin)	
Used desktop	Binary variable equal to 1 if the applicant used a desktop computer to fill out the digital assister application; 0 if the applicant used another type of device ⁶	
Household size	Number of people reported in the household	
Zip code	Zip code of residency that the applicant reported in the application	
Single source of income	Binary variable equal to 1 if the applicant reported having a single source of income; O else	
Not self-employed	Binary variable equal to 1 if the applicant reported that their income is not from self-employment; 0 else	
Stable income	Binary variable equal to 1 if the applicant reported that their income is not expected to fluctuate; 0 else	
Marked for inclusion	Binary variable equal to 1 if the applicant was marked for inclusion in the randomized evaluation; 0 else	
Gets SSI	Binary variable equal to 1 if applicant reported receiving SSI; 0 else	
Over 60 or disabled	Binary variable equal to 1 if applicant reported being over 60 or disabled; 0 else	
Applicant has cookies turned on	Binary variable equal to 1 if CfA recorded that the applicant has its cookies turned on; 0 else	
Responses to questions on the application	Raw application data for each question in the digital assister application	

Outcomes to be analyzed:

- Application completion
- Income reporting option selected
- Income module completion
- Uploading of any SNAP verification documents
- Predicted SNAP eligibility

⁶ Baseline data shows that approximately 40% of applicants use a desktop computer, providing a large sample for both desktop computer and another type of device.

- Total time spent on the application
- Total time spent to submit the application
- Total amount of income reported
- Screen of application drop-off

Imported variables:

Not applicable, unless we are able to gain access to application outcomes from the state.

Transformations of variables:

See table below describing the construction of the variables that are not already in the raw data.

Transformed variable	Description	How it is constructed		
Outcome variables				
Application completed	Binary variable equal to 1 if the applicant completed the full SNAP application; 0 else	Coded based on whether applicant entered in a signature on the last page of the application and hit "sign and submit application"		
Time to submission	Time in seconds from the start of the SNAP application until the timestamp on submission; Missing if the applicant did not submit the application	Continuous variable created by taking the difference in seconds between application_end_time and applicaiton_start_time		
Time to drop-off	Time in seconds from the start of the SNAP application until the timestamp on the last screen completed; Missing if the applicant completed the application	Continuous variable created by taking the difference in seconds between final_time_stamp and application_start_time		
Total time spent on application	Time in seconds from the start of the SNAP application until the timestamp on submission or on the last screen completed	Continuous variable that takes value of "time to submission" for applicants who completed the application and "time to drop-off" for applicants who did not complete the application.		
		Note that for timer variables, we will look at the distributions and assess		

		whether any transformations are needed to handle outliers.		
Completed the income module	Binary variable equal to 1 if the applicant completed the income module; 0 else, including if the applicant dropped off before this screen	Transformed based on whether application data indicates applicant made it through the income module		
Predicted to be ineligible for SNAP	Binary variable equal to 1 if CfA's software automatically determines that the applicant is likely ineligible for SNAP as of the end of the income module; 0 else, including if the applicant dropped off before this screen	Transformed from CfA coding applicant as predicted ineligible		
Selected the "I can give an estimate" reporting option	Binary variable equal to 1 if final_controller is equal to estimated_earned_incomes; 0 else, including if the applicant dropped off before this screen	Transformed from estimated_earned_incomes		
Selected one of the three other reporting options (i.e., not the "I can give an estimate" option)	Binary variable equal to 1 if final_controller is equal to exact_earned_incomes, hourly_earned_incomes, and salaried_earned_incomes; 0 else, including if the applicant dropped off before this screen	Transformed from estimated_earned_incomes		
Uploaded verification documents	Binary variable equal to 1 if the applicant uploaded at least 1 income verification document to the application; 0 else, including if the applicant dropped off before this screen	Transformed from back-end data		
Completion of each application screen (one variable for each screen of the application)	Binary variables equal to 1 if an applicant provided responses up to and including the given application screen; 0 else, including if the applicant dropped off before this screen	Transformed from raw responses to application questions		
Demographic variables				
English language	Binary variable equal to 1 if applicant opted to complete the language in English; 0 else	Transformed from Language where language selected = 1 if English was selected		
County	A vector of binary variables, one for each county, indicating if an applicant lives in that county	Transformed from zip code		

Transformations of data structure:

Not applicable; data is already at the level of the applicant session. We anticipate having to merge data from the randomized evaluation module itself (including treatment assignment and selected reporting option) with the rest of the application. We will merge using the unique digital assister ID that CfA generates, which is available in both datasets.

Data exclusion:

We will include all applicants who were sorted into the randomized evaluation module, regardless of indicated eligibility. The mismatch between evaluation eligibility and inclusion in the randomized evaluation module arises from use of the back button – applicants can enter in one set of answers to the eligibility screener questions, move forward through the application, and then reverse and change those answers. Therefore, some applicants whose final, recorded answers indicated eligibility (ineligibility) were excluded from (included in) the randomized evaluation. We measure inclusion in the randomized evaluation module by whether an applicant has a value for the treatment_group variable and include all applicants who have a value for this variable in our study.

We will exclude all applicants who were excluded from the randomized evaluation, as we will not have randomized evaluation data for them, even if their final, recorded responses to the eligibility questions indicate that they should have been eligible.

Treatment of missing data:

Missing data could arise when applicants do not complete a portion of the digital assister application. The transformations above describe how we will code missing data for each of the outcomes so that we can include applicants who dropped off in our causal analyses. We may also run descriptive analyses focused on the applicants who completed the entire application.

We also note that there are eligible applicants who were excluded from the randomized evaluation due to their original form response but were eligible to be included based on changes to their response. These applicants were never randomized to an intervention arm and thus excluded from our evaluation. The fact that these applicants are missing from our evaluation has implications for the external validity of our evaluation.

Descriptive statistics, tables, & graphs

Primary Analysis

We will create bar charts for the primary analyses to be included in the project abstract:

- We will create a bar chart showing the probability of submitting the digital assister SNAP application for each of the intervention arms, with 95% confidence intervals.
- We will create a bar chart with one bar each showing the proportion of applicants who selected each reporting option in Intervention Arm 2 (the arm that includes all reporting options). Depending on the results, we may display a bar chart showing the proportion of people who selected each of the individual reporting options for each of the intervention conditions.

Understanding Compliance

We will produce the following summary statistics, tables, and graphs to understand treatment assignment and treatment compliance. Randomization was conducted by Vanity, so we cannot assess a priori treatment assignment probabilities. Instead we will:

- Show the proportion of sample applicants who were assigned to each of the three intervention arms.
- Conduct balance checks between each of the intervention arms for the vector of applicant-level covariates that are included in our adjusted model. We will conduct this balance test by:
 - Testing for the significance of the difference between each of these based on an omnibus F-test of all applicant-level covariates. We will conduct the F-test through three regressions, which will allow us to regress assignment to each intervention arm on the baseline variable measures.

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 model for confirmatory analyses:

1) Causal analysis of choice set on application outcomes

To assess H1, we will conduct an F-test of joint significance to assess whether any of the treatment arms have meaningfully different submission rates from one another. If the F-test detects significant differences in outcomes between intervention arms, we will estimate pairwise comparisons in the form of a post-hoc test using the *emmeans* (estimated marginal means) package in R.

We will report the results of the F-test as our focal confirmatory test, and estimate the ITTs for the intervention arms using an unadjusted OLS linear probability model. As a robustness check, we will also estimate a covariate-adjusted model, where we include one variable that indicates membership in our key demographic subgroup (desktop computer completion) for which we will test for treatment effectiveness separately as a <u>secondary analysis</u>, an additional set of covariates, and fixed effects for the applicant's reported county of residence. The regression models are presented below.

$$Y_{i} = \beta_{0} + \beta_{1} Arm_{A_{i}} + \beta_{2} Arm_{B_{i}} + e_{i}$$
(1) Unadjusted Model

$$Y_{i} = \beta_{0} + \beta_{1} Arm_{A_{i}} + \beta_{2} Arm_{B_{i}} + X_{i} + e_{i}$$
(2) Fully-Adjusted Model

Where *i* indexes applications:

- *Y*_i is the outcome of interest
- β_0 is the intercept
- Arm_A_i is a binary indicator indicating that the applicant was randomly assigned to receive Intervention Arm A.
- Arm_B_i is a binary indicator indicating that the applicant was randomly assigned to receive Intervention Arm B.
- X_i is a vector of application-level baseline characteristics collected during the application (prior to assignment), where the vector of covariates includes:
 - *language_english*, is a binary indicator for whether an applicant reported preferring English (versus Chinese or Spanish, the other two languages offered for the digital assister)
 - *hh_size*_i is an integer for the total number of people in the household reported in the initial application
 - on_ssi, is a binary indicator for whether an applicant reported being on SSI
 - used_desktop_computer, is a binary indicator for whether an applicant used a desktop computer to complete the application (versus using a mobile device, tablet, etc.)
 - over_60_disabled, is a binary indicator for whether an applicant reported being over 60 or disabled; and
- e_i is the idiosyncratic error term.

We will estimate each of these statistical models with heteroskedastic-consistent standard errors (HC2), per <u>OES guidance</u>. For the covariate-adjusted models, we will use Lin-adjusted covariates.⁷

We will use *emmeans* in R to test the pairwise comparisons between each of the conditions. The *emmeans* package will take the regression model specified in the <u>Statistical Models</u> section and output post-hoc pairwise comparisons for:

- Control versus Intervention Arm 1
- Control versus Intervention Arm 2
- Intervention Arm 1 vs. Intervention Arm 2

⁷ See Winston Lin. 2013. Agnostic Notes on Regression Adjustment to Experimental Data: Reexamining Freedman's Critique. *The Annals of Applied Statistics* 7(1): 295-318.

This method will be used to test the following hypotheses:

$$H_{Control.Arm1}: \mu_{submission.Control} = \mu_{submission.Arm1}$$
$$H_{Control.Arm2}: \mu_{submission.Control} = \mu_{submission.Arm2}$$
$$H_{Arm1.Arm2}: \mu_{submission.Arm1} = \mu_{submission.Arm2}$$

Multiple comparisons

As we only have one primary outcome, we will not adjust for the number of outcome-related hypotheses, but we do need to adjust for multiple comparisons generated from the three post-hoc pairwise comparisons. We will use the Tukey adjustment in *emmeans* to adjust these comparisons.

Robustness checks

The unadjusted model will be our primary statistical model and the basis for our confirmatory test. We also plan to conduct the following robustness checks to the confirmatory model:.

- Estimate the fully adjusted OLS model with the covariates listed above.
- Estimate the fully adjusted OLS model only for applicants who are eligible for the randomized evaluation (stable income = 1, single job = 1, self-employed =0) based on their final responses, marked for inclusion, and included in the randomized evaluation.
- [If available] Estimate the fully adjusted linear model only for applicants who have cookies turned on.⁸
- 2) Descriptive analysis of reporting option selection

We will not rely on a statistical model for the descriptive analysis of reporting option selection.

Our analysis of the selection of individual reporting options will be to calculate the proportion of all applicants who selected each reporting option in Intervention Arm 2 choice set (including "give an estimate" and no selection). We will not conduct a hypothesis test for differences in the share of applicants selecting each individual reporting option.

Exploratory analysis:

<u>Regression analysis for secondary outcomes:</u> For each secondary outcome, we will use the same approach of taking the F-test of joint significance before measuring pairwise differences if the F-test is significant for that outcome. We will run the same OLS model specified for our primary analysis for each of the secondary outcomes. The list of secondary outcomes can be found in the <u>Secondary Hypotheses</u> section of the analysis plan. Note that for the continuous outcomes (total amount of income reported and time spent on the application), we may run two separate regressions for each outcome – one that includes and one that excludes people who did not submit the application.

⁸ This would effectively only include applicants for which Vanity was able to re-assign to the same intervention condition upon returning to the application.

- 2) <u>Subgroup analysis:</u> As with the primary model, we will use the unadjusted model as our primary regression specification. We will run a regression model that regresses our primary outcome of application submission on treatment, the indicator for filling out the application using a desktop computer, and an interaction term. We will use the same approach of taking the F-test of joint significance as our confirmatory test, and conducting a test of pairwise differences if the F-test is significant.
- 3) <u>Common application drop-off points:</u> We will analyze the data indicating which application screens were completed to learn when applicants are more or less likely to abandon the application. We will analyze drop-off rates only for questions that occur during or after the income module.
- 4) <u>Instrumental variables analysis:</u> We will run a first stage analysis to measure the extent to which treatment assignment induced applicants to select one or more reporting options. If we find a strong first stage then we will run a two-stage least squares regression of predicted likelihood for selecting that reporting option on application submission.

Inference criteria, including any adjustments for multiple comparisons:

Our confirmatory test of pairwise differences adjusts for multiple hypothesis corrections, as we discuss in the section in Statistical Model for Confirmatory Analyses, sub-section <u>Multiple</u> <u>Comparisons</u>. We will reject the null hypothesis that there is no significant difference in outcomes between intervention arms at a significance level of p = 0.05 to determine statistical significance using two-tailed tests. We will reject the null hypotheses of pairwise comparisons at a significance level of p = 0.05 and we will use the Tukey adjustment for three comparisons.

Limitations:

The primary limitation arises from the fact that people can change their responses and eligibility for the evaluation by using the back button. This issue leads to two limitations:

- Eligible applicants may be excluded from the randomized evaluation (which poses an external validity challenge) and ineligible applicants may be included in the randomized evaluation (which introduces noise).
- Applicants' initial answers are rewritten after they have been assigned to (or completed) the randomized evaluation module when they use the back button.

Link to an analysis code/script:

N/A