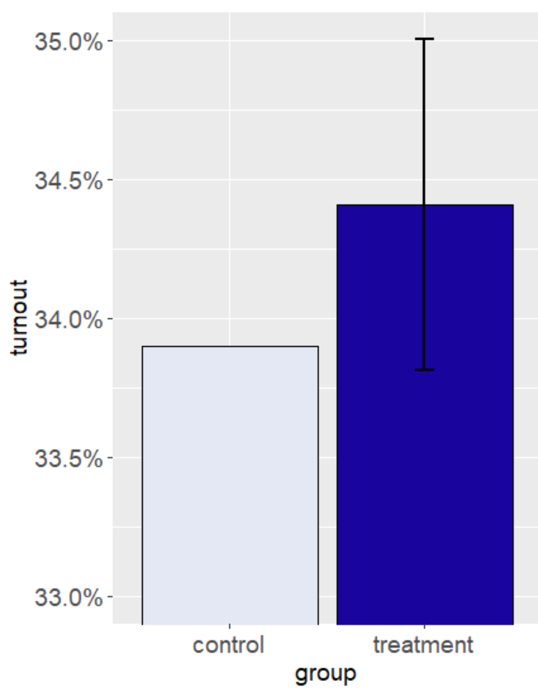
High traffic on-the-spot relational turnout: 2022 GA runoff experiment

Analysis Report, Vote Rev Action Fund, October 2023

Executive Summary

In the 2022 Georgia Senate runoffs, Vote Rev Action Fund ran a randomized controlled trial of high-traffic on-the-spot relational turnout (HTOTS). Canvassers in public places asked pedestrians ("mobilizers") to send text messages to 5 friends or family members ("friends"), reminding them to vote. To test effectiveness, friends were randomized at the household level to receive texts or not.



n=33,478
90% confidence interval
p (1-tailed)=.085

Friends reminded by mobilizers voted at a rate 0.51pp higher than in the control group (p=.085, one-tailed). Independent analysis by One Minus Beta analytics had similar findings. **HTOTS is an effective way to increase voter turnout**, even in a high-salience federal election.

Friends reached with reminders were predominantly Black (48%), female (55%), and young (median age 35), with many high-potential voters (half had runoff turnout scores under 60).

Reminders appeared more effective among friends under 35 years old, and a model including age indicated that **HTOTS on college campuses could increase voter turnout by 0.74pp**.

Vote Rev provides free support to organizations implementing HTOTS in 2023 and 2024! Please contact Marisa Kanof, Director of Partner Success: marisa@voterev.org.

Contents

Related documents

- See protocol documents[1]
- Analysis code available upon request
- An independent analysis was carried out by Dr. Kate Duch of One Minus Beta Analytics. In general, Dr. Duch agrees with our findings though her effect estimate is slightly lower. We discuss the reasons in this section.

Background

HTOTS

In high-traffic on-the-spot relational turnout (HTOTS), canvassers station at locations with high foot traffic and ask passersby ("mobilizers") to send a message to friends or family members ("friends") to remind them to vote in an upcoming election.

In the 2022 Georgia senate runoff elections, Vote Rev Action Fund worked with GRSG, a paid canvassing firm, to run a large-scale randomized controlled trial (RCT) of HTOTS. Canvassers logged ~4,200 work-hours across six urban areas in Georgia[2] over 10 days, starting on November 27 and ending on election day (December 6).

The RCT

The strength of relational turnout programs is that mobilizers use their own social networks, contacting friends and family that organizations may never be able to reach. However, this means we do not learn who the friends are, and can't easily check their voting outcomes. For this trial, we asked mobilizers to show the canvasser who their friends were, using a mobile voter file lookup app (Grassroots Unwired, GRU). If a mobilizer was unable to find their friend, or if they found multiple matches in the voter file and could not disambiguate them, the canvasser asked them to look up a different person.

Prior to the trial, we pre-randomized the full voter file so that every potential friend would be marked as "ok to text" or "do not text" in the app[3]. The canvasser then instructed the mobilizer to either text that friend a reminder about voting in the election, or to not text

[1] HTOTS originated from a tactic called "Vote tripling" in which voters were asked to remind 3 friends to vote. We now ask for 5 or more relational reminders per interaction. In related documents we continue to use the legacy terms. "Vote tripling" refers to relational turnout, "tripler" refers to the mobilizer, and "triplee" refers to the friend.
[2] The canvassing regions were, in descending order of friends collected, College Park (southern Atlanta), Athens, DeKalb County (eastern Atlanta), Augusta, Marietta, and Savannah.
[3] Randomization was carried out at the household level, in order to prevent concerns about within-household spillover contaminating control group friends who live with a treatment group friend.

them and move on to the next friend.[4] Thus, we developed comparable treatment and control groups – both composed of people that a mobilizer expressed a willingness to text about the election – and were able to compare their voting outcomes in the election. To reduce participant burden, we did not collect any information about the identity of the mobilizer.

Our full protocol, including precise details on our sample universe and randomization, was described in three stages: a pre-registered protocol written prior to data collection; a post-data-collection addendum registered before any outcomes analysis, and an unregistered addendum covering minor decisions that we needed to make in the process of analysis. The protocol docs can be viewed (combined into a single document) here. The pre-registration is viewable on the Open Science Framework.

Implementation

Implementation issues

We encountered several technical and logistical issues during the execution of the study. Issues with impact on the analysis or interpretation of results are described in the section of the study protocol titled "Post-data-collection registered addendum". Issues of particular note were:

1. We changed our randomization protocol shortly before the study began, but canvassers did not receive the new randomized assignments until several days into the study. This led to a small number of friends being excluded from the study because they were named by multiple mobilizers and assigned to different treatment conditions. See the protocol link above for details.
2. Canvassers required retraining on a number of issues. In general these were addressed during training or in the first few days of the study. Most prominently, they initially seemed to consider friends randomized to the control group as "failures" or as "not counting" and needed encouragement to record them in the same way as those in the treatment group. Field observers also noted that canvassers sometimes recorded friends as having been messaged when the mobilizer either pretended to message them or promised to message them later. Internal Vote Rev users can read more about these issues in the implementation log.

[4] One of the biggest challenges of executing this study was convincing mobilizers to *not* text a person they had just said they were willing to text. Based on piloting we encouraged canvassers to simply ask the mobilizer not to text the person and move on quickly, rather than giving the mobilizer time to think about the situation. Some canvassers instead gravitated towards telling mobilizers that the recipient was on a "do not call" list or just not on their organization's list of people to contact.

3. Our final dataset had a significantly higher proportion of friends in the treatment group than in the control group. See the Concerns and limitations section for further discussion of this.
4. The Vote Rev Action Fund field team detected a substantial number of canvassers who submitted fraudulent records, often to appear that they had been working at times when they were not on location at all. See the Fraudulent data section for details.

Internal Vote Rev users can also view the implementation log for details on minor incidents.

Data acquisition

Voter file matching used a copy of the Georgia voter file obtained from TargetSmart shortly before the runoff, containing the latest data available from the state of Georgia. TargetSmart also provided data used for covariates such as modeled race/ethnicity and gender, from the same time period. Modeled demographic and turnout scores did not include results from the general election.

Data exclusions

Exclusion criteria

Exclusion was carried out in four stages:

1. Exclude records with markers indicating that they are bogus, corrupted, definite fraud, or from the training period. For a full list, see the data inclusion section of the protocol addendum.
2. For friends who were named more than once, exclude all but one of their records. For a full description of the exclusion algorithm, see the "handling duplicate friends" section of the protocol addendum.
3. Exclude friends who voted early prior to the first day on which a mobilizer contacted them.
    a. We specified we would do this only if the friends it would exclude do not show a bias towards treatment vs control based on the normalized differences criterion in the balance checks section of our preregistered protocol. The normalized difference was .013, easily meeting our criterion of <.25.
4. Exclude friend records that the Vote Rev field team believed, but were not certain, were fraudulent ("medium-confidence fraud", flagged in the dataset as `bad_fraud_medium`). Do this only if the remaining non-excluded data passes balance check criteria. See Fraud exclusion below for details.

Fraud exclusion

VRAF does not pay canvassers per mobilizer but occasionally canvassers provide fraudulent data anyway, typically because they're not on-site or not working at all. 414 records (0.9% of all data) were excluded because the VRAF field team was able to show that they were submitted fraudulently. There were also 1,362 otherwise-usable records (2.9% of all data) that our field team believed – but were not fully confident – were fraudulent. We intended to exclude these records unless this would create a treatment : control imbalance in the data on either covariates, or proportion of treatment vs control cases. See the protocol documents (the pre-registered addendum and post-registration addendum sections) for details on this process.

Fraudulent data covariate balance

No variable approached our criterion of normalized difference <.25, either before or after excluding medium-confidence fraud data.

Fraudulent data condition assignment balance

The registered protocol said "we will check for balance between treatment and control friends in the data that would remain if [the data tagged as medium-confidence fraud] were removed. If they are unbalanced, we will retain this data." Our intention was to retain the medium-confidence fraud data as long as it would not make the remaining data *more* unbalanced (ie, farther from a theoretical 50/50 split). This misstatement was caused by our failing to envision the possibility that treatment and control would *already* be unbalanced before excluding the data.

Prior to medium-confidence fraud exclusions, the sample was 47.4% control and 52.6% treatment, which is already unbalanced ($p<.001$).[5] If medium-confidence fraud is excluded, the sample is 47.2% control, not a significant change (chi-squared=0.11, $p=.74$). As a result, the composition of the final dataset differs depending on whether we interpret the exclusion criterion as intended, or strictly as written.

We believe that the written interpretation, in this case, is not sensible: There's no reason to believe fraudulent data would be the sole cause of any imbalances in the data, so it would not make sense to make its removal contingent on fixing existing imbalances. Our analyses below remove medium-confidence fraud records, and we re-run the primary analysis with that data included as a robustness check.

Numbers excluded

[5] See the Treatment balance section for more on this issue.

Our raw dataset contained 41,871 records after removing data from the training period, which we specified would not be used.[6] After all other exclusions, 33,484 (80%) of the data remained. Most of the excluded data was from friends who voted early prior to being named by their mobilizer. We pre-specified that these individuals would be excluded because their outcome was predetermined before treatment was delivered.

The table below shows how many records were removed at each stage of this process.

	sample size	number removed	% removed
Total records	45936	-	-
Remove training days	41871	4065	9.7%
Remove bogus, corrupted, or clearly fraudulent data	40806	1065	2.6%
Remove duplicates	39396	1410	3.6%
Remove early voters prior to contact	34846	4550	13.1%
Remove medium-confidence fraud (final dataset)	33484	1362	4.1%

Data exclusion funnel

Final dataset descriptives

The descriptives labeled "analysis sample" below refer to the data that met all criteria to be used in the main analysis. This includes removing those who voted prior to being contacted, or who were classified as medium-confidence fraud.

Excluding friends who voted prior to being contacted alters the sample by removing individuals who are more likely to vote and to vote early. The "including early voters" column shows descriptives with those friends included.
The "including early voters" column is appropriate for generally characterizing the type of friend reached by HTOTS. The "analysis sample" column is more appropriate for characterizing the population of people who are eligible to have their voting rates

[6] We specified that training data would not be used because canvassers would still be learning the intervention and the software, and might not deliver the intervention effectively. This data has been included in some other reports, which focus on characterizing the friend sample and not on voting outcomes.

affected by HTOTS, when run with timing similar to the current study and in a similar electoral context (length of early voting, time before election, etc).

		Analysis sample	Including early voters
Treatment condition	% control	47.2%	47.3%
Campus status	% recruited at campus location	38.6%	38.9%
Race/ethnicity[7]	African-American	47.7%	48.1%
	Caucasian	40.0%	40.0%
	Asian	4.7%	4.6%
	Hispanic	4.3%	4.1%
	Other / Uncoded	2.6%	2.4%
	Native American	0.7%	0.7%
Gender	Female	54.2%	54.8%
	Male	45.7%	45.1%
	Unknown	0.1%	0.1%
Age	25th %ile	24	24
	Median	34	35
	75th %ile	50	52
Modeled runoff turnout score	25th %ile	29	31.5
	Median	54	59.6
	75th %ile	85	88.1
Ideology (higher = more likely Democrat)	25th %ile	49	50
	Median	71	72
	75th %ile	82	82
Past election	2022 general	41%	48%

[7] The taxonomy and category names used here are taken from the available voter file data. Although they are not ideal in terms of accurately representing real-world race and ethnicity identities, the dataset has high coverage and we believe it to be accurate.

turnout	2020 general	61% 64% of those old enough to have voted	63% 68% of those old enough to have voted
	2020 runoff	48% 52% of those old enough to have voted	52% 56% of those old enough to have voted
	2018 runoff	10% 13% of those old enough to have voted	13% 16% of those old enough to have voted

Final dataset descriptives

Balance checks

Preparing variables for analysis

As described in our protocol, categorical variable categories with <50 records were grouped into one category and, if still <50 records, were assimilated into the largest category. This resulted in:

- For race/ethnicity, individuals with missing data or coded "Uncoded" were combined into "African-American"
- For gender, individuals who were missing data or were coded as "Unknown" were combined into "Female".[8]

Covariate balance

Following our protocol, we computed normalized difference scores for treatment vs. control on all covariates. For categorical variables, we treated every level as a separate variable. We ran this process once for the full dataset and once for the subset from the Atlanta metro area.

No variable approached our threshold of 0.25. The largest normalized difference was campus recruitment (0.04 for both the Atlanta subsample and the sample overall). All other variables were ≤0.02.

Outcomes

For all tests of treatment effect on voting, we prespecified that the significance criterion would be p=.10, one-tailed, because we had a directional hypothesis that treatment would

[8] Male and Female were the only genders present in our dataset, likely due to limitations imposed by the sources reporting this information to our data provider.

be higher than control.[9] Other tests, including treatment interaction effects, are two-tailed because we did not have directional hypotheses for these.

Main analysis: Friend voting outcomes

Our primary analysis was an OLS linear regression predicting **friend probability of voting in the runoff** based on treatment condition and a large set of covariates. Standard errors were clustered by household.

For details see the statistical approach of the protocol document. As noted in the post-registration section, we added one covariate that was unintentionally omitted (Atlanta metro residence) and one that reflects an unexpected technical issue (first vs. second version of the randomization list, described in the randomization issues section of the protocol). Both inclusions were approved by our external evaluator.

Results

After controlling for covariates and adjusting for household clustering, **friends in the treatment condition voted at a rate 0.51pp higher than those in control** ($p=.085$, one-tailed). Other covariates are reported in the appendix.

Robustness checks

We ran several alternate analyses that were either prespecified, or that check the influence of choices that were made during data analysis. The following changes had minimal effect on the results, with treatment effects ranging .49-.52, all $p<.10$:

- Removing the covariates for Atlanta metro area and version of the randomization file, which were added post-hoc
- Excluding friends who were missing values on ideology and turnout score, instead of using mean imputation
- Using Dr. Duch's list of exclusion variables instead of our own[10], while still excluding medium-confidence fraud.

[9] Vote Rev Action Fund is generally moving towards one-tailed tests when assessing the effectiveness of a tactic vs. no treatment. This is because the purpose of our RCTs is to determine whether organizations should invest in the tactic. If the tactic is not *more* effective than doing nothing, they should invest no resources in it; showing that it's worse than nothing would not be of any additional use. We continue to use two-tailed tests when comparing active treatments against each other.
[10] Differences are described in the protocol addendum. In brief, Dr. Duch did not exclude friends based on our flags for questionable names, lack of assignment in the appropriate randomization file, mangled data where a voter ID was still recoverable, or being submitted by a canvass lead.

The final robustness check included medium-confidence fraud data instead of excluding it. Including data tagged as medium-confidence fraud causes the treatment effect to become nonsignificant (p=.12), though the estimated effect is still meaningfully large (.42 vs .51 with fraud excluded) and the change in p-value is small.

As a post-hoc analysis, we ran the primary outcome model with an interaction between treatment and medium-confidence fraud status. The coefficient for treatment when data is not marked as fraud is comparable to when medium-confidence fraud data is excluded entirely (.50pp, p=.09), and the interaction term between treatment and fraud status is very large yet nonsignificant (-2.00pp, p=.30). This corroborates our interpretation that adding medium-confidence fraud data does not significantly change the outcome; the treatment effect was already close to the significance threshold and adding the medium-confidence fraud data was just enough to move it across.

Comparison with external evaluation

Our external evaluator, Dr. Kate Duch at One Minus Beta Analytics, ran an independent analysis of the study. We agreed on most data cleaning and analysis principles and methods ahead of time, but diverged in a few ways:

- Dr. Duch used a more stringent criterion for covariate imbalances and concluded that our treatment conditions differed on several covariates. She addressed this by adding interaction terms for those covariates to her primary analysis.[11]
- Dr. Duch used the more restrictive interpretation of our medium-confidence fraud exclusion process, and retained this data in her primary analysis.
- Dr. Duch declined to use several other exclusion variables, believing them to be too prone to false exclusions. As noted in the Robustness checks, this made effectively no difference.

Dr. Duch's primary outcome treatment effect estimate (+0.44pp, p=0.11) was very similar to our estimate with medium-confidence fraud data not excluded. She carried out a robustness check that did exclude the medium-confidence fraud data (p. 8, Table 3, specification #3) and obtained a result of +.52pp, p=.08, very similar to our main analysis. Thus, apart from the decision about excluding medium-confidence fraud data, her analysis agrees with ours. We describe above our reasoning for our exclusion protocol.

Overall, Dr. Duch substantively agreed with our conclusions. Her report states that although she found the treatment effect to be just above the significance threshold, it was "probably not due to chance."

[11] The treatment effects reported in Dr. Duch's writeup and quoted in this document are estimated values for the main effect of treatment at average levels of other covariates with interaction effects, as produced by Stata's "margins" command

Secondary analyses

Differential effectiveness: Race/ethnicity

As described in the exploratory analyses section of the pre-registered protocol we repeated the main outcome analysis, adding interactions between race/ethnicity categories and treatment effect. The reference category was set to "Caucasian" in order to focus on detecting scenarios in which the intervention works well for non-Hispanic white participants and poorly for some other groups.

Results

The main effect for treatment, which here reflects the effect on white friends only, was considerably larger and still significant (0.96pp, p=.05). Interaction effects were all nonsignificant, indicating that there is no strong evidence for differential effectiveness now but we may want to investigate this topic in future design work.

Differential effectiveness: Campus, age, and date

As described in the exploratory analyses section of the pre-registered protocol, we repeated the main outcome analysis, adding interactions between treatment effect and each of the following: age, age squared, campus status (ie, whether the mobilizer was recruited in a campus area), and number of days prior to the election.

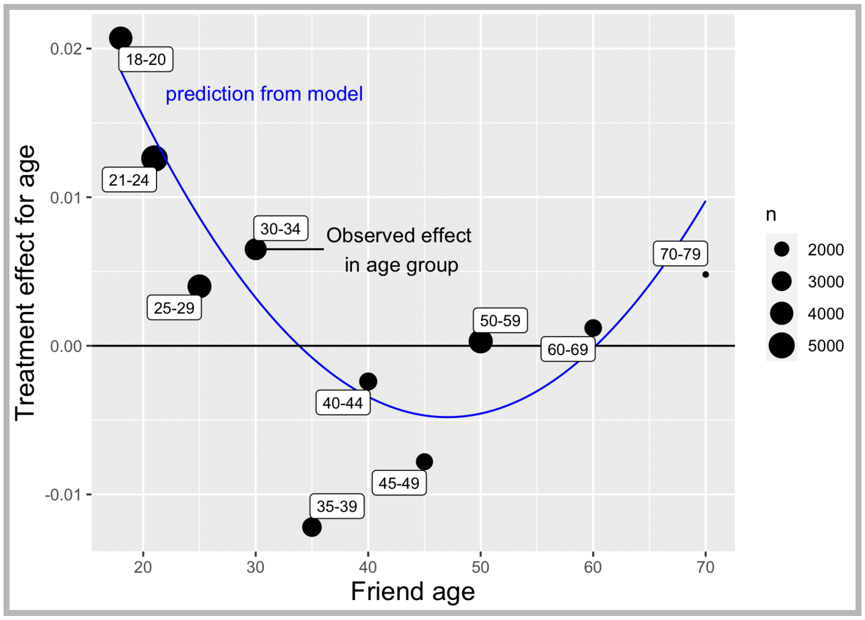
Results

The interaction effects for campus status and days prior to election did not approach significance (p=.78 and p=.16 respectively) and are not discussed further.

The interaction effects for age and age squared were meaningfully large and significant (age: -.06pp per year, p=.06; age^2: .+0027pp per year, p=.02), suggesting that effectiveness is higher for young and very old friends.

To elucidate this unexpected finding, we re-ran the primary analysis repeatedly within a variety of age buckets.[12] The graph below plots these age groups against the curve from the regression model. This analysis is underpowered, and was not pre-registered. It is used as a tentative illustration only.

[12] The age buckets used here were originally chosen to display meaningful age ranges to mobilizers when doing voter file lookups. We retain them here because 1) they also represent socially meaningful age breakpoints, and 2) due to having a relatively young sample, each bucket has a reasonable number of friends (minimum 2,347 for age 45-49, maximum 5,087 for age 21-24).



Treatment effects by age. Blue line represents the predicted effect from the regression model; points represent the observed treatment effect within each age group. Y-axis is treatment effect in percent (eg, .01 = 1pp).

This suggests that the treatment is strongly effective (>1pp) for friends ages 24 and younger, somewhat effective for friends ages 25-34, ineffective for those ages 35-69, and possibly effective again for those 70 and older. **These individual results are mostly not significant and have large uncertainties,** especially in the older age groups, but they generally coincide with the age interaction effect from our pre-registered analysis.

As an additional step, we re-ran these analyses with an interaction effect for campus status, to check whether younger friend ages were a proxy for their mobilizers being students. This caused results to swing wildly and with no apparent pattern, suggesting that our sample isn't powered to assess this hypothesis.

Effectiveness in campus contexts

HTOTS doesn't have a way to directly target younger friends because mobilizers choose the friends they reach out to. However, mobilizers we recruited on college campuses reached out to younger friends on average (median age 27, compared to median age 37 for non-campus mobilizers).

We created a model that predicts treatment effect based on friend's age and mobilizer's campus status, then used that model to predict the mean effect based on the actual distribution of ages for friends named by campus mobilizers. This suggested that running HTOTS exclusively on campuses could produce a voting boost of around **0.74pp** on friends.

Long-term voting outcomes

Our protocol specified that we would examine voting outcomes in the 2024 general election. As of 2023, this data is unavailable.

Discussion

Potential to increase voting

HTOTS produced a statistically significant increase in voting rates of more than 0.5pp, and suggested a difference of more than 0.7pp if run in a campus context**.** This occurred in a high-salience election with a great deal of messaging competing for voters' attention. Friends reached with reminders were predominantly Black (48%), female (55%), and young (median age 35), with many high-potential voters (half had runoff turnout scores under 60). This provides strong evidence that a mainstreamed version of HTOTS can meaningfully increase voting rates among historically disenfranchised communities.

We are now working to optimize the version of HTOTS tested in the runoff. We believe there are **levers for increasing votes per relational contact:**

1. Younger friends appear to have a stronger response. As described above, an HTOTS program on college campuses would have substantially stronger effects. We are also investigating whether canvassers can coach mobilizers of all ages to preferentially contact younger friends.
2. Our field observations and input from experts suggest that we may be able to coach mobilizers to send more helpful or motivating messages to their friends. Future design research can help us develop options.

Real-world impact

This trial treated 17,666 friends, causing an estimated 0.51% to vote when they otherwise would not have. Thus, the trial itself generated **90.1 additional votes** and would have generated 170.1 total votes if all friends had been treated.

In the general elections we ran a non-research version of HTOTS in Arizona (ie, not slowed down by voter file matching), and found that it could generate **28.6 relational reminders per canvasser-hour**. A small A/B test found that improving canvasser training and increasing the ask from 5 friends to 10 friends could raise efficiency in the final days of voting as high as **43.2 relational reminders per canvasser-hour**.

Additionally, we believe **HTOTS may be more effective than shown in this study** for the following reasons:

1. It is likely that some mobilizers, having had a positive experience reminding their friends to vote, reminded other friends later – including some in the control group.
2. Staff observed that mobilizers sometimes failed to message treatment group friends, or did send messages to control group friends before the canvasser could stop them. We asked canvassers to record when this happened, and used this information in an instrumental variable regression (see appendix), which provided very similar results to the main analysis. However, based on observation we suspect that some canvassers may have misunderstood what to record, or been reluctant to report "failures" and not provided accurate information. Therefore, there may be greater noncompliance issues than we were able to measure.

These phenomena would have spuriously reduced our treatment effect, with no relevance to real-life HTOTS.

Concerns and limitations

Treatment balance

The final sample was 47.2% control and 52.8% treatment. This difference is meaningfully large and statistically significant, and we don't know why it happened. We have *eliminated* the following hypotheses:

- Underlying imbalance in randomization: There is no such imbalance in the full voter file randomization and no condition differences on any covariates large enough to explain this.
- Canvasser unawareness: We repeatedly retrained canvassers on appropriate procedures, including the importance of consistently recording control group friends. However, the imbalance did not improve over time. The anomaly could still be caused by canvasser behavior, but it was not a simple lack of awareness.
- Biased exclusion criteria: The imbalance is present, at a similar size, in the raw data.
- Issues with individual sites or canvassers: There was variability in the treatment/control balance across sites and canvassers, but we were not able to find any prominent subset driving the effect.

We are not aware of any way this difference could have affected the outcomes of the study, but recognize that it is a concern. Vote Rev Action Fund will continue to investigate this phenomenon both retrospectively and when preparing future canvassing operations.

Fraudulent data

Fraud detection and exclusion of fraudulently submitted data are critical to the integrity of future RCTs and future HTOTS implementations with paid canvassers. We are reviewing our methods for identifying fraudulently submitted data and working to improve detection and reduce ambiguities in the future.

Fortunately, fraud is likely to be less of an issue for partners implementing the program with volunteer canvassers, who will presumably not have financial incentives to fake being at work or to make their performance look better.

Appendix: Main analysis covariates

Covariates from our Main analysis are shown below.

Coefficients have been translated into percentage points. p values are two-tailed because we did not have directional hypotheses for tests other than the main effect for treatment.

We have removed coefficients for individual canvassers because they are not of interest. We have removed the coefficient for "days before election day when named" because it is an artifact; as election day gets closer more friends are excluded from the data because they already voted, so friends appear to vote less when named closer to the election.

These are not adjusted for multiple comparisons and are not intended as generalizable outcomes.

	percentage point effect on voting	p value
Voter history		
Voted 2018 general	4.3	<.001
Voted 2018 runoff	10.8	<.001
Voted 2020 general	*0.7*	*0.15*
Voted 2020 runoff	8.8	<.001
Voted 2022 general	50.6	<.001
Modeled turnout probability, 2022 runoff (percentage point score)	0.15 (per point)	<.001
Race (reference class: African-American)		
Asian	*-0.5*	*0.59*
Caucasian	2.5	<.001
Hispanic	1.8	0.07
Native American	*1.6*	*0.43*
Other	-2.3	0.05
Other demographics		
Gender (male relative to female)	-1.1	0.005
Age	0.5	<.001

Age^2	0.005	<.001
Lives in Atlanta metro area	-1.6	<.001
Voters in household	-0.25	0.07
Modeled ideology score (out of 100; higher = more likely Democrat)	*0.02*	*0.13*
Study-related variables		
Mobilizer recruited on college campus	1.5	0.01
First randomization file	*0.3*	*0.62*

Appendix: Instrumental variable regression

Not all mobilizers complied with instructions to text or not text a given friend. Canvassers recorded whether or not each friend was actually texted in their presence. Texted status is indeed a better predictor of voting outcomes than condition: +0.62pp, p=.046. However, there were likely non-random determinants of which friends were actually texted vs. not, so we conducted a (non-pre registered) instrumental variable (IV) regression.

We calculated predicted values for texted status based on treatment condition and all covariates, and then re-ran our primary outcome analysis with the instrumented version of texted status as the independent variable, in place of treatment condition. Results were nearly unchanged: +0.55pp, p=.085 in the main specification and +0.45pp, p=.12 if medium-confidence fraud data is included.