High Traffic On-the-Spot Vote Tripling LargeCT

(Nov-Dec 2022)

Pre-registered Protocol

Trial Info

Team: Research
Partners: Kate Duch (external evaluator)
Vendors: GRSG (canvassing firm)
Funding: Paid for by the Vote Tripling Action Fund, with c3 funds
Partisanship: Non-partisan voter turnout work
Target election: 2022 Georgia Runoff Election

PICOS Statement

Trial Launch Date:	Data collection begins November 27, 2022 and ends December 6, 2022. (subject to change due to weather events, canvasser training needs, and early voting policy changes)
What is the problem?	Eligible voters frequently do not turn out for elections. We believe this is in part because people they know don't talk to them about voting.
Population:	**Triplers**[1] are adult passersby in high-traffic places who agree to immediately remind up to 5 people they know to vote. Triplers do not

[1] The terms "tripler" and "triplee" were coined in an earlier version of this intervention when we requested the names of 3 friends. We made the change to 5 friends very recently, and most of Vote Rev Action Fund's mainstreamed programs still use 3. For consistency, this document will continue to use the legacy terms "tripler" and "triplee".

Who are the participants?	need to be registered voters and do not need to be identifiable in the voter file. **Triplees** are the friends[2], family members, or other contacts whom triplers might remind to vote. Triplers select their own triplees. In this study, only registered voters are eligible to be named as triplees, and the tripler will only send a given triplee a reminder if they were pre-randomized into the treatment condition.
Sample details	See section 07 / Sample & Setting for details on sample size.
Intervention: What are we doing?	We will recruit triplers in high traffic locations, and triplers will encourage randomly selected triplees in the moment (“on the spot”) to vote in the upcoming runoff election.
Intervention details	Triplers will be asked to only remind a randomized subset of their triplees. The subset they are asked to remind will make up the treatment group, and those they are not asked to remind make up the control group. Triplees are not contacted by study staff, only by triplers.
Comparison:	The turnout rate of control triplees (whose triplers are assigned to receive no reminders to encourage them to vote) compared to that of treatment triplees
Comparison details	The comparison between groups will have reduced validity if triplers text friends who are in the control group. Because triplers are asked to remind their friends on the spot, canvassers can usually notice this when it happens. However, it is possible that the experience of sending a reminder might inspire triplers to send their control triplees a reminder later that day. This could potentially lead to an underestimate of the true effect of HTOTS, but a significant difference between the groups provides

[2]For simplicity, in the rest of this document we refer to triplees simply as "friends", but in all cases this also includes family members, co-workers, classmates, or anyone else the tripler chooses. Similarly, "housemates" includes all residents at the same address such as spouses and adult children.

	definitive evidence that HTOTS increases voting rates above the counterfactual baseline.
Outcome(s):	Primary outcome: Voter turnout rate for triplees in the Georgia runoff election
Setting:	The trial will take place in Georgia (primarily in Augusta, Athens, College Park in Atlanta, Decatur, Marietta, Kenneaw and Savannah) in the days leading up to the Georgia runoff election as well as Election Day (locations subject to change based on voting dates, logistics, and recruitment success)

Contents

Background and Intervention

01 / Purpose

The purpose of this study is to test the effectiveness of High-Traffic On-the-spot vote tripling (HTOTS) as a tactic for increasing turnout in elections. Below we provide an overview of how HTOTS works in "real life", outside of an evaluation context. In order to be able to evaluate HTOTS, we have to adapt the tactic; see more detail under Intervention Design.

What is High-Traffic On-the-spot

High-traffic on-the-spot (HTOTS) is a vote tripling tactic in which canvassers station at locations with high foot traffic and ask people to send a message then and there (on the spot!) to their friends to remind them about an upcoming election. Vote Rev has experimented with variations of different elements of this tactic:

- The number of friends that people are asked to remind
- Incentives given to people who participate
- What canvassers carry and wear to grab people's attention
- Canvasser script
- Canvassing location

The trial will run in Georgia during the period preceding the Georgia runoff election as well as on Election Day.

02 / Approach & Challenge

One challenge of measuring the impact of vote tripling tactics is that the impact we want to measure (triplee turnout) is one step removed from the actual point of contact (pledge collection with and reminder text messages to the tripler). In typical practice, HTOTS canvassers would not collect any information about the triplees.

In this trial, canvassers will ask triplers to help them look up the triplees in the voter file, based on name, age, and address. We will record triplees' voter IDs so we can later look up their turnout.

In our recent study of vote tripling pledge collection, the only way to construct a comparison group was to have all participants pledge to encourage their friends to vote. The pledge was collected weeks or months before election day, and just before election day triplers were randomized to either receive a reminder about their pledge or not. This relied on the assumption that triplers who were not reminded were much less likely to encourage their friends to vote, thus making them a valid comparison group. In contrast,

in the current study of HTOTS,triplers are told to encourage their friends to vote immediately and in view of the canvasser. This allows us to know with certainty whether the tripler followed through on the request, and whether the tripler followed the request not to text control triplees (at least, not at that time).

An additional advantage of this tactic is that at least some people who pledge in advance to remind friends will not do so, diluting the treatment effect. When triplers are asked to send messages on the spot, the canvasser can verify that they have done so, making our measured effect closer to the full treatment-on-treated effect.

03 / Intervention Design

Paid canvassers will station themselves at high traffic locations such as college campuses, shopping centers, and DMVs. They will operate in Augusta, Athens, College Park in Atlanta, Decatur, Marietta, Kenneaw and Savannah, in the days leading up to the 2022 Georgia runoff election and on the day of the election as well (see Timeline). Canvassers will be recruited and managed by our canvassing vendors in each state.

In a typical HTOTS implementation, canvassers would simply ask passerby to encourage their friends to vote, and urge them to do so right then and there. In this trial, people who agree to encourage their friends to vote will then be asked to match their friends to the voter file. The canvasser will collect data on their phone and will ask the tripler to help identify their triplees using a voter file lookup app (Grassroots Unwired; GRU). Based on random assignment, the app will then give the canvasser instructions to either ask the tripler to follow through with texting that person, or to ask the tripler to move on and select another person[3]. This process will repeat until the tripler has named 5 people total, regardless of the number of them who are assigned to be texted, or until the tripler decides to stop.

Only triplees who can be found in the state voter file will be eligible to be selected for the study. Pilot tests indicate that canvassers are able to achieve a match rate of >70%, meaning that most individuals a tripler wants to encourage will be matched. If a potential triplee can't be matched, the canvasser will ask the tripler to name a different person until they do find a match.

CONDITION	DESCRIPTION
Overall sample	Individuals in the GA voter file who the tripler selects as potential message recipients

[3] In pilot testing we have found that triplers are usually willing to go along with this process, either assuming that there's some reason behind which people they're asked to message or because they believe our control group list is some kind of "do not text" list.

Control	Triplees who were randomly assigned NOT to receive a message from the tripler, after the tripler selected them
Treatment	Triplees who were randomly assigned to receive a message from the tripler, after the tripler selected them

Evaluation Design

04 / Study Design

This will be a randomized controlled trial. Note that although triplees can be grouped based on their tripler, random assignment is not carried out at a tripler level; most triplers will have a mix of treatment and control triplees.

We chose to randomize at the household level, in order to reduce complications associated with treatment spillover. The entire set of households in the state of Georgia will be randomly assigned to treatment or control (see 06 / Assignment for details of randomization), and this assignment will affect all of the household members.

An example of the spillover scenarios we wish to avoid: Ernie and Bert live together. Ernie is randomly assigned to treatment and Bert to control. Oscar is recruited as a tripler and names both Ernie and Bert. When Ernie receives a message from Oscar, he might also remind or otherwise encourage Bert to vote, meaning that Bert's outcomes no longer reflect a true untreated counterfactual. Assigning all household members to the same condition prevents this.

We assume that the effect of receiving *multiple* treatments is small compared to receiving just one, and treat these individuals as regular treatment group participants. These include: 1) people who receive a message and live with someone who receives a message, and 2) people who are friends with multiple triplers and receive messages from more than one of them.

Food incentives

Throughout the data collection period, we will offer free food (such as tacos, donuts, or ice cream) across locations to people who agree to encourage friends to vote. The canvassing firm will determine when in the interaction between a passer-by and a

canvasser it is best to offer the free food. This is a feature of the overall HTOTS tactic but is not related to treatment assignment.

Canvassing sites

Canvassing sites will be chosen by the canvassing firms based on local area knowledge and in collaboration with VRAF's field team. In keeping with our mission to amplify the power of historically disenfranchised communities, we will preferentially target areas that have high proportions of young people and people of color.
In our analysis we will include a covariate term for triplees whose triplers were recruited on a college campus vs any other location type. Other location types will not be differentiated.

Potential issues

- Canvassers are unable to recruit enough triplers
    - The VRAF field team and research teams will be monitoring recruitment rates continuously and are experienced at working with canvassing firms to understand and fix canvasser issues. We will also test recruitment rates during a run-in period before Thanksgiving and will not proceed with the study unless we are able to get close to the required However, if we are still not able to reach our recruitment goal, we will collect as much data as possible and run the analyses below as described.
- The electoral environment in Georgia is exceedingly noisy, leading people to ignore canvassers or tripling messages.
    - This is expected, given that this race is attracting national attention and money. Because we are researching HTOTS as a tactic to use in the 2024 presidential election, we consider the difficulties involved to be realistic.
- After completing the interaction with the canvasser, triplers later text their control group triplees and remind them too, at a higher rate than the would have done otherwise.
    - We are exploring ways to check whether this occurs.

05 / Research Question

Does prompting a person to encourage friends to vote increase the probability that those friends will vote?

06 / Assignment

METHOD:	The random assignment process will be done by research team staff using the R "randomizr" package.
STRUCTURE:	The entire set of registered voters in the state of Georgia will be considered potential triplees for this intervention. Potential triplees will be randomized at the household level, stratifying as described below.
ARMS:	Potential triplees will be randomized into 2 arms: • Treatment (50%): If selected by a tripler, the canvasser will ask the tripler to message them on the spot • Control (50%): If selected by a tripler, the canvasser will ask the tripler to move on to naming the next triplee, without sending a message.
UNIT OF ASSIGNMENT:	Households. All members of a household will be assigned to the same treatment condition. See below for details. Note that only individuals who are selected by triplers will be in the actual study sample. This means that many triplees will be part of a household for assignment, but will be analyzed as single-person clusters (if no one else in their house is selected as a triplee).
UNIT OF MEASURE:	Outcomes are measured at the triplee level.

Sampling and assignment process

Householding

Households will be identified using TargetSmart's Exact Address Track label, which is a proprietary identifier that uses voter registration data and other consumer data to identify the true residence of an individual. They then attempt to combine address values that refer to the same location (based on minor variations in address formatting or city name, etc).

We will stratify households for random assignment using the following variables:

- Number of registered voters in the household: 1, 2, 3, 4, 5-6. Households with more than 6 registered voters will be assumed to be anomalous (apartments lacking apartment numbers, bad address data, institutions, etc) and each individual will be treated as a separate household.
- Whether the household is in the Atlanta metro area[4]
- Number of registered voters who voted in the 2020 runoff: 1, 2, 3, 4, 5-6
- Number of registered voters who identify as white (1, 2, 3, 4, 5-6)
- Number of registered voters who identify as Black (1, 2, 3, 4, 5-6)

Note that this process determines the condition a triplee *would* be in, but the triplee only becomes part of the study sample after a tripler names them. We will not analyze outcomes for individuals not named by a tripler, even if they have housemates who are named.

Internal validity

- Triplers and triplees will be unaware that any randomization will occur. Triplers will be aware that they may be asked to message some of the friends they name and not others based on a "list". Canvassers may inform them that the list is randomly generated for a study, but will be instructed to do this only if directly asked what the list is.

Balance check process

Balance checks will occur twice, once after the pre-randomized list is generated and once after the study is complete.

Balance checks will use the normalized differences approach of Imbens and Rubin (2015), which is calculated by taking the difference in averages by condition, scaled by the square root of the average of the variances. This is a scale-invariant measure.

For compatibility with the method proposed by our external evaluator, we will also use a second method for balance checks: Running a logistic regression, clustered at the household level, with all balance check variables as predictors and condition assignment as the outcome. We will consider the sample balanced if no variable (or variable level in the case of categorical variables) has $p<.1$.

[4] This is defined as being in one of the 28 counties of the Atlanta-Sandy Springs-Marietta, GA Metropolitan Statistical Area: Atlanta-Sandy Springs-Marietta, GA Metropolitan Statistical Area (2000 Census). County of residence is based on TargetSmart's county name field rather than the FIPS county code field. The counties are:
Barrow County, Bartow County, Butts County, Carroll County, Cherokee County, Clayton County, Cobb County, Coweta County, Dawson County, DeKalb County, Douglas County, Fayette County, Forsyth County, Fulton County, Gwinnett County, Haralson County, Heard County, Henry County, Jasper County, Lamar County, Meriwether County, Newton County, Paulding County, Pickens County, Pike County, Rockdale County, Spalding County, Walton County.

First check

Following stratified randomization, we will test for balance between the pre-randomized arms on age, gender, and race/ethnicity, using the values and categories provided in the TargetSmart voter file. Variables with multiple categorical values will be recoded to one dichotomous dummy variable per category. We will also check for balance on 2018 general elections turnout, 2020 general elections turnout, 2018 runoff turnout, TargetSmart ideology model, residence in the Atlanta metro area, and likelihood of 2022 turnout in the runoff election based on TargetSmart's proprietary turnout model.

Because we are able to repeat this randomization as needed in order to achieve balance, we will re-randomize until all variables have normalized differences <0.05 *and* a logistic regression outcome with p>.1 for all predictors. Because most canvassing will take place in the Atlanta area, we will also require that the subset of individuals listing their city as Atlanta also meet these criteria.

Final check

After all data collection has completed, we will compare the treatment and control group triplees on the same variables as in the first check, plus date and location of canvass (treating these variables the same way they are treated in the primary analysis). Note that this sample will differ from the first check; it will include only individuals who *were* named by a tripler as opposed to all individuals eligible to be named.

Following Imbens and Rubin's recommendations, we will consider our sample balanced if all normalized differences are <0.25. Gender and race/ethnicity categories that occur fewer than 50 times in the sample will be combined into a single category for this check; if there are still fewer than 50 cases they will be combined with the largest single category.

If imbalances are identified in the final check, it will be impossible to rectify them. Our analyses will include all balance check variables as covariates, but because any imbalance implies a higher probability of imbalance on unobserved variables, we will not be able to rule out the possibility of confounds.

07 / Sample & Setting

Sample characteristics:

See the Assignment process section for details on which individuals will be included in our samples.

Exclusion criteria

- Triplees will be removed if the canvassing firm or the VRAF field team determines it is likely that the data was fabricated by a canvasser. The team will use metrics such as interaction duration, location, and conversion rate to find anomalous pledges, and is also experimenting with requiring canvassers to collect signatures from triplers.
    - If there is data that we suspect, but are not highly confident, is fraudulent, we will check for balance between treatment and control triplees in the data that would remain if it were removed. If they would be unbalanced, we will retain this data.
    - If there is data that we are highly confident is fraudulent, we will remove it regardless of balance.
    - Note that we do not expect there to be balance between treatment and control triplees in the data being removed, because canvassers might tend towards fabricating data with treatment vs. control individuals.
- Individuals who are not registered to vote, or who cannot be found in the voter file by their tripler, will not be eligible for inclusion in this study.
- As early and absentee vote data comes in, we will update the voter file used by canvassers to indicate which triplees have already voted. The app will instruct them to not accept this person as a triplee and ask the tripler to name a different friend instead.
    - Due to lag between voting and being updated in our voter file, we will also check after data collection is complete, and remove any triplees from the dataset who had voted prior to when their tripler was recruited. Before doing this, we will check for balance between control and treatment in the data to be removed; if there is an imbalance we will keep the data.

Handling noncompliance

A tripler may disregard instructions and message a control group tripler or refuse to message a treatment group tripler.

If a canvasser notices this before the tripler finishes doing it, they will reiterate the instructions to text or not text the person. If the tripler insists on doing so, or if the canvasser only notices afterwards, they will record the triplee's actual messaging status along with their intended treatment assignment.

We will use these recorded deviations for QA. We will not exclude these triplees from analysis because there are likely nonrandom reasons for both types of noncompliance and this would create systematic differences between the treatment and control samples.

Bias in representation:

- This study will collect a substantial number of pledges on college campuses. Collecting at colleges is consistent with our mission of amplifying the power of historically disenfranchised communities, as young people generally vote at lower

rates than other groups and colleges are more demographically diverse than surrounding areas. However, colleges select for higher-SES populations, so collecting data from demographically diverse non-college areas is also critical.
- VRAF and the canvassing firm will do mapping activities and interview local residents to identify non-college high traffic areas that will ensure demographic diversity in our non-college subsample.
- Voter turnout will only be measured for triplees who we can match to exactly one record in the voter file. For example, if a tripler wants to remind a friend named Maria Sanchez, but there are 5 people by that name, in the same age bracket, in the relevant city, and the tripler can't disambiguate them by street address, that person will not be eligible and the tripler will be asked to name a different person. Because name frequency differs between race and ethnicity groups, these exclusions may occur more for some groups of people than others. In a name matching equity analysis based on Virginia voter file data, we found that white individuals' names had the *lowest* match rate, likely because there is a larger pool of other white people to compare against (assuming statistical trends towards name similarity within groups).
    - Note also that this is only an issue for evaluation, not for the effectiveness or generalizability of HTOTS as a voter mobilization tactic. Because triplers contact their friends themselves, it works whether or not the organization canvassing can identify those friends.
- There are likely to be differences between individuals who are willing to speak to a canvasser and those who are not, so inferences about the effectiveness of pledge collection might not generalize to the entire population. However, they do accurately represent the population of people who would be reached by a real-life, non-experimental vote tripling intervention.

Sample size and Power

We calculated power for a cluster randomized trial using the R package "CRTSize". Based on previous studies we believe that in the final sample the average cluster (household) size will be ~1.2 triplees per household. However, CRTSize requires integer cluster sizes and so we specified clusters of 2. This means that the required sample size is likely an overestimate, and the true minimum sample size is closer to what would be calculated for an unclustered study.[5]

Parameters:
- Minimum detectable effect size = 1pp increase in triplee turnout
- alpha = 0.1, 1-tailed (treatment higher than control only)
- power = 0.8

[5] Example code for clustered and unclustered analysis (the final multipliers account for 2 conditions, 2 triplees per cluster, and the 10% inflation factor)
n4means(delta = .01, sigma = .4, m = 2, ICC = .4, alpha = .1, power = .8, two.tailed = F)$nC * 2 * 2 * 1.1
pwr.t.test(d = .01/.4, sig.level = .1, power = .8, alternative = "greater")$n * 2 * 1.1

- Intra-class correlation between triplees within households = 0.37 (based on our analysis of a variety of state voter files)
- Standard deviation of outcome = 0.4 (based on a 2018 turnout rate of ~80% among *registered* voters, who are the population for this study).
- Triplees per pledge=3
- Triplee household cluster size = 2
- Inflation factor: 10% (arbitrary increase to account for fraud, canvasser error, and other causes of data loss)

We find that we need 40,391 usable triplees across all arms of the study, which equates to 14,811 triplers. Should we experience recruitment problems, a sample as small as 19,747 (6,582 triplers) will be sufficient to detect a 1.5pp increase.

Analytical Strategy

08 / Outcomes

All analyses described here will follow the inclusion/exclusion criteria described in sections 06 / Assignment and 07 / Sample & Setting.

Independent oversight of statistical analysis

VRAF staff will run all the analyses in this protocol. Kate Duch of One Minus Beta Analytics will independently analyze the Primary Effectiveness Measures based on this protocol and produce her own report.

Kate Duch will also review this protocol prior to pre-registration and state whether she approves of the final version.

Primary Effectiveness Measures

Triplee voting rate

- **Measure**: Voting rate among triplees in the GA senate runoff elections. Data will be taken from state voter files available after the election.
- **Point of collection**: Data used for inclusion will be taken from the latest version of the voter file we have available. Data used for voting status and early voting dates will be taken from the file available after the election.
- **Type**: Binary (voted vs did not vote)
- **Rationale**: Voting is the ultimate target of voter turnout interventions.

Other Effectiveness Measures

Long-term voting rates

We will also measure voting rates for triplees in future elections, including the 2024 general elections, 2 years after the intervention. There will be no further intervention delivered during this time.[6]

Individuals who are no longer in the voter file in the 2024 generals (e.g., because they moved, changed names, or lost the right to vote) will be treated as non-voters.[7]

09 / Statistical Approach

Triplee voting rates

We are testing whether to reject the null hypothesis of no difference in triplee voting rates in our sample based on whether the triplee was selected to receive a text from a friend reminding them to vote or not.

Note that even though randomization occurs at the household level, individuals will only be included if they are directly named by a tripler; our analyses will not include housemates of triplees. This means that in many cases, even individuals who were randomized as part of a multi-voter household will be analyzed as a cluster of size 1.

Our primary analysis will use an OLS linear regression with clustered standard errors by household.

- Main (fixed) effects:
    - Treatment condition (main outcome of interest)
    - College campus vs not
    - Triplee race/ethnicity
        - Categories with fewer than 50 observations will be combined into a single category for analysis; if still under 50 then they will be added to the largest category used in the analysis.
    - Triplee age
        - Represented with one predictor for age and a separate one for age squared
    - Triplee gender

[6] Alternatively, we may decide to follow up by contacting some or all of these individuals. If we do that, it will be covered under a separate experimental protocol, and the evaluations described here will not be run.

[7] We discussed removing unfindable individuals from the analysis, but this may cause issues because voting history is sometimes used by states as a *reason* for removing individuals.

        - Categories with fewer than 50 observations will be handled in the same way as described for race/ethnicity
    - Number of registered voters in household
        - Treated the same way as in balance checks, except that number of registered voters will be treated as a continuous variable ranging from 1-6.
    - Vote history in the 2018 general election, 2020 general election, 2018 runoff, 2020 runoff, and (if available) 2022 general election
    - TargetSmart turnout score for the 2022 runoff election
    - TargetSmart Ideology score
    - Number of days before election day when the interaction took place
    - Canvasser who recruited the tripler
- Clustered errors:
    - Household (defined as described in Householding)

All covariates will be taken from the latest version of the voter file that uses only pre-runoff data (ie, we won't use scores calculated post-election that may incorporate the dependent variable into their calculation; however, we can use general election turnout data that isn't available until after the runoff).

We will consider the test an unequivocal rejection of the null hypothesis if the fixed effect parameter for treatment is significant at $p<0.1$ (one tailed, higher than control only).

Exploratory Analyses

Race/ethnicity interactions

We will run exploratory analyses to test for race/ethnicity interactions that modify the effectiveness of the intervention. These are considered exploratory because we don't anticipate being powered to detect these reliably.

We will test these effects by adding an interaction between treatment effect and race/ethnicity to the primary analysis. We will set white as the reference group[8].

Other interactions

We will rerun the model above with interaction terms between treatment condition and:

- campus vs. non-campus

[8] We set white as the reference group because our primary, equity-based concern is to make sure we don't broadly recommend HTOTS if it works primarily for white participants and significantly less well for any other group. If there are large differences between any other pairs of interaction variables, we will investigate this in post-hoc testing.

- triplee age and triplee age squared
- Days prior to election day when the interaction took place

These are considered exploratory because we don't anticipate being powered to detect these reliably.

Multiple comparison adjustment

To address multiple comparisons, we will use the Benjamini-Hochberg step up procedure, aiming to maintain a FDR of .1 (one tailed, higher than control only) for the entire set of exploratory analyses under the Exploratory Analyses heading.

Long-term carryover effects

We hypothesize that the treatment in this election may carry over to future elections. After the 2024 November general elections, we will assess whether there were carryover effects on triplee voting in those elections. We will re-run the models above, replacing the turnout outcome in the 2022 runoff elections for the 2024 general elections. These are not included in the multiple comparison adjustment because these outcomes will not have occurred at the time we are analyzing the other study outcomes.

As noted above, we will not run these analyses if we instead choose to recontact triplers from this study as part of either a separate study or an impact-oriented program.

Implementation

10 / Trial Procedure

Independent oversight of implementation

Kate Duch, an independent researcher, has reviewed this protocol and will meet with VRAF approximately weekly to advise them on implementation issues.

Timeline

The timeline below is subject to changes due to vendor issues, weather, etc.

Task	Done by	Date / deadline
Voter file pre-randomization	Vote Rev staff	11/25

Collect pledges	Canvassing Firms	11/27-12/6
Intervention (sending triplees encouragement to vote)	Triplers	11/27-12/6
Election Day		12/6
Analyze data/results	Vote Rev staff + Kate Duch	Likely in the first half of 2023; Depends on when state voter file information becomes available

11 / Risks and Ethical Considerations

Implementation checks and mitigation strategies:

VALIDITY CONCERN	WHO WILL CHECK?	HOW WILL THEY CHECK?
Did triplees receive the intervention as intended?	Canvassers, canvassing leads, and VRAF field team	Canvassers will monitor that triplers actually send texts to their triplees on the spot. Canvassing leads and the VRAF field team will monitor that canvassers are correctly monitoring the delivery.
Canvassers fake pledge data	Canvassing firm, VRAF field team	Will use anomaly detection metrics computed by our canvassing software (GRU), our canvassing vendors, and/or the VRAF field team.

		These metrics include implausible timestamps, locations, and conversion rates. The field team is also experimenting with having canvassers collect physical signatures from triplers to verify that they participated.

Risk of harm

Reduced turnout: There is no theoretical reason to believe that this type of intervention would reduce turnout, and reduced turnout has not been seen with similar interventions. In pilots HTOTS produced a large number of relational reminders, which usually lead to increased turnout.

Canvasser interactions: Potential triplers may suffer anxiety, stress, or regret due to their interaction with canvassers, or may feel intimidated or coerced by them. We address this risk by training canvassers to be outgoing and firm, but also to respect triplers' boundaries. Canvassing firm staff will observe canvassers in the field to check for inappropriate behavior.

Discomfort from political conversations: Triplers and triplees could both experience discomfort due to talking about politics or being encouraged to vote. In our field research, when asked, triplers did not express concerns about this. Triplers typically select triplees whom they believe hold similar political views and would not mind receiving a message about voting. Therefore, the risk to both parties is not much higher than they might experience in their ordinary, pre-existing relationship.

COVID safety: Interactions between canvassers and triplers pose a non-zero risk of transmitting COVID or other infections. We believe the risk is not high: Canvassers will follow mask mandates when relevant and most pledge collection will take place outdoors, which reduces risk by an estimated 95% relative to indoor interactions.

Canvasser safety: Canvassers face some risk of harassment, violence, or discomfort while doing their jobs. To minimize this risk, canvassers are sent to sites in pairs, and they all receive a 1-pager document with safety information as well as the numbers of people they should reach out to if anything goes wrong. We will be monitoring any reported incidents on a daily basis and following up with canvassers.

Stopping rules:

Rule	Monitoring	Who is responsible
We will stop canvassing in a given area if: 1) canvassers are experiencing hostile reactions more than very rarely, 2) canvassers receive threats or implications of violence, 3) weather or other environmental conditions make the location hazardous to canvassers, or 4) canvassers are told to move by an appropriate authority (police, security, store owner, etc).	Canvassing firm team leads	Canvassing firm

12 / Data Requirements

Data sources

Canvassing app (GRU) ticksheet and pledge data

PURPOSE:	Voter file lookup, treatment assignment.
OWNER:	Primarily collected by the canvassing app vendor (GRU)
REQUEST:	Data will be available to Vote Rev staff at any time via the app's administrative functions
ACCURACY / RELIABILITY:	Data is collected and saved automatically when canvassers take pledges. The voter file matching process will help ensure that canvassers are recording information accurately
HISTORICAL:	We have seen pulls of similar data in our Michigan and Florida pilots in the 2022 primary elections and our HTPC RCT in Michigan, North Carolina, Michigan, and Arizona in the 2022 general elections..
PII / CONSENT:	Data contains PII (first and last names, and voter IDs) for triplees. Data security and storage is set out below.

COUNTRY:	Collected: USA Stored: Google Drive

Voter file

PURPOSE:	Voter file matching, outcome measures and covariates
OWNER:	Vote Rev will obtain a copy of the voter files from a data vendor
REQUEST:	Available for purchase any time from vendor
ACCURACY / RELIABILITY:	The data originates with the state Secretary of State's office and is believed to be as accurate as possible
HISTORICAL:	We will have access to versions of the voter files, excluding the study outcomes, for the state of Georgia before the intervention as they need to be uploaded in the data collection (GRU) app.
PII / CONSENT:	This dataset contains a large amount of PII, including first and last name, age and address. This dataset is public record so we do not need consent to process it. Data will be stored on vendors with enterprise-level security processes, including Google Cloud Platform, Civis, and Google Drive.
COUNTRY:	Collected: USA Stored: Google and Civis servers

Variable construction

VARIABLE	TYPE	COLLECTION POINT	SOURCE	MEASUREMENT

Triplee matched to voter file	Inclusion	After manual matching is completed	Voter file	Binary - 1/0, triplee level
Voted in election	Dependent	After the relevant election	Voter file	Binary - 1/0, tripler and triplee level
Treatment arm (control / reminder / reminder with letter)	Treatment assignment	Random allocation	Generated	categorical
Demographics and turnout scores for triplees (age, sex, race/ethnicity)	Covariate	Before randomization	Voter file	Continuous (age, turnout score) or categorical (sex, race/ethnicity); tripler and triplee level data
Pledge collection location	Covariate	Before canvasser interaction	Canvassing data	Binary (Campus vs not campus)
Canvasser	Covariate	Before canvasser interaction	Canvassing data	Categorical
Triplee voter IDs	Used in implementation	During pledge or during separate voter file matching process	State voter file	ID number

Data security and storage

Vote Rev will store data on several vendors with appropriate security policies: Google Cloud Platform, AWS, Civis (running on AWS), or Google Drive. Data access will be limited to team members working on the project, including VRAF staff and staff at canvassing firms.

Voter file data and pledge data will also be stored on servers operated by our canvassing software vendor (GRU).

Post-data-collection registered addendum

February 2023

Background

This document describes changes and updates to our analysis plan for the HTOTS LargeCT.
Note: Link goes to our OSF pre-registration. **This link allows the user to view documents that are currently embargoed for the general public so be thoughtful about sharing it**.

Purpose

During program implementation, we discovered a number of canvasser errors and data anomalies that we hadn't predicted when writing our initial protocol. This means that we need to make data cleaning and inclusion decisions that were not part of our original pre-registration. In this document we will pre-register, as much as possible, all added data cleaning and inclusion decisions **before** Vote Rev Action Fund (VRAF) has voting outcome data in hand.

For clarity:

- Our original pre-registered protocol was registered before any study data was available, except for canvassing results from early training runs that were used to refine the study process
- This pre-registered update was registered after the study was complete, and after Georgia released its voter file results, but before VRAF purchased that data. At the time of registration, VRAF has access to data on canvassing outcomes (eg, contacts per hour) and demographic data on message recipients ("triplees"), but individual-level outcome data is not available to anyone at VRAF and there is no way we could know what effect any given change or analysis decision would have on our outcomes.

Background

Protocol changes

Randomization issues

In protocol section 6 - Assignment, we describe our process for randomizing households and balance-checking. During training, canvassers used an early randomization file (referred to as "R1") that was generated by a similar but simpler version of the process. When we attempted to replace R1 with a new randomization file based on the full process in the protocol (referred to as "R2"), we encountered a number of technical errors (internal VRAF users, see the HTOTS GA LargeCT implementation log for details) that caused R1 to be used for the first days of the study, and on some days there was a mix of canvassers using R1 and canvassers using R2. This issue would have been invisible to canvassers (unless two canvassers looked up the same person on their phones and compared results, which should never have happened in practice), so it only affects how we handle randomization.

Differences between randomizations

R2 is as described in the main protocol. R1 differed in that:

- The only household stratification variable was number of registered voters in the house, treated as 1 vs 2 vs 3 vs 4 vs 5 vs 6-10. Individuals in households larger than 10 were treated as their own household.
- Residence in Atlanta was determined by having "Atlanta" as the city name (in R2 we used county instead).
- The full list of variables used in balance checks was:
    - triplee age, gender, and race/ethnicity
    - Triplee turnout in the 2018 general election, 2020 general election,
    - Ideology score
    - Residence in the city of Atlanta
    - Modeled 2022 turnout score
- Balance checks used the normalized differences approach only, instead of combining that approach with a logistic regression.
- Categorical variable levels with fewer than 50 observations were combined into a single category; if that category was still smaller than 50, it was not used in balance checks.

Additionally, we received an update to the state voter file data between R1 and R2, with new registrants added and purged registrants removed. This means that each randomization contains a small number of individuals not present in the other.

Re-randomization

R2 was constructed as described in the main protocol. However, if a triplee name had already been collected by a canvasser whose device was using R1, we constrained that triplee's R2 condition to be the same as R1. This ensured that repeat triplees would be in the same study arm both times.

There were still a few cases of a triplee being recruited twice in different arms – this could occur if they were collected by a canvasser who was stuck on the R1 list after the R2 list had already been created. This occurred for ~242 triplees (0.6% of the total sample). These triplees cannot be used in the analysis.

Likely effects

Our raw data indicates whether a given canvass interaction used R1 or R2, so we always know with certainty which condition assignment the canvasser was actually told to use (see the *noassignment* flag below for a very small number of exceptions).

The R1 and R2 processes were very similar, and both lists passed balance checks. The partial use of R1 gives us a slightly higher risk of finding a chance imbalance on one of the

R2 balance check variables during our post-data-collection balance checks, but if that does not occur, there should be absolutely no effect on our results.

Data quality and inclusion

Canvassing quality

During field observation we noticed a number of canvasser behaviors that reduced the quality and validity of our data. Some canvassers, especially early in the study, failed to record control triplees, believing that only triplees who were eligible to receive messages "counted."[9]

We discussed a potential protocol for excluding canvasser-days of data with too large a treatment/control imbalance, but in consultation with Kate Duch decided that we could not detect erroneous data accurately enough. Dr. Duch recommended that we instead just continue with our stated intention to exclude all data from 11/25 and 11/26, which were intended as training days.

Therefore, we will include all canvasser data as originally intended, with the exception of specific exclusions described below.

This means that our final dataset will show some imbalance, with a larger number of treatment group triplees than control group triplees. Based on field observations and canvasser interviews, we do not believe this presents any risk of spurious findings. The errors we observed were generally occurring at a canvasser level, meaning that none of the missing data would have been associated with properties of the tripler or the triplee.

Data inclusion

We will exclude data based on the following computed flags. Some of these were specified in advance, and some were specified after reviewing canvassing data and noticing patterns that we believed strongly indicated invalid data.

- training_data - Data that was collected on 11/25 or 11/26, which we specified in our original protocol as a training period.
- questionable_name - Participants with the names "John Doe" and "John Doehla", which occur many times in the dataset and are likely to be from trainings or examples

[9] Internal VRAF users, see HTOTS_LargeCT_data_quality_addendum for details on these issues

- canvass_lead_name - triplees who share a first and last name with one of our canvassing team leads. We know that at least some of these individuals looked themselves up or had canvassers look them up frequently as part of training.
- noassignment - Raw data indicates that no random assignment is present for the individual in the randomization file. This appears to be due to a software error that recorded triplees who were present in R2 but not R1 as somehow receiving an R1 assignment. This affects only 4 triplees (0.01% of all records).
- mangled_data - Indicates 177 tripler records (0.39% of all records) where data is in the wrong columns (eg, there's no voter ID and instead the city name is in the voter ID field). We don't know what caused this but most cases also contain lowercase text, indicating manual data entry instead of voter file lookup.
- canvass_lead_submission - this data was submitted by a site lead, likely as training or a demonstration. Leads were not supposed to canvass at all.
- duplicate_within_interaction - This triplee was already named once in this tripling interaction, eg, when asked for triplee 1 the tripler selected Michael Cohn with voter ID #1111, and when asked for triplee 3 they identified the exact same individual. We would keep the first instance of Michael Cohn and discard the others. [10]
- crossed_assignment - All of the following occurred:
    - This triplee was named by more than one tripler
    - At least one of those triplers was asked to message them (ie, the triplee was in the treatment group)
    - At least one other tripler was asked not to message them (ie, the triplee was in the control group).

  Therefore, their treatment assignment is ambiguous (see Re-randomization for an explanation of how this happened).

Fraud

We will handle data identified by the field team as likely or confidently fraudulent as described in the protocol, without changes. The flag variables used are:

- fraud_high - this data was confidently identified as fraud and will be excluded.
- fraud_medium - the field team saw signs that this data was fraudulent, but they were not strongly confident about this. The original protocol said "we will check for balance between treatment and control triplees in the data that would remain if [the data tagged as fraud_medium] were removed. If they would be unbalanced, we will retain this data." This was a misstatement; our intention was to exclude the

[10]More commonly, a triplee is duplicated because more than one tripler knew them and decided to message them. As described in the protocol, we will use the same process of keeping the first record for that triplee and excluding the others (with the exception of crossed assignments; see below). However, in our descriptive data we will note the number of triplees named by more than one tripler, whereas data with the duplicate_within_interaction flag will be treated as an artifact, and the triplee in question will be treated as if they were named only once.

data only if it would introduce *new* imbalances between treatment and control. That is, if the data is unbalanced on any set of variables, and excluding the cases tagged as fraud_medium would leave that set unchanged or make it smaller, we will exclude them.[11]
As a robustness check, we will also check our primary analysis with the fraud_medium individuals included.

Consistency with external review

Our external reviewer, Kate Duch, has said she plans to exclude triplees using only the *canvasser_name, duplicate_within_interaction,* and *crossed_assignment* flags, along with the fraud protocol. She will include all other individuals who have an entry resembling a voter ID in any field. As a robustness check, we will re-run our analysis using her exclusion scheme and will note any changes that are large or statistically significant. For the 4 triplees with the *noassignment* flag we will use their treatment assignment according to the randomization file in which they do appear.

Campus location

Coding

Canvass firms did not provide us with the data needed to determine which pledges were collected on college campuses. In order to classify the data, we visually examined a map of pledge location data and identified all campuses where pledges were collected. For each campus, we then selected numerous reference points across all parts of the campus and whose coordinates were within the campus boundaries. We then categorized all data points recorded within 500 meters of a campus reference point as college campus pledges. We then created a visualization of the locations of every pledge and their proximity to a campus. On a few occasions, we saw that there were pledges that were very close or adjacent to a campus boundary but were not close enough to a reference point to qualify as a campus pledge. It is normal canvassing practice for a campus canvassing site to sometimes include streets very close to a college campus that are frequented largely by students. To cover those situations, we added additional reference points for the campus so that those pledges would be within 500 meters of a reference point.[12]

---

[11] In the unexpected event that excluding the fraud_medium cases improves balance on some variables and harms it on others, we will exclude the data only if the new set of imbalanced variables is strictly smaller than the original set.
[12] This is essentially a lower-tech version of creating a shape object exactly reflecting the boundary of the campus and including points within 500 meters of it.

Unlocalized records

Some records have missing or malformed GPS data. These records will be treated as non-campus data for analysis.

Other data quality issues

- 10 triplees who were part of the initial randomization were missing data on all covariates and we were not able to recover it. We suspect this is due to their leaving GA in between R1 and R2, combined with an error in how we archived earlier versions of the voter file. These triplees will be excluded from the analysis.
- The early vote data includes a small number of triplees with early vote dates prior to the start of early vote in their county (≤2.5%[13]). In our experience, dates in early vote data are often off by several days and this is not an indication that the data is otherwise questionable. We decided to exclude early vote data from two individuals whose runoff early vote dates were recorded as 10/25 and 11/07, prior to the general election. These individuals were treated as if they had not voted early.
    - Note that the uncertainty about precise early vote dates introduces some random error into our process of excluding individuals who voted prior to the day on which their tripler contacted them.
    - We also call attention to the fact that this exclusion process 1) is biased towards excluding voters relative to non-voters and 2) creates a negative correlation between the date the tripler sent their message and the triplee's likelihood of voting (because those contacted later had a longer window to vote early and thus be excluded).
    - Both of these processes should behave identically across the two study conditions, so they do not bias our results.

Analysis updates

Interaction analyses

The "Other Interactions" section is ambiguous about whether we intend to run a separate model for each of the 3 interaction variables being tested, vs one combined model with all 3. We intend to use one combined model.

Instrumental variable analysis

For each triplee, canvassers were asked to record whether the tripler messaged them in the canvasser's presence. We were surprised by the number of discordances that were

[13] Early vote dates varied by county and we did not check every county to determine their start date, so the number may be substantially below 2.5%

recorded (messaging a control triplee in front of the canvasser, or declining to message a treatment triplee). Therefore, we will add an instrumental variable analysis, re-running the "primary analysis" described in the protocol with the critical independent variable being whether the canvasser observed the triplee being messaged, instrumented by treatment assignment and the covariates specified in the primary analysis.

This analysis will not be included in our adjustments for multiple comparisons.

If we do not find significant effects in our primary analysis but do find them in the instrumental variable analysis, we will not report this as unambiguous support for the HTOTS intervention as delivered. We will consider it a positive signal for 1) further refinement of our experimental approach and 2) re-testing with alterations to improve tripler adherence.

Post-data-collection unregistered addendum

March 2023

Background

This document describes changes and updates to our analysis plan for the HTOTS LargeCT. Unlike the first addendum it records decisions and events made after pre-registration and after outcome data was available.
Note: Link goes to our OSF pre-registration. **This link allows the user to view documents that are currently embargoed for the general public so be thoughtful about sharing it**.

Purpose

This document records any decisions we needed to make that deviate from or were underspecified in our pre-registered trial and its pre-registered addendum. It also contains various other errata. Decisions in this document were not pre-registered and were made after we had access to full voting outcome data.

Protocol changes

Stratified Randomization

During randomization (R2 only) we added another stratification variable: *number of members of the household who voted in the 2022 general election* (we hadn't expected to have access to this before study launch, but it arrived just in time). We neglected to update the protocol, so our pre-registration doesn't list this along with the other stratification variables used in the "Sampling and assignment process" section.

Access to outcome data

In our pre-registered protocol we stated that all decisions had been made before we had access to outcome data. We realized that we were mistaken: Regular study operations required us to download lists of early voters so they could be excluded from the list of

valid triplees; some of these early voters had been named by triplers earlier in the study, so this constituted outcome data for them. Therefore, the first addendum to our protocol was written while we did have access to partial outcome data, though we pledge that we did not run any analyses on this data, or otherwise look at how it was affected by treatment condition, during that time. This only affects the addendum; the main protocol was written before canvassing and before VRAF had any early vote data.

Triplee/tripler awareness

We stated that "Canvassers may inform [triplers] that the list is randomly generated for a study, but will be instructed to do this only if directly asked what the list is." In fact, canvass leads at different sites took different approaches to this, some encouraging canvassers to be more forthcoming about this information than others. We don't have a theoretical reason to believe this would bias outcomes and there is no data relevant to this question.

Analysis updates

Balance checks

Significance testing

We stated that we would use a one-tailed test for treatment effects on voter turnout (treatment higher than control only). We did not explicitly state this, but we will use the standard two-tailed test for all other comparisons (such as balance checks and interaction effects) because we did not state, and do not have, a directional hypothesis for any other tests.

First and final check methods

In the original protocol, we state that we will use both the normalized differences method and our external evaluator's regression method for balance checks. However, under the "first check" and "second check" subsections we state that we will use both as balance criteria for the initial (randomization file) check but only mention the normalized differences metric for the (post-data-collection) check. This organization leads to some ambiguity; we should have put the paragraph "for compatibility with the method..." under the "First check" heading instead of above it.

Fraud-related covariate balance checks

The same ambiguity is present when we state "If there is data that we suspect, but are not highly confident, is fraudulent, we will check for balance between treatment and control triplees in the data that would remain if it were removed." The intention was to use

the same balance check method and criterion used in the main balance check for the same dataset, meaning the normalized differences method (threshold of .25 for imbalance) and not the regression method.

After data analysis began, it was also pointed out to us that the stated procedure could be read as referring to balance between the proportion of treatment and control cases, ie, that we would not exclude the data if it would result in the proportion differing significantly from 50/50. This criterion would have the same issue as the covariates balance check: The dataset without medium-confidence fraud removed was already imbalanced (47.25% control, 52.75% treatment), and our intent was to remove medium-confidence fraud unless doing so would make the dataset *more* imbalanced. We will test for this by chi-squared test for the proportion of treatment participants in the full dataset vs. the dataset with medium-confidence fraud removed; if the two differ with $p<.05$, and the post-exclusion data is farther from the intended 50/50 split, we will retain the medium-confidence fraud data.

Fraud-related balance check treatment of categorical variables

In the pre-registered addendum to our protocol we clarified that we intended to exclude this medium-confidence fraudulent data as "if the [full dataset including medium-confident fraud] is unbalanced on any set of variables, and excluding the cases tagged as fraud_medium would leave that set unchanged or make it smaller". However, we did not specify whether we would treat the number-in-household variables (eg, number of White voters) as a continuous variable or as a categorical variable. We chose to treat them as categorical because of how they were treated during random assignment (used as categorical variables with size 5-6 treated as a single category).

For balance checks, each categorical variable is expanded into a set of dummy variables. This raises ambiguity about what to do if the list of variables in our dataset with imbalances is unchanged, but the list of dummy variables with imbalances has grown longer. For example, there could be a treatment-control imbalance on the number of 3-voter households; after removing medium-confidence fraud records the groups could then be imbalanced on the number of 3-voter households and the number of 4-voter households. It is ambiguous whether this meets the criterion of "leave [the set of variables with imbalances] unchanged" or not. We choose to treat it as meeting the criterion, but recognize that this is a slightly more lax approach that was not properly established ahead of time.

Checking data excluded for early vote

In the protocol we state that we will "remove any triplees from the dataset who had voted prior to when their tripler was recruited. Before doing this, we will check for balance between control and treatment in the data to be removed". Two clarifications:

1. To be precise, we marked a triplee for removal if the date shown in the voter file for their early vote was before the date the canvasser interacted with their triplee.
2. As with the other post-data-collection balance checks, this referred to the normalized differences test with a criterion of ≤.25.

Location of canvass

We stated that the final balance check would include "location of canvass." This was intended to refer to campus vs. non-campus, not to the canvassing office / region.

Handling duplicate triplees

If a particular person was named as a triplee by more than one tripler (or one tripler who pledged multiple times), we need to keep only one of their records, so their outcome is only counted once. In the preregistered protocol we stated that we would preferentially keep (not exclude) records that were not excluded for any other reason. When constructing exclusions we needed to add several specifications:

- We treated medium-confidence fraud and early voting (prior to tripler contact date) as exclusions for this purpose, even though their status as exclusion variables depends on balance checks performed later in the process.
- We did not originally specify how we'd choose between records that may differ on other covariates, which include canvassing site, canvasser, date, and on vs. off campus. We have no reason to systematically prefer any values for these variables. At Dr. Duch's advice we opted to select the earliest record.

Predictor variables

Number of voters in household

We stated that this would be treated the same as in balance checks except as a continuous variable. This implies but did not explicitly say that it would also use the same adjustment used for randomization: If a "household" consisted of >6 registered voters at the same address, we would treat each voter as an individual household.

Additional predictors

Our external evaluator, Dr. Kate Duch, pointed out that:

1) Atlanta metro area status was used in balance checks but not specified as a covariate in our main analysis, and it would make sense to add it.
2) An indicator variable for first vs second randomization should be added to our main analysis; it was not pre-specified because the use of two separate randomizations resulted from an unanticipated technical problem.

We agreed with Dr. Duch that both variables should be added to our analyses as covariates. As a robustness check, we repeated the analysis without them.

Missing data

We unexpectedly learned that two of our covariates, ideology and modeled runoff turnout score, were missing data for 1,564 triplees. Since these variables were of much lower importance than the main treatment and outcome variables, we imputed the mean value for these individuals so that they could be included in our analyses. As a robustness check, we repeated the analysis with those who were missing data dropped.

Instrumental variable analysis

For each triplee, canvassers were asked to record whether the tripler messaged them in the canvasser's presence. We were surprised by the number of discordances that were recorded (messaging a control triplee in front of the canvasser, or declining to message a treatment triplee). Therefore, we will add an instrumental variable analysis, re-running the "primary analysis" described in the protocol with the critical independent variable being whether the canvasser observed the triplee being messaged, instrumented by treatment assignment and the covariates specified in the primary analysis.

This analysis will not be included in our adjustments for multiple comparisons.

If we do not find significant effects in our primary analysis but do find them in the instrumental variable analysis, we will not report this as unambiguous support for the HTOTS intervention as delivered. We will consider it a positive signal for 1) further refinement of our experimental approach and 2) re-testing with alterations to improve tripler adherence.