Analysis Plan

Project Name: Reducing filing errors via outreach and education to tax preparers

Project Code: 2505

Date Finalized: 4/15/2025

Project description

In 2023, the Internal Revenue Service (IRS) estimated over $24 billion dollars in overpayments from three high-priority refundable tax credit programs: the Additional Child Tax Credit, American Opportunity Tax Credit, and Earned Income Tax Credit.[1] This evaluation examines the impacts of education and outreach — including webinars, letter outreach via mail and online delivery, and phone calls — to tax preparers who potentially made errors on their clients' returns. One specific aim is to understand how the type of outreach — for instance, letters versus phone calls — affects compliance. Another aim is to understand whether the effects of letters differ when mailed or distributed electronically. The IRS will use evidence from this evaluation to make data-informed decisions to continuously improve their education and outreach under the Return Integrity and Compliance Services (RICS) Return Preparer Strategy (RPS) to enforce tax compliance among tax return preparers.

Evaluation design

This randomized evaluation was designed to understand the effects of outreach to tax preparers. Figure 1 shows the randomized design. Preparers were assigned to one of four combinations of pre-filing season and filing season interventions:

1. **Control:** Control group (no intervention) during both the pre-filing season and the filing season
2. **Webinar:** An invitation to a webinar during the pre-filing season and no intervention during the filing season
3. **Pre-filing season letter and filing season call:** A letter during the pre-filing season and a call during the filing season
4. **Pre-filing season letter and filing season letter:** A letter during the pre-filing season and a second letter during the filing season

[1] "Agency Financial Report, FY 2023," Department of the Treasury, accessed November 2024, https://home.treasury.gov/system/files/266/Treasury-FY-2023-AFR-111523.pdf, p. 50.

Additionally, IRS has the option to distribute letter outreach via the National Distribution Center (NDC) for all tax preparers and by NDC or Online Accounts (OLA) for tax preparers who have set up online accounts (approximately four out of five preparers in the evaluation sample) . Letters are delivered by mail for both distribution approaches and in the future letters will also be uploaded to preparers' online accounts when distributed by OLA.

Since distribution via OLA was a new operational approach available to IRS, another objective of the evaluation is to examine whether the effects differ when pre-filing season and filing season letters are distributed via NDC or distributed via OLA. In practice, a few key operational differences between NDC and OLA for the pre-filing season letter included the timing of when letters were sent, the personalization of the letters, and whether the letters were translated into Spanish by default, among other differences.

Among preparers with online accounts, the evaluation design included randomization of the distribution method (NDC or OLA) for the pre-filing season letter (Figure 2). All filing season letters and webinar invitations were distributed through NDC.

Figure 1. Intervention assignment flow-process

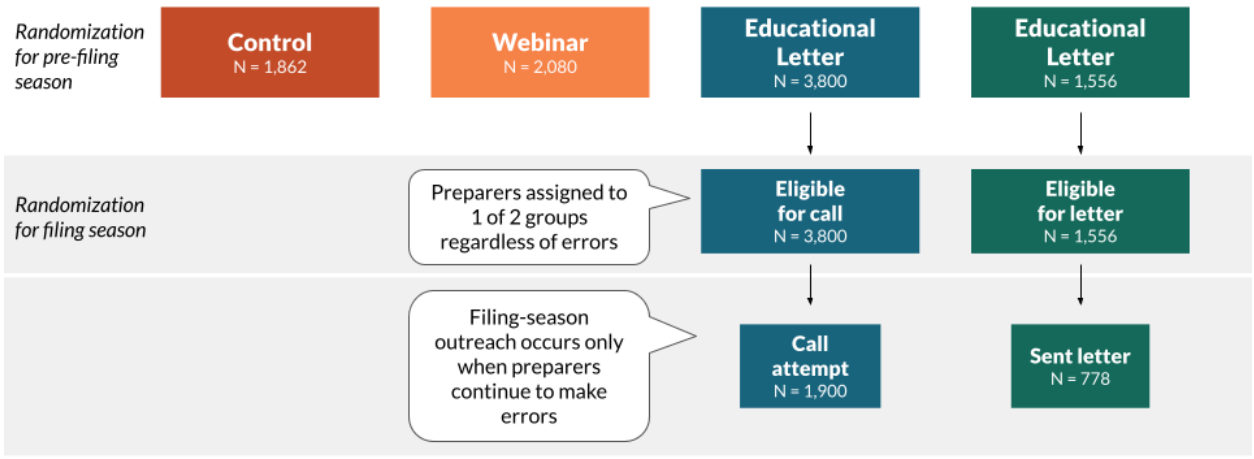
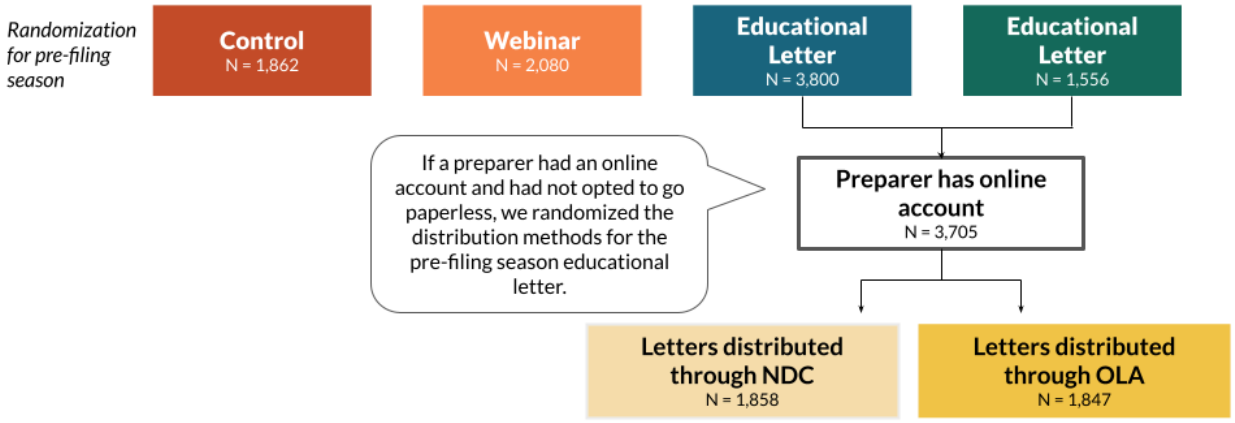


Figure 2. Random assignment among preparers with online accounts[2]



Randomization occurred with blocks of preparers who were similar at baseline. One dimension by which blocking occurred was the frequency in which preparers received recent outreach. This is an important dimension for blocking, since the IRS makes data-informed decisions on future outreach for these groups separately and the treatment effects likely differ for each group. The groups are defined as:

- New RPS preparers: This group includes preparers who have not received compliance-focused outreach from the IRS during the previous three filing seasons. Based on results from a previous OES evaluation, we expect the treatment to have a larger effect on this group of preparers.[3]
- Returning RPS preparers: This group includes preparers who have received compliance-focused outreach from the IRS during the prior three filing seasons. Based on the results from a previous OES evaluation, we expect the treatment to have a smaller effect on this group of preparers.[4]

Preregistration details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before outcome data are analyzed.

[2] Note that 55 preparers with online accounts opted to go paperless. These preparers are excluded from the NDC vs. OLA analysis, since their distribution method via OLA was determined by their decision to go paperless rather than random assignment. OLA distributed a total of 1902 pre-filing season letters.
[3] https://oes.gsa.gov/projects/return-preparer-tax-compliance/.
[4] *Ibid.*

Hypotheses

For Research Questions 1 and 2, we will examine effects for *infrequent outreach preparers* separately from *frequent outreach preparers*. For Research Question 3, we pool across the two preparer groups.

Research question 1 (primary): Does outreach to tax preparers improve tax compliance?

> **Hypothesis 1:** Tax preparers who were invited to attend a pre-filing season webinar will have a lower total refund amount and lower total erroneous dollars than those who were assigned to the control group.
>
> **Hypothesis 2:** Tax preparers who were sent a letter during the pre-filing season and were randomized to receive a filing season call if they reached the error threshold will have a lower total refund amount and lower total erroneous dollars than those who were assigned to the control group.
>
> **Hypothesis 3:** Tax preparers who were sent a letter during the pre-filing season and were randomized to receive a filing season letter if they reached the error threshold will have a lower total refund amount and lower total erroneous dollars than those who were assigned to the control treatment.

Research question 2 (primary): Are there differences in effects between filing season outreach methods on tax compliance?

> **Hypothesis 4:** Tax preparers who were sent a pre-filing season letter and were randomized to receive a filing season call will have a different total refund amount and a different total erroneous dollars than those who were sent a pre-filing season letter and were randomized to receive a filing season letter.

Research question 3 (primary): How does the distribution method for letters/notifications affect tax compliance?

> **Hypothesis 5:** Within the set of tax preparers who have registered for an online account (an OLA account), those sent a pre-filing season letter via OLA will have a different probability of reaching the error threshold for receiving the filing season treatment compared to those sent a pre-filing season letter via NDC.

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):

Our primary data source will be data that the IRS pulls for return preparers and clients during the filing season. OES will access this source in the form of processed, return-level data that Taxpayer Services, a division of the IRS, pulls for return preparers and clients (i.e., returns) at the end of the 2025 filing season (summer 2025).[5]

The primary time periods that will be used for analysis are returns filed during the 2024 filing season (for pre-treatment covariates and blocking) and returns filed during the 2025 filing season up until the end of June 2025 (for outcomes).

Outcomes to be analyzed:

For the purposes of this evaluation, we will examine benefits that include: the earned income tax credit (EITC), child tax credit/additional child tax credit/credit for other dependents (CTC/ACTC/ODC), American opportunity tax credit (AOTC) and head of household (HOH) filing status. We measure likely errors in claiming these benefits using IRS algorithms used to identify common errors, since only audits can identify errors with certainty.

We anticipate analyzing the following primary outcomes:

Primary outcomes (Research questions 1 and 3)

1. **Sum of erroneous dollars:** This is a continuous measure that equals the numeric value for the sum of erroneous dollars that may have been claimed for certain benefits aggregated across the return preparer's returns filed for their clients.

   Note that this measure includes the total credit amount for EITC, CTC/ACTC/ODC and AOTC when these credits have been claimed with likely error. That is, this measure does not distinguish between the portions of the credit claimed with likely error (erroneous dollar amount claimed) and portions of the credit claimed without errors (non-erroneous dollar amount claimed). Since there is not a credit amount associated with claiming HOH with likely error, this measure excludes dollars claimed erroneously by claiming HOH with likely errors.

2. **Total refund amount:** A numeric variable representing the refund amount aggregated across the return preparer's returns filed for their clients.

Primary outcome (Research question 2)

[5] This corresponds to approximately cycle 23.

3. **Whether the preparer makes a sufficient number of errors during the filing season to qualify for a filing season intervention:** this is a binary measure of whether the preparer meets the threshold to qualify for a filing season intervention.

   We focus on this outcome for the analysis of the distribution method of the pre-filing season letter, since preparers sent the pre-filing season letter could also receive filing season treatments, which introduce a source of post-treatment bias for isolating the effect of the pre-filing season letter on end-of-filing season outcomes. More specifically, if there are different effects of the distribution methods on the accrual of enough filing season errors to qualify for a filing season treatment, we risk conflating the effects of the distribution method with the effects of the filing season treatment. Additionally, the sharpest contrasts between the two distribution methods occurs for the pre-filing season letter.

Additional secondary outcomes include:

4. **Proportion of tax returns that may contain errors when claiming Refundable Tax Credits (RTCs)[6]:**

$$\frac{Number\ of\ returns\ prepared\ that\ may\ contain\ one\ or\ more\ errors\ when\ claiming\ certain\ benefits}{Total\ number\ of\ returns\ prepared},$$

   Where, for the purpose of this study, RTCs include: the earned income tax credit (EITC), child tax credit/additional child tax credit/credit for other dependents (CTC/ACTC/ODC), American opportunity tax credit (AOTC) and head of household (HOH) filing status.

5. **Whether the preparer filed any returns:** A binary indicator for whether the preparer filed any returns during the 2025 tax season. The aim of this secondary outcome is to capture whether the treatment caused preparers to cease filing returns for any clients (since we'd expect the underlying risks of attrition to be balanced across randomization conditions at baseline).
6. **Number of returns filed**: A numeric variable representing the number of tax returns filed by the tax preparer (note that this count excludes the tax preparer's own tax return). We impute this value to 0 if the preparer filed no returns for clients in filing season 2025.
7. **Average sum of erroneous dollars per client:** this measure divides the sum of erroneous dollars measure by the number of clients a preparer filed returns for in filing season 2025[7] We impute this value to 0 if the preparer filed no returns for clients in filing season 2025.

[6] We refer to Refundable Tax Credits (RTCs), but note that we will also measure likely errors in claiming head of household (HOH) filing status.
[7] This helps us disentangle whether drops in the aggregate outcomes are relatively uniform across the preparers or whether they are driven by a few preparers with high numbers of returns.

8. **Average total refund amount per client:** this measure divides the sum of refund amount measure by the number of clients a preparer filed returns for in filing season 2025. We impute this value to 0 if the preparer filed no returns for clients in filing season 2025.
9. **Change in erroneous EITC dollars**: this measure is calculated as the EITC dollars for EITC credits claimed erroneously on returns of clients of a preparer during the endline year minus the EITC dollars for EITC credits claimed erroneously on returns of clients of a preparer during the baseline year.[8,9]
10. **Change in erroneous ACTC dollars:** this measure is calculated as the ACTC dollars for ACTC credits claimed erroneously on returns of clients of a preparer during the endline year minus the ACTC dollars for ACTC credits claimed erroneously on returns of clients of a preparer during the baseline year.
11. **Change in AOTC dollars:** this measure is calculated as the AOTC dollars (regardless of errors present) on returns of clients of a preparer during the endline year minus the AOTC dollars (regardless of errors present) on returns of clients of a preparer during the baseline year.
12. **Change in refund dollars:** this measure is calculated as the refund dollars (regardless of errors present) on returns of clients of a preparer during the endline year minus the refund dollars (regardless of errors present) on returns of clients of a preparer during the baseline year.

Imported variables:

N/A

Transformations of variables:

N/A (See definitions above.)

Transformations of data structure:

Many outcomes are based on data at the level of individual returns. For these we will aggregate return-level data up to the return preparer level. For our primary analysis, this process will be performed for returns associated with the following clients:

- **Clients from the 2025 filing season (endline clients):** A second version of return-preparer-level outcomes will be constructed based on the pool of clients a given return preparer served during the 2025 (post-treatment) filing season.[10]

In an exploratory analysis, we will examine the following group:

[8] Our analysis with the "change variables" as outcomes will not include lagged baseline variables as covariates.
[9] Note that the data does not distinguish between portions of a refundable tax credit which are flagged with likely errors and portions which are not flagged. As such, the entirety of the refundable tax credit is treated as erroneous dollars.
[10] We will try to limit this group to those for whom returns are filed for tax year 2024, as opposed to late returns from tax year 2023.

- **Clients from the 2024 filing season (baseline clients):** One version of return-preparer-level outcomes will be constructed based on the pool of clients a given return preparer served during the 2024 (pre-treatment) filing season.

Data exclusion:

All preparers randomized will be included in the analysis of the effect of pre-filing season interventions on errors accrued during the filing season. Similarly, all preparers randomized will be included in the analysis of the combined interventions on end-of-filing season outcomes (total refund amounts and sum of erroneous dollars).

Treatment of missing data:

For the purposes of this study, our analysis will rely on data processed by the end of July 2025 for end-of-filing season outcomes, when we expect the majority of returns to be fully processed. Until then, there may be individuals who have filed their returns, but their return has yet to be processed fully. In this case, outcomes data will be missing for these returns until they are fully processed.

Descriptive statistics, tables, and graphs

We will have six figures, each corresponding to one of the primary outcomes:

- Figure 1A and 1B: Impact of combined treatments on the sum of erroneous dollars. This will have a bar for the following conditions, with standard error bars constructed from the below regression:
    - Control during both pre-filing season and filing season
    - Webinar invitation during pre-filing season
    - Letter during pre-filing season and call during filing season
    - Letter during pre-filing season and another letter during filing season
    - P-value for difference in effects between two letter during pre-filing season group

  We will show these results in two figures, one for *infrequent-contact preparers (Figure 1A)* and another for *frequent contact preparers (Figure 1B)*.

- Figure 2A and 2B: impact of combined treatments on the total refund amount. This will have a bar for the following conditions, with standard error bars constructed from the below regression:
    - Control during both pre-filing season and filing season
    - Webinar invitation during pre-filing season
    - Letter during pre-filing season and call during filing season
    - Letter during pre-filing season and another letter during filing season
    - P-value for difference in effects between two letter during pre-filing season group

  We will show these results in two figures, one for *infrequent-contact preparers (Figure 2A)* and another for *frequent contact preparers (Figure 2B)*.

- Figure 3: Impact of distribution method on hitting threshold for filing season treatment. This will have a bar for the following conditions, with standard error bars constructed from the below regression:
    - Pre-filing season letter distributed via NDC
    - Pre-filing season letter distributed via OLA

Statistical models and hypothesis tests

This section describes the statistical models and hypothesis tests of 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 rely on the following key regression specifications. All analyses examine the intent-to-treat effect of being randomized to the condition. For our primary analysis for Research Questions 1 and 2, we will run separate regressions using Specification 1 for each of our samples of interest: *infrequent contact preparers and frequent contact preparers*. For this analysis the reference group is preparers assigned to the control group. For our primary analysis for Research Question 2, we will run one regression where the sample is limited to preparers who have online accounts and who were assigned to one of the pre-filing season letter groups.

We will run all models using OLS with Lin-adjusted covariates, and we will use heteroskedastic robust standard errors (HC2).[11] We use OLS even for the binary outcomes for better interpretability of the treatment effect estimates.

Specification 1 (Research questions 1 and 2):

$$Y_{ibt} = \alpha_0 + \alpha_1 W_{ibt} + \alpha_2 LC_{ibt} + \alpha_3 LL_{ibt} + \alpha_4 Y_{ib,t-1} + \gamma Z'_{ibt} + \varepsilon_{ibt}$$

where i indexes return preparers in block b in tax return year t and:

- Y_{ibt} is our primary or secondary outcome of interest, as defined above (for instance, proportion of tax returns that may contain errors in claiming certain benefits);
- W_{ibt} is one if return preparer i had the combined treatment status of a webinar invitation during the pre-filing season and no intervention during the filing season
- LC_{ibt} is one if return preparer i had the combined treatment status of a pre-filing season letter and eligibility for a filing season call
- LL_{ibt} is one if return preparer i had the combined treatment status of a pre-filing season letter and eligibility for a filing season letter
- $Y_{ib,t-1}$ is the lagged outcome measure from the 2024 filing season;

[11] 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.

- Z'_{ibt} are the categorical variables used to generate the blocks; and
- ε_{ibt} is an error term.

For research question 1, we will test three null hypothesis $\alpha_1 = 0$, $\alpha_2 = 0$, and $\alpha_3 = 0$ on two primary outcomes (refund amount and sum of erroneous dollars) for two samples (infrequent contact preparers and frequent contact preparers). For Research Question 2, we will test one null hypothesis $\alpha_2 = \alpha_3$ on two primary outcomes (refund amount and sum of erroneous dollars) for two samples (infrequent contact preparers and frequent contact preparers).

Specification 2 (Research question 3):

$$Y_{ibt} = \beta_0 + \beta_1 OLA_{ibt} + \beta_2 Y_{ib,t-1} + \delta Z'_{ibt} + \varepsilon_{ibt}$$

where i indexes return preparers in block b in tax return year t, restricted to preparers (1) randomized to the pre-filing season letter condition and (2) who have an online account:

- Y_{ibt} represents whether return preparer i accrued 10+ errors during the window for sending out filing season interventions;
- OLA_{ibt} is one if return preparer i was randomized to the OLA distribution method (with the reference category being NDC);
- $Y_{ib,t-1}$ is the lagged outcome measure from filing season 2024;
- Z'_{ibt} is a vector of categorical variables used to generate random assignment block; and
- ε_{ibt} is an error term.

For research question 3, we will test one null hypothesis $\beta_1 = 0$ on one primary outcome (whether the preparer hit the threshold for making errors) on one sample of preparers (preparers with online accounts assigned to the pre-filing season letter group).

Confirmatory analyses:

We will treat the following tests as confirmatory, also specifying the family of tests for the purpose of adjusting for multiple testing. Each focuses on the filing season 2025/tax year 2024 clients of preparers for the confirmatory analyses:

Table 1A. Families of tests for research question 1

Hypothesis: Outcome	**Test (H_0)** Specification 1	**Sample and Family**	
		Infrequent	**Frequent**
H1: Sum of erroneous dollars	$\alpha_1 = 0$	1	2
H1: Refund amount	$\alpha_1 = 0$	1	2
H2: Sum of erroneous dollars	$\alpha_2 = 0$	1	2
H2: Refund amount	$\alpha_2 = 0$	1	2
H3: Sum of erroneous dollars	$\alpha_3 = 0$	1	2
H3: Refund amount	$\alpha_3 = 0$	1	2

Table 1B. Families of tests for research question 2

Hypothesis: Outcome	**Test (H_0)** Specification 1	**Sample and Family**	
		Infrequent	**Frequent**
H4: Sum of erroneous dollars	$\alpha_2 = \alpha_3$	3	4
H4: Refund amount	$\alpha_2 = \alpha_3$	3	4

Table 1C. Families of tests for research question 3

Hypothesis: Outcome	**Test (H_0)** Specification 2	**Sample and Family**
		Pre-filing season letter group with online accounts
H5: Hit likely error threshold for filing season outreach	$\alpha_2 = \alpha_3$	5

Exploratory analysis:

<u>Additional comparisons:</u>

- *Research question 2:* We plan to test for differences in effects between the webinar invitation and the pre-filing season letter and filing season letter outreach and test for

difference in effects between the webinar invitation and the pre-filing season letter and filing season call outreach.
- *Research question 3:* Among preparers with online accounts, we plan to compare the control group to the NDC pre-filing season letter group and compare the control group to the OLA pre-filing season letter group.

Additional outcomes:

For *Research Question 1 and 2:* We plan to run our primary regression models on secondary outcomes as described above.

For *Research Question 3:* We plan to examine the impact of the distribution method on the primary end-of-filing season outcomes (sum of erroneous dollars and total refund amount). For this analysis, to avoid the risk of post-treatment bias if the distribution method affects who qualifies for a filing season treatment, we will estimate the distribution method effects as part of a bundled treatment alongside the filing season intervention, estimating the following among preparers with online accounts and including the control group:

$$Y_{ibt} = \beta_0 + \beta_1 OLA_{ibt} * FS\ call_{ibt} + \beta_2 OLA_{ibt} * FS\ letter_{ibt} + \beta_3 NDC_{ibt} * FS\ call_{ibt} + \beta_4 NDC_{ibt} * FS\ letter_{ibt} + \beta_5 Y_{ib,t-1} + \delta Z'_{ibt} + \varepsilon_{ibt}$$

Note that we anticipate weaker effects of the distribution methods on these outcomes relative to the filing season outcome and treat this as exploratory.

Additional sample:

We plan to run our primary regression models on primary outcomes using different sample specifications:

- For *Research Questions 1 and 2*, pooling across *Infrequent Contact Preparers* and *Frequent Contact Preparers.*
- For *Research Question 3*, running the analysis separately for *Infrequent Contact Preparers* and *Frequent Contact Preparers.*
- For *Research Questions 1-3*, we plan to conduct our primary analysis using the baseline client aggregation approach as described above.

Sensitivity analysis and robustness checks

- Analyzing the sensitivity to exclusion of outlier preparers with high sums of erroneous dollars and total refund amounts: we will plot the distribution of these outcome variables and test the sensitivity of results to excluding preparers at the 99th percentile of each outcome for the respective regression. We will also estimate a rank regression that transforms the outcome from raw values into percentile ranks.

- Survival approach to examining the accrual of filing season errors: our primary outcome for the distribution methods analysis examines a binary indicator for whether the preparer accrued 10+ errors during the window for qualifying for filing season interventions (end of

February 2025). Since the treatment may influence not only whether a preparer accrues errors, but how quickly they accrue these errors, an exploratory analysis will use a survival model to examine the effect on "days or weeks (depending on granularity of measurement) until qualifies."

Complier effects

Complier effects for webinar: The main analysis looks at the intent-to-treat (ITT) of receiving an invitation to the webinar, regardless of whether the preparer attends. To examine complier effects, we will use the 2SLS method to: (1) first regress webinar attendance on the treatment indicator and (2) use the fitted values from step one to examine the impact on the primary outcomes of sum of erroneous dollars and total refund amounts. This will help us understand the impact on invited preparers who attended the webinar. Equations are here:

First stage: $AttendWebinar_{ibt} = \beta_0 + \beta_1 WebinarInvitation_{ibt} + \delta Y_{ib,t-1} + \gamma Z'_{ibt} + \varepsilon_{ibt}$

Second stage: $Y_{ibt} = \beta_0 + \beta_1 Predicted\ AttendWebinar_{ibt} + \delta Y_{ib,t-1} + \gamma Z'_{ibt} + \varepsilon_{ibt}$

Complier effects for comparing filing season letter to a filing season call: preparers randomized to be eligible for a filing season letter or call all received the same pre-filing season treatment: a letter. Therefore, and since the randomization was done without respect filing season error rates (i.e., unconditional upon filing season errors), we will subset this comparison to preparers who reached the threshold for getting a filing season treatment. Then, we will conduct two comparisons:

- **Letter sent (omitted category) versus call attempt:**

$$Y_{ibt} = \beta_0 + \beta_1 CallAttempt_{ibt} + \delta Y_{ib,t-1} + \gamma Z'_{ibt} + \varepsilon_{ibt}$$

- **Letter sent (omitted category) versus call pickup/successful call (2SLS approach):**

First stage: $Call\ pickup_{ibt} = \beta_0 + \beta_1 Randomized\ to\ call_{ibt} + \delta Y_{ib,t-1} + \gamma Z'_{ibt} + \varepsilon_{ibt}$

Second stage: $Y_{ibt} = \beta_0 + \beta_1 Predicted\ call\ pickup_{ibt} + \delta Y_{ib,t-1} + \gamma Z'_{ibt} + \varepsilon_{ibt}$

Inference criteria, including any adjustments for multiple comparisons:

We will apply multiple hypothesis corrections within the above-defined families of OES-reported primary outcomes, across all relevant tests of interest (see table above). Because some of the outcomes within a family may be highly correlated, we will run simulations to control the family-wise error rate, in line with point #7 in Alex Coppock's guide.[12] We will then use a FWER = 0.05 as the cutoff for statistical significance.

[12] https://egap.org/resource/10-things-to-know-about-multiple-comparisons.

If reporting point estimates unadjusted for familywise error rate, we will use a cutoff of p = 0.05 to determine statistical significance (with stars according to +p=0.10, *p = 0.05, and **p=0.01). All t-tests will be two-tailed.

Limitations:

We anticipate the following limitations:

- **Measurement error in likely errors:** Our measurement of errors is based on likely errors, since true errors can only be identified by audit.
- **Difficulty measuring end-of-year filing season outcomes for distribution method:** Since preparers sent pre-filing season letters were also eligible for filing season treatments that were dependent on their filing season behaviors, our evaluation design makes it difficult to disentangle effects of the pre-filing season distribution from the filing season treatments. We account for this limitation by measuring the effects of the distribution method on an intermediate outcome—hitting the threshold for a filing season treatment. Future research could further disentangle the effects of pre-filing season treatments from filing season treatments by randomizing a proportion of the preparers in a pre-filing season outreach group to a control group during the filing season.
- **Difficulty interpreting complier effects of the webinar:** our proposed method for analyzing the effect of the webinar among those who attended the webinar relies on the exclusion restriction assumption: that the only path through which the webinar invitation affects potential outcomes among preparers is via their probability of attendance at the webinar. This assumption is violated if the webinar invitation affects preparer behavior via a pathway other than through webinar attendance—for instance, if a preparer doesn't attend the webinar but the language of the invitation (which mentions errors) produces changes in compliance. Because of this limitation, we will interpret this analysis in conjunction with the ITT effects (if the webinar invitation has no observed direct impacts on preparer outcomes then there is a lower risk of an exclusion restriction violation). We also note the complier analysis as an exploratory one that should receive follow up via an RCT aimed at testing this pathway more directly (e.g., randomized interventions to encourage attendance).

Appendix A. Statistical power

We have also analyzed the statistical power of our primary specifications. We modify the sample parameters in this analysis to calculate minimal detectable effects (MDEs) for the New RPS Preparer sample, the Returning RPS Preparer sample, and the pooled sample that includes New RPS Preparers and Returning RPS Preparers.

For *Research Questions 1 and 2*, we run this analysis using total refund amount as the outcome of interest.[13] For *Research Question 3*, we run this analysis using the binary outcome of meeting the threshold for the filing season treatment.

Summary of power analysis for Research Question 1

For our power analysis of the effect of each outreach method on total refund amount (Table 2), in our pooled approach, we are able to detect between a $56,188 MDE for the comparison between a pre-filing season letter and a filing season call versus the control group and a $68,110 MDE for the comparison between pre-filing season and filing season letter group and the control group. The webinar group falls in between with a $63,292 MDE.

For our primary analysis, the MDEs are similar in range ($59,913 up to $72,791) when analyzing the effect among the New RPS Preparer group. The 2021 OES evaluation of outreach to tax preparers found significant effects among this preparer segment and reductions of the total refund amount by $36,351. While this is smaller than the MDEs we observed in prior evaluation, we expect improvements in power from blocking and the inclusion of covariates, and believe we are reasonably well powered among this group. In contrast, for the Returning RPS Preparer group, the MDEs are significantly larger due to the smaller sample size, and results from the past evaluation found a much smaller effect ($4,625) so we expect to be underpowered when analyzing this subgroup.

[13] The other primary outcome, sum of erroneous dollars, is highly correlated with this outcome so we do not conduct a separate power analysis.

Table 2. Minimal Detectable Effects (MDE) for Research Question 1 on total refund amount

Treatment group	***Preparer sample***					
	New RPS Preparer Group (Mean: $777,948; SD: $637,665)		**Returning RPS Preparer Group** (Mean: $1,111,499; SD: $799,216)		**Pooled** (Mean: $874,816; SD: $704,909)	
	N	**MDE**	**N**	**MDE**	**N**	**MDE**
Control	1,328	NA	534	NA	1,832	NA
Webinar	1,478	$67,555	602	$133,216	2,080	$63,292
Pre-filing season letter and filing season call	2,695	$59,913	~1,105	$118,053	3,800	$56,188
Pre-filing season letter and filing season letter	~1,104	$72,791	~452	$143,248	1,556	$68,110

Summary of power analysis for Research Question 2

Turning to Research Question 2, we show the MDEs for this question in Table 3. When comparing the effect of the filing season letter and call treatments to each other (as opposed to comparing each to the control group) we see that we are similarly well powered as we are for Research Question 1 when considering the New RPS Preparer group sample (MDE: $63,864) and pooled sample (MDE: $59,446). We are again underpowered for the Returning RPS Preparer group sample.

Table 3. Minimal Detectable Effects (MDE) for *Research Question 2* on total refund amount

Treatment group	***Preparer sample***					
	New RPS Preparer Group (Mean: $777,948; SD: $637,665)		**Returning RPS Preparer Group** (Mean: $1,111,499; SD: $799,216)		**Pooled** (Mean: $874,816; SD: $704,909)	
	N	**MDE**	**N**	**MDE**	**N**	**MDE**
Pre-filing season letter and filing season call	2,695	N/A	~1,105	N/A	3,800	N/A
Pre-filing season letter and filing season letter	~1,104	$63,864	~452	$125,096	1,556	$59,446

Summary of power analysis for Research Question 3

For our power analysis of the effect of distribution method on qualifying for a filing season treatment, where we contrast the new distribution method (OLA) against the comparison distribution method (NDC), we are powered to detect a 5.6 percentage point difference for the pooled sample, a 5.31 percentage point difference among New RPS Preparers, and an 8.7 percentage point difference among the Returning RPS Preparers. We do not have results from a previous evaluation to benchmark these MDEs against, but expect that we are powered to detect effects for the pooled sample and the New RPS Preparer sample, but underpowered for the Returning RPS Preparer sample.

Table 4. Minimal Detectable Effects (MDEs) for Research Question 3 on qualifying for a filing season treatment

Group	***Preparer sample***					
	Pooled Base rate: 0.53		**New to Preparer Program Base rate: 0.6**		**Returning to Preparer Program Base rate: 0.5**	
	N	**MDE**	**N**	**MDE**	**N**	**MDE**
Comparison group: NDC	1,858	NA	1,338	NA	520	NA
Treatment group: OLA	1,847	0.051	1,331	0.053	516	0.087