

Comparing levels of service in tax filing (Part II)

Date implemented: June 2022

Date written: January 2023

Abstract: This paper uses random assignment to compare the experiences of new and intermittent tax filers in using different tax filing services—and, especially, using simplified filing (GetCTC) versus full filing (GetYourRefund, Facilitated Self-Assistance) services. In June 2022, we messaged 41,475 clients who had used GetCTC in 2021, encouraging them to use one of four different tax filing services. We find that clients complete simplified filing at a far higher rate—about eight times more clients complete simplified versus full returns ($p < .0001$). This difference stems both from clients' greater difficulty getting through full filing, as well as their strong preference for simplified filing, as a meaningful number of clients assigned to full filing switch over to simplified filing instead. Simplified filers finish their returns far faster than full filers and, in fact, appear to claim similarly-sized refunds in cases we are able to compare. We do not detect a difference between completion rates of the two full filing services studied. Directing clients to a tool allowing them to pick the filing option that is best for them simply shows yet again that far more clients elect and complete simplified filing; among clients who were given a choice of services, 15-30 times more file a simplified versus a full return. There is no evidence of any benefit to allowing clients to pick the service that is best for them and, if anything, there is a penalty relative to sending them straight to simplified filing. Overall, this paper proves rigorously for the first time that simplified filing very meaningfully breaks down barriers for new and intermittent filers. Note that these findings are from an experiment among traditional non-filers, during the traditional filing off-season, and may or may not generalize to other populations and contexts.

Authors: Gabriel Zucker, Maximilian Hell

Other experimental results and research from GetCTC 2022 are available [here](#).

1. Research Questions

This study was designed to measure and compare the experience of new and intermittent filers using a variety of different tax preparation services. Specifically, we sought to compare filers' experiences using GetYourRefund Full Service (GYR—Code for America (CFA)'s virtual VITA service), Facilitated Self-Assistance (FSA—a TaxSlayer product provided through VITA to file a full return), and GetCTC (CFA's simplified filing service).

In 2021 and 2022, CFA offered GetCTC, a simplified tax filing service, so called because it allowed taxpayers to file a limited-scope return without reporting their income data, which is generally the most difficult aspect of filing a return. Meanwhile, GYR and FSA were “full filing” services, meaning they supported all aspects of tax filing but with the additional complexity of income reporting. While plenty of indirect evidence—and common sense—suggested that simplified filing broke down barriers relative to full filing, we did not have any direct evidence. This study was designed to provide that evidence, specifically:

1. Are new and intermittent filers more likely to finish filing a return using a simplified service, versus using a full filing service? How much more likely?
2. Are there other relevant differences in the experience of using a simplified versus a full service?

Secondarily, we also sought to learn more about the behavior of new and intermittent filers using our various full filing services. Again, as with the comparison between full and simplified filing, we had intuitions about which services would be preferable to such clients, but no direct evidence. We sought to learn:

3. Are new and intermittent filers more likely to finish a return using GYR versus FSA?

Finally, we sought to learn about the choices of new and intermittent filers given an election of filing choices, and the impact of offering that choice:

4. Given a tool to help choose between the three options, which service are new and intermittent filers most likely to elect?
5. Does providing this choice make them more or less likely to complete a return than assigning them to one at random?

Keep in mind that the study sought to answer these questions specifically for a population of new and intermittent filers—low-income households with limited or no past interaction with the tax system. The results should be understood in the context of that population.

2. Study Design and Implementation

2.1 Design and implementation

Our sample consisted of clients who had used GetCTC to file a return in 2021. Specifically, we selected 41,475 clients who:

- Successfully used GetCTC in 2021, defined as:
 - Filed an accepted return with GetCTC, *OR*
 - Submitted a return with GetCTC, received an IRS reject because their dependent had already been claimed, and received no other reject reasons.¹

¹ Our assessment was that these clients effectively showed they could successfully file a return, but they happened to have filed too late in the year for the return to be successful. That rejection was not dispositive regarding their potential success in 2022.

- Had not yet started a return using our GetCTC or GetYourRefund in 2022 (as identified by phone number and name matches).
- Used phone as their contact method in 2021.
- Filed in English in 2021.
- Claimed a dependent in 2021.²

By virtue of the fact that these clients used GetCTC in 2021, we classify them as likely “new and intermittent filers”—that is, the types of people who do not regularly file a return and are likely to miss out on refundable credits. Their 2021 GetCTC filing suggests their income is very likely below the standard deduction amount for their filing status and that they do not have an obligation to file a full return.

Clients were assigned to one of four treatment groups, receiving a text message pointing them to one of the four following services:

- GYR—The clients were assigned to participate in GYR Full Service. Clients landed on a GYR home page assigned to a specific VITA partner (Urban Upbound). Upon clicking Get Started, clients were routed straight to the Full Service option. The landing page contained a banner alerting clients of the GetCTC filing option, as well. In Full Service, clients complete an online intake that is routed upon completion to a VITA site. Volunteers at the VITA site perform two phone calls with the client to confirm their information, and they file the return through VITA software.
- FSA—The clients were assigned to participate in FSA. Clients landed on a page describing the FSA option. If they clicked Get Started, they were asked for their email address, and on the next page could continue to TaxSlayer FSA. FSA is a version of TaxSlayer’s standard tax filing software; upon completion, clients efile their return directly with the IRS. If clients need assistance during the process, they are able to access the CFA client support team using the live chat function on the GetYourRefund website or via email.
- GetCTC—The clients were assigned to participate in GetCTC. Clients landed on the standard GetCTC home page. If they clicked Get Started, they went through the GetCTC flow. In the GetCTC flow, like all GetCTC clients, clients had an option to switch instead to GYR; they were explicitly advised to do so if they were ineligible to use simplified filing. In GetCTC, clients complete their simplified return online, notably not including any income information, and efile it directly with the IRS.
- Triage—The clients were assigned to the standard home page at getyourrefund.org. Upon clicking Get Started, clients were routed into CFA’s triage tool, which asked a few questions and then recommended one or more of the three tax filing services (GYR, FSA, GetCTC). It is expected that most clients in this population would have been recommended to use GetCTC or GYR, in that order. The landing page contained a banner alerting clients of the GetCTC filing option, as well.

² Due to changes in the tax law, clients who had filed in 2021 and who were not claiming dependents in 2022 would be unlikely to have any money to claim.

This study was a follow-up to [a similar study executed in April 2022](#). That study included both a cohort of 2021 GYR clients and a cohort of 2021 GetCTC clients, and it assigned users to a GYR group, FSA group, or Triage group. (GetCTC was not yet open in April, so no clients in Cohort #1 were assigned to GetCTC.) Implementation issues resulted in that study largely not having interpretable results. 2021 GetCTC clients who had been included in the April 2022 outreach and had not started a return in the interim are referred to as Cohort #1; they inherited the same treatment status they had in the April study. To make the groups the same size, a random set of clients in two of the groups were dropped. Only a randomly selected subset of 2021 GetCTC clients had been included in the April 2022 study. The remainder—Cohort #2—were assigned for this study to one of the four possible treatments, with 50% assigned to GetCTC and 16.6% each to the other treatments. As such, the final counts in each treatment were:

	Cohort #1	Cohort #2
GYR	5,899	3,963
FSA	5,899	3,963
Triage	5,899	3,963
GetCTC	—	11,889

Note: GetCTC was not open in April for Cohort #1 outreach.

The scripts were constant across all groups, with only the links changing. The messages read:

Hello [FirstName], this is GetCTC — you used our service to claim CTC or stimulus payments last year. If you haven't filed yet this year, you might be eligible to claim cash benefits. It's free and easy to file to get your money. Visit [Link] (Automated message, do not reply)

Messages were sent on Tuesday, June 21, 2022. Messages were sent via text message to the phone number on record from 2021 GetCTC use. Because the sample was restricted to clients who completed GetCTC in English, all messages were sent in English. The message came from Code for America's five-digit shortcode, from which the client would have received numerous messages in 2021 about their return status. As such, it can be considered a trusted phone number.

To ensure all clients ultimately had an opportunity to claim their tax benefits, those who had not filed received additional messages later in the tax season encouraging them to use GetCTC, regardless of their treatment status in this experiment.

3. Measurement issues

The unique design of this study raises a variety of measurement issues, which are discussed in this section, before presenting results in the following section.

3.1 Defining return completions

As in other studies on CFA products this year, outcomes can be accounted in two different ways: (1) by matching the Social Security Number (SSN) of the individual receiving the outreach to the Social Security Number on returns filed through CFA products, or (2) by looking at returns filed using specific source URLs (e.g., tracking all returns filed by clients who navigated to getctc.org/assigned-source-url).

Measuring outcomes was a particular challenge in this study, however, because CFA observes relatively little direct data from FSA. CFA is able to track when a user opts to begin FSA but cannot resolve this user back to a given identity. Moreover, CFA does not know if (or when) a given return is completed using FSA. CFA receives only aggregate return completion statistics for clients sent to FSA.

To partially ameliorate the FSA issue, CFA secured two dedicated FSA URLs for clients in the FSA treatment group (one unique link per cohort). The FSA link was substituted into the GYR website for clients who arrived at the FSA page using the unique URL associated with the FSA subtreatments. (That is, if the client clicked the dedicated link in the outreach message, e.g., getyourrefund.org/22-benefits, the FSA page would route them to the dedicated link associated with their cohort, rather than to CFA's generic FSA link.) The dedicated FSA URL allowed us to measure the number of returns in the FSA treatment group that were actually completed, in the aggregate. (Because CFA does not observe the SSN of anyone using FSA, FSA outcomes cannot be measured using the SSN method.) This method, however, does not allow us to observe the FSA completion rate for clients in any other subtreatments; their FSA returns were routed into a general FSA link that was used by clients outside the study and cannot be disaggregated.

Use of source parameter tracking to measure outcomes has other issues in this study, as well: Clients assigned to a given source parameter may lose that source parameter before filing, for several reasons. First, they may simply navigate to getyourrefund.org or getctc.org manually, without clicking the link with a source parameter at all. Second, and more significantly, source parameters may be overwritten before filing. Specifically, source parameters do not persist across the getyourrefund.org-getctc.org boundary. (Triage, GYR, and FSA were all housed on getyourrefund.org.) Source parameters assigned on getyourrefund.org are lost if a client clicks to GetCTC (even if they click back and complete GYR or FSA). Conversely, source parameters assigned on getctc.org are lost if a client clicks to GYR. This means that any metrics calculated using source parameters are undercounts—including FSA metrics, which relied on dynamic replacement of the FSA link based on source parameters assigned for the experiment.

As a result, some comparisons are not possible in the results, and others should be taken with a grain of salt. Reliable comparisons of GYR and GetCTC use are possible using SSN matching. Any outcomes including FSA must use source URL matching and may be incomplete. And FSA returns themselves are not witnessed for any clients outside of the FSA subtreatment.

The variation in tax filing services being studied raises one final consideration: Only accepted tax returns are a consistent metric across services. Other studies of GetCTC look at both submitted and accepted tax returns, with the former often quite larger than the latter. Counts of submitted returns on

GetCTC, however, are not comparable to counts of submitted returns on GYR. Many returns that would be submitted and rejected on GetCTC are not submitted at all on GYR; the reason that would have led to the rejection is flagged by a volunteer screener earlier in the process, who resolves it or advises the client not to file. GYR submitted counts are mechanically lower than GetCTC submitted counts, for the same set of successful filers. As a result, the outcomes below look only at accepted returns.

3.2 Non-compliant return completions

Another measurement question considers what services should be counted in the headline outcome metrics.

Recall that clients in any treatment group—and especially those in full filing treatments—*could*, in principle, override their treatment assignments and opt to complete any of the tax filing services. If a client filed using a different service, should this be included as an outcome? That is, suppose we had the following data:

	Assigned to GYR	Assigned to GetCTC
Total	100	100
Completed GYR	20 (Cell A)	5 (Cell B)
Completed GetCTC	25 (Cell C)	40 (Cell D)
Total completed returns, “either service” specification	45	45
Total completed returns, “assigned service” specification	20	5

Should we score these results as 45-45 or 20-5? As we will see below, large numbers of GYR-assigned clients *did* indeed file using GetCTC, so this is not merely an academic question.

One answer is that returns from either service should be included in the outcome measure. (This is loosely related to, but not properly the same as, the notion that treatment compliance should be ignored in intent-to-treat specifications.) We call this the “either service” specification. In the example above, then, both groups have 45 returns. But it would clearly be an error to conclude from this that both services were equally easy; over twice as many clients completed GetCTC than GYR. This formulation does not actually directly measure a research question of interest.

The other answer is to look only at returns completed using the assigned service—the “assigned service” specification. In the data above, this would show GetCTC strongly outperforming GYR by a factor of four. This specification, too, seems to leave something out.

It helps to consider the assigned service specification as being comprised of four effects: (1) the rate at which clients click the link at all (which should be roughly balanced across treatments); (2) the rate at which clients elect, before starting the process, to switch services, in violation of their treatment assignment; (3) the rate at which clients start the process, find the process too hard, and stop filing

outright; and (4) the rate at which clients start the process, find the process too hard, and switch services, in violation of their treatment assignment.

Differences in the “assigned services” specification essentially measure differences in all effects; because (1) is balanced, it measures differences in (2), (3), and (4). Differences in the “either service” specification are, roughly, measuring only differences in (3); to the degree there are differences in (2) or (4), they are ignored, because the other-service return is included. The difference between the specifications, in other words, stems from (2) and (4).

But it would be a mistake to say, then, that the “either service” specification measures difficulty and the “assigned service” measures difficulty plus client choice. Switching during the process (Effect 4) may well dominate switching before starting (Effect 2); if so, then a meaningful fraction of the difference between specifications should be ascribed to difficulty. So the “either service” specification can be seen as measuring some—but not all—of the difficulty differences between the services.

In the results below, we take the view that the “assigned service” specification is cleaner for several reasons. First, given the argument above, the “either service” specification is hard to interpret. Second, given the information in Section 3.1, “either service” specifications are not possible when FSA returns are in the picture. Third, the “assigned service” specification may have more policy legibility in a context where easily switching between multiple different filing options is not possible—though it greatly depends on the details of the system. However, for headline results we also show the “either service” specification. If the two meaningfully diverge, like in the illustrative results above, it suggests that a meaningful fraction of the effect comes from effects (2) and (4), rather than (3). This raises the *possibility* that the effect is driven by (2)—that is, abstract preference, rather than actual difficulty—but does not prove such a result.

3.3 Timing of return completions

An additional thorny question concerns what deadlines should be imposed in the measurement of completed returns.

One option is simply to look at all returns completed at any point in 2022. However, for SSN-match specifications, this would bias results strongly toward simplified filing. Throughout the year—and accelerating into the fall—clients in the sample population were subjected to considerable amounts of outreach pointing them to GetCTC, but effectively no outreach pointing them to GYR. This dynamic is especially compounded by the fact that we ourselves sent further rounds of outreach to the sample population, beginning the second week in September, pointing only to GetCTC, without regard for previous treatment status. The longer the measurement period—and especially if it extends into September, when GetCTC outreach took off—the more the results may be biased toward simplified filing. (Notably, outreach—and GetCTC activity in general—were very quiet through July and August, meaning that such interference was likely minimal in these months, though nonzero.)

Another option would be to restrict only to returns completed within a very brief window, when the treatment nudges are guaranteed to be dominant. However, because full filing—and especially

GYR—takes longer than simplified filing, such a strategy would *also* bias results toward simplified filing. It is important to set a large enough window to witness all full returns prompted by the experiment.

Meanwhile, the use of FSA data places an additional restriction on measurement: Aggregate data from FSA is provided only at a monthly cadence, on the first of each month. Realistically, given that messages were sent on June 21, we could perform analysis with a completion date of July 1 (one week to complete returns), August 1 (5-6 weeks to complete returns), or September 1 (11 weeks to complete returns).

Appendix A shows the headline results under all of these different deadline specifications, as well as cumulative distribution functions showing the timing of submissions for the different treatments. The results suggest that July 1—unsurprisingly—is too early to show results, but September 1 is too late; the secular increase in simplified returns strongly dominates throughout the month of September. As a result, all results below are shown for acceptances completed by August 1.

For FSA returns, this means showing differences between the number of returns that appeared on June 1 and August 1 reports for the specified URLs. These figures could be slight overestimates for two reasons. First, we used the same URLs for the earlier round of this experiment in early April, and some FSA returns filed in June *could* have been stragglers from the earlier round of outreach—though the timing of returns in this experiment suggests this is unlikely. Second, while these links were *supposed* to be unused links from other VITA partners that saw no other outside use, it is technically possible that this was an error, and one or more VITA sites were using these links independently. Again, there is no evidence for this type of leakage.

4. Results

4.1 Full versus simplified filing

We consider first the results of full filing compared to simplified filing. Because nobody in Cohort #1 was sent directly to GetCTC, all of the results in this section use only Cohort #2. Refer to Section 3.1 for limitations on possible pairwise comparisons. Also as noted in Section 3.1, we use only accepted returns as outcomes. As noted in Section 3.3, all results are restricted to returns accepted before August 1.

Table 1: Accepted return rates for simplified vs full filing (Cohort #2 Only)

	GYR vs GetCTC Assigned service SSN match (1)	GYR vs GetCTC Assigned service Source match (2)	FSA vs GetCTC Assigned service Source match (3)	FSA+GYR vs GetCTC Assigned service Source match (4)	GYR vs GetCTC Either service SSN match (5)
Simplified filing	2.38% (283)	1.77% (211)	1.77% (211)	1.77% (211)	2.41% (287)
Full filing (FSA/GYR)	0.30% (12)	0.33% (13)	0.23% (9)	0.28% (22)	1.31% (52)
P(simp = full)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Notes: Number of returns shown in parentheses; percentage is of the treated population. Outcome variable is an accepted tax return, with acceptance date strictly before August 1, 2022; later acceptances are not included (see Appendix). Submitted

returns are not used as an outcome measure because of low comparability between GYR and GetCTC submissions. Columns (3) and (4) count FSA completed returns using SPEC reports for the specified FSA links assigned to the treatment sources, comparing differences between June 1 and August 1 reports. Column (5) shows returns accepted on either service, regardless of assignment. No clients in the GYR full service group saw an “at capacity” page, so GYR capacity would not have impacted any results, with the possible exception of Column (2), which could underestimate the number of GYR returns from the GetCTC treatment group. FSA returns could be slight overestimates because some of the returns could be late effects of the April 2022 outreach round, using the same link. p -values from χ^2 tests.

The results show clearly that simplified filing far outperforms full filing. In the principal Column 1 specification, clients complete simplified filing at a rate eight times higher than full filing. Using source matching (Column 2), looking at FSA rather than GYR (Column 3), and pooling GYR with FSA (Column 4) all change the results only slightly—simplified filing outperforms full filing by a factor of 5-8. All of these results are highly significant.

Column 5 shows the results of using the “either service” specification, rather than the “assigned service” specification. The difference between simplified and full filing is still large and highly significant, but it is now a little less than a factor of two, rather than a factor of seven. Indeed, as in the illustrative example from Section 3.2, many GYR assignees used GetCTC. Recall from Section 3.2 that this raises the possibility that some—though not all—of the enormous difference between GYR and GetCTC completions could stem from client preference rather than sheer client inability to complete the service. Saying that only a seventh as many clients are *able* to complete full filing is probably an overstatement.

To further delve into this point, Table 2 explores in more detail the causes of the difference in completion rates across GYR and GetCTC. Row 3 shows a meaningful discrepancy in the rate of clients who start each service, suggesting perhaps that a meaningful number of GYR assignees see the GetCTC banner and switch services before starting a return. But GYR is also losing more clients at various steps in the process. 61% of GetCTC assignees who enter an SSN go on to complete their return, compared to just 39% of GYR assignees completing their intake. And, in GetCTC, completing the return is nearly coterminous with submitting it; in GYR, clients must also go through days or weeks of back and forth with volunteers, during which 67% drop off the process. In total, then, only about 1 in 10 of the GYR starters get to an accepted return, compared to 34% of the GetCTC starters.

Table 2: Aspects of dropoff in simplified versus full filing (Cohort #2 only, assigned service only)

		GYR			GetCTC		
		Number	Fraction of total	Fraction of prev step	Number	Fraction of total	Fraction of prev step
1	Total in treatment group	3963	100%	—	11889	100%	—
2	Home page visit	799	20.16%	20.16%	2093	17.60%	17.60%
3	Starts intake & enters SSN	127	3.20%	15.89%	619	5.21%	29.57%
4	Complete intake / return	49	1.24%	38.58%	377	3.17%	60.90%

5	Submit return	16	0.40%	32.65%	363	3.05%	96.3%
6	Accept return	13	0.33%	81.3%	211	1.77%	58.1%

Notes: Actions only included prior to August 1. Results are calculated based on source matching, since SSN matching does not allow us to witness home page visits. p -values from χ^2 tests.

Table 3 examines other differences between the simplified and full filing experiences, beyond merely whether a return was actually completed. Note that these results are impacted by our inability to record certain aspects of certain returns, and as such may make the results look more favorable to simplified filing. We would expect that GYR may generate higher refunds than FSA and that FSA may take less time than GYR. However, we can only observe the refund amount from FSA and the time to completion of GYR.

Columns 1-3 look at the amount of the refund. (The columns compare only GetCTC and FSA because the structure of GYR is such that VITA partners observe the exact refund amount, but Code for America does not.) One common objection to simplified filing is that, while it may increase the rate at which clients complete a return, it does so at the cost of denying them their full benefits, since simplified filing included only CTC and RRC, leaving out withholding, EITC, and other credits. Surprisingly, Column 1 shows that the average refund on an accepted GetCTC return was *larger* than the average refund on an accepted FSA return. (We cannot put a p -value on this comparison because the FSA statistics are only in the aggregate, so the standard deviation of the refund amount is not known.) This counterintuitive result would suggest that there *is* no trade-off between simplified and full filing as hypothesized: More clients complete simplified filing *and* they receive as much (or more) than full filing clients. Column 2, meanwhile, shows the average refund amount per message sent, regardless of whether a return was submitted and accepted; this is a function of rate of return completion *and* the refund amount. This throws into sharper relief the contrast here: Due to the differential completion rate, the average GetCTC outreach message generated \$63 in refunds, compared to just \$7 for FSA messages.

At face value, it is hard to understand the result that the average GetCTC refund exceeds the average FSA refund. One possible explanation is that RRC functionality was reported to be harder to find in FSA than in GetCTC—FSA clients had to actively opt to claim RRC on a dense page of many options, whereas GetCTC clients were all prompted to report their EIP3 amount. Column 3 tests this hypothesis and suggests it is probably a significant part of the story. The average return filed through GetCTC claimed \$900 more in RRC than the average FSA return. In other words, the average FSA return claimed about \$600 more in other payments (e.g., EITC and withholding)—but this effect was overwhelmed by the lower RRC claim rate. Of course, it is possible that some clients claimed too much RRC, by understating the amount they received in advance payments. Even if all RRC payments were *entirely* illusory, though, FSA returns would only be claiming \$600 more than simplified returns, or 29% more. This is not a great tradeoff for a service with one seventh the completion rate.

Table 3: Other aspects of simplified versus full filing (Cohort #2 only)

	Refund amounts, FSA vs GetCTC (Source match)	Median time to acceptance, GYR vs GetCTC (SSN match)
--	---	---

	Avg total refund, accepted returns (1)	Avg total refund, all msg recipients (2)	Avg RRC amount, accepted returns (3)	From start of return (4)	From message send (5)
Simple	\$3,576 ³	\$63.46	\$1,503	40.5 hours	42.3 hours
Full	\$3,289 ⁴	\$7.47	\$622	138.3 hours	148.5 hours

Notes: FSA figures are based on the difference between June 1 and August 1 SPEC reports. Time of completion for GetYourRefund is the time of the last status transition for the return in the GYR app, which may not exactly reflect when the return was accepted. *p*-values not calculable for Columns 1-3 because the variance in refund amounts is not observed.

Columns 4 and 5 look at another aspect of the different experience across simple and full filing: the time to complete a return. In this case, there is not a hypothesized trade-off—the quicker completion time of GetCTC is another plausible advantage of simplified filing, in addition to the higher completion rates. Columns 4 and 5 bear this out; the median GetCTC acceptance came in less than two days after messages were sent (and after the process began), whereas the median GYR acceptance took about a week. Note that this is not measuring time in the app per se, merely time to overall completion.

4.2 Comparing GYR and FSA

Do new and intermittent filers finish a particular full filing service, GYR or FSA, more often? We hypothesized that such filers, who are unfamiliar with the tax system, may have more success using GYR, which provides more hands-on assistance than FSA. But the results do not bear this out; there is no significant difference in the completion rates across GYR and FSA, in either cohort, or with the results pooled. One caveat to this result is that the comparison had relatively little statistical power: With such low baseline filing rates, one service would have to greatly outperform the other to be detected with this sample size. For example, if the true FSA completion rate were 0.3%, GYR would have to have a completion rate of around 0.7% for the difference to be likely detectable.

Table 4: Full service GYR vs FSA (Source match, accepted returns, assigned treatment only)

Assignment	Both cohorts (pooled)	Cohort 1	Cohort 2
GYR	0.22% (22)	0.15% (9)	0.33% (13)
FSA	0.19% (19)	0.17% (10)	0.23% (9)
P()	0.64	0.79	0.30

Notes: Outcome is accepted returns. Source matching for GYR. Returns only included if completed before August 1. FSA returns are from the SPEC report and are only observed for clients who have the assigned source parameter. There are three clients who saw an at-capacity page. No adjustment is made for these clients, since on average, three clients who start the process would lead to far less than 0.5 completed returns. Results for submitted returns are not included because submission is not comparable across FSA and GYR. FSA results are the difference between August 1 and June 1 SPEC reports. FSA returns

³ There is a total of \$754,431 in refunds on source 'ctc-22' submitted before August 1, for 211 such returns. This refund amount includes RRC.

⁴ \$62,100 in refunds across 19 accepted returns.

could be slight overestimates because some of the returns could be late effects of the April 2022 outreach round, using the same link. p -values from χ^2 tests.

4.3 Triage

If given a tool to help people choose between the three services, which option do clients pick and how successful are they? The results from the Triage condition are somewhat difficult to parse because of challenges tracking Triage condition clients across the services. (See notes to Table 5 for details.) Recall also that most clients in the sample, given their income and tax filing history, would not have been recommended to start FSA and would only have had GYR and GetCTC as options.

Table 5 shows that far more clients in the Triage condition start GetCTC than GYR, by a factor of about four (Column 1). The difference in successful completions is even more extreme; 15-30 times as many clients submit an accepted return through GetCTC as through GYR (Columns 2-3). (The ambiguity is due to clients who hit the at-capacity page; see notes.) These numbers suggest a striking rate of drop-off through GYR, even larger than in the other cohorts. It is possible that some of the clients who started GYR were “comparison shopping”—checking out what GYR and GetCTC each, respectively, would be like, before electing which one to use. Regardless, it is clear that clients in this group overwhelmingly prefer GetCTC, given the option.

The comparison between FSA and GYR among this group is even harder to parse because of data issues (see notes). Column 4 suggests that perhaps slightly more Triage clients made it through GYR than FSA—but this could also be a simple function of the fact that more were recommended to use GYR. There is little we can actually infer from this comparison.

Table 5: Outcomes from the Triage condition (Triage condition only, pooling Cohorts #1 and #2, August 1 cutoff, N = 9862)

Service used	Starts process and enters SSN, SSN match (1)	Accepted return, SSN match (2)	Accepted return, adjusting for at-capacity, SSN match (3)	Accepted return, adjusting for at-capacity, Source match, FSA constructed (4)
GYR	1.06% (105)	0.05% (5)	0.09% (9)	0.09% (9)
FSA	N/A	N/A	N/A	0.03-0.06% (3-6, est.)
GetCTC	4.59% (453)	1.64% (162)	1.64% (162)	N/A
P()	< .0001	< .0001	< .0001	—

Notes: Service starts across GYR and FSA cannot be shown; GYR start is recorded at the top of the triage process and cannot be disaggregated as a GYR start until SSN provision, whereas FSA starts do not witness SSN provision. 41 clients selecting GYR from the Triage group saw an at-capacity page; if these clients completed GYR at the rate of others making it to that point in the process, approximately four would have gotten accepted returns. FSA completions cannot be definitively determined because Triage group clients were routed to the generic FSA link. We can impute the number of FSA completions based on the number of clicks into FSA from this group (83). The overall national conversion rate from an FSA click suggests 5.7

accepted FSA returns; the conversion rate in the more comparable FSA group in this experiment suggests 2.8 accepted FSA returns. p -values are from χ^2 tests.

What about the import of giving clients choice in general? Are clients in the Triage group more likely to complete a return, since they are given the choice and freedom to select the service that works best for them? Or is the choice itself paralyzing, and clients would be better off being sent straight to a service? Tables 6 and 7 explore this question, for accepted returns (Table 6) and started returns (Table 7). Keep in mind that many direct comparisons are impossible or compromised because of the tracking issues. The key story in Table 6 is that clients finish GetCTC at vastly higher rates than other services—and so, the more directly clients get to GetCTC, the more clients file returns at all. As a result, sending clients straight to GetCTC generates more returns than sending them to Triage, which in turn generates more returns than sending them to GYR. In all these cases, though, the vast majority of the returns are actually coming from GetCTC. If there is any beneficial effect of Triage allowing clients to select the service that is actually best for them, it is so dwarfed by the vastly higher completion rates on GetCTC that the effect cannot really be seen at the level of outcomes.

Table 6: Accepted returns for clients assigned to Triage compared to specific services

	Both cohorts (pooled)				Cohort 2			
Assnmt.	GYR (SSN match)	FSA (source match; imputed)	GetCTC (SSN match)	Total (sum of prior 3)	GYR (SSN match)	FSA (source match; imputed)	GetCTC (SSN match)	Total (sum of prior 3)
GYR	0.21% (21)	0	1.00% (99)	1.22% (120)	0.30% (12)	0.000% (0)	1.00% (40)	1.312% (52)
Triage	0.05% (5)	0.05% (5)	1.64% (162)	1.74% (172)	0.050% (2) ⁵	0.101% (4)	1.59% (63)	1.741% (69)
GetCTC					0.034% (4)	Unknown	2.38% (283)	2.414% (287)
P(GYR = Triage)	0.002	0.026	0.0001	0.003	0.008	0.045	0.020	0.119
P(GetCTC = GYR)					< 0.0001	N/A	< 0.0001	< 0.0001
P(GetCTC = Triage)					0.65	N/A	0.003	0.013

Notes: Accepted returns on each service, by treatment assignment. FSA completions are imputed based on clicks to the FSA service. In Panel A, both cohorts are shown, so GetCTC treatment is not included. p -values are from χ^2 tests. Assignment to FSA is not included because of lower tracking comparability to other services. Actions included through August 1.

⁵ The Triage condition saw one accepted GYR return, and another eight people saw the At Capacity page, yielding an estimate of two accepted returns.

Still, we can try to explore the question of whether offering choice is beneficial or paralyzing by looking at start rates. If start rates were far higher for the Triage group, it would suggest clients appreciate some amount of choice, even if the ease of completing GetCTC far outweighs this effect at the level of accepted returns. Table 7 presents these results. Keep in mind that we do not have entirely consistent measures of starts across all services—specifically, the measure of starting FSA is upstream of the other measures, and as such, FSA starts should be considered an overcount. The evidence suggests that assignment to the Triage group does not generally encourage more starts. The GYR and Triage groups have similar start rates across services; and, since FSA starts may be an overestimate, it is more likely that starts in the GYR group outnumber starts in the Triage group. The GetCTC group, meanwhile, starts returns at a far higher rate than the Triage group, suggesting that even the appeal of the simpler filing process far exceeds any advantage from giving clients a choice.

Table 7: Started returns for clients assigned to Triage compared to specific services

	Both cohorts (pooled)				Cohort 2			
Assnmt.	GYR: starts & enters SSN (SSN match)	FSA: starts (source match)	GetCTC: starts & enters SSN (SSN match)	Total started (sum of prior 3)	GYR: starts & enters SSN (SSN match)	FSA: starts (source match)	GetCTC: starts & enters SSN (SSN match)	Total started (sum of prior 3)
GYR	2.94% (290)	0.09% (9)	3.40% (335)	6.43% (634)	3.31% (131)	0.08% (3)	3.31% (131)	6.69% (265)
Triage	1.06% (105)	0.88% (87)	4.59% (453)	6.54% (645)	0.96% (38)	0.83% (33)	4.31% (171)	6.11% (242)
GetCTC					0.36% (43)	Unknown	7.03% (836)	7.39% (879)
P(GYR = Triage)	< .0001	< .0001	< .0001	0.75	< .0001	< .0001	0.020	0.36
P(GetCTC = GYR)					< .0001	Unknown	< .0001	0.099
P(GetCTC = Triage)					< .0001	Unknown	< .0001	0.006

Notes: Started returns on each service, by treatment assignment. FSA starts are based on clicks to the FSA service and are earlier in the funnel than starts on the other services. GYR and GetCTC starts are defined as entering an SSN. FSA starts were not tracked for the GetCTC condition, and the Total Started is necessarily an undercount for that cohort. Results shown are based on Cohort 2 only. p -values are from χ^2 tests. Assignment to FSA is not included because of lower tracking comparability to other services. Actions included through August 1.

5. Discussion

Overall, this study bears out in detail the finding that had already been suggested by ample circumstantial evidence: new and intermittent filers have a far easier time completing simplified filing (GetCTC) than full filing (GYR, FSA). In the standard specification, seven times more clients complete returns using GetCTC than using GetYourRefund full service. Some (but not all) of this discrepancy is due to preference—clients who are encouraged to use full service actively elect instead to switch over and use simplified filing. But much of the discrepancy is simply due to much higher dropoff throughout the GYR full service process.

Surprisingly, compared to FSA returns, clients appear to claim slightly larger refunds from GetCTC returns. This difference seems to be driven by RRC claims, which are lower on FSA returns. This effect outweighs the effect of additional payments on full returns, like withholding and EITC. Unsurprisingly, GetCTC returns take far less time than GYR returns.

This study does not detect a significant difference between the completion of GYR and FSA—although, given the low baseline completion rates, we would not have been powered to detect modest differences.

When clients are given a triage tool to help select the resource that is best for them, they overwhelmingly elect to try simplified filing, and about 15-30 times more clients go on to complete simplified filing versus full filing. The higher completion rates of simplified filing overwhelm any other effect of offering choice per se. The evidence seems to suggest, though, that offering choice itself does not carry any particular benefits on the margin. Clients assigned to the Triage group start returns at about the same or a slightly lower rate than clients assigned to the GYR group, and at a significantly lower rate than clients assigned to the GetCTC group. In this specific context, at least, it is not necessarily productive to offer clients a choice of products in the hopes they will pick one that is right for them; it is better to offer the service that they are empirically most likely to use.

Keep in mind, again, that this experiment was run within a very specific sample of new and intermittent tax filers. As clients who had used GetCTC in 2021, the households in this study do not have a filing obligation, and they very likely have rarely or never interacted with the tax system. They are also at least familiar with GetCTC and thus may have a baseline inclination to trust it. The findings may not generalize to other populations, who may be far more able and willing to complete a full return. That said, the population of new and intermittent filers is exactly the population of interest to our program and to others seeking to close the refundable credits participation gap. The findings also may not generalize to outreach campaigns run at other times of the year, including during the regular filing season.

These findings serve as rigorous confirmation of what we have already reported circumstantially in other forums: Simplified filing absolutely makes a difference for new and intermittent filers. This quicker and easier process not only saves clients time but literally makes the difference for many between getting thousands of dollars, and getting no money at all. Extending the availability of simplified filing processes like GetCTC is critical to closing the tax benefits coverage gap.

Appendix A. Date Cut-offs

Appendix Table 1: Robustness Check: Accepted Returns, Cohort 2 only, different timeline cutoffs

		Message Sent to	GYR vs GetCTC, Assigned service, SSN match (1)	GYR vs GetCTC, Assigned service, Source match (2)	FSA vs GetCTC, Assigned service, Source match (3)	FSA+GYR vs GetCTC, Assigned service, Source match (4)	GYR vs GetCTC, Either service, SSN match (5)
Finished by July 22 (1 month after msg sent)	Simplified	11889	2.29% 272	1.74% 207	—	—	2.31% 275
	Full filing	3963	0.28% 11	0.30% 12	—	—	1.19% 47
Finished by Aug 1	Simplified	11889	2.38% 283	1.77% 211	1.77% 211	1.77% 211	2.41% 287
	Full filing	3963	0.30% 12	0.33% 13	0.23% 9	0.28% 22	1.31% 52
Finished by Sep 1	Simplified	11889	2.84% 338	1.91% 227	1.91% 227	1.91% 227	2.89% 344
	Full filing	3963	0.38% 15	0.35% 14	0.25% 10	0.30% 24	1.74% 69
No restriction	Simplified	11889	5.66% 673	2.15% 256	2.15% 256	2.15% 256	5.71% 679
	Full filing	3963	0.38% 15	0.35% 14	0.28% 11	0.32% 25	5.07% 201

Appendix Figure 1-3: Cumulative distribution function of filing dates

