Can Nudges Increase Take-up of the EITC?: Evidence from Multiple Field Experiments

Elizabeth Linos, Allen Prohofsky, Aparna Ramesh, Jesse Rothstein, and Matt Unrath*†

University of California, Berkeley

July 2, 2020

Abstract

The Earned Income Tax Credit (EITC) distributes more than $60 billion to over 20 million low-income families annually. Nevertheless, an estimated one-fifth of eligible households do not claim it. We ran six pre-registered, large-scale field experiments to test whether "nudges" could increase EITC take-up (N=1 million). Despite varying the content, design, messenger, and mode of our messages, we find no evidence that they affected households' likelihood of filing a tax return or claiming the credit. We conclude that even the most behaviorally informed low-touch outreach efforts cannot overcome the barriers faced by low-income households who do not file returns.

*Contact: elinos@berkeley.edu; allen.prohofsky@ftb.ca.gov; aparna.ramesh@berkeley.edu; rothstein@berkeley.edu; unrath@berkeley.edu

†This research was conducted through the California Policy Lab at UC Berkeley. We thank Evan White, Christopher Campos, Konrad Franco, John Iselin, Sree Kancherla, and Charles Davis, who were instrumental in making this project a success. We also thank Professor Jacob Goldin and Taylor Cranor from Stanford Law School and Akshay Ravikumar Kalose from Stanford University who were critical to launching our CalFresh experiments in 2018.

We are grateful to our partners for their extraordinary work in administering these interventions and their commitment to informing policy through rigorous research. At the California Franchise Tax Board, we thank Sean McDaniel, Monica Trefz, Ian Kiltz, Chad Angaretis, Silvano Guitierrez, Bob Schlie, and Xudong Chen, and especially Allen Prohofsky and Julie Moreno. At the California Department of Social Services, we thank Alexis Fernandez, Kim McCoy-Wade, Jianjun Chen, Brittney Gossard, Dionne Evans-Dean, and Akhtar Khan. We also thank our multiple CalFresh county partners, Alberto Garcia, Deanna Abrahamian, Todd Blair, Ashley Arnold, Gladys Deloney, Yunny Tai, Margareta Hodzic, and Idelle Villarreal. We also thank Trevor Rogers, Sebastien Chaubard, Timothy Merritt and Josh Fryday at Golden State Opportunity for all of their help in administering these studies. And we thank Pete Manzo and Nalleli Sandoval at the United Way of California, along with staff from United Way Bay Area, United Way of the Wine Country, and Orange County United Way for their input along the way.

Finally, we are grateful for financial support from the Abdul Latif Jameel Poverty Action Lab, Arnold Ventures, Tipping Point Community, the University of Wisconsin Institute for Research on Poverty, The James Irvine Foundation and the Horowitz Foundation for Social Policy. All errors should be attributed to the authors. The findings reported herein were performed with the permission of CDSS and FTB. The opinions and conclusions expressed herein are solely those of the authors and should not be considered as representing the policy of the collaborating department, agency, or any department or agency in the California government.

This study is registered in the AEA RCT Registry as: "AEARCTR-0005734"

1 Introduction

The Earned Income Tax Credit (EITC) is a critical income support for working families. In 2019, 25 million households received about $63 billion nationwide, with an average benefit of approximately $2,500. Numerous studies have documented the EITC's beneficial effects on work, income, and poverty; children's educational achievement and attainment; and adult and infant health (see reviews in Hoynes and Rothstein, 2016; Nichols and Rothstein, 2016). Despite these benefits, the IRS estimates that one in five eligible households do not take up the program (IRS, nd). For eligible families with the lowest incomes, take-up may be much lower, with approximately one in two households foregoing their cash benefit (Jones, 2014; Plueger, 2009).

The academic literature in various disciplines has proposed that learning, compliance and psychological costs can explain incomplete take-up of government benefits (Finkelstein and Notowidigdo, 2019; Herd and Moynihan, 2019; Currie, 2006; Moffitt et al., 1983). To claim the EITC, households must first overcome the learning costs associated with discovering the credit exists and determining whether they are eligible. Second, for families aware of the credit, the filing process can be confusing, complex, and burdensome. Previous research suggests that the direct and indirect compliance costs of filing, difficult for the average tax filer to navigate, may be especially burdensome for low-income families (Herd and Moynihan, 2019; Goldin and Liscow, 2018; Bhargava and Manoli, 2015; Currie, 2006).

Third, even if learning and compliance costs are low, the target population may face psychological barriers that inhibit take-up. Although the EITC is thought to carry less stigma than other benefit programs (Halpern-Meekin et al., 2015), a potential recipient may nevertheless distrust government or face psychological stresses that prevent them from carrying out plans to file a return (Hovland and Weiss, 1951; Pornpitakpan, 2004). Trust may be a unique challenge for EITC outreach. Outreach messages often include promises of free cash that can be hard to distinguish from scams to which families are frequently exposed. The relative importance of all these explanations remains largely unexplored.

"Nudges" – small changes to the choice architecture surrounding a decision that aim to alter people's behavior without meaningfully changing incentives – have been used to address many of these barriers, with substantial impacts on enrollment decisions across a wide array of policy contexts (Benartzi et al., 2017; Hallsworth et al., 2017; Thaler and Sunstein, 2009; Thaler and Benartzi, 2004; Madrian and Shea, 2001). Nudges have been used to increase take-up of the Supplemental Nutrition Assistance Program (SNAP) (Finkelstein and Notowidigdo, 2019), the EITC (Bhargava and Manoli, 2015), and even college enrollment through increasing completion of the Free Application for Federal Student Aid (FAFSA) (Bettinger et al., 2012). A recent review of over 100 nudge interventions in a range of contexts, with over 20 million people nudged, found that the average effect of a nudge intervention in government is 1.4 percentage points (Dellavigna and Linos, 2020).

However, the evidence is not unambiguous; more recent studies suggest that nudges may be ineffective in other settings or may fail to scale (Camerer et al., 2018; Bird et al., 2019; Castleman et al., 2019; Bergman et al., 2019).

Studies on EITC take-up also point to mixed results. One study of IRS outreach to EITC-eligible non-filers found that this outreach had positive but limited effects on tax filing rates (Guyton et al., 2017). Kopczuk and Pop-Eleches (2007) find that the availability of tax preparation software increased EITC claiming. Chetty et al. (2013) similarly show that the availability of nearby paid tax preparation services in a neighborhood predict knowledge about the program and usage. However, Cranor et al. (2019) find that mandating employers to inform employees about their potential eligibility for the EITC had no effect on EITC take-up.

It is noteworthy that many of the early successful nudge studies focused on populations who already had some interaction with the government. For example, both the Bhargava and Manoli (2015) and Bettinger et al. (2012) interventions were conducted among taxpayers who had already filed a return and only needed to be nudged to complete additional forms covering similar material. Finkelstein and Notowidigdo (2019) contacted seniors who were already enrolled in Medicaid, but though eligible, failed to enroll in SNAP. For those who do not already file or who may have limited positive interactions with government, the learning, compliance, and psychological costs associated with EITC take-up may be much higher. More research focused on this population is needed.

To contribute to the growing evidence on whether nudges "work" and for whom, and to tease out potential theoretical mechanisms that may explain barriers to take-up, we conducted six large-scale, pre-registered randomized controlled trials in California in 2018 and 2019. These trials were carried out in collaboration with the California tax agency (the Franchise Tax Board, or FTB), state and local agencies that administer SNAP in California (CalFresh), and a large NGO dedicated to statewide EITC outreach (Golden State Opportunity). All three were interested in increasing take-up of California's state EITC (CalEITC), introduced in 2015.

The CalEITC supplements the federal credit, as do similar programs in 25 other states (Nichols and Rothstein, 2016), though unlike in other states the California credit does not simply magnify the federal schedule but is more concentrated at the lowest incomes, where it is worth as much as $3,000 per year. Figure 1 shows how the federal EITC and the CalEITC relate to family income for a single parent with two children; schedules vary for other family types but have similar shapes. For families with earnings around $7,000, the combined federal and state credits can be worth as much as $5,500, increasing family resources by close to 80%.

To claim the federal and state EITCs, families must file income tax returns, a potentially complex process. Many families who qualify for the credit are otherwise not required to file returns because their incomes fall below the thresholds that trigger filing requirements. To claim the credit for the first time, families might have to engage with a tax system with which they have never interacted. Policymakers in California continue to be concerned that, absent outreach and education, many families will not claim the credits for which they are eligible.

We used six randomized trials, several with multiple treatment arms, to test a range of outreach messages that our partners sent to potentially eligible households. Our messages aimed to inform recipients of their potential eligibility for the EITC and encourage them to file a tax return. They were delivered by postal mail or by text message. Other outreach efforts administered by our partners or other organizations, such as billboards, public events or traditional advertisements, were much lower-touch and would have affected our treatment and control groups similarly.

Subjects for our studies were drawn from participant lists for the CalFresh program and from a marketing database that included people with little or no existing interaction with government. Each study was randomized, and we linked treatment and control rosters to FTB tax records. Our outcomes are filing a tax return or claiming the EITC, each measured at the household level.

Table 1 describes the six studies; the Appendix includes copies of each of the treatments. Each arm in our studies was designed to test a set of hypotheses, drawn from the literature on administrative burdens and ordeals (discussed above) about why people may fail to take up this available benefit. Specifically, the experiments aimed to reduce learning, compliance, and/or psychological costs associated with EITC participation via scalable, low-touch nudges. All studies included a control group that received no message,[1] and all treatment arms provided information about the program and its value. If eligible households did not know about the program or did not know about the potential amounts they were likely to receive, we hypothesized that receiving this information would reduce learning costs and therefore increase take-up.

Treatment arms in Studies 3, 4 and 5 tested the impact of additional information about how to obtain help in filing a return, which targeted compliance costs. Specifically, if people knew about the EITC and understood the potential benefits, but the process was too burdensome, we hypothesized that providing individuals with detailed information about how to obtain help with tax preparation would reduce compliance hurdles and increase take-up. We pointed individuals to existing support services, as these represented the most scalable interventions. In Study 3, we directed people either to online resources, text-based assistance, or to a hotline. In Studies 4 and 5, we provided detailed information about a local Volunteer Income Tax Assistance (VITA) site that provides free, in-person tax preparation assistance to low-income households.

To test whether psychological costs associated with source credibility might affect take-up decisions, treatment arms in Studies 2 and 4 delivered the same information in different formats and via different messengers. Since the notifications informed households that they were eligible to

[1] It is possible that a member of the control group for one study was in the treatment group for another. The randomization design, discussed below, explicitly stratified on other treatment assignments where possible, to ensure a precisely zero correlation between treatment statuses in the different studies. We present analyses of each study separately. Analyses that use the overlapping samples to explore potential interactions among the different treatments show no evidence of any such interactions.

receive free money, we were concerned that households might distrust the information or assume it was a scam. We hypothesized that communication from a government agency and information presented in more formal formats would increase source credibility and reduce distrust. In Study 2, we varied the messenger to test whether receiving information from a government agency was more effective than receiving the same information from a non-profit. We made the difference in messenger salient by changing the logos, signatures, and return addresses on the letters. In both Study 2 and Study 4, we also explored source credibility by sending the same information in both formal and informal formats. The formal treatments adopted the design used by the federal government and in other EITC nudging experiments to communicate with taxpayers about the EITC (e.g., Bhargava and Manoli 2015). In the "informal" format, we used similar styles, images, and colors used in marketing materials employed by other statewide outreach efforts. In both cases, we used communication designs that could plausibly be scaled by a government agency. Studies 5 and 6 also aimed to enhance source credibility by delivering messages from the local CalFresh agency with which recipients regularly interacted, using similar wording as the agency's other outreach messages.

Despite testing a range of interventions designed to leverage many of the behavioral explanations for incomplete take-up, we found that none of our interventions had substantively or statistically meaningful effects on tax filing or EITC claiming. Our messages were received – we find relatively high engagement with websites linked to in the messages. Nevertheless, in each of our trials, we can reject effects as large as the average nudge effect in the studies reviewed by Dellavigna and Linos (2020). We conclude that even behaviorally-informed well-designed outreach aimed at increasing EITC take-up may not be enough to overcome the burdens faced by low-income households who do not file taxes.

We believe it is important for the field to grapple with null findings in the same way it grapples with negative and positive findings. Our study improves our understanding of both the promise and limitations of behavioral interventions for low-income populations. It thereby makes a major contribution to the scholarship on behavioral science and the literature on incomplete take-up of means-tested programs; it should also inform potential policy reforms related to outreach efforts and tax administration. While nudges are a potentially valuable part of the policy toolkit, outreach to hard-to-reach populations will often need to include higher-touch interventions that simplify the underlying processes.

2 Methods

This paper encompasses six distinct but partially overlapping randomized controlled trials. The studies were implemented in spring 2018 (Studies 1 and 5) and spring 2019 (Studies 2, 3, 4, and 6), and focused on tax filing for the 2017 and 2018 tax years, respectively. United States income tax returns are typically filed between February and April, and are based on income received during the previous calendar year (the "tax year").

Interventions were delivered by public agencies, the California Franchise Tax Board and the California Department of Social Services, and by a non-governmental organization (Golden State Opportunity) which receives funding from the state to conduct EITC outreach. Some features were chosen to meet agency needs rather than those of researchers.[2]

Sampling frames

A major challenge for EITC outreach is that non-claimers are unlikely to appear in tax records, which are used for most tax-related outreach. Our samples of potential non-claimers drew from two sources. Studies 1-4 used a database purchased from a private marketing firm, TargetSmart. Records were purchased for California low income households, first in spring 2018 and then updated in spring 2019. This yielded approximately 1.3 million records.

From the original sample, we removed individuals younger than 18 and those apparently living in group residences (identified by more than four records at the same address). Our eventual sample consisted of 1.2 million individuals in 1 million households.

Many of the individuals in studies 1-3 would have filed taxes even without being nudged. For Study 4, we focused on those households that had not filed taxes in the past three years. To do so, in early 2019 we merged the TargetSmart data to FTB tax filing records for tax years 2015-2017. This was a fuzzy merge, based on names, addresses, and dates of birth, with allowance for apparently legitimate differences between records in the two databases (e.g., misspelled names, alternative ways of recording addresses, potential local moves). The universe for Study 4 was limited to a subset of approximately 200,000 TargetSmart records that did not appear in the FTB filing records in any of the three preceding tax years.

The second original source of potential non-filers, used in Studies 5 and 6, was administrative records of participants in the CalFresh program, the California instantiation of the federal Supplemental Nutrition Assistance Program (SNAP). We began with approximately 6 million individuals enrolled in CalFresh at any point in calendar year 2017 (Study 5) or 2018 (Study 6), grouped into case (household) units. We excluded cases containing only seniors. We then linked adults to their quarterly earnings records for 2017:Q1 through 2017:Q3 for Study 5 and 2018:Q1 through 2018:Q3 for Study 6 from the California Employment Development Department, which administers the state's Unemployment Insurance program.

We used CalFresh case compositions and earnings records to simulate federal and CalEITC eligibility. Our simulations assumed that the CalFresh case corresponded to the potential tax household; that all children in the CalFresh case would qualify as children for purposes of EITC

[2] Our analysis of the data from these experiments was overseen by the California State Committee for the Protection of Human Subjects (protocols 2019-021, 2019-002, 2018-037, 2018-194).

eligibility; that earnings in 2017:Q4 and 2018:Q4, which were not yet available when we administered treatments, would equal one-third of total earnings over the previous three quarters; and that there would be zero self employment earnings or other income not covered by the earnings records. Based on this simulation, we selected cases with EITC eligibility above $50. This excluded those with zero or very low earnings and those with earnings too high to qualify for the EITC.

Last, we restricted to participating counties: Sacramento and San Diego in Study 5, and those counties plus San Francisco, San Mateo, and Santa Clara in Study 6. Welfare offices in these counties use text messages to communicate with CalFresh participants about their cases. We limited our sample to individuals with valid phone numbers who had consented to receiving these text messages.

The two data sources each have advantages and limitations. The TargetSmart sample provides a broad cross-section of low-income Californians, including those who interact with government infrequently. However, these data have limited information about earned income or family structure and may contain outdated or incorrect records. The CalFresh enrollee contact information is updated regularly, with detailed and reliable information about income and family composition. However, the CalFresh database only contains households already connected with state social assistance programs, meaning they have exhibited an awareness of and are willing to enroll in these forms of assistance.

Randomization

All six studies were implemented using stratified random assignment. Each included a control group that received no treatment and one or more treatment arms that received text messages or letters. Where there were multiple treatment arms, all were assigned with equal probability, though in several cases the control group was larger than any single treatment arm. Randomization was at the household level, defined by the address in the TargetSmart data and by the case number in the CalFresh data. A single representative was selected from each household to receive the treatment.

Appendix Table 2 provides details about sample sizes and randomization strata. Studies 1-4 were implemented sequentially, with assignment in one used as a stratification variable for the next, as discussed below. This ensured that treatments in each study were perfectly orthogonal to those in the other prior studies. Studies 1 and 2 used varying assignment rates across strata. In Study 1, assignment rates varied across counties and zip codes to meet GSO's needs. In Study 2, rates were set to yield 5,000 observations per treatment arm in Riverside county and 5,000 in the other counties combined. Other studies used constant assignment probabilities across strata.

Procedures

Treatments consisted of sending a single letter or a single text message to an individual chosen from each household. There was no other interaction with subjects. Control group members did not receive the letters or texts, though they may have been exposed to other outreach. Examples of each treatment can be found in the Appendix.

Study 1. Text messages were sent manually by GSO volunteers in March and April 2018, with observations sequenced randomly. Texts informed recipients of their potential eligibility and of the need to file a return in order to claim the credit, and included a link with more information to reduce learning costs. The exact wording of the texts varied over the course of the study.

Study 2. There were four treatment arms, delivered as different letters, that addressed learning costs and psychological costs related to potential mistrust of the messenger or message. Letters varied in two dimensions that both addressed source credibility: the source (GSO or FTB) and the formality. Each sender's letters used the relevant logos, signatures and return addresses. In addition, half were structured as formal letters and half as informal flyers. The front of each letter was printed in English; the back contained the same information in Spanish. Letters were mailed in February 2019.

Study 3. There were four treatment arms, each consisting of a single text message. To target learning costs, each message informed recipients about potential eligibility and the need to file taxes to claim the credit. Treatment arms 2 and 3 also targeted compliance costs by offering assistance through a hotline or via text respectively. Treatment arm 4 included additional information on the average benefit amount to further address learning costs. Texts were sent manually between February and April 2019.

Study 4. There were eight treatment arms, delivered as different letters. Letters came from FTB and contained one of four different messages: a simple message about the credit; a simple message that also included information about the average value of the credit (addressing learning costs); a message that added information about the location, hours, and contact information of the nearest in-person free tax preparation assistance site (addressing compliance costs); and a message that included both the average value of the credit and tax assistance information. Each message was delivered in a formal and an informal version, with the idea that formal letters might signal more source credibility (addressing psychological costs). The front of each letter was printed in English; the back contained the same information in Spanish. In addition, each letter contained a URL at which recipients could find the letter translated into Korean, Vietnamese, Chinese, or Russian. Letters were mailed in February and March 2019.[3]

Study 5. All treated individuals received the same sequence of text messages, designed to ad-

[3] Seven of the eight treatment arms were mailed on February 15, 2019. Due to a mailroom error, letters for one arm (a formal letter with benefit amount and in-person free tax preparation site information) were not sent until March 25, 2019.

dress both learning and compliance barriers. The first message included a personalized benefit amount estimated using the recipients' household composition and quarterly earnings data (see methods for more information). If the recipient texted "1" for more information, they were provided the URL to an online free tax-preparation software. If they responded "1" again, they were provided the address and hours of the closest VITA site to the client's 9 digit zip code. When that site required appointments, the text also included a link for registration. Texts were sent in English or Spanish, depending on the language indicated in the CalFresh record, and were delivered over two days in March 2018.

Study 6. There were three treatment arms, each delivered by text message. The first treatment arm was a simple text, informing recipients of their potential eligibility, and provided a URL to calculate their credit and a hotline to learn where to file for free. The second treatment arm provided the same information as the first text, along with the average benefit amount. The third treatment arm, as in Study 5, included a personalized credit amount. The three treatments did progressively more to address learning costs. Moreover, the fact that they came from the local CalFresh program should have increased source credibility and reduced psychological costs. Texts were delivered in March 2019 in the language indicated in the recipient's CalFresh record: English, Spanish, Chinese, or Vietnamese. Speakers of other languages received the English message.

Data and outcomes

This study is made possible by unprecedented access to an array of administrative data from California, including income tax, wage, and social service records. Our use of social service and wage records is discussed above. We measure impacts of our interventions from actual California income tax filings.

For each study, we attempted to match each member of each household to FTB records, using the same fuzzy match discussed above. We measured whether each individual successfully matched to a tax return and whether that return included a claim for either the federal EITC or CalEITC. Our primary outcome measures are an indicator for the presence of a matched return and an indicator for the presence of any EITC claim.

We analyze the data at the household level. We consider a household to have filed a return and to have taken up the EITC, respectively, if any member appears on a return and if any member's return includes an EITC claim.

We also track website visits in Studies 2, 4, and 6. Distinct URLs were used for each study and treatment arm. In Studies 2 and 4, we count all hits to these URLs; in Study 6, we count unique users. These measures are not available for the control groups.

Appendix Table 1 presents summary statistics for our main samples. The TargetSmart sample is relatively old, with a mean age of 60, though nearly half are indicated as having children in the house. 55% are white, 58% are female, and 41% filed taxes in the previous year. In the CalFresh data, individuals are younger (mean age 37) and less likely to be white (23%). A larger share have children, and 74% filed taxes in the prior year. In our CalFresh sample, we also have access to earnings records covering three quarters of the tax year, which we use to simulate EITC eligibility. We include in the study only families eligible for a federal or state EITC of at least $50. In our sample, the average family's estimated annual income is just over $14,000, and the average total (federal and state) EITC eligibility is $2,715. Across all observable characteristics to which we have access, our trial arms are balanced.

Statistical analysis

Our primary analyses examine the effect of any treatment versus control. We construct a single observation for each simulated tax filing unit (household), and estimate regressions of the form:

$$Y_{is} = \alpha + T_{is}\beta + \mu_s + \varepsilon_{is} \tag{1}$$

Here, Y_{is} is the outcome for household i in stratum s – either the presence of a tax return for some member of the household or the presence of an EITC claim. T_{is} is an indicator for treatment, and μ_s is a vector of stratum fixed effects. The impact of treatment is β. These are reported as the first estimates for each study (in black) in Figure 3.

A second set of analyses examine the effect of each treatment arm separately, where relevant. These are similar, but replace the single treatment effect with a series of separate effects:

$$Y_{is} = \alpha + \sum_j T_{isj}\beta_j + \mu_s + \varepsilon_{is} \tag{2}$$

Here, T_{isj} an indicator for assignment to treatment arm j, and β_j is the impact of that treatment relative to control. These are reported as the second and subsequent estimates for each study in Figure 3. P-values correspond to the hypotheses that each of the β_js, considered individually, equal zero. We have also tested the joint hypotheses that all of the β_js in a particular study equal zero. Across the four studies for which this is relevant and for both outcomes, we never reject the null hypothesis.

We also use a version of this model to test for baseline balance. For each baseline covariate, we estimate Equation 2 separately for each study and report β_js and the p-values for the joint hypotheses that all β_js equal zero in Appendix Tables 4-9. In Appendix Table 1, we report a single p-value that aggregates across all studies (1-4 in Panel A and 5-6 in Panel B). This is based on a sample that stacks all observations from the relevant group of studies and includes study-by-stratum and study-by-treatment-arm effects. The p-value is based on an F-test of the joint hypothesis that all study-by-treatment-arm effects equal zero.

Last, in Studies 2 and 4, treatment arms were identified by the presence or absence of particular features. We estimate a separate set of models that examines the impact of each feature.

In Study 2, these take the form:

$$Y_{is} = \alpha + Formal_{is}\gamma_F + FTB_{is}\gamma_M + \mu_s + \varepsilon_{is} \tag{3}$$

where $Formal_{is}$ and FTB_{is} are indicators for whether the letter was more formal (vs. informal) and came from the FTB (vs. GSO).

In Study 4, these take the form:

$$Y_{is} = \alpha + Formal_{is}\gamma_F + Amount_{is}\gamma_A + VITA_{is}\gamma_V + \mu_s + \varepsilon_{is} \tag{4}$$

where $Formal_{is}$, $Amount_{is}$, and $VITA_{is}$ are indicators for the presence of formal, credit amount, and VITA information features on the letters. Estimates for Equations 3 and 4 are reported in Table 2.

3 Results

Figure 2 presents the main findings for two types of engagement indicators. In Studies 2 and 4, paper letters included URLs unique to each treatment arm, which allowed us to measure total website visits by each arm. In Study 6, we measure unique click-throughs to URLs embedded in the text messages. Our engagement measures capture only those who click on (or type in) the links included in our messages, and likely underestimate the number of recipients who received and read the letters. Engagement is high compared to other estimates of successful online engagement: The average click through rate for our text messages was 10%. Even letters, which required users to type URLs by hand, generated click through rates of around 1%.[4]

Engagement patterns are in line with our main hypotheses: Engagement was higher when the messenger was the government (and therefore perhaps better known); when the message provided useful, personalized, new information (i.e. the location of a VITA site or a personalized credit amount), and when more formal presentation was used to increase source credibility. In the Study 6 text messages, more information about the value of the credit increased engagement (messages including a personalized credit amount exhibited the highest engagement), but letters in Study 4 that listed the average credit amount did not elicit more pageviews.

Figure 3 presents our main findings regarding effects on tax filing and EITC claiming. We present the estimated effect of any treatment relative to control for each study, then for each sep-

[4] Irvine (2020) reports that click-through-rates (CTRs) for Facebook ads, the behavioral measure most similar to clicking on an unsolicited text message, are around 1%, while a study by Wozney et al. (2019) obtains a CTR around 0.1%.

arate treatment arm. Across all trials and each treatment arm, our interventions did not have significant effects on either outcome. Point estimates are uniformly close to zero. Because our sample sizes are large, our estimates are highly precise. In Studies 1, 3, and 4, we can rule out effects of 0.5 percentage point or larger; in the smaller studies, we can rule out effects larger than 1 or 1.5 percentage point.

Studies 2 and 4 were cross-classified to enable us to examine the effects of different features in isolation. We present estimated effects of letter features in Table 2: a formal letter format vs. a flyer; a credible messenger (the FTB) vs. an unknown NGO; the inclusion vs. omission of the average value of the credit; and information about the closest VITA site vs. none. As explained above, these features were designed to test particular hypothesized costs. The format and messenger features were meant to reduce psychological costs by increasing credibility. Specifying the credit's value should have reduced learning costs. Finally, VITA information should have reduced compliance costs. We find no evidence that any feature generated non-zero effects.

4 Discussion

As has been reinforced by the substantial challenges governments have faced in delivering economic relief during the COVID-19 crisis, the difficulty of accessing public benefits can be a major limitation to the effectiveness of government policy. Yet, it remains unclear whether the low take-up rates for many public programs reflect design choices, lack of awareness, or other factors. In the absence of good understanding of the determinants of take-up, many efforts to increase take-up begin and end with low-cost informational interventions, sometimes called "nudges".

Early successes demonstrated that these interventions can yield significant effects on enrollment decisions (Thaler and Sunstein, 2009; Hummel and Maedche, 2019; Allcott, 2016; Gerber and Rogers, 2009; Beshears et al., 2015). These successes led to the creation of over 200 "nudge units" working in and with governments across the world (OECD, 2017). However, our understanding of what types of nudges work, in what settings, and for whom remains underdeveloped. Most public nudges focus on people who are already interacting with government programs. This makes intuitive sense: It is both practically and theoretically easier to conduct effective outreach to individuals who already have relationships with the government agency providing the nudge. Understanding how to reach people who have not had previous interactions with government is crucial to improving equity in government service delivery and helping the most vulnerable populations escape poverty. This study aimed to do just that. Based on extant theories about behavioral reasons for non-take-up of benefit programs, our messages should have raised participation.

Similar to many behavioral studies, our experiments show substantial engagement, measured by click-throughs or visits to a website. These show expected patterns, indicating that messages were received and that many recipients engaged with the material they received. Yet all of our messages had null effects on the intended behavioral outcomes, filing taxes and claiming the EITC. Our sample sizes are large enough and the effect sizes small and consistent enough that we are confident that our results can be interpreted as precise zeros.

There are several potential explanations for the failure of our outreach efforts.

First, it is possible that the recipients of our messages were already exposed to the relevant information, and that outreach would have been effective in previous years or in other contexts where there are fewer additional outreach efforts. As with any randomized controlled trial, we cannot make strong claims about external validity. Second, the value of the EITC may not be large enough to warrant filing a return and claiming it. Non-filers in our sample are eligible for smaller credits than are filers, on average (see Appendix Figures 3 and 4). Twenty-five percent of non-filers are eligible for credits of $300 or less, as compared with 9% of filers. For most non-filers, however, potential credits are large relative to plausible financial costs. Three-quarters of non-filers are eligible for credits in excess of $300 – enough to outweigh the potential financial costs of paying a tax preparer to file returns (U.S. Government Accountability Office 2014, National Society of Accountants n.d.).

Third, our messages – one-time communications that inform low-income households that they are eligible for large sums of cash – may simply feel too good to be true. We tested various message framings that aimed to get past this resistance, varying formatting of the messages and their sources to increase credibility. Nevertheless, it is possible that none of these broke through, but that alternative messages or messengers would be more successful.

Fourth, some eligible filers may face additional direct or psychological costs to interacting with the federal tax agency that may not be addressed simply through information. For example, a family concerned that immigration authorities would use tax returns or claiming behavior to target enforcement efforts or deny citizenship applications might be willing to leave free money on the table. Our interventions were carried out in 2018 and 2019, when such mistrust may have been heightened.

Finally, our treatments do not directly reduce compliance or psychological costs. Instead, we test the effect of messages that use information to reduce perceived learning, compliance and psychological costs. Our evidence suggests that nudges alone may be insufficient, but does not indicate that such costs are absent. The real compliance costs faced by our target population may simply be too high for our messages to overcome. While some trial arms pointed recipients to phone and online support as well as free tax filing assistance, actually receiving that help would have involved seeking out that assistance, going to a VITA site at specified times, compiling all of the required documents, and potentially making an advance appointment. Expanding the availability of this type of free tax support by extending the hours and availability of VITA sites, by increasing access to and quality of free online resources, or by bundling tax services with other services for low-income families may be crucial in helping people overcome real compliance hurdles. Similarly, simplifying the tax filing process itself by using existing administrative data to pre-fill tax returns could reduce real compliance costs.

Nudges can provide a low-cost way to achieve many public purposes. They are particularly well suited to bringing marginal people over the threshold into participation, but are less likely to be successful for more inframarginal potential participants. For these populations, higher-touch interventions, particularly design choices that make programs accessible from the beginning, are likely to be required.

References

Allcott, H. (2016). Paternalism and energy efficiency: an overview. *Annual Review of Economics*, 8:145–176.

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., and Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8):1041–1055.

Bergman, P., Denning, J. T., and Manoli, D. (2019). Is information enough? the effect of information about education tax benefits on student outcomes. *Journal of Policy Analysis and Management*, 38(3):706–731.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Milkman, K. L. (2015). The effect of providing peer information on retirement savings decisions. *The Journal of Finance*, 70(3):1161–1201.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment. *The Quarterly Journal of Economics*, 127(3):1205–1242.

Bhargava, S. and Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, 105(11):3489–3529.

Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lamberton, C., and Rosinger, K. O. (2019). Nudging at scale: Experimental evidence from fafsa completion campaigns. Technical report, National Bureau of Economic Research.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9):637–644.

Castleman, B. L., Patterson, R., and Skimmyhorn, W. (2019). Benefits left on the table: Evidence from the servicemembers' civil relief act. *Economics of Education Review*.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review*, 103(7):2683–2721.

Cranor, T., Kotb, S., and Goldin, J. (2019). Does informing employees about tax benefits increase take-up?: Evidence from EITC notification laws. *National Tax Journal*, 72(2):1–8.

Currie, J. (2006). The take-up of social benefits. In Alan Auerbach, D. C. and Quigley, J., editors, *Poverty, the Distribution of Income, and Public Policy*, pages 80–148. Russell Sage.

Dellavigna, S. and Linos, E. (2020). RCTs to scale: Comprehensive evidence from two nudge units.

Finkelstein, A. and Notowidigdo, M. J. (2019). Take-up and targeting: Experimental evidence from SNAP. *The Quarterly Journal of Economics*, 134(3):1505–1556.

Gerber, A. S. and Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191.

Goldin, J. and Liscow, Z. (2018). Tax benefit complexity and take-up: Lessons from the earned income tax credit. *Tax Law Review*, 72.

Guyton, J., Langetieg, P., Manoli, D., Payne, M., Schafer, B., and Sebastiani, M. (2017). Reminders and recidivism: using administrative data to characterize nonfilers and conduct EITC outreach. *American Economic Review: Papers & Proceedings*, 107(5):471–75.

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31.

Halpern-Meekin, S., Edin, K., Tach, L., and Sykes, J. (2015). *It's not like I'm poor: How working families make ends meet in a post-welfare world*. Univ of California Press.

Herd, P. and Moynihan, D. P. (2019). *Administrative burden: Policymaking by other means*. Russell Sage Foundation.

Hovland, C. I. and Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4):635–650.

Hoynes, H. and Rothstein, J. (2016). Tax policy toward low-income families. *NBER Working Paper*.

Hummel, D. and Maedche, A. (2019). How effective is nudging? a quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80:47–58.

IRS (n.d.). EITC participation rates by state. https://www.eitc.irs.gov/eitc-central/participation-rate/eitc-participation-rate-by-states. Accessed: 2020-03-20.

Irvine, M. (2020). Facebook ad benchmarks for your industry. https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks. Accessed: 2020-03-20.

Jones, M. R. (2014). Changes in EITC eligibility and participation, 2005-2009. *Center for Administrative Records Research and Applications Working Paper*.

Kopczuk, W. and Pop-Eleches, C. (2007). Electronic filing, tax preparers and participation in the earned income tax credit. *Journal of Public Economics*, 91(7-8):1351–1367.

Madrian, B. C. and Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, 116(4):1149–1187.

Moffitt, R. et al. (1983). An economic model of welfare stigma. *American Economic Review*, 73(5):1023–1035.

Nichols, A. and Rothstein, J. (2016). The earned income tax credit. In *Economics of Means-Tested Transfer Programs in the United States, Volume 1*, pages 137–218. University of Chicago Press.

OECD (2017). Behavioural insights and public policy: Lessons from around the world. *OECD*.

Plueger, D. (2009). The EITC participation rate for tax year 2005. *Internal Revenue Service Bulletin*.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2):243–281.

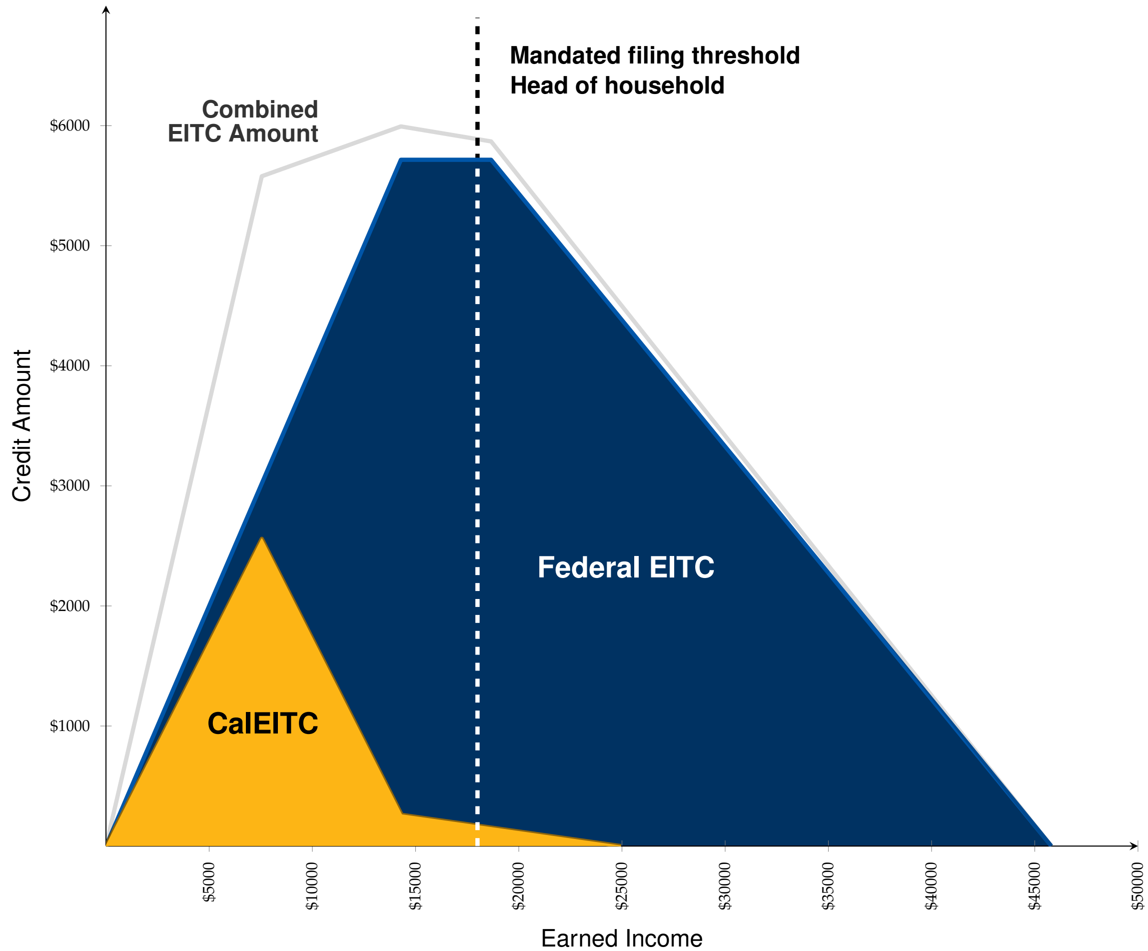
Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1):S164–S187.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Wozney, L., Turner, K., Rose-Davis, B., and McGrath, P. J. (2019). Facebook ads to the rescue? recruiting a hard to reach population into an internet-based behavioral health intervention trial. *Internet interventions*, 17:100246.

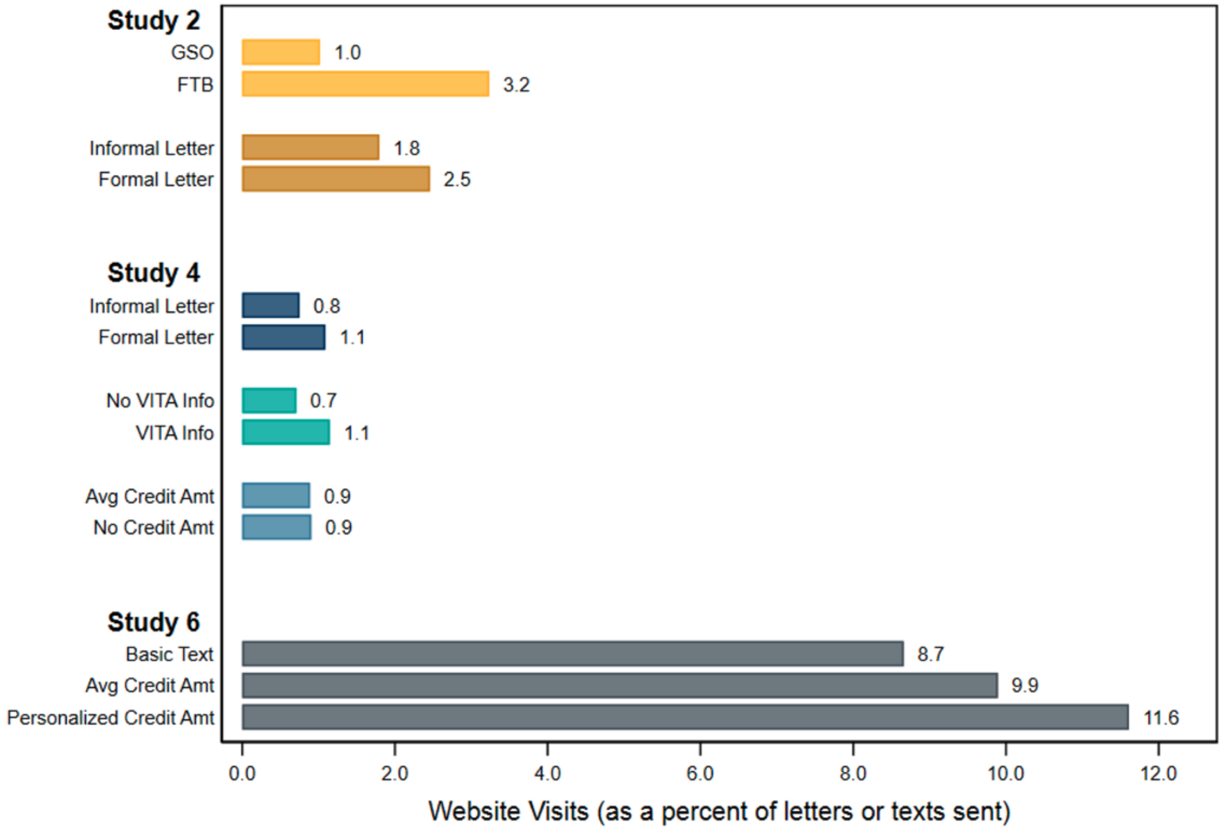
Figures

Figure 1: Federal and California EITC schedules for a single-parent family with two children, tax year 2018



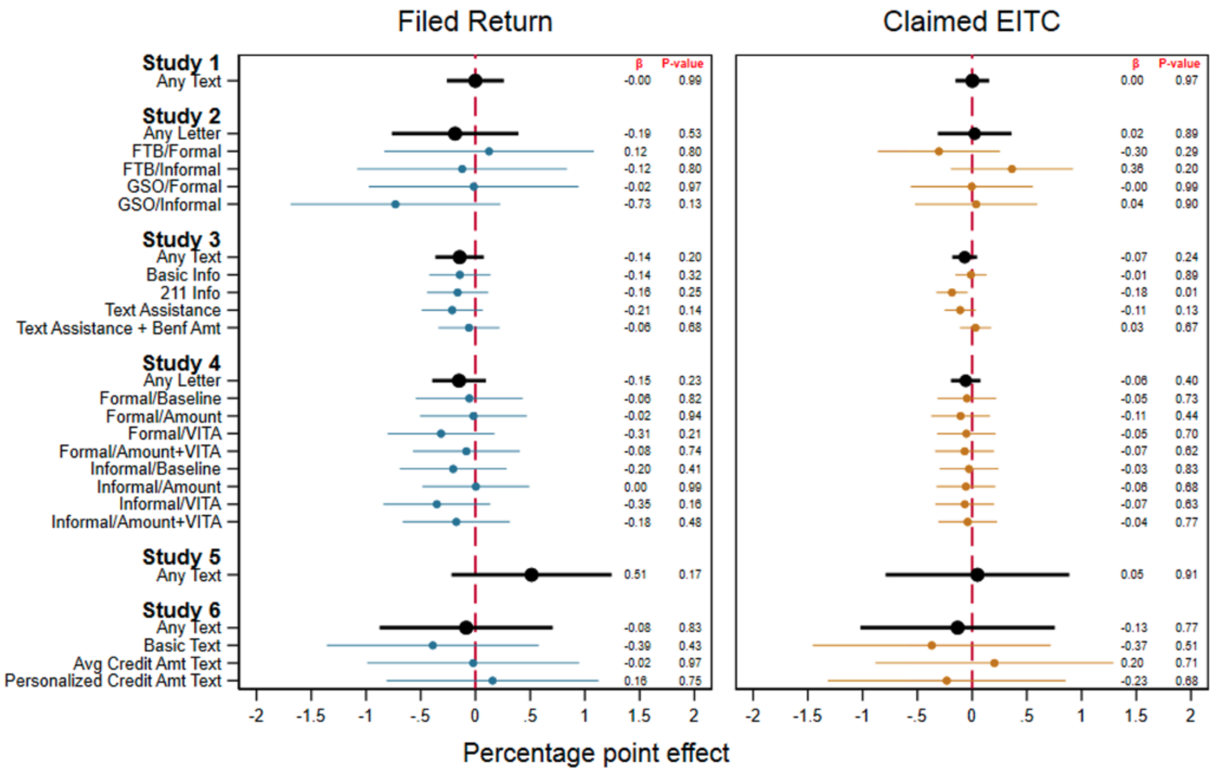
Note: This diagram illustrates the federal (blue) and state (gold) EITC schedules for a head of household with two children. The gray line illustrates the combined value of the EITC for a filer who claims both credits. The dotted line denotes the filing threshold for a head of household in tax year 2018, which was $18,000; families with incomes below this threshold are generally not required to file returns.

Figure 2: Web traffic by pooled treatment features, tax year 2018



Notes: Figure shows measures of engagement as a share of the number of messages sent, by study and treatment feature. In Studies 2 and 4, the features are shared across several treatment arms, and traffic estimates aggregate over all relevant arms. In Studies 3 and 4, engagement measure is visits to a website, hand-entered from treatment-arm-specific URLs included in letters, divided by the number of letters sent. In Study 6, engagement measure is the number of unique visitors to a URL included in the text messages, divided by the number of texts sent.

Figure 3: Effects of outreach treatments on tax filing and EITC claiming, by study



Notes: Figure shows estimated treatment effects on tax filing (left) and EITC claiming (right), with 95% confidence intervals. P-values reflect two-sided significant tests. Larger black dots and lines show effects of any treatment vs. control, while smaller, colored dots and lines show effects of each treatment arm individually.

Tables

Table 1: Study descriptions

Study	Treatment arms	Treatment mode	Costs addressed
Study 1 N=639,244	1	Text messages sent by NGO	Learning •Simple message
Study 2 N=96,370	4	Letters sent by state government and NGO	Learning •Simple message vs. average benefit amount Psychological •Government vs. NGO messenger •Formal vs. informal
Study 3 N=1,084,018	4	Text messages sent by NGO	Learning •Simple message vs. average benefit amount Compliance •Web vs. text vs. phone-based assistance
Study 4 N=204,285	8	Letters sent by state government	Learning •Simple message vs. average benefit amount Compliance •Local in-person free tax preparation information Psychological •Formal vs. informal formatting
Study 5 N=38,093	1	Text messages sent by county welfare office	Learning •Personalized benefit amount Compliance •Free tax preparation website •Address of local in-person free tax preparation assistance
Study 6 N=47,104	3	Text messages sent by county welfare office	Learning •Average vs. personalized benefit amount

Table 2: Estimates of pooled treatment features, identified from across-treatment arm variation

	Filed Return		Claimed EITC	
	Study 2	Study 4	Study 2	Study 4
Baseline	0.377	0.089	0.076	0.024
	(0.002)	(0.001)	(0.001)	(0.000)
Formal letter	0.002	-0.000	-0.004	-0.000
	(0.004)	(0.001)	(0.005)	(0.001)
Messenger: FTB	0.000		0.001	
	(0.004)		(0.002)	
Benefit amount		0.001		-0.000
		(0.001)		(0.001)
VITA referral		-0.002		-0.000
		(0.001)		(0.001)
N	96,370	204,285	96,370	204,285
p-value, $\gamma = 0$	0.89	0.45	0.53	0.88

Notes: These estimates correspond to γ coefficients in Equations 3 and 4.