High-traffic pledge collection

2022 Midterms RCT, Results Writeup, May 2024

Executive Summary

High-traffic pledge collection (HTPC) is a site-based relational turnout tactic. Canvassers in public places ask pedestrians ("mobilizers") to pledge to remind up to five people ("friends") to vote. These pledges are collected weeks or months before voting begins. After voting begins, mobilizers receive messages reminding them of their pledge and prompting them to remind their friends to vote.

In the leadup to the 2022 Midterm General Election, Vote Rev Action Fund ran a large randomized controlled trial to test the effect of HTPC on both mobilizer and friends' turnout. Between August 25th and October 19th, HTPC canvassers collected pledges from 30,838 mobilizers in Arizona, Florida, Michigan, and North Carolina. Mobilizers were randomized to receive the usual text message reminders or to receive no reminder (presumably leading most to forget about their pledge). We also tested supplementing text message reminders with a volunteer-written letter via our partner, Vote Forward.

There was no significant difference in voting rates between friends whose mobilizers received reminders and those whose mobilizers did not. There was also no significant difference in voting rates between treatment mobilizers who received reminders and those who did not. There was a nonsignificant trend[1] towards *lower* voting rates for mobilizers who received text message reminders without a letter, but for reasons discussed in the paper we suspect this was

---

[1] The p value for this finding was .05 but we had declared a one-tailed test in the opposite direction. Nonetheless, we take the potential for a negative mobilizer effect seriously and the paper discusses possible interpretations in detail.

chance. [Independent analysis](#)[2] by Professor Melissa Michelson's lab found similar results.

We conclude that the current version of HTPC did not increase turnout in the 2022 Midterm General Election. An early-cycle relational tactic could still have tremendous value and we are pursuing refinements that might increase mobilizer commitment and follow-through.
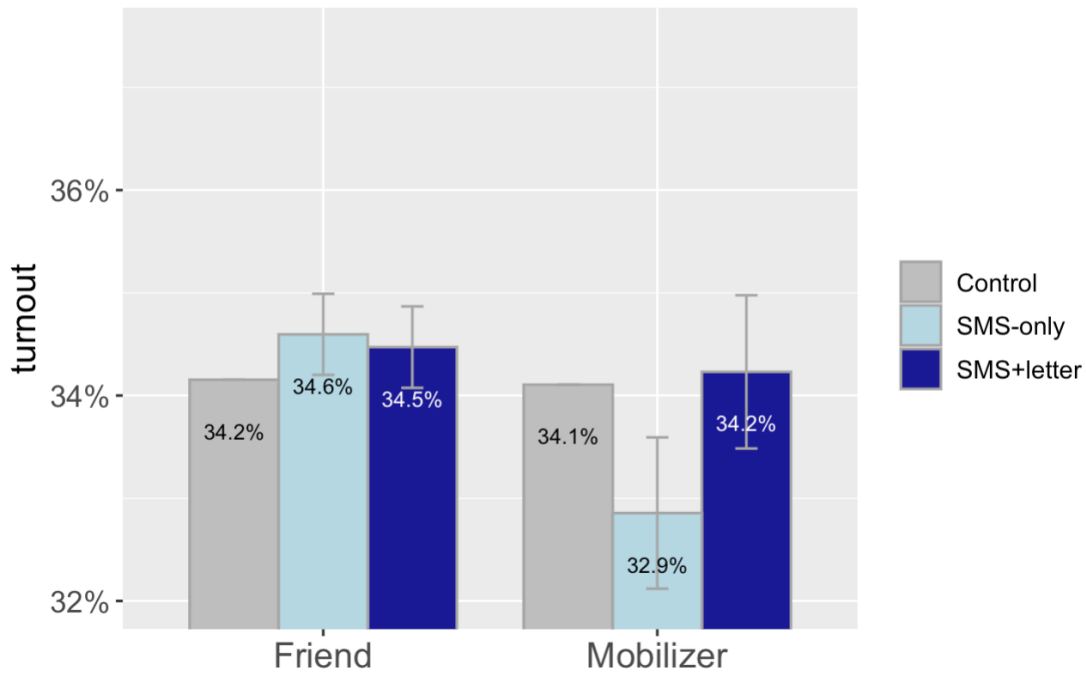
Friends
n=67,880 (after all exclusions)
SMS reminders: +0.3pp
(p=.18 one-tailed, 95% CI = [-0.3, 1.2])
SMS + letter reminders: +0.3pp
(p=.23 one-tailed, 95% CI = [-0.5, 1.1])

Mobilizers
n=17,659 (after all exclusions)
SMS reminders: -1.3pp
(p=.96 one-tailed[1], 95% CI = [-2.7, 0.2])
SMS + letter reminders: +0.1pp
(p=.42 one-tailed, 95% CI = [-1.3, 1.6])

Error bars represent standard error of the relevant treatment effect, so no bar is shown for the control group. For a full explanation of this figure and the accompanying statistics, see the [primary analysis](#) section.

---

[2] If the link requests a password, enter *engine*

Table of Contents

Background

The tactic: high-traffic pledge collection

The purpose of this study was to generate evidence on and definitively test the effectiveness of high-traffic pledge collection (HTPC), for increasing turnout in medium- to high-salience elections.

High-traffic pledge collection is a two-part relational turnout tactic in which a canvasser in a high foot-traffic location (e.g., a college campus, outside a mall) asks passersby to pledge to remind a certain number of friends to vote. This interaction takes place weeks or months before voting begins. People who pledge (dubbed "mobilizers") provide their name, cell phone number, and the first names of their friends. This information is then used by turnout organizations to send the mobilizer text messages closer to Election Day, prompting them to remind their friends to vote.

As a relational program, HTPC has several major strengths:
- It can be carried out far in advance of the usual GOTV period
- It can reach into friend networks that may not be in the usual contact universes
- Mobilizers are very likely to have accurate, up-to-date contact information for their friends
- Friends are much more likely to notice, read, and trust a message from someone they know, compared to a cold contact.

The study

HTPC is a difficult tactic to study because the organization collecting pledges does not know the identity of the friends whom the mobilizer will remind. In this study, we asked mobilizers to match friends to the voter file so that it was possible to examine their voter turnout.

In the treatment group, mobilizers received text message reminders and prompts to talk to their friends, as described above. We compared this turnout to a control condition in which mobilizers weren't reminded of their pledge, and presumably forgot that they had made it in the weeks or months between the canvasser interaction and the start of voting.

This was a cluster-randomized trial (CRT; friends nested within mobilizers) that ran in multiple states (FL, MI, NC, AZ) during the period preceding the midterm elections, from August 25 to October 19.

The study included three treatment arms:
- Control: Mobilizers pledged to remind friends, but were not reminded about their pledge.
- SMS: The standard version of HTPC: Mobilizers pledged to remind friends, and received SMS reminders during early voting and just before election day (text in appendix)

- SMS + Letter: Mobilizers pledged to remind friends and received the same reminders as the SMS arm. Additionally, they received a handwritten letter from our collaborator Vote Forward (text in appendix), in which a volunteer hand-wrote a message describing why they value talking to friends about voting. The letters also contained printed information about voting in the General Election and were mailed out to mobilizers on October 29th.

Additionally, we tested the effects of providing small food incentives, such as donuts or candy bars, to mobilizers. We used stratified randomization to ensure that the treatment arms and the food incentive were orthogonal to each other.

For full details on the HTPC process and how randomization was carried out, please see the pre-registered study protocol in the appendix.

Supporting documents

Note: Not all docs are visible to non-Vote Rev personnel. Contact michael@voterev.org if you would like access.

Trial protocol and protocol addendum - The trial protocol details the intervention design and evaluation design that were pre registered prior to the start of evaluation of the 2022 HTPC LargeCT. The protocol addendum documents any deviations from the trial protocol. It also provides the justifications and context for said deviations

External analysis results[3] - This document is provided by our external evaluators from Dr. Melissa Michelson's lab and details the results of their independent analysis of our trial as well as any deviations between Vote Rev's evaluation results and their own.

Implementation log - This document details all of the things that actually happened during the 2022 HTPC LargeCT implementation including message scripts and message send-dates as well as any issues with implementation that arose.

VCI 2022 Retrospective Report - This report by our Voter Contact Innovations teams has insights on implementing different field programs in 2022, including the HTPC LargeCT. This document contains a lot of learnings about the logistics regarding running a scaled implementation of HTPC.

Data Collection and Reporting Retrospective - This report summarizes the experience and learnings of the non-evaluative data aspects of different field programs in 2022, including the HTPC LargeCT.

---

[3] If the link requests a password, enter *engine*

Sample

Data quality and final dataset criteria

In reviewing our data, we found a wide variety of edge cases, errors, and other anomalies that led to mobilizers or friends not receiving the intended treatment. These arose from three main sources:

1. errors in programming and data management (eg, a small number of mobilizers received messages with the wrong friend names)
2. bugs or quirks in our texting vendor (eg, unpredictable results occurred when two mobilizers shared the same phone number), and
3. inherent complexities in running a networked study with rolling stratified randomization (eg, two different mobilizers might name the same friend, but be assigned to different conditions).

We used the following principles in developing exclusion criteria.

1. **Friends or mobilizers who did not receive the intended treatment should be removed from the data if, and only if, one of the following conditions applies:**
   - the exclusion can be applied systematically across treatment and control groups,
   - the conditions leading to the exclusion all occurred prior to randomization, OR
   - we have good reason to believe the issue occurred at random.

     If none of these conditions is satisfied, then we retain the data for analysis in order to avoid creating group differences that may be correlated with voting behavior.
2. **Each voter should be counted in the analysis only once**, even if present as both a mobilizer and a friend.
3. **We always remove data that is unambiguously invalid** due to being nonsensical (eg, a mobilizer listing themselves as their own friend), corrupted such that it can't be used, or identified as fraudulent by our field team.
   - Records removed in this way do not count towards exclusion criteria for other records. Example: If person *A* appears in the dataset as the friend of one real mobilizer and one fraudulent mobilizer, the fraudulent record will be removed and *A* will not be treated as a multitriplee).
4. When an individual has not been matched to the voter file, we will assume that they are unique from every other individual in the study.

For details on all exclusion criteria used, see the Data quality and inclusion section of the preregistered addendum to the protocol.

Data Acquisition

We sourced our data from TargetSmart. This included voter registration information, race, gender, modeled turnout scores, modeled partisanship scores, and vote history (including 2022 Midterm General Election turnout results).

Sample descriptives

These refer to the final sample remaining after all exclusions.

		Mobilizers	Friends
Total N		17,659	67,880
Treatment condition	% control	49.7%	49.7%
Campus status	% recruited at campus location	33.6%	36.7%
Race/ethnicity[4]	African-American	30.1%	30.9%
	Caucasian	48.0%	49.6%
	Asian	2.7%	2.9%
	Hispanic	13.4%	11.2%
	Other / Uncoded	4.7%	4.4%
	Multiracial	0.7%	0.6%
	Native American	0.4%	0.4%
Gender	Female	52.8%	53.7%
	Male	47.2%	46.2%
	Unknown	0.02%	0.02%
Age	25th %ile	24	26
	Median	32	34
	75th %ile	43	48
Modeled midterm turnout score	25th %ile	20.5	20.8
	Median	49.2	51.3
	75th %ile	73.8	76.5
Partisanship score (higher = more likely Democrat)	25th %ile	48.5	42.6
	Median	84.0	82.4
	75th %ile	94.3	94.1

---

[4] 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	2022 general	33.8%	34.9%
	2020 general	55.9% 56.1% of those old enough to have voted	57.0% 57.2% of those old enough to have voted
	2018 general	28.6% 31.1% of those old enough to have voted	30.8% 33.0% of those old enough to have voted

Table 1: Sample descriptives split by mobilizers and friends

Data exclusions

We excluded mobilizers and friends from our sample for various reasons. Some records could not be used for evaluation purposes because the voter was not matched to the voter file or had already cast a ballot prior the intervention. Some mobilizers had invalid or malformed phone numbers so the mobilizer and their associated friends could not have received the intervention and as result were not used for evaluation. These exclusions were specified in our pre-registered trial protocol. We also added additional exclusions for specific circumstances that we did not foresee. We excluded records with suspicious names as well as mobilizers and friends who were present in the sample multiple times across treatment conditions (multitriplers and multitriplees) or were present as both a mobilizer and as a friend (tripleers).

Exclusion Criteria	Mobilizers	Friends	Total
Full Sample	30,838	126,692	157,530
Exclude unmatched	27,991	121,314	149,305
Exclude already voted	23,711	100,469	124,180
Exclude bad phone	20,013	84,261	104,274
Exclude suspicious names[5]	19,927	83,447	103,374
Exclude multitripler and multitriplee[6]	18,700	71,429	90,129
Exclude tripleers[7]	17,659	67,880	85,539

---

[5] We noticed an unusually large number of names in our sample that start with 'Aa'. This almost certainly represents some kind of unintended behavior by canvassers. We decided that if a person's full name begins with "Aa", and is present in the sample more than five times, all people with that full name will be excluded.
[6] Multitriplers are mobilizers who were present in the sample more than once and multitriplees are friends who were present in the sample more than once.
[7] Tripleers are people who were present in the sample both as a mobilizer and as a friend.

Table 2: Table of exclusion criteria and number of excluded and remaining records

Analyses

Terminology

We refer to both mobilizers and friends as being in the "SMS only" or "SMS plus letter" conditions. Without exception, this refers to *the type of reminder sent from VRAF to the mobilizer* (eg, a friend referred to as "in the SMS only arm" is one whose mobilizer received SMS reminders). Friends never received any communication from VRAF.

Planned analyses

Planned analyses are laid out in the [preregistered protocol](#) and amended in the [preregistered addendum](#) and the [post-registration changes document](#) (all documents are in appendices to this paper). A brief overview:

- Primary analysis: Effect of treatment on voting outcomes. Estimate separate effects for SMS and letter, interacting with role (mobilizer vs friend).
  - Alternate model: predict friend voting rates only, treating mobilizer ID as a random effect instead of a cluster
  - Instrumental variable analysis: Attempt an unbiased assessment of the direct effect of the mobilizer receiving the reminder message, rather than just being assigned to receive it
- Secondary analyses:
  - Interactions between treatment and age, campus, food incentives, and number of matched friends
  - Interactions between treatment and race/ethnicity
- Robustness checks
  - Do not exclude [improbably common names](#) or [duplicate mobilizer phone numbers](#)
  - Do not exclude individuals who voted prior to being included in the study
  - Exclude letter-ineligible individuals

Analyses not run

We originally specified that we would also analyze voting outcomes for household members of friends, in order to check for treatment spillover effects. Obtaining these records would have incurred substantial extra costs, so after finding a null result for a direct effect on friends we chose to assume that the spillover effect would also be null.

Significance testing

We use 1-tailed p values when the independent variable is treatment arm for reminders or treatment arm for Vote Forward mail (treatment higher than control; This was chosen because we were not strongly interested in distinguishing between a tactic with a null effect and a tactic with a harmful effect, since we would not recommend either one). We did not have directional

hypotheses for any other variables or for any interactions between treatment and other variables, so these tests used 2-tailed p values. In all cases, alpha = .10.

Coding for letter arm

An important update to our procedure is described in Denoting letter-ineligible individuals: Mobilizers who could not have received a letter (typically due to not having an address available), and their friends, were coded as "ineligible" for their letter arm assignment, such that their outcomes were used to estimate the effects of SMS vs. no SMS but not the effects of being randomized to receive a letter vs. not.

Covariates

All analyses used the same set of covariates, except when indicated. See Missing covariates for a description of how missing values were handled.
- Role (mobilizer or friend)
- State in which registered to vote
- Food incentives vs not
- College campus vs not
- Demographic variables for both friends and mobilizers:
  - race/ethnicity
  - age and age-squared
  - gender
  - TargetSmart turnout score for midterm elections
  - TargetSmart ideology score
  - Individual's census tract demographics
    - average income
    - proportion with Bachelor's degree or higher
  - Friend and mobilizer voting history in 2018 and 2020 general elections
  - Number of days prior to the election the pledge was collected
  - Canvasser ID[8]
  - Number of matched friends in the pledge

Additionally, standard errors were clustered by mobilizer (ie, a cluster contained a mobilizer and all their friends).

Balance Checks

We tested for balance on covariates using the normalized differences method of Imbens and Rubin (2015) (see Balance check process). We tested for balance between SMS and no-SMS assignments and between letter and no-letter (ignoring letter-ineligible individuals). All

---

[8] Canvasser is conceptually more like a random effect, but due to technological limitations we have not been able to find a way to run mixed-effects models with clustered standard errors. This should not affect the accuracy of our model: The number of canvassers is much smaller than the number of triplers or triplees, and individual canvasser coefficients will not be interpreted as outcomes of interest.

covariates fell well below the threshold of 0.25, with the largest standardized difference being 0.016 (campus vs. non-campus for letter assignment arm).

Primary analysis: Mobilizer and friend voting outcomes

Implementation

Regression with clustered analysis used the R function `miceadds::lm.cluster` (which is built on the `sandwich` package). Significance tests of combined parameters (eg, testing whether the effect of SMS + SMS*mobilizer_status is different from zero) used the function `car::linearHypothesis`.

Results

For friends, we found a nonsignificant effect of reminders on voting rates (+0.3pp), with no meaningful difference between SMS-only and SMS plus letter. For mobilizers, we found a large *reduction* in voting rates (-1.3pp) when receiving SMS reminders only, which entirely disappears when receiving both SMS reminders and a Vote Forward letter.

Figure 2 shows modeled voting rates in each condition. Full parameters and p-values are reported in Table 3. Full model output showing all covariate parameters can be found in an appendix.

Due to our one-tailed testing procedure the reduction for mobilizer SMS-only can't be interpreted as significant, but the p value for the interaction term would be .05 under a two-tailed test and the magnitude is large enough to be of practical importance. See the Discussion section for more on this.
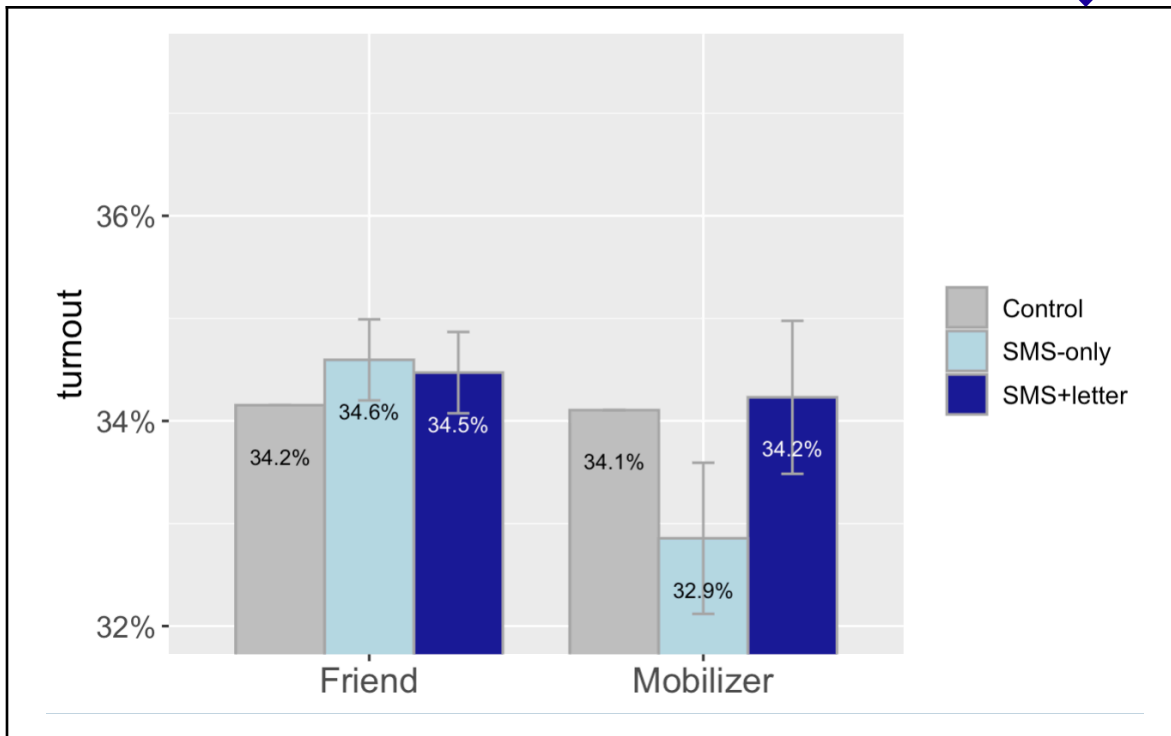
Friends

n=67,880 (after all exclusions)

SMS-only: +0.3pp
(p=.18 one-tailed, 95% CI = [-0.3, 1.2])

SMS+letter: +0.3pp
(p=.23 one-tailed, 95% CI = [-0.5, 1.1])

Mobilizers

n=17,659 (after all exclusions)

SMS-only: -1.3pp
(p=.96 one-tailed[1], 95% CI = [-2.7, 0.2])

SMS+letter: +0.1pp
(p=.42 one-tailed, 95% CI = [-1.3, 1.6])

Figure 2: Voting rates by condition. To equalize effects of covariates across groups, values for each group are predicted from the model based on actual covariate values in the full sample. Letter-ineligible individuals were excluded.

Error bars represent standard error of the relevant treatment effect, so no bar is shown for the control group.

	total treatment effect	Regression coefficient	SEM (coefficient)	p (total effect)	p (coefficient)
SMS only - on friends	+0.3pp	.003	.004	.18	.37
SMS with letter - on friends	+0.3pp	-.001	.005	.23	.91
SMS only - on mobilizers	-1.3pp	-.016	.008	.96[1]	.05
SMS with letter - on mobilizers	+0.1pp	.015	.010	.42	.13

Table 3: Analysis results by group. The first column ("total treatment effect") shows the combination of regression parameters relevant to that group (for friends, the difference from control group friends; for mobilizers, the difference from control group mobilizers). The remaining columns show the unique regression parameter for that group.
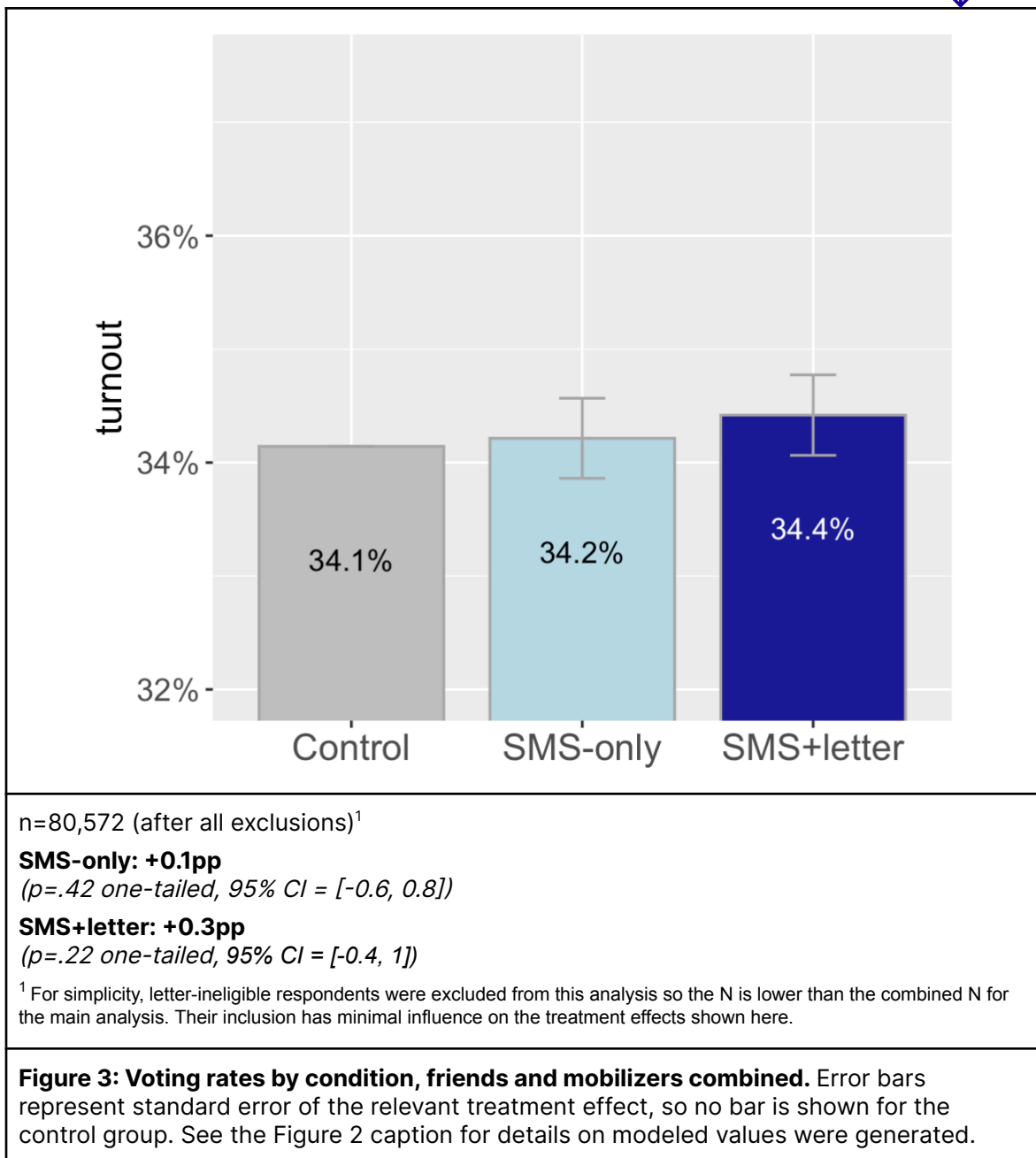
For example, in the row labeled "SMS with letter - on friends" the first column shows that the estimated effect for SMS and letter combined is +0.3pp. The second column shows that the unique coefficient for adding letter to SMS is very small, as evidenced by the fact that the total treatment effect is nearly the same as in the row labeled "SMS only - on friends".

[1] Because we prespecified a one-tailed test with a positive effect for SMS, the observed negative difference between groups is assigned a very large p value instead of a small one. Later in the paper we discuss the potential meaning of a negative effect for SMS reminders (but not SMS reminders plus a letter) and the evidence for and against this being a causal effect.

Effects and total impact

The possibility of a negative mobilizer effect could create concerns about further testing of this tactic. Figure 3 combines the friend and mobilizer results from the previous figure and shows that the overall best estimate of the intervention's impact is essentially zero – because each mobilizer affects ~4 friends, the nonsignificant positive friend effect becomes large enough to balance out the negative mobilizer effect.

This also indicates that the RCT itself generated essentially zero votes.

n=80,572 (after all exclusions)[1]

SMS-only: +0.1pp
(p=.42 one-tailed, 95% CI = [-0.6, 0.8])

SMS+letter: +0.3pp
(p=.22 one-tailed, 95% CI = [-0.4, 1])

[1] For simplicity, letter-ineligible respondents were excluded from this analysis so the N is lower than the combined N for the main analysis. Their inclusion has minimal influence on the treatment effects shown here.

Figure 3: Voting rates by condition, friends and mobilizers combined. Error bars represent standard error of the relevant treatment effect, so no bar is shown for the control group. See the Figure 2 caption for details on modeled values were generated.

Independent reanalysis

We worked with an independent academic research team, including Prof. Melissa Michelson (Menlo University) and Dr. Stephanie DeMora (University of Pennsylvania) to replicate our analysis. Working from the analytic strategy described in the preregistered protocol, they coded their own analysis and reached independent conclusions about the outcomes. Their report can be viewed here. Their effect estimates show some small differences from the ones reported in this paper, caused by different choices about how to carry out the analysis protocol and test total combined effects. The pattern and significance of results is exactly the same as those presented here.

Note that the independent analysis used a dataset that we had already processed and cleaned in accordance with the protocol described in this document. Therefore, their work reinforces the appropriateness and accuracy of our hypothesis testing but does not protect against any technical or conceptual errors in our implementation of the study or our data processing and validation. Any such errors would remain the responsibility of Vote Rev Action Fund.

Random effects variant

We estimated the effects of SMS and SMS+letter reminders on friends only, using a model without clustered errors and instead treating mobilizer ID as a random effect. Parameter estimates were almost identical to those from the main analysis.

Robustness Checks

In order to check the robustness of our results, we ran multiple variations of our analysis using slightly altered exclusion criteria, person type assignments, or added interactions.

- Mobilizers, friends, and friends whose mobilizers had suspicious names were not excluded
- Mobilizers with duplicate phone numbers and friends whose mobilizers had duplicate phones not excluded
- Mobilizers and friends who voted prior to the intervention not excluded
- All letter 'ineligible' people excluded
- Tripleers not excluded
- All tripleers excluded
- All tripleers are evaluated as 'triplers/mobilizers' instead of 'triplees/friends'

These alterations did not meaningfully change the results of our analyses. All of them showed small increases in the voting effect for friends, and in the version without exclusions for duplicate mobilizer phone numbers the friend effect became significant (+0.46pp, p=.0095). Mobilizer effects were unchanged. The increased friend effect was counterintuitive because we suspected that highly duplicated phone numbers were bogus and would result in the intervention not actually being delivered. In any case, the change is consistent with random noise and we do not believe it indicates issues with our main analysis.

Instrumental variable analysis

We created an instrumental variable model by using actual message delivery (the putative mechanism for any effects) as predicted by condition assignment. Parameter estimates were almost identical to those from the main analysis, except that the negative interaction between SMS treatment and mobilizer status increased from -1.3pp to -1.8pp.

Postdiction analysis

At the recommendation of a colleague we ran an unregistered postdiction analysis, replicating the main analysis but replacing 2022 turnout with 2018 or 2020 turnout. If treatment assignment predicts voting behavior prior to when the study was run, this suggests that post-treatment differences in voting behavior may have been caused by a failure of random assignment.

This analysis used all exclusion variables and, to support comparison between years, included only individuals who were old enough that they were eligible to vote in 2018[9]. All three of the models (2018, 2020, and 2022) removed past voting behavior as a covariate, as well as removing modeled turnout score because it is partly based on past voting behavior.

Results are shown in Figure 4. Friends showed similar results in all 3 years: An increase of 0.5-0.7pp in the SMS-only condition (p<.1, one tailed, for 2018 and 2020), while the effect of adding a letter remained small.

Mobilizers showed a larger effect for 2022 than in the main analysis, due to the slightly different sample used. The negative SMS*Mobilizer parameter increased to -2.1pp, and the positive Letter*Mobilizer parameter increased to 2.2pp, so that the net effect of SMS+letter for mobilizers remained near zero. For 2018 they showed the same pattern, with a smaller magnitude (SMS*Mobilizer: -1.1pp, Letter*Mobilizer: +0.7pp). However, for 2020 these parameters were both ~-0.5pp and far from significance.
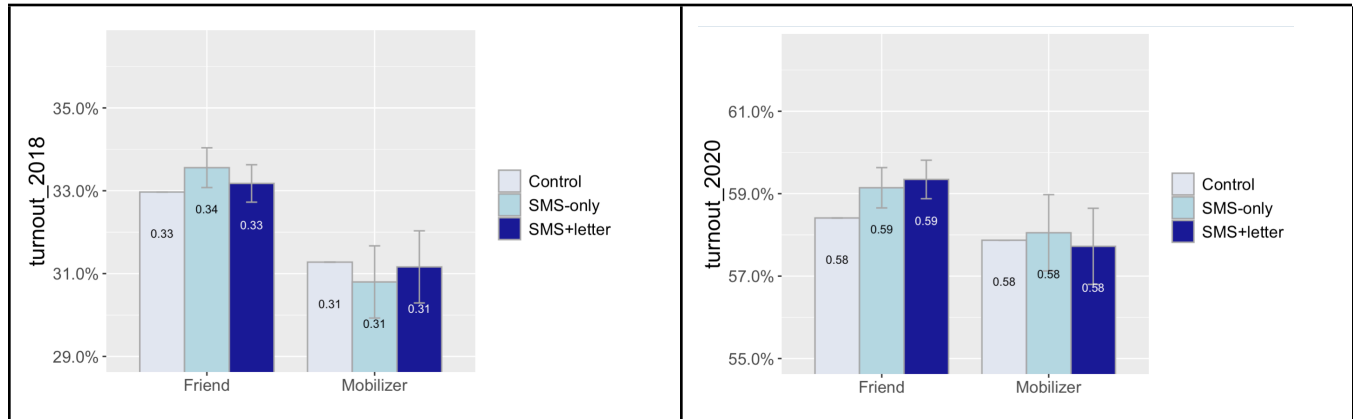

Figure 4: Postdiction analysis. Modeled voting outcomes for 2018 and 2020, following the same method as Figure 2. Error bars represent standard error of the relevant treatment effect, so no bar is shown for the control group. We do not present p-values for these analyses because they are post-hoc and highly speculative.

The 2018 results suggest that **at least some portion of the observed group differences, for both friends and mobilizers, may have been due to failures of randomization and not to an actual treatment effect.**

Note that the main analysis should have controlled for any randomization failures, by including a covariate for 2018 voting. This raises the possibility that there was both a randomization failure *and* a true negative effect of receiving SMS reminders only. However, a pre-existing trait difference that influences voting behavior would only be partially controlled for by any particular year's voting behavior, and so it is still possible that a failure in initial randomization accounts for a large portion of the observed negative effect.

---

[9] Note that this also leads to estimates for 2022 voting that are slightly different from the ones reported in the main outcomes analysis.

Secondary Analysis: Differential effectiveness

Because a large chunk of our sample was recruited from college campuses, we wanted to explore the particular impact that the interventions may have had on voters of different ages as well as voters in our sample who were recruited on campus. In order to investigate this, we ran a version of the analysis with additional interactions for both age and campus status (referring to whether the mobilizer was recruited on a campus or not).

Age

Age interaction effects are visualized in an appendix. Interactions between age and SMS reminders did not approach significance. For friends, they directionally suggested a stronger effect for older friends, with peak effectiveness around age 65. For mobilizers the interaction effect sizes were much smaller.

Interaction effects between age and letter reminders did not approach significance for friends or mobilizers.

Person type	Interaction	Interaction effect	P Value
Friend	Treatment x Age	+0.1pp	0.23
Friend	Treatment x Age sq	-1.1E-5pp	0.40
Friend	Letter x Age	+0.1pp	0.50
Friend	Letter x Age sq	-1.8E-5pp	0.24
Mobilizer	Treatment x Age	+0.2pp	0.41
Mobilizer	Treatment x Age sq	-3.5E-5pp	0.24
Mobilizer	Letter x Age	+0pp	0.97
Mobilizer	Letter x Age sq	+9.8E-6pp	0.78

Table 4: Table of interaction effect sizes (not combined effects) of SMS and letter conditions interacted with age and age squared for friends and mobilizers

Campus recruitment

When using an interaction term to split treatment effects by campus status, some large differences emerged (see table 5). This analysis was run with and without age included as a covariate, and the results were nearly identical, suggesting that campus status is not simply a proxy for age; it is either a primary driver of the effect or a proxy for some other difference. Results reported here, for simplicity, were calculated for mobilizers and friends using separate models and do not include age interactions.

We found that for friends, SMS reminders did not have a significant effect, just as in the overall model without campus status. When a letter was added to SMS reminders, the total treatment effect became large and significantly positive for non-campus friends and large and significantly negative for campus friends. For mobilizers, SMS reminders alone had a large negative effect, just as in the model without campus status. When a letter was added to the SMS reminders, the

treatment effect became large and positive for non-campus mobilizers, but did not change for campus mobilizers. In other words, the letter appeared to have opposite effects depending on campus status: Positive for mobilizers and their friends recruited in non-campus areas, and negative or null for those recruited in campus areas.

One speculative explanation for the campus mobilizer null effect is that deliverability may have been poor for mobilizers living on campuses: dorm addresses might be inaccurately recorded in the voter file; students might still be listed at their parents' address and receive mail inconsistently; or they may move frequently and have other old addresses in the file. This doesn't explain why the effects of a letter might be negative for their friends, but they make it more plausible that the true effect for friends is zero and the dip into negative values was a fluke.

In all, these results suggest that the Vote Forward letter may have had a large and meaningful positive effect on voting rates for non-campus mobilizers and their friends.

Person type	Campusness	Interaction	Total treatment effect	P Value
Mobilizer	Campus	SMS only	-1.4pp	0.14
Mobilizer	Campus	SMS + letter	-1.5pp	0.13
Mobilizer	**Non Campus**	**SMS only**	**-1.2pp**	**0.09**
Mobilizer	**Non Campus**	**SMS + letter**	**+1.0pp**	**0.15**
Friend	Campus	SMS only	+0.4pp	0.27
Friend	**Campus**	**SMS + letter**	**-0.9pp**	**0.10**
Friend	Non Campus	SMS only	+0.3pp	0.28
Friend	Non Campus	SMS + letter	+0.9pp	0.03

Table 5: Table of total treatment effects for mobilizers and friends in both SMS and letter conditions interacted with campusness

Race

We ran the main model with added interactions for an individual's race/ethnicity.[10] This was framed as an exploratory analysis as we did not expect to have large enough sample sizes to achieve significance, especially with adjustment for the large number of individual race x treatment parameters tested. Some interaction effects are of notable size and have a $p < 0.1$; we discuss these below but also acknowledge a strong likelihood that at least some occurred by chance.

---

[10] This analysis uses the voter file categories, which treat "Hispanic" as a category exclusive with options for race. For simplicity we will refer to this variable as "race" throughout.

First, the dip in turnout among treatment mobilizers who did not receive a letter was largely limited to white mobilizers. Caucasian mobilizers make up the largest share of the sample (approximately 48%) and show a dip in turnout almost double that of the next largest race group (African-American which makes up around 30% of the sample). This is not conclusive and we continue to interpret our results as if there were a dip for mobilizers regardless of race (see discussion). However, we do take this as weak evidence reassuring us that the tactic does not have disparate negative effects on voters of color.

Second, we should keep an eye on Hispanic voters in future relational tactics similar to HTPC. Even though Hispanic voters made up less than 15% of our sample, the interaction effect for Hispanic friends suggested a large increase in voting rates and a comparatively small dip in the SMS-only condition. Hispanic mobilizers also showed increases in turnout for those in the letter condition.

Person type	Race Group	Treatment Group	Total treatment effect	P Value
Friend	Caucasian	SMS only	+0.3pp	0.31
Friend	**African-American**	SMS only	+0.9pp	**0.07**
Friend	**Hispanic**	SMS only	+1.2pp	**0.11**
Friend	Asian	SMS only	+0.1pp	0.49
Friend	Caucasian	SMS + letter	+0.1pp	0.43
Friend	African-American	SMS + letter	+0.4pp	0.28
Friend	**Hispanic**	SMS **+ letter**	**+1.8pp**	**0.04**
Friend	Asian	SMS + letter	+1.1pp	0.31
Mobilizer	**Caucasian**	SMS only	**-1.4pp**	**0.05**
Mobilizer	**African-American**	SMS only	-0.8pp	0.19
Mobilizer	Hispanic	SMS only	-0.4pp	0.35
Mobilizer	Asian	SMS only	-6.7pp	0.23
Mobilizer	Caucasian	SMS + letter	-0.1pp	0.47
Mobilizer	African-American	SMS + letter	+0.2pp	0.42
Mobilizer	**Hispanic**	SMS + letter	**+1.6pp**	**0.09**
Mobilizer	Asian	SMS + letter	+0.9pp	0.35

Table 6: Table of treatment effects interacted with race for friends and mobilizers

Discussion

Effects of treatment

The main model results suggest that SMS reminders reduce voting for mobilizers and provide preliminary but unreliable evidence that they may increase it for friends. SMS reminders plus a handwritten reminder letter sent to the mobilizer appear to eliminate the reduction in voting for the mobilizer while maintaining the possibility of a beneficial effect on friends. These outcomes are unchanged under a variety of alternate analysis methods and robustness checks.

Letter mobilizer effects

We hypothesized that adding a Vote Forward letter to SMS reminders might improve voting outcomes, and this appears true for mobilizers: The SMS-only condition voted less than control, while the SMS-plus-letter condition showed no decrease. This could indicate that the letter has a large positive effect and SMS reminders have an equally large negative effect, or that the letter does not independently affect voting but does somehow neutralize the apparent negative effect of SMS. Because we did not anticipate this pattern of results, and because we were concerned that letters might in some cases not reach mobilizers in time, we did not include a letter-only condition that would have allowed us to tease apart the different possibilities.

Interaction analysis suggests that the positive effects of the letter – including its ability to undo the apparent negative mobilizer effect – occurred only for off-campus mobilizers and their friends. Future studies should test handwritten letters further, likely concentrating in non-campus populations.

Measurement validity

Vote Rev's past research on relational tactics were predominantly SMS-based. In these studies we were able to measure effects on mobilizer voting, since mobilizers are usually targeted based on voter file records. We were not able to measure the full impact on friends' voting, because mobilizers selected their own friends and we usually did not know who they were. An important strength of this study is that we were able to measure the effect on all friends and mobilizers.

However, this study was not able to measure the true counterfactual effect of a HTPC canvassing interaction, because all mobilizers spoke to a canvasser and named friends they pledged to remind. This was necessary in order to construct an equivalent control group: Among mobilizers, general population voting rates couldn't be compared to voting rates among people who self-selected into stopping to talk to a canvasser. Among friends, the people a mobilizer was willing to talk to about voting couldn't be compared to either the general population or a list of friends generated in any other way. Therefore, this study measured **the effect of the mobilizer being reminded about their relational turnout pledge, versus not being reminded.** It is very likely that some control group mobilizers remembered, or were primed, to encourage their friends to vote anyway. This would reduce the observed treatment-control difference and could

produce false null results when the experiment is interpreted as testing the full effect of running an HTPC program. However, we did not expect this reduction to be large a priori, and so we consider it at most a small contributor to the lack of a significant effect in this study.

Pre-existing differences

A postdiction analysis using voting behavior prior to treatment provides uncertain evidence that both friend and mobilizer effects could be due to a pre-existing difference between conditions. This could be due to a technical error that assigned mobilizers recruited at different times or places unevenly between conditions; there is no direct evidence for that but our data pipeline did suffer a variety of technical snags (discussed further in our internal postmortem documents). The pre-existing differences could also be due to random chance. We do not believe any intentional features of our experimental design could have produced this bias.

Negative mobilizer effects

VRAF, and various advisors we have consulted, find it plausible that many tactics are ineffective but less plausible that a seemingly benign tactic would have a real negative effect on voting rates. It's possible to invent explanations for the negative mobilizer effect, such as:
- A moral licensing effect (I've fulfilled my civic duty by reminding others, so I don't need to go to the trouble of voting)
- A negative emotional reaction to the SMS reminders
- Unsubscribing from SMS reminders and feeling inspired to then unsubscribe from other groups' reminders as well.

However, these explanations – and all others we've been able to imagine – have the following weaknesses:
- They were created post hoc and were not considered plausible before seeing the result
- There's no apparent reason they would be entirely negated by the handwritten reminder letter
- They apply equally well to other relational and/or SMS-based tactics, but have not been observed in other studies

Also note that *all* mobilizers experienced the in-person parts of the tactic: speaking to a canvasser, thinking of friends to remind, and matching themselves and the friends to the voter file. Any explanation for a control-vs-SMS difference must be based on only experiences that took place after that: Receiving a welcome SMS message, receiving a reminder message weeks later, and (possibly) reminding friends to vote.[11]

During the Georgia Senate runoff election, before results of this trial were available, we ran a trial of High-traffic on-the-spot relational turnout (HTOTS). HTOTS is a similar site-based tactic that

---

[11] Internal VRAF users can consult a more detailed breakdown of the differences between treatment and control mobilizers' experiences

involves sending reminder messages immediately instead of after a delay. HTOTS increased friends' turnout but the design of the study made it impossible to measure mobilizer turnout[12,13].

Between the reasoning above and the postdiction results suggesting that findings may be partially spurious, Vote Rev will proceed under the assumption that SMS treatment had no effect on mobilizer voting rates. However, we take the possibility of a negative effect seriously and future studies of both HTPC and HTOTS must prioritize measuring mobilizer effects to ensure we detect and limit any harms that do exist.

Future work on HTPC

Based on positive results from our study of on-the-spot relational reminders, and other research on relational turnout, we continue to believe that a relational pledge collection tactic can be effective. Furthermore, there would be major benefits to the progressive movement if we could develop a tactic that can supplement door-to-door canvassing during the months prior to GOTV, in order to put early money and volunteer hours to better use. Therefore, we intend to continue investing in design and piloting research to develop a version of HTPC that is more effective than the one studied in this RCT.

Areas for improvement

We currently speculate that the tactic we tested was ineffective because the canvassing interaction didn't create sufficient motivation for mobilizers to follow through on reminding friends, or to increase their own propensity to vote. Our subsequent design work has focused on 1) increasing initial mobilizer commitment and 2) strengthening the canvasser-mobilizer relationship, so that the reminder messages have more impact. We've done preliminary testing on ideas including 1) a commit-to-vote style pledge card that we mail back to the mobilizer; 2) asking mobilizers to affirm their commitment in a text message sent shortly after their pledge; and 3) sending other communications to "keep the relationship warm" during the time between the pledge and the election. See this presentation (Analyst Group membership required) for some preliminary results.

We attempted to validate our hypothesis that mobilizers weren't following through on reminding friends by sending out phone and text message surveys shortly following the election. Response rates were extremely low (well under 1%) which we take as an indicator of 1) low mobilizer commitment and 2) general exhaustion and burnout following a high-salience election with heavy political text messaging.

---

[12] The internal-only treatment/control differences document linked above also diagrams divergences between the HTPC and HTOTS experiences.

[13] The HTPC study relied on the long delay between providing friends' names and being able to remind them to vote, which plausibly gave mobilizers time to forget about their pledge. HTOTS took place during early voting and so it would have been impossible to ask mobilizers to provide the names of friends they would remind but not remind them. In the future we may be able to study HTOTS with other approaches that do allow us to measure mobilizer effects.

Operational learnings and observations for researchers

This RCT was a massive logistical and design achievement, with over 9,200 canvasser hours across multiple paid canvass teams in AZ, MI, FL, and NC. Canvassers collected 32,071 pledges at ~3.5 pledges/hour, demonstrating good cost efficiency. We dramatically improved our canvassing best practices (e.g., using food incentives) and canvasser management capacity, which was crucial to our ability to rapidly implement a successful test of HTOTS when the GA Senate runoff was announced.

We also developed an extensive set of automated and manual review techniques for fraud detection. Even though canvassers were not paid per pledge or given quotas, there were still many instances of canvassers not working at all but entering false pledges. In one case we detected an organized group of canvassers who were all skipping work together while entering hundreds of fabricated pledges. Organizations running similar projects are welcome to reach out to discuss our methods further.

The large-scale voter file matching required for the HTPC RCT allowed us to demonstrate that our site-based relational tactics reach a diverse and high-potential group of voters. The median friend reached by HTPC was 34 and median midterm turnout score was 51.3; they were 50% white, 31% Black, and 11% Hispanic. A subsequent analysis by the Democratic Data Exchange indicated that HTPC friends and mobilizers were considered valued targets by the progressive ecosystem in general and often difficult to reach by other means.

Other learnings included
- Mobilizers are willing to identify friends in the voter file, which unlocks the ability to test all kinds of site-based relational tactics
- Canvassers found HTPC to be a positive canvassing experience, suggesting that it has high mainstreaming potential in both paid and volunteer contexts.

We also took some cautionary learnings from the issues with our data pipeline, which proved to be overly complex and prone to serious technical failures. We were not able to successfully carry out batched randomization with balance checks and automatic processing of network crossover issues (such as mobilizers receiving reminders from other mobilizers). In the future we will work with our vendors to enable immediate random assignment that feeds back to a shared database, so that people who are enrolled repeatedly are always assigned to the same condition (in the HTOTS study we were able to accomplish this by pre-randomizing the full voter file for the state).

Acknowledgements

This project was a massive undertaking, beginning with a pilot in the Virginia 2021 Gubernatorial election and culminating in our 2022 LargeCT. Vote Rev Action Fund gives our thanks to the following colleagues, collaborators, and friends:

- Our canvassers and canvassing team leads for their inventiveness, tenacity, and commitment in persuading thousands of ordinary people to become relational mobilizers. Michigan canvassing lead Donyetta Hill built an especially amazing team that tested out many tactics for attracting mobilizers, including shark costumes!
- The Vote Rev Action Fund 2022 Voter Contact Innovations field team: Chris MacNeil, Royal Shepherd, Donna Davis, and Derek Giamundo
- Our 2021 pilot study field director, Laura Leavitt
- Our Vote Forward research collaborators: Emily Wasserman, Phong Le, and former Executive Director Scott Forman
- Vote Rev Research Team alumni Duja Michael and Pedro Alberto Antenucci Benesch
- The Analyst Institute, especially former AI researchers Saul Cunow and Juan Machado for methodological discussion and Deputy Director of Research Jeff Ferguson for help with analysis methods and data interpretation
- Professor Melissa Michelson and team for methodological advice and review
- Dr. Stephanie DeMora for conducting an independent replication of our analysis
- Dr. Kate Duch for methodological advice during our initial pilot study
- The anonymous donors who supported every stage of this work

Appendix: Initial preregistered study protocol

Trial Info

Date & Version: Preregistration, August 30th, 2022
Team: Research
Partners: Vote Forward (mail reminders), Professor Melissa Michelson (external evaluator)
Vendors: Revolution Field Strategy (canvassing firm - MI), La Machine (canvassing firm - AZ), GRSG (canvassing firm - FL, NC), Grassroots Unwired (canvassing app), SMS texting platform TBD
Funding: Paid for by the Vote Tripling Action Fund, with c3 funds
Partisanship: Non-partisan voter turnout work
Target election: 2022 midterm elections in FL, MI, NC, and AZ.

PICOS Statement

Trial Launch Date:	Pledge collection begins August 25, 2022 and ends Oct 21, 2022. For pledge collection, no differential treatment between conditions occurs until the first reminders are sent out (see [Reminder process](#))
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: Who are the participants?	**Triplers**[14] are adult passersby in high-traffic places in our chosen states who take a pledge to encourage 5 people they know to vote in the upcoming election. **Triplees** are the friends[15], family members, or other contacts whom triplers will reach out to. Triplers select their own triplees.

---

[14] 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".
[15]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 co-residents such as spouses and adult children.

Sample details	See section [07 / Sample & Setting](#) for details on sample size.
Intervention: What are we doing?	We will recruit triplers, and triplers will encourage[16] triplees to vote in the upcoming midterm election[17]. Triplers will be randomly assigned to receive reminders of their pledge prior to election day, or to receive no reminders of any kind.
Intervention details	Triplers in the treatment group are sent text message reminders to encourage their triplees to vote. Half of the treatment group triplers will also be sent a handwritten letter to remind them to encourage their triplees to vote (see the full [Reminder process](#)). Triplees are not contacted by study staff, only by triplers.
Comparison:	There are two main comparisons that we are interested in: 1) The turnout rate of control triplees (whose triplers receive no reminders to encourage them to vote) compared to that of treatment triplees (whose triplers are sent reminders, either via text message only or text message and handwritten letter). 2) The turnout rate of control triplers compared to that of treatment triplers.
Comparison details	Constructing a list of triplees requires asking triplers to name and voter-file-match people they will encourage to vote. This means that even without reminders, control group triplees might receive encouragement to vote from their triplers. As a result, the contrast between the treatment and control arms will actually tell us whether sending triplers reminders about their pledge increases voting rates. This means that the effect of HTPC that we estimate in this trial will likely be an underestimate of the true effect of HTPC, but a significant difference between the groups provides definitive evidence that HTPC increases voting rates above the counterfactual baseline.

---

[16] Terminology can become confusing because there is a chain of multiple communications. The initial interaction between the canvasser and the tripler is a "**pledge**." The message that study staff send to the tripler is a "**reminder**" (reminding them to talk to their triplees). Any message that the tripler then sends to their triplees is "**encouragement**" (encouraging them to vote).

[17] For triplers, the intervention is being reminded of their pledge, while for triplees, the intervention is having their tripler reminded to encourage them.

Outcome(s):	Primary outcomes: 1) voter turnout rate for triplees in the upcoming midterm elections; 2) voter turnout for triplers in the midterms.
Setting:	The cluster randomized trial (CRT) will take place in Florida, Michigan, North Carolina, and Arizona[18] prior to the 2022 midterm general election.

Contents

Background and Intervention

01 / Purpose

The purpose of this study is to generate evidence on and definitively test the effectiveness of High-Traffic Pledge Collection vote tripling (HTPC) as a tactic for increasing turnout in medium to

---

[18] A large majority of the population in Arizona votes early or by mail, which starts on October 12th. Pledge collection in the state will end Oct 11th to avoid encountering/getting pledges from triplers who have already voted (or whose triplees have voted).

high salience elections. Below we provide an overview of how HTPC works in "real life", outside of an evaluation context. In order to be able to evaluate HTPC, we have to adapt the tactic; see more detail under Intervention Design.

> ### What is High-Traffic Pledge Collection?
>
> High-Traffic Pledge Collection vote tripling is a two-part tactic in which a canvasser in a high foot-traffic location (e.g., a college campus, outside a mall) asks voters to pledge to remind a certain number of friends to vote. Voters who pledge (dubbed "triplers") fill out pledge cards containing their name, number, and the first names of their friends. This information is then used by turnout organizations to send the tripler text messages and/or mail closer to Election Day, prompting them to remind their friends to vote.

The CRT will run in multiple states (FL, MI, NC, AZ) during the period preceding the midterm elections. Two enhancements to the HTPC tactic will be tested: First, some locations will provide food incentives, such as donuts or candy bars, to pledgers. Second, volunteers managed by our partner Vote Forward will send handwritten reminder letters to some triplers reminding them to encourage their triplees (see Reminder Process for more details). We will use randomization with stratification to ensure that these enhancements are evenly distributed between the two treatment arms and do not invalidate the primary comparison, between triplers who received any reminder and triplers who received none.

This study is powered to definitively test HTPC, but the analysis of add-ons (food incentives and handwritten letters) will be exploratory.

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).

Previous attempts at evaluation have relied on tracking a tripler's household's turnout to measure impact. However, we know that when triplers select triplees, they only select co-residents around 10% of the time. Thus, the impact on co-residents who received a relational contact is watered down by a large number of co-residents who did not.

In this CRT, instead of relying on housemate turnout, we will ask for more detailed information on triplers and triplees so that we can match them back individually to the voter file. This allows us to directly track the turnout of the intended intervention recipients.

03 / Intervention Design

Paid canvassers will collect vote tripling pledges at high traffic locations such as college campuses, shopping centers, and DMVs. They will operate in large cities (FL- Tampa, MI - Detroit and Ann Arbor, NC - RDU area, AZ - Phoenix/Tempe/Tucson) in the two months prior to the 2022
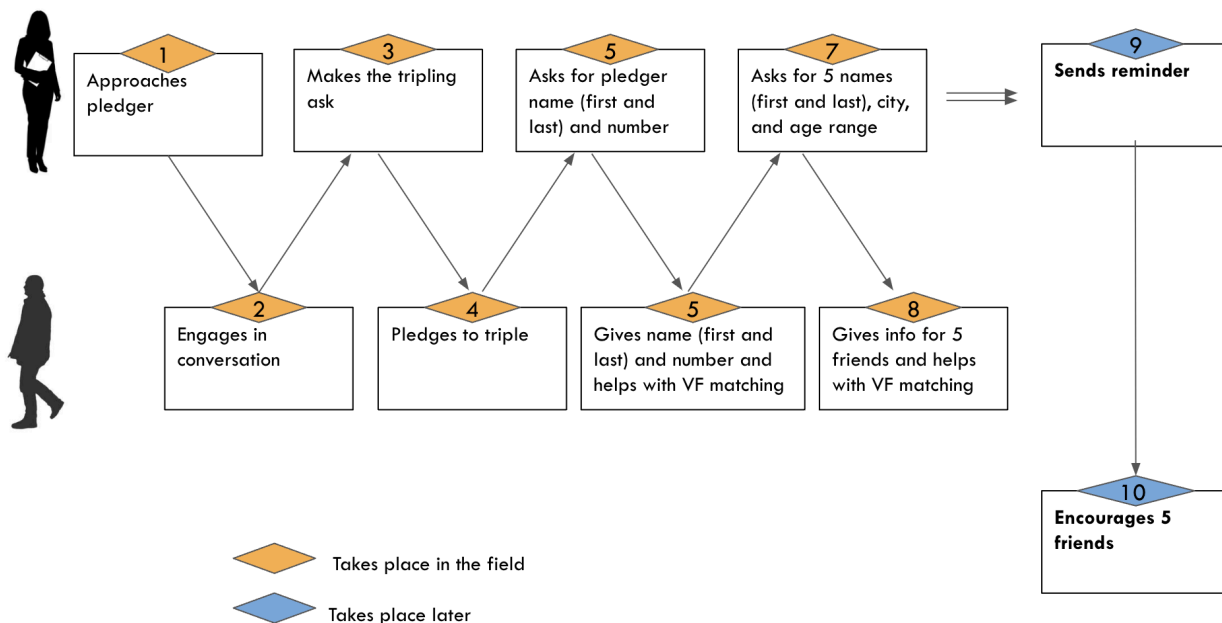
general elections (see Timeline). Canvassers will be recruited and managed by our canvassing vendors in each state.

In the "real life" version of HTPC (see above) that will be used by other organizations, canvassers would typically use a paper pledge card to collect the tripler's phone number and the first names of their triplees, which are used to send a personalized SMS reminder just before the election. For this trial, the canvasser will collect data on their phone and will ask the tripler to help identify themselves and their triplees using a voter file lookup app (Grassroots Unwired; GRU). This will allow us to look up voting records after the election and test for differences between treatment and control arms.

The voter file matching process was tested in pilot studies during the primaries (MI and FL) and we were able to match over 70% of individuals to a unique record in the voter file. In cases where the tripler declines to help with voter file matching (either due to time or to privacy concerns), canvassers will manually record the required information and we will later attempt to match triplers and triplees to the voter file (see Appendix: Linking individuals to the voter file).

Triplers in the treatment group will receive messages reminding them to encourage their triplees to vote (for messaging schedule and details, see section 04 / Study Design). This friend-to-friend encouragement is the proximate mechanism for affecting voter turnout.

Figure 1. The step-by-step process of collecting pledges

CONDITION	DESCRIPTION
Overall sample	Triplees who were matched uniquely (1:1) to the voter file.

Control	Triplees whose tripler does not receive a reminder to encourage them to vote. Note that these triplers *did* make a pledge to encourage their triplees, they just won't receive a reminder to do so.
Treatment	Triplees whose tripler receives reminders to encourage them to vote

Evaluation Design

04 / Study Design

This will be a multi-site cluster randomized controlled trial: in each state, triplers will be randomly assigned to treatment or control, and this assignment will affect all of their triplees. Assignment is not carried out at the triplee level because of the high risk of spillover (even if we only remind a tripler to encourage friend #1 to vote, that may cause them to remember that they also wanted to encourage friends #2- #5).

Food incentives

Throughout the period of pledge collection, we will offer free food (such as tacos, donuts, or ice cream) in some locations in Michigan to people who agree to encourage friends to vote and fill out the pledge form. This does not change the nature of the treatment, but because it may attract different populations or change the feel of the canvasser interaction, we will differentiate these pledges in our analysis (see [Analytical Strategy](#)).

Canvassing sites

Canvassing sites, including those that will provide food incentives, will be chosen by the canvassing firms based on local area knowledge and in collaboration with VRAF's field team, with a substantial proportion taking place on college campuses. 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.

Reminder process

All treatment triplers will be sent a series of SMS text messages reminding them to encourage their friends to vote. Half of the treatment triplers will be randomized to also receive handwritten letters reminding them to encourage their triplees to vote[19]. The handwritten letters will be composed and sent out by volunteers with Vote Forward, a nonprofit collaborating with VRAF on

---

[19]Due to the logistics of randomization, possible schedule changes, and sending the letters by post, it is possible that some triplers who pledge late will not be sent letters on time before Election Day. In this case, we will use ITT (intent to treat) analysis.

this trial. Control group triplers will neither receive SMS texts nor handwritten letters. <mark>The text of the messages below is not final.</mark>

Timing	Message type	Possible content
1-6 days after pledging (based on how quickly randomization can be implemented)	SMS: Initial confirmation	Hi {{recipient_firstname}}! It was great meeting you on {{pledge_weekday}} + thank you for agreeing to remind {{triplee_string}} to vote in [STATE]'s midterm election! [STOP to opt out] Want to learn more about the election? -VoteRevActionFund
1 day before end of early voting	SMS: Early voting reminder	Hi {{recipient_firstname}}! It's VoteRev following up. Tomorrow, [INSERT EV end DAY & DATE], is the last day to vote early or request an absentee ballot. Right now, can you remind {{triplee_string}} to vote? Text back an emoji to confirm!
1-2 days before election day	SMS: Election day reminder	Hi {{recipient_firstname}}, Election Day is tomorrow - Tues 11/08. Will you ask {{triplee_string}} to vote before polls close at 8pm? [For NC and MI:] If you're not registered, you can register right at the polls!
12-28 hours after prior message	SMS: Followup	Just checking in, did you get a chance to ask {{triplee_string}} to vote?
Sent: Oct 31 Received: ~1 day before election day	Postal mail: written reminder	Dear {name} Thank you for pledging to remind **{triplee_string}** to vote in the **midterm election** in a recent conversation with a volunteer about voting! Will you get in touch with them right now? **Reminders make a big difference** especially when they come from someone we know and trust - and it only takes a minute! I volunteer to write letters like this because [volunteer handwrites personal message here]

		Important dates: ● The last day to vote early is **[enter state last day]** ● Vote on election day: **Tuesday, Nov 8** before [enter polling place closing time in state] Thank you for being a voter, and for encouraging **{triplee_string}** to join you at the polls! Sincerely, [Volunteer name]

Potential issues
- Due to technical and regulatory issues, some triplers won't be able to receive some, or any, SMS messages. Our Primary Analysis will be intent-to-treat, treating these individuals the same as other treatment group triplers.
- Regulatory issues may also lead to delays in sending messages to some participants; in this case we will update the text of the reminders as appropriate (eg, to say that election day is today instead of tomorrow).
  - If delays prove to be common, we may add some buffer by targeting reminder texts to arrive a day earlier than shown above.

05 / Research Question

Given that someone has pledged to encourage specific friends to vote, does sending that person a reminder to do so increase the probability that those friends will vote[20]?

We interpret this as an externally valid estimate of the effect of vote tripling, though it will be an underestimate since some triplers will remember to encourage their friends even if they don't receive a reminder, reducing the difference between groups.

06 / Assignment

---

[20] This trial focuses on outcomes in the upcoming midterm general elections (see section 08 / Outcomes). We may contact all triplers who pledge in this study about future elections. Future contacts post this midterm election will constitute a separate study and are thus not part of this protocol.

METHOD:	The random assignment process will be done by research team staff using the R "randomizr" package.
STRUCTURE:	Triplers will be randomized (post exclusions) in batches every 1-3 days. Randomization will be stratified on: • geographic state • college campus vs. not • food incentives offered vs. not • canvasser • whether the tripler was matched to the voter file "live" by the canvasser vs. later We will regularly check covariates for possible failures of randomization (see internal validity below).

ARMS:	The triplers (and all their triplees) will be randomized into 2 arms:
	• 50%: Treatment: tripler receives reminder(s) to encourage their triplees to vote, composed of:
	
	    ○ 25%: SMS reminders
	
	    ○ 25%: SMS reminders plus handwritten letter
	
	• 50%: Control (tripler does not receive a reminder)
	
	The primary analysis will pool both treatment group subgroups together; exploratory analyses will test for marginal benefits of receiving the handwritten letter.

UNIT OF ASSIGNMENT:	Triplers (see details in the section above). All of a tripler's triplees will receive the same treatment as one another (but they won't be aware of this).
UNIT OF MEASURE:	Triplee outcomes are measured at the triplee level. Tripler outcomes are measured at the tripler level.

Assignment process

All triplers who do not meet any exclusion criteria will be assigned as described above. The triplees who are linked to those triplers, and who could be identified in the voter file, will form the sample to be used for the triplee turnout outcome. Triplees who cannot be matched to the voter file are still included in the treatment (since their triplers receive their names in the reminder messages), but they are not a part of the sample to be analyzed because we cannot measure

their outcomes. The subset of triplers who could be matched to the voter file will be used for our tripler turnout outcome.

Note that we will randomize triplers in batches every few days; this is so that we are able to send the introductory message (see Reminder process) to treatment group triplers only.

Internal validity

- Triplers and triplees will be unaware that any randomization will occur.
- To avoid contamination, we will, prior to batch randomization:
    - exclude any triplee who was listed by more than one tripler (because their triplers could be in different treatment conditions)
    - assign any tripler who pledges more than once to the same treatment condition for all their pledges
    - exclude any triplers who are also listed as triplees because the encouragement to vote that they could receive as triplees could also serve as a reminder to encourage their own triplees, even if they're in the control condition. These triplers will be excluded from the sample and will not receive any messages as triplers, so their information as triplees will still be used without risking contamination.
    - Checking for these conditions above will happen during each randomization batch and both the tripler and triplee voterIDs will be checked against all previous batches.
- New batches of participants will be randomized every 1-3 days. Balance checks will be run on the entire cumulative sample[21], and if any balance checks fail we will re-randomize the current batch and test again. Balance variables are:
    - The average triplees per pledge in each treatment arm
    - Triplee age, sex, and race/ethinicity
    - Tripler age, sex, and race/ethnicity
    - Triplee and tripler turnout score, as predicted by TargetSmart, for midterm elections
- In the case that an individual who was previously randomized becomes disqualified (eg, a tripler from a previous batch is listed as a triplee of a new tripler), that individual will be retroactively dropped and future balance checks will exclude them.

Balance check process

We will use the normalized differences approach to balance checks of Imbens and Rubin (2015), which is calculated by taking the difference in averages by treatment status, scaled by the square root of the sum of the variances. This is a scale-invariant measure, and a value of 0.25 or less indicates good balance.

For race/ethnicity and sex, we will create dummy codes and run balance checks on each one, disregarding categories that do not yet have 50 individuals in the cumulative sample (treating triplers and triplees separately).

For race/ethnicity we will collapse together the "other" and "uncoded" categories. For individuals not matched to the voter file we will code both gender and race/ethnicity as "unmatched".

---

[21] We will use the normalized differences approach to balance checks of Imbens and Rubin (2015), which is calculated by taking the difference in averages by treatment status, scaled by the square root of the sum of the variances. This is a scale-invariant measure, and a value of 0.25 or less indicates good balance.

07 / Sample & Setting

Sample characteristics:

See the [Assignment process](#) section for details on which individuals will be included in our samples.

Exclusion criteria

Prior to randomization

- Triplers and triplees will be removed if the canvassing firm or the VRAF field team determines there is a high probability that their pledge was fabricated by a canvasser.

    - If the fabrication reports come in after the batch has been randomized, we will drop those entries from the randomized sample and run all future balance checks excluding them[22]

- Triplers and triplees will be removed from the sample based on voter file matching and contamination criteria, as described under "internal validity" above.

- Triplees and triplers who are not registered to vote, or who cannot be found in the voter file, will not be usable for some analyses:

    - Triplees who can't be matched won't be used in the analysis of triplee voting outcomes, but their tripler is still eligible to be used in analyses of tripler outcomes

    - Triplers who can't be matched won't be used in the analysis of tripler outcomes, but their triplees are still eligible to be used in analyses of triplee outcomes.

After randomization

- We will use our texting provider's phone number validation service on triplers, even those allocated to the control condition who will not receive messages. This allows us to exclude those who gave us numbers not capable of receiving texts, and their triplees, without introducing imbalance between conditions[23].

    - We expect to process these exclusions post-randomization for technological reasons, but we will do them prior to randomization if reasonably possible.

    - We will not exclude triplers who give us incorrect but valid phone numbers, since we will only discover this for triplers in the treatment condition (ie, when the recipient responds telling us we have a wrong number).

---

[22] Because we stratify by canvasser when randomizing, and because most fraudulent entries are created en masse by a small number of canvassers, removing these entries post-randomization should not introduce any imbalance.
[23] We will check that the number of triplers excluded due to invalid phone numbers is roughly balanced across treatment arms.

- At the time data is analyzed, our organization may or may not have access to voter records with precise early voting dates. If we do, we will exclude any triplers or triplees whose early or absentee ballot was *received* prior to the date on which the tripler gave their pledge[24].

    - This is because if they voted prior to the pledge, the pledge can't have directly or indirectly influenced their vote.

    - We will check whether early voting rates were significantly different between treatment conditions. If they were, we will not use this exclusion criterion.

    - Note that if a tripler is excluded from the tripler sample for this reason, we *can* still use their triplees in analyses of triplee outcomes (in other words, a person who has already voted can still encourage others to vote).

Reaching our sample:

- As detailed above in the intervention design, we will contract with paid canvassing firms who will deploy canvassers to locations they judge to have sufficient traffic, and sufficient interest in pledging. Canvassers will induce people passing by at these locations to pledge to encourage five friends to vote.
- Canvassers will have mobile voter file access and are instructed to ask triplers to help match them and their friends to the voter file. We expect around 70% of triplers and triplees to be matched at the time the pledge is taken. In cases where in-field matching was not attempted, we will do the matching afterward using a combination of automated algorithms and human matchers (see Appendix: Linking individuals to the voter file).
- For details on how we will deliver reminders, see section 04 / Study Design.

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.
- We will do mapping exercises and ask our canvassing firm 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 and triplers who we can match to exactly one record in the voter file. It is possible that matching rates will differ based on factors such as name frequency, which vary between race and ethnicity groups. 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 (we did not use race as a matching criterion, but we assume that there are statistical trends towards name similarity within groups). Therefore, this is not an immediate concern.
    - Note also that this is only an issue for evaluation, not for the effectiveness or generalizability of vote tripling HTPC as a voter mobilization tactic. Because triplers contact their friends themselves, it works whether or not the organization canvassing can identify those friends.

---

[24] We will also run the analysis with the sample that includes them as a robustness check.

- 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 are powering this study to detect a 1pp change in the outcome (turnout rate). We used the R package "CRTSize" and the parameters listed below:

- alpha = 0.1, 1-tailed (HTPC higher than control)
- power = 0.8
- baseline voting rate = approx. 50% based on turnout among registered voters in the 2018 midterms ([source](#))
- Intra-class correlation between triplees within one tripler cluster = 0.37 (based on recent Texas primaries data)
- Standard deviation of outcome = 0.5 (SD for a Bernoulli distribution with a success rate equal to the baseline voting rate).
- Triplees per pledge=4
- Triplee match rate=70%

We find that we need 78,437 usable (voter file-matched) triplees across all arms of the study, requiring us to collect a total of 112,053 triplees. This equates to 28,014 pledges.

In practical terms, a 1pp increase in voter turnout for a sample size of 112,053 triplees translates into 560 marginal votes cast in the treatment group.

SAMPLE SIZE	We will collect over 28,000 pledges. We expect an average of 4 triplee names per tripler, and to match ~70% of triplees to the voter file, for an estimated N of 112,000 triplees.
SAMPLE SIZE FOR EACH ARM:	~56,000 triplees, grouped into ~14,000 clusters, in each of the treatment and control groups
CLUSTERS:	Triplees will be clustered within pledges (1-5 triplees per tripler)
JUSTIFICATION:	Under the assumptions above, this study is powered to detect a 1pp increase in turnout. Due to budgetary constraints we are unable to

collect a larger sample, even though HTPC could still be a viable tactic at smaller effect sizes (perhaps a minimum of 0.5pp).

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

VRAF staff will run all the analyses in this protocol. Dr. Michelson will independently analyze the Primary Effectiveness Measures based on this protocol. Any discrepancies between methods will be resolved by consultation between the two groups. In case consensus can't be reached, VRAF will clearly state in the writeup that Dr. Michelson independently analyzed the data and will provide reasons for not following her recommendations.

Dr. Michelson will also review VRAF's writeup of other analyses and their interpretation of the outcomes. She will offer recommendations on the drafts of the study writeup and will state whether she approves of the final version. VRAF will prominently state her involvement and her approval/nonapproval in their writeup.

Primary Effectiveness Measures

Triplee voting rate

- **Measure**: Voting rate among triplees in the midterm general 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 prior to when we send the first SMS message. Data used for voting status 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.

Tripler voting rate

- **Measure:** Voting rate among triplers in the midterm elections of the selected states. Data will be taken from public voter files available after the election.
- **Point of collection**: Same as triplee voting rate above.
- **Type**: Same as triplee voting rate above

- **Rationale**: Same as triplee voting rate above. In this case we are testing whether being reminded of one's pledge to encourage a friend to vote also increases one's own probability of voting.

Other Effectiveness Measures

Long-term voting rates

We will also measure voting rates for triplees and triplers in medium salience 2023 elections as well as the 2024 general elections, 2 years after the intervention. There will be no further intervention delivered during this time.[25]

Individuals who were matched in the midterms but are no longer in the voter file in the 2023 elections or the 2024 generals (eg, because they moved, changed names, or lost the right to vote) will be excluded from the main analysis. We will also check for differences in match rate between the two conditions, and check the robustness of our results by running an analysis with these individuals treated as non-voters for the election in question.

Household member spillover effect

We will measure voting rates for household members of triplers to see if there are spillover effects. A household member is defined as any registered voter living at the same address as the tripler[26]. We will run a separate analysis for household members of triplees.

We will exclude housemates whom the tripler listed as a triplee, since they will already be contributing to triplee outcomes. We will also exclude housemates who are themselves triplers.

For triplers, the most likely spillover mechanism is that being reminded to encourage their triplees to vote leads to also talking to other people close to them about voting. Another mechanism for both triplers and triplees is leading by doing, where people are more likely to vote if a co-resident models the behavior.

09 / Statistical Approach

---

[25] Alternately, 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.
[26] We define "living" based on the addresses available in our voter file data; for some states this consists of the voter registration mailing address and for other states it is the registration mailing address supplemented by other commercial data. We are aware that some people, especially college students, may be registered at a family address instead of their current residence, but because such people are likely to be in contact with their family, we consider spillover in either direction to be plausible.

Triplee voting rates

We are testing whether to reject the null hypothesis of no difference in triplee voting rates based on which treatment the tripler received: receiving reminders to encourage their triplees to vote, or not.

Our primary analysis will use an OLS linear regression with clustered standard errors by pledge.
- Main (fixed) effects:
    - Treatment condition (main outcome of interest)
    - State
    - Food incentives vs not
    - College campus vs not
    - Triplee Race/ethnicity
    - Triplee age
    - Triplee sex
    - TargetSmart Triplee turnout score for midterm elections
    - Number of days prior to the election the pledge was collected
    - Canvasser ID[27]
    - Number of matched triplees in the pledge
- Clustered errors:
    - Tripler ID

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).

Tripler voting rates

This outcome will be analyzed using the same set of models as triplee voting rate described above, replacing triplee covariates such as age and voting propensity with the values for the tripler. There will be no clustered standard errors because triplers are independent of one another.

Alternate models

Mixed model

To further explore the structure of the data, we will run an alternate analysis that represents correlated errors among triplers using a mixed effects model instead of clustered errors. In this model, we nest triplees within triplers within canvassers within states. We will estimate random intercepts per tripler, state, and canvasser. This model has the advantage of explaining variation at the canvasser level rather than "controlling for it". Additionally, we will add the following

---

[27] Canvasser is conceptually more like a random effect, but due to technological limitations we have not been able to find a way to run mixed-effects models with clustered standard errors. This should not affect the accuracy of our model: The number of canvassers is much smaller than the number of triplers or triplees, and individual canvasser coefficients will not be interpreted as outcomes of interest.

canvasser-level covariates: canvasser age, gender, race/ethnicity, education level, and past canvassing experience. Using a random effect for canvassers, rather than a fixed effect, also makes the estimation more efficient. We do not hypothesize what or how much variance the canvasser covariates will help explain.

If model assumptions hold, then both approaches to analyzing the outcomes will give similar results. If the two differ, we will treat the first model (clustered standard errors with fixed effects per canvasser) as definitive for determining whether the data supports the intervention.

Instrumental variable analysis

We expect that some triplers who are assigned to receive reminders will not receive them, due to 10DLC restrictions or other SMS deliverability issues.[28] Triplers and their triplees will be coded as "treated" if the tripler received at least 1 of the 2 reminder messages (early voting or election day). Note that we may make changes to this criterion based on the actual issues observed when sending messages, but we will settle on it before running the analysis.

We will then perform an instrumental variable analysis, re-running the "primary analysis" described above with the critical effect being the treatment "treated" instrumented by treatment assignment and the covariates specified in the primary analysis.

Exploratory Analyses

Food incentive effects

We will re-run the models described above with an interaction between treatment condition and food incentive status (dichotomous, incentive vs none). A significant effect will be taken to indicate that triplers who receive a food incentive respond to the treatment differently from those who did not.

If the primary analysis shows no treatment effect, but the interaction effect from this analysis suggests that the treatment effect may be meaningfully large only for those who received (or did not receive) the food incentive, we will treat this as suggestive but not definitive evidence that the intervention is effective in that subgroup.

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.

---

[28] Triplers may also fail to receive messages because the message is delivered, but marked as spam or "unknown sender" on their device. This is impossible for us to detect, and will be ignored.

We will test these effects by adding triplee race/ethnicity to the model described in the primary analysis, using the race/ethnicity category data available in the voter file[29]. We will add both a main effect for race/ethnicity and an interaction with treatment effect. We will set white as the reference group[30].

Multiple comparison adjustment

To address multiple comparisons, we will use the B-H 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 2023 and 2024 November general elections, we will assess whether there were carryover effects on tripler and triplee voting in those elections: We will re-run the models above, replacing the turnout score covariate with the turnout score right before those elections. Note that 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.

Spillover effects

We will analyze voting behavior of the household members of triplers and triplees (as described in Household member spillover effect).

In the triplers' housemate model, we replace all triplee covariates, as well as the outcome variable, with those of the housemates. We will exclude from the analysis any housemates that meet the exclusion criteria listed under "housemate spillover effect". We do not expect to find any significant effects due to power considerations, but we will run this exploratorily.

We will run a very similar model for the triplees' housemate spillover effect, but we will take into account the extra level of clustering that occurs at the triplee level.

Note that similarly to the long-term carryover effects, the analysis of spillover effects will not be included in the multiple comparison adjustment because the analysis is run on a different sample.

Implementation

10 / Trial Procedure

---

[29] These categories are (verbatim): African-American, Asian, Caucasian, Hispanic, Native American, Other, and Uncoded.
[30] We set white as the reference group because our primary, equity-based concern is to make sure we don't broadly recommend HTPC 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.

Appendix: _Preregistered study protocols_

Independent oversight

Dr. Melissa Michelson, an independent academic researcher, has reviewed this protocol and will meet with VRAF approximately biweekly to advise them on implementation issues. She will also provide [independent oversight](#) of VRAF's data analysis.

Approximate timeline

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

Task	Done by	Date / deadline
Collect pledges	Canvassing Firms	08/25 - 10/21[31]
Randomization	Vote Rev staff	Ongoing, to be done as quickly as feasible following automatic voter file matching.
Intervention (reminder messages to triplers)	Text messages: texting vendor or Vote Rev staff Written mail: Vote Forward volunteers	See [Reminder process](#)
Election Day		11/08
Analyze data/results	Vote Rev staff + Michelson lab	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:

---

[31] Pledge collection will stop more than 2 weeks before election day. We believe this allows the control group to "forget" about pledge collection and provide better contrast for our comparison. It also allows time for composing and sending out the handwritten reminders by post.

VALIDITY CONCERN	WHO WILL CHECK?	HOW WILL THEY CHECK?
Did recipients receive the intervention as & when intended?	Texting vendor & Vote Forward	Will report on deliverability, error, and opt-out rates for text messages, and opt-out rate for written letters
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.

Risk of harm

Reduced turnout: In a small 2021 pilot study, triplers in the treatment condition voted at a lower rate, which trended towards significance ($p=.19$). On balance we consider it unlikely that this is a causal effect, because: 1) there is no theoretical reason to believe that this type of intervention would reduce turnout, 2) reduced turnout has not been seen with similar interventions like Commit to Vote, and 3) an internal meta-analysis of SMS-based vote tripling interventions suggested a *positive* effect on triplers. Nonetheless, the observations are concerning and we have responded in the following ways:

- Increasing the emphasis on tripler voter file matching in our protocol, in order to increase our power to detect tripler effects
- Promoting tripler effects from a secondary to a primary outcome in the current study
- Added a focus on measuring tripler outcomes to other large RCTs VRAF is running in 2022

Canvasser interaction costs: 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.

Deception towards control group: Triplers in the control group will be told that they'll receive a reminder before the election, but will not receive one. We have not been able to find an acceptable way to implement this research method without the use of deception, and believe the potential harm to be low.

Discomfort from political conversations: Triplers and triplees could both experience discomfort due to talking about politics or being encouraged to vote. During piloting, 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.

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

Interactions between canvasser and triplers also pose a non-zero risk of canvassers facing harassment, violence, or discomfort. 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, sending reminders

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 which took place in July and August 2022.
PII / CONSENT:	Data contains PII (first and last names, cell numbers, and voter IDs) for triplees and triplers. 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 all the states that the CRT will run in before the analysis 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
Tripler matched to voter file	Inclusion	After manual matching is completed	Voter file	Binary - 1/0, tripler 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 triplers and triplees (age, sex, race/ethnicity)	Covariate	Before randomization	Consumer data file (if available to our org)[32], otherwise voter file; this may differ by state	Continuous (age, turnout score) or categorical (sex, race/ethnicity); tripler and triplee level data
Tripler and triplee internal ID numbers	Covariate	Before randomization	Assigned in our dataset	ID number
Pledge collection location	Covariate	Before randomization	Canvassing data	Categorical

---

[32] Commercial data files such as those from TargetSmart typically use voter file information combined with information from other consumer data sources

Canvasser	Covariate	Before randomization	Canvassing data	Categorical
Tripler phone number	Used in implementation	During pledge	Canvassing data	Phone number
Tripler and triplee voter IDs	Used in implementation	During pledge or during separate voter file matching process	State voter file	ID number
Canvasser demographic information	Covariate	Before randomization	Canvasser self-report	various

Data security and storage

Vote Rev will store data on several vendors with appropriate security policies: Google Cloud Platform, 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.

Pledge data will also be stored on servers operated by our canvassing software vendor (GRU).

Appendix: Linking individuals to the voter file

This describes the process we will use for looking up triplers or triplees in the voter file.

- Canvassers will attempt to do matching in the field, with the assistance of the tripler. They will look up possible matches using the information given (full name, age range, and city) and ask triplers to help disambiguate names with multiple matches. If the tripler leaves before or during this step, canvassers will not do matching on their own.

- If we know that a canvasser attempted but was unable to find a match for a given individual, we will consider that person definitively unmatched and will not attempt further matching.

- For individuals that canvassers did not attempt to match in the field, we will query the state's voter file after their data has been uploaded to a centralized table. We will go through the following matching steps in order, stopping anytime a unique match is found.

  - Query for people who match exactly on first and last name, age range, and city

  - Translate common nicknames to full names (eg, "Mike" to "Michael"), using a publicly available corpus of names[33]

  - Remove the city criterion (eg, a tripler might list their triplee as living in a large city, when the triplee's postal address is actually in a suburb of that city)

  - Expand the age range by adding 2 years to either end (eg, a tripler might believe their friend is about 40 and record their age as "40-49", when they're actually 39).

  - If any step leads to more than one valid match, we will terminate the process and flag the person as "ambiguous". If no valid match is available we will flag the person as "unfindable".

- For those still flagged as ambiguous or unfindable, we will use one or more proprietary matching algorithms provided by vendors such as Civis and TargetSmart. These algorithms use information about recency of registration and other behaviors to find high-probability matches.

- For those still flagged as ambiguous or unfindable, we will add them to a list of names for human matching. We will hire a team of workers to look them up manually in the voter file and use their own judgment as to whether they can match them. Human matchers will use their knowledge of nicknames, abbreviations, and typos, as well as indicators that individuals may be related or live at the same address. (note that address matching may also be added to our automated checks, if it proves feasible).

Note that there will be some false positive matches. This occurs when the true triplee can't be found with the information given, but exactly one other person does match that information. Because matching and condition assignment have no influence on one another, false positive matches will add noise but not bias.

---

[33] We haven't settled on a source at this time; it will likely be this list or a similar one.

Appendix: Preregistered protocol addendum

July 2023

Background

This document describes changes and updates to our analysis plan for the HTPC LargeCT. The original document can be seen in the current pre-registration, or the original pre-registration (datestamped prior to data collection) can be viewed at this private link.

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. We have also had conversations with design consultants and other collaborators that have led to some changes in our analysis plans.

In this document we will pre-register all decisions made **after** recruitment and intervention delivery but **before** Vote Rev Action Fund (VRAF) has voting outcome data in hand.[34] This means that when VRAF made these decisions we knew about issues that came up during canvassing, and about the size and demographics of our final sample. However, we have no information about outcomes (ie, voting rates) and could not have made any of these decisions with foreknowledge of how they will affect the study outcomes.

---

[34] VRAF needed to obtain data on early voting only, for other purposes. Our organization does own this data at the time of writing this document, but it has never been joined to our data from this study.

Protocol changes
Stratified randomization

Due to difficulty receiving timely data from our canvassing firm, we were not able to use food incentive or campus status as randomized stratification variables in most cases.

Randomization overlap and clustering

In our original protocol, were aware that it was possible people could appear in the data multiple times, in two different ways:

- **Multi-triplees** (named as a triplee by more than 1 tripler)
    - example: Pearl is recruited as a tripler and names 5 triplees, including Connie. Later, Steven is recruited as a tripler and names 5 triplees, also including Connie (verified using the voter file to be the same "Connie" both times)
    - Potential problem: Steven is randomized to treatment and Pearl is randomized to control. Connie receives a reminder from Steven so she can't be treated as a control group triplee anymore.
    - An incorrect solution would be to remove Connie's connection to Pearl and treat her as a treatment group tripler only. This would lead to an imbalance between conditions: the treatment group would grow larger than the control group, and we would tend to sort more socially connected people into treatment and less connected people into control.

- **"Tripleers" (both a tripler and triplee)**
    - Example: Pearl is recruited as a tripler and names 5 triplees, including Rose. Later, and unrelatedly, Rose herself is recruited as a tripler.
    - Potential problem: Suppose Pearl is randomized to treatment and Rose is randomized to control. When Pearl encourages Rose to vote, that might remind Rose that she was supposed to encourage her own triplees to vote.

Initially we said we would drop all multi-triplees and tripleers. However, we later determined that we could in some cases randomize them interdependently, avoiding the problems above and increasing our available sample size.

Solution

We discussed a number of possible solutions and settled on randomizing interdependently. That is, if two triplers name the same triplee, make sure both triplers are assigned to the same condition (if one was randomized earlier than the other, automatically assign the second one to the first one's condition). Assign tripleers to the same condition as the person who named them.

This could create pathological situations with very large, interdependent clusters. Therefore, we made the choice to only handle clusters involving two triplers, and exclude triplees whose situation is more complex. See the section Tripler cluster overlap for the exact process to be used. Due to complexities in the data we are not certain that our process caught all individuals to whom this should have been applied, but there should be no potential issues with biased application.

Data quality and inclusion

General principles

In reviewing our data, we found a wide variety of edge cases, errors, and other anomalies that led to triplers or triplees not receiving the intended treatment. These arose from three main sources:

1. errors in programming and data management (eg, a small number of triplers received messages with the wrong triplee names)
2. bugs or quirks in our texting vendor (eg, unpredictable results occurred when two triplers shared the same phone number), and
3. inherent complexities in running a networked study with rolling stratified randomization (eg, two different triplers might name the same triplee, but be assigned to different conditions).

We catalog the issues we have detected below. In general, our principles are as follows:

2. **Triplers or triplees who did not receive the intended treatment should be removed from the data if, and only if, one of the following conditions applies:**
   - the exclusion can be applied systematically across treatment and control groups,
   - the conditions leading to the exclusion all occurred prior to randomization, OR
   - we have good reason to believe the issue occurred at random.

If none of these conditions is satisfied, then we retain the data for analysis in order to avoid creating group difference that may be correlated with voting behavior.

3. **Each voter should be counted in the analysis only once**, even if present as both a tripler and a triplee.

4. **We will always remove data that is unambiguously invalid** due to being nonsensical (eg, a tripler listing themselves as a triplee), corrupted such that it can't be used, or identified as fraudulent by our field team.
   - Records removed in this way will never count towards exclusion criteria for other records. Example: If person *A* appears in the dataset as the triplee of one real tripler and one fraudulent tripler, the fraudulent record will be removed and *A* will not be treated as a multitriplee).

5. When an individual has not been matched to the voter file, we will assume that they are unique from every other individual in the study.
   - For triplees this is largely irrelevant because we can't use their voting outcomes in any case.
     i. It does mean that in a small number of cases a triplee might be incorrectly excluded for being named by too many different triplers, when in fact they were just named repeatedly by the same unmatched tripler.
   - For triplers this is relevant only in that if multiple unmatched pledges could be attributed to the same person, we would exclude triplees from their later pledges. As it is, triplees from both pledges will be included. However, they will still be

subject to the exclusion criteria described for duplicate phone numbers, which prevents most pathological consequences of including them.

Definitions and notes

- When we say that a record will be retained or not be excluded, this refers to the specific exclusion criterion under discussion. The record could still be excluded from the dataset if it meets a different exclusion criterion.
- Exclusion refers to exclusion from analyses of outcomes. Information from that record can still be used for other purposes. Notably, triplers are sometimes removed for reasons that allow us to keep their triplees, and in these cases the tripler's features can still be used as covariates for analysis of their triplees.

Early voting

In the original protocol we said we would exclude from analysis of outcomes anyone who voted early prior to when they appeared in the dataset. After discussion (and before seeing any early vote data) we decided instead to exclude anyone who voted early prior to the first treatment message going out to triplers (which took place near the end of each state's early voting period). This is based on the premise that triplers are unlikely to have encouraged their triplees to vote prior to receiving the first text from us reminding them to do so.[35]

This change excludes a larger number of participants than the original criterion did, and slightly reduces the study's generalizability – any effect we detect will apply only to registered voters who were not going to vote early, as opposed to all registered voters. However, it does not create any systematic group differences that could lead to spurious findings.

As before, we will not use this exclusion if early voters are unevenly distributed between treatment and control groups, and we will run a robustness check that does not use the exclusion.

Questionable data

Unreasonably common names

We examined the most common full names in the dataset to determine if canvassers might have been using certain names as tests, or selecting them for other arbitrary reasons. For these purposes we excluded fraud, unusable data, and duplicate signups by the same person. Examples:

- "Greg DeMayo" and "Greg Diamond" are two different names
- One voter named "Greg DeMayo" who is in the data three times because he was named by multiple triplers (ie, they all identified the same voter file record for him) only counts as one occurrence of that name

---

[35] This is also the premise of the study's treatment-control comparison, so if it is incorrect then we are unlikely to detect an effect in any case

- Two different voters named "Greg DeMayo" count as a name occurring twice

There are 108 names that apply to >5 distinct people. Of these, 77 start with "Aa." For example, there are 24 different people named "Aaron Jones" and 13 named "Aaron Walker." In all the 77 names comprise 64 "Aaron" names, 12 "Aaliyah" names, and 1 Aakash. This almost certainly represents some kind of unintended behavior by canvassers. We have a number of speculations on this topic but for the purposes of this protocol we establish the rule that **if a tripler or triplee's full name begins with "Aa", and more than 5 people in the dataset have that full name, all people with that full name will be excluded.** If a tripler is excluded on these grounds, their triplees will also be excluded.

The frequent names that do not begin with "Aa" appear to simply be common names (eg, 10 "Taylor Smith", 6 "Maria Lopez") and will not be excluded.

Other data quality investigation

At the time of registering this protocol addendum, we have done an initial examination of our data for signs of canvasser fraud or data mis-entry. However, investigations are ongoing and it's possible that we will find convincing evidence of data validity issues in the future. We reserve the right to add new exclusions and re-analyses of data in the future, with the following stipulations:

1. We will clearly indicate when communicating results if we've made any data validation or inclusion changes *after* the initial run of our main analysis.
2. We will never made data inclusion decisions with knowledge about how they will affect our analysis of outcomes.

Message delivery

Our intention was that triplers assigned to the control condition would receive no contact following their initial canvasser interaction, and triplers assigned to treatment would receive 4 messages:

1. **Welcome:** A welcome message in the week following their signup
2. **EV:** A message near the end of their state's early voting (EV) period with relevant information
3. **Pre E-day:** A message 1-2 days before election day reminding them to remind friends to vote soon
4. **E-Day:** A message on election day asking if they'd reached out to their triplees yet

We refer to the latter three messages collectively as the "treatment messages."

Below we address different issues that arose with message delivery. In order to avoid ambiguity, we will describe the full range of issues, and explicitly indicate which ones were handled as specified in our original protocol and which involve changes.

Welcome messages

We were aware that the welcome message would go out to some control group participants and not to some treatment group participants. This was necessitated by the updates to our randomization protocol described in [Randomization overlap and clustering](#), which allows triplers to sometimes change study arms based on subsequent tripler signups. Additionally, some triplers had issues with their pledge data that prevented us from adding them until more than a week had passed. We decided it would be confusing to receive a welcome message so late.

How we deal with it: We are updating the protocol to add receipt of the welcome message as a covariate, both a main effect and an interaction with treatment arm.

Undelivered messages

We were aware that some messages would be invisibly dropped by the recipient's carrier or marked as spam. We consulted with vendors and determined there was no way to detect or quantify this issue.
How we deal with it: We do not do anything in response to this issue (ie, we treat all properly sent messages as if they were delivered).

Unreachable numbers

Some triplers provided numbers that we could detect as nonexistent or malformed (collectively referred to as "invalid") through observation or using automated testing through our texting vendor. This testing was applied to both treatment and control triplers.
How we deal with it: In keeping with our original protocol these individuals will be excluded from the study, as long as our balance check criterion shows that the treatment and control groups do not differ on frequency of such numbers.

Wrong numbers and opt outs

In some cases we received responses to our messages saying that we'd reached a wrong number, or asking to not receive further messages.
How we deal with it: We did not message these individuals again, but we retain them and their triplees in the data. We were only able to detect wrong numbers and receive opt-outs in the treatment condition, so excluding these individuals would likely lead to treatment-control imbalance on unobserved variables.

Message omissions

Some messages were not sent as intended[36] due to data management errors or glitches in our third-party texting platform. In some cases people in the treatment condition did not receive reminder messages, despite having valid phone numbers and not having opted out or indicated a wrong number. In other cases, they received some but not all reminders (eg, they did not receive the early voting reminder but did receive the election day reminder). We refer to these below as message "omissions" to avoid confusion with any of the other issues described above.

---

[36] The final writeup will provide precise numbers; they are not relevant to the pre-analysis protocol decisions in this document.

How we deal with it: At the time of registering this protocol addendum we are still investigating the exact causes of message omissions. It is of high importance to understand whether message omissions can be treated as random, ie, unlikely to be related to any tripler traits except for ones that were measured and included in the statistical model. If our team concludes that any subset of triplers who received none of the treatment messages were omitted essentially at random, we will analyze those triplers as if they were in the control group. If we do recode them as control, we will run a robustness check analysis keeping them in the treatment group.

In any case, triplers who received some but not all of the treatment messages will be retained in the treatment group.

Vote Forward letters: Due to errors in our process, a small number of triplers in the control condition received Vote Forward letters, which should have been confined entirely to triplers in the treatment group. The recoding process described above, in which we treat triplers for whom all messages were omitted as control group triplers, will create an additional set of controls who received letters.

How we deal with it: Because our model estimates effects for text treatment and letter treatment separately, this does not cause problems, though the comparatively very small cell of people who did not receive texts but did receive letters will add some degree of error.

Tripler cluster overlap

Above we described changes we made to random assignment to accommodate multitriplees (triplees named by more than one tripler) and tripleers (triplers who were also named as triplees). Regardless of how condition assignment was altered, we will use the following criteria to determine if they can be used in the final dataset.

Multitriplees

If a triplee was named by exactly two triplers, and both of those triplers were assigned to the same condition, they will be retained. Triplers who were never randomized due to exclusion criteria don't count towards this total. Otherwise, they will be excluded.

For multitriplees who can be included in the data, select only one of their records to include. Choose a record that does not meet any other exclusion criteria if possible, and among those choose the earliest one chronologically.

Tripleers

Suppose that tripler A was named as a triplee by tripler B.:

- Tripler A's triplees: if B is the only tripler who named A as a triplee, and A's tripler record was randomized to the same condition as B, retain tripler A's triplees. Otherwise, exclude

them. Do not count triplers who named A but were never randomized to a condition due to exclusion rules.
- Tripler A's records: Retain only one of their records. Choose a record that does not meet any other exclusion criteria if possible, and among those choose a triplee record if possible, and among those choose the earliest one chronologically.
- Tripler B: tripler B and their other triplees are unaffected.

Multiple pledges by the same tripler

Some triplers pledged more than once, generally on different days or with different canvassers. We attempted to add subsequent triplees to their initial pledge (meaning that they could receive reminder messages with 10 names or more). However, due to vendor limitations we were usually unable to do this except for when they made multiple pledges in the same day. In order to deal with this consistently and systematically across groups, we will **exclude all of that tripler's records except for one from their first day.** We will **exclude all of their triplees not named on that first day.**

Duplicate tripler phone numbers

In a number of cases, multiple triplers appeared in the dataset with the same phone number. This is likely accurate for at least some cases (two people who share a phone line), so we will not exclude triplers on this basis. However, problems arise if the people were randomly assigned to different conditions because a reminder is likely to affect both of them.

Example: Suppose that Dave and Rose are both triplers and sign up with the same phone number. Dave is assigned to control and Rose to treatment. Dave sees a message arrive on his phone for Rose reminding her to encourage her friends to vote. This may cue him to reach out the the people he named, even though he didn't receive a reminder in his own name.

We did not anticipate this being a very common problem when designing the study and didn't put a protocol in place for it. Our plan now is to **exclude** all triplers who share phone numbers with other triplers unless all of the following are true:
- No more than 2 triplers total share that number *and*
- All triplers with that phone number were randomly assigned to the same condition.

Incorrect names used

In a very small number of cases, triplers received messages referring to them or to their friends by the wrong names. We believe this primarily occurred when multiple triplers were listed under the same phone number and the system confused their records with each other. Because this occurred in an unpredictable fashion and only in the treatment group, these records cannot be excluded.

Data exclusion summary

See for a summary of all exclusions and the order in which they will be applied.

Data processing

Small categories

For consistency with other projects, we will change our coding process for small categorical variable levels as follows. If a level has <50 instances:
- If the level represents missing, uncoded, or unknown values, we will combine it into any other levels representing missing/uncoded/unknown values.
- Otherwise we will combine it with all other levels that have <50 instances and do not represent missing/uncoded/unknown values.
- If any composite level created this way still has fewer than 50 instances, we will combine it into the largest level.

This only applies to inferential tests, not descriptives.

College campus coding

As a clarification, all uses of "college campus" anywhere in the document refer to whether the canvassing site where the pledge was collected was a college campus.

Although the vast majority of data was able to be traced to a specific canvasser shift and location, there were approximately 6,000 records that could not be connected back to a particular canvass shift for various reasons. Using a pre-existing list of canvassing locations and pledge location data, we were able to categorize the outstanding records. We used two criteria to categorize these pledges:
1. If the nearest campus canvassing site was less than 2,000 meters away and was also the closest canvassing site, then it was categorized as campus
2. If the nearest campus site was less than 500 meters away it was categorized as campus

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; our criteria were an attempt to reflect this.

Unlocalized records

Some records had missing or impossible location data, which we believe primarily occurred when the canvasser's phone malfunctioned or when canvassers turned off location services (which they were instructed not to do). These pledges will be treated as non-campus data for our primary outcome analysis.

Analysis updates

Householding

For our analysis of within-household spillover, we will find household members of the target individual based on TargetSmart's Exact Address Track ID if we are able to obtain it. This proprietary identifier uses a variety of methods to normalize postal addresses and reduce issues with city name variants and the like. If this is unavailable, we will use street address and zip code.

The model run for household members will be the same as for the main analysis, ie, we are moving to a single analysis combining tripler household members and triplee household members.

If a person shows up in the household dataset more than once (eg, they live with one tripler and one triplee, or with two different triplers) we will use the following criteria. We use the term "study member" to refer to any triplee or tripler but not other household members:

- If a household member lived with more than 4 study members, exclude them
- If a household member lived with study members from more than 2 tripler clusters, remove them[37]
- If a household member lives with more than 1 study member, exclude all but 1 of their records to avoid double-counting them. Prefer a record that does not meet any exclusion criteria, and secondarily is a tripler rather than a triplee, and tertiarily has the earliest timestamp.

Variables added to analysis

After pre-registering our protocol, we learned about other results that suggest that it would be valuable to add the following covariates. Based on conversations with design consultants, and results from other studies:

- We will represent age as two continuous variables, one for age and one for age^2.
- We will include as covariates the individual's (not tripler's) vote history in the 2018 general election and, 2020 general elections
    - We note that the modeled turnout score likely incorporates vote history already. This creates multicollinearity issues, but we are not attempting to interpret the value or significance of specific covariate parameters. The precision of the treatment effect estimate should still benefit.
- We will add modeled ideology score as a covariate. This is a number between 0 and 100, generated by a proprietary TargetSmart algorithm, with higher values indication greater likelihood of being a Democrat.
- We will add tripler age, race, gender, ideology, and turnout score as predictors.
- We will add census-tract level variables for:

---

[37] Triplers often choose to remind family members, so it's likely that many household members will live with a tripler and one or more triplees from the same cluster

- ○ average income
- ○ proportion of population with a bachelor's degree

In our initial description of the analysis we referred to "treatment condition" but were not explicit that we were referring to two treatment variables. These will both be included as main effects:
- Texting arm(tripler was sent a reminder vs. was not sent a reminder) – this is what is usually meant when we refer to the "treatment group"
- letter arm (tripler was sent a written letter vs. was not)

There will not be a separate interaction term for these, because only treatment group triplers were randomized to receive Vote Forward letters. We will code being sent a letter as the reference group, so that the coefficient for the texting arm reflects the effect of both treatments combined. If the coefficient for letter arm is significant and negative, that will indicate that the letter has a marginal effect above and beyond the text reminder.

Missing covariates

Some individuals may be missing data for certain covariates. This occurs primarily for the following reasons: 1) the datasets for ideology, past voter turnout, and certain other variables do not overlap perfectly with the voter file used for matching, 2) for voting history variables, some individuals are too young to have voted in the relevant election, and 3) in some cases a triplee is matchable to the voter file and can be analyzed, but their tripler is not; this means that the triplee will have missing data for covariates sourced from the tripler (eg, tripler's ideology score). We will handle this as follows:
- Continuous variables: Use mean imputation (for age^2, use the square of the mean age)
- Turnout variables: Add a code for "too young to vote" for individuals who could not have voted in the relevant election. When this is ambiguous – eg, someone who was 20 in mid-2022 may or may not have been over 18 on the day of the 2020 general election – we will assume that they were eligible to vote. If we are missing data for another reason, treat the individual as not having voted.
- Categorical variables: Use "Uncoded" or the equivalent category already present in the voter file.

Significance testing

We specified that we would use 1-tailed p values; this applies anytime the independent variable is treatment arm for reminders or treatment arm for Vote Forward mail. We do not have directional hypotheses for any other variables or for any treatment interaction effects, so these will use 2-tailed p values. In all cases we will use alpha = .10.

Multiple comparisons analysis

We initially said that we would adjust our secondary and exploratory analyses for multiple comparisons. However, in our experience secondary and exploratory analysis findings are not interpreted in the same definitive, binary way as main outcomes, and are primarily used to inform future research ideas. Therefore, we have decided that we will not adjust our secondary and exploratory analyses for multiple comparisons, with the following stipulations:

1. Regardless of what we find in secondary and exploratory analyses, **we will not claim demonstrated effectiveness for HTPC unless we find significant results in the [Main Analysis](#)**.
2. When discussing results and making recommendations for implementation, **we will explicitly differentiate between our pre-registered analyses and any further, unregistered analyses.**

Main Analysis structure

Single analysis for triplers and triplees

Based on conversations with stakeholders and statistical consultants, we chose to change our primary outcomes analysis to use a single model for triplers and triplees both, as opposed to separate models. The revised, full primary analysis is detailed below.

We are testing whether to reject the null hypothesis of no difference in voting rates based on which treatment the tripler received:
- receiving text message reminders to encourage their triplees to vote, or not ("text arm")
- receiving Vote Forward mail reminders to encourage their triplees to vote, or not ("letter arm")

Because we consider the intervention to have an effect on both triplers and triplees, we choose to include both in the same analysis, with indicator variables and interaction effects to distinguish tripler and triplee effects.

Our primary analysis will use an OLS linear regression with clustered standard errors by pledge.
- Main outcomes (fixed effects):
  - Texting arm: reminder or no reminder (reference group: no reminder)
  - Letter arm: letter or no letter (reference group: letter)
- Interaction effects:
  - texting arm x role
  - texting arm x welcome message
  - letter arm x role
  - letter arm x texting arm
  - letter arm x texting arm x role
- Covariates (fixed effects)
  - Role (tripler or triplee)
  - State in which registered to vote
  - Food incentives vs not
  - College campus vs not
  - Demographic variables for both triplees and triplers
    - race/ethnicity
    - age and age-squared
    - gender

- - ■ TargetSmart turnout score for midterm elections
    - ■ TargetSmart ideology score
  - ○ Individual's census tract demographics
    - ■ average income
    - ■ proportion with Bachelor's degree or higher
  - ○ Triplee and tripler voting history in 2018 and 2020 general elections
  - ○ Number of days prior to the election the pledge was collected
  - ○ Canvasser ID[38]
  - ○ Number of matched triplees in the pledge
  - ○ Welcome message[39]
- ● Clustered errors:
  - ○ Tripling cluster (a tripler and all their triplees)

Our main outcomes will be for the effect of SMS treatment (across letter conditions) 1) for triplers and 2) for triplees. We will generate and test these coefficients using the R functions `margins::margins()` or `alr::deltaMethod()`. We will consider the test an unequivocal rejection of the null hypothesis for the relevant contact type if the fixed effect parameter for treatment is significant at p<0.1 (one tailed, higher than control only).

Exploratory and secondary analyses

Random effects model

Our alternate model using random effects to predict voting rates will remain limited to triplee voting outcomes only.

Long-term carryover effects
We stated in error that we would examine results from 2023 elections. We intend to examine results from the 2024 general elections only.

Alternate model with additional interaction effects
Since developing this protocol, we've learned about other results that point to a potential interaction between relational mobilization and age and/or college campus status and age. We will remove the analysis labeled "Food incentive effects" and instead run an analysis with multiple interaction terms. The full list of interactions will be:
- ● Texting arm * tripler age
- ● Texting arm * tripler age^2
- ● Texting arm * triplee age

---

[38] Canvasser is conceptually more like a random effect, but due to technological limitations we have not been able to find a way to run mixed-effects models with clustered standard errors. This should not affect the accuracy of our model: The number of canvassers is much smaller than the number of triplers or triplees, and individual canvasser coefficients will not be interpreted as outcomes of interest.
[39] See the section on Welcome messages for details. We will treat receipt of a welcome message as the reference group, meaning that all estimated parameters for treatment condition will refer to the full, intended treatment.

- Texting arm * triplee age^2
- Texting arm * college campus status
- Texting arm * presence of food incentive
- Texting arm * number of matched triplees in the pledge

These models will *not* focus on interpreting main effects of treatment assignment, but rather on whether any of the interaction effects appear to modify the effects seen in the main analysis. If any interactions are significant, we will investigate the size of the treatment effect within each level of age range / campus status / food incentive used in the study.

We may run additional exploratory analyses investigating interactions with letter arm, but will not consider these primary study outcomes.

Alternate model without exclusions based on tripleer assignment
In our primary model, triplees will be excluded if their tripler was listed as a triplee by another tripler, *and* the condition assignments for the tripler and the tripler's tripler are different. We will run an alternate, exploratory model, for triplee outcomes only, in which we retain these triplees and add a dichotomous variable indicating whether the tripler was assigned to receive a reminder (0 = tripler has no tripler, or their tripler is in the control arm; 1=tripler has a tripler, and their tripler is in the treatment arm). This will be used as both a main effect and an interaction with study arm.

Instrumental variable analysis
We specified that the key "treated" predictor in our IV analysis would be receipt of either the early vote message or the election day message. To clarify, we will accept any of the three treatment messages as a sufficient criterion for "treated": early vote, pre-e-day, or e-day.

Robustness checks
We will run several robustness checks to determine whether the data processing choices described in this protocol have a pivotal effect on any results detected. In general these models are all more conservative than the main model we use, ie, they are more likely to fail to reject the null hypothesis that the treatment has no effect.

If our main analysis detects an effect of treatment, but any of these models does not, we will report this in our writeup.

Robustness checks include:
- We will rerun the primary analysis without several of our exclusion criteria:
  - improbably common names
  - duplicate tripler phone numbers
- We will rerun the primary analysis without excluding Individuals who voted prior to being included in the study

- If we determine that we can recode any subset of treatment group triplers as control based on message omissions, we will rerun the primary analysis and instead retain their original assignment.

Appendix: Data exclusion process

This operationalizes the data exclusion process. Unless otherwise noted, triplers excluded under a given criterion will also have all their triplees excluded.

1	Exclude triplees and triplers who are not registered to vote, or who could not be matched to the voter file. Do not exclude triplees of triplers excluded this way.
2	Triplers and triplees will be removed if the canvassing firm or the VRAF field team determines there is a high probability that their pledge was fabricated by a canvasser.
3	if a tripler or triplee's full name begins with "Aa", and more than 5 different people in the dataset have that full name, all people with that full name will be excluded.
4	Exclude triplers with phone numbers that are invalid (based on ThruText) or malformed (based on simple checks like string length)
5	Exclude all triplers who share phone numbers with other triplers unless only 2 triplers total share that number and both were assigned to the same condition. Do not count triplers who were never randomized.
6	If possible exclude triplers who voted prior to the first EV message. Do not exclude triplees of triplers excluded this way.
7	multipledgers: If a tripler pledges multiple times, exclude all of that tripler's records except for one from the first day, and all triplees except ones from their first day. Example: If tripler A names 3 triplees on September 1, then another 4 triplees later on September 1, then 2 triplees on September 2, keep the first 7 and exclude the last 2.

8	multitriplees: If a triplee is named by exactly one or two triplers, and all of those triplers were assigned to the same condition, they will be retained. Otherwise, exclude them. Do not count: - Triplers who were never randomized - Triplers who do not have valid phone numbers _Do_ count all other cases, including but not limited to: - Multi-pledgers for whom this triplee was excluded because the triplee was from one of the later pledges
9	Tripleers: Suppose that tripler A was named as a triplee by tripler B. if B is the only tripler who named A as a triplee, and A's tripler record was randomized to the same condition as B, retain triplee A _and_ tripler A _and_ tripler A's triplees. If not -- that is, if A was named by a second tripler, OR if A and B are in different conditions, exclude both of A's records and all of A's triplees. Do not count triplers who named A but: - who were never randomized - who do not have valid phone numbers
10	Tripleers - removing duplicates: If person A is in the data as both a tripleer and a triplee, exclude all but one of their records, as described in the section on tripleers in the protocol Do not exclude triplees of triplers excluded for this reason
11	Multitriplees - removing duplicates: If person A is in the data as a triplee of two triplers, keep only their first valid record

Appendix: post-registration procedure changes

Updated exclusions

Clustering issues
~8000 individuals in the data had their conditions reassigned due to clustering issues (see the discussion of multitriplees and tripleers in the addendum). However, we were unable to change their letter treatment assignment, meaning that a nonrandom subset of individuals received letters in the control condition, or did not receive letters in the treatment condition (and were ineligible to do so, making them different from those randomly assigned to not receive letters). This creates issues with estimating the effects of either treatment arm, and so we have chosen to exclude these individuals from the main analysis. We will run a robustness check in which they are included but marked as ineligible to receive letters so that they do not contribute to the estimation of the effect of letters.

Invalid letter assignments
A small number of participants (<200) had condition assignments that should not have been possible (no SMS but did receive a Vote Forward letter, or SMS but did not have a random letter assignment), for reasons we are unable to determine. These individuals are excluded from the main analysis but will be included in a robustness check.

Denoting letter-ineligible individuals

Aside from the complications described in Updated exclusions, some triplers were ineligible to receive letters for other reasons that do *not* prevent them and/or their triplers from being included in the estimate of the effect of SMS. It would be incorrect to code them as not receiving letters, because they would create a non-random imbalance between letter recipients and nonrecipients. Instead, they and their triplees have their arm_letter variable set to a third level, "ineligible". We maintain "no_letter" as the reference level for arm_letter, meaning that the parameter estimate for "yes_letter" gives an unbiased estimate of the effect of letter. The parameter estimate for "ineligible" is ignored.

The following are the criteria that can lead to an individual being marked as ineligible for estimating the effects of receiving a letter:

- In most cases (n=~5000), this happens because the tripler was not matched to the voter file and therefore had no mailing address available.[40] These triplers are excluded from all analyses because we were unable to obtain voting outcomes for them, but their triplees can still be included for estimating the effects of SMS.[41]
- In a small number of cases (n=~100) a triplee was able to be included in the data even though their tripler was excluded, for reasons other than those listed above. We code all these triplees' arm_letter as "ineligible."

In a small number of cases (n=~300), some triplees had two triplers, both in the treatment group, but one tripler was randomized to receive a letter and the other was not. Their actual arm_letter coding is randomly split between yes_letter and no_letter based on which of their multitriplee records was retained. This creates a mild bias towards underestimating the effect of letter treatment (because some no_letter individuals did have one tripler who received a letter) but we choose to keep this unchanged because we're uncertain whether any possible fixes would create imbalances.

Main analysis

Welcome message receipt

In our protocol, we stated that we would include receipt of the welcome SMS message as a covariate and model an interaction with SMS treatment arm. We later realized that, because welcome message receipt is likely nonrandom, this would not merely function as a covariate but would have unpredictable effects on our main outcomes of interest. Therefore, we chose to remove this variable from the analysis. We consulted with two independent statistical analysts who had no access to the data and they agreed with our decision that it should not have been used.

Main analysis structure

Our analysis estimates main effects for SMS (variable name: arm), letter (variable name: arm_letter), and tripler vs triplee status (variable name: persontype). Persontype is interacted with both arm variables. SMS and letter do not have an interaction because letters could only be sent to individuals in the SMS arm.

Data imputation

Missing value imputation for covariates was carried out as described previously. In cases where mean-imputation was used, we used the mean of the sample for that specific analysis (eg, the main analysis used the mean of the sample following all exclusions, and the robustness check that re-included early voters calculated a new mean that included their values)

---

[40] It was possible for triplers to self-report their address but in practice this almost never happened
[41] We sent SMS using phone numbers that triplers self-reported, so they were able to receive messages even if they couldn't be matched to the voter file.

Canvassers with low numbers of records

Some analysis code had difficulty running because it was trying to estimate parameters for canvassers who had triplers in the dataset but no triplees or vice versa (this could occur due to matching failures or exclusion criteria). We addressed this by recoding canvassers with fewer than 2 of either role into a single "low-frequency canvassers" category.

Appendix: Main model full results

This appendix shows all coefficients from the main model, except for individual canvasser fixed effects.

	Estimate	Std. Error	t value	p (two tailed)	total effect	p total effect (one tail adjusted)
(Intercept)	-0.055	0.025	-2.184	0.03		
SMS	0.003	0.004	0.893	0.37	0.003	0.185822391
mobilizer	0.007	0.005	1.480	0.14		
letter=ineligible	-0.010	0.017	-0.582	0.56		
letter=yes	-0.001	0.004	-0.114	0.91	0.003	0.235
SMS : mobilizer	-0.016	0.008	-1.959	0.05	-0.013	0.957
letter=yes : mobilizer	0.015	0.010	1.533	0.13	0.001	0.4239
stateFL	-0.064	0.023	-2.767	0.01		
stateMI	0.147	0.050	2.940	0.00		
stateNC	-0.045	0.027	-1.683	0.09		
foodTRUE	-0.005	0.005	-0.964	0.34		
pledge_days_before_eday	0.000	0.000	0.476	0.63		
centered_age	0.002	0.001	4.225	0.00		
centered_age_sq	0.000	0.000	-4.383	0.00		
centered_campus_analysis	0.011	0.004	2.664	0.01		
race_analysisAsian	0.006	0.010	0.599	0.55		
race_analysisCaucasian	0.050	0.005	10.905	0.00		
race_analysisHispanic	0.015	0.006	2.479	0.01		
race_analysisMultiracial	0.038	0.019	2.035	0.04		
race_analysisNative American	0.021	0.019	1.094	0.27		
race_analysisOther	-0.002	0.012	-0.188	0.85		
race_analysisUncoded	0.021	0.009	2.255	0.02		
gender_analysisMale	0.004	0.003	1.228	0.22		
gender_analysisUnknown	-0.100	0.017	-5.960	0.00		
partisanship_score_imputed	-0.080	0.006	-13.626	0.00		
midterm_turnout_imputed	0.008	0.000	82.390	0.00		

vote_g2020_agestoo young	-0.011	0.027	-0.423	0.67		
vote_g2020_agesyes	-0.032	0.005	-6.615	0.00		
vote_g2018_agestoo young	0.056	0.007	8.050	0.00		
vote_g2018_agesyes	0.082	0.005	18.030	0.00		
mobilizer_race_analysisAsian	-0.004	0.011	-0.368	0.71		
mobilizer_race_analysisCaucasian	0.008	0.005	1.619	0.11		
mobilizer_race_analysisHispanic	-0.005	0.007	-0.697	0.49		
mobilizer_race_analysisMultiracial	0.006	0.018	0.344	0.73		
mobilizer_race_analysisNative American	0.030	0.027	1.106	0.27		
mobilizer_race_analysisOther	-0.005	0.014	-0.371	0.71		
mobilizer_race_analysisUncoded	0.004	0.009	0.458	0.65		
mobilizer_age_imputed	0.000	0.001	-0.253	0.80		
mobilizer_age_sq_imputed	0.000	0.000	0.541	0.59		
mobilizer_gender_analysisMale	-0.002	0.003	-0.528	0.60		
mobilizer_gender_analysisUnknown	0.002	0.014	0.148	0.88		
mobilizer_partisanship_score_imputed	0.005	0.006	0.880	0.38		
mobilizer_midterm_turnout_imputed	0.000	0.000	0.397	0.69		
mobilizer_vote_g2020_agestoo young	-0.030	0.023	-1.289	0.20		
mobilizer_vote_g2020_agesyes	0.001	0.005	0.221	0.83		
mobilizer_vote_g2018_agestoo young	0.010	0.007	1.411	0.16		
mobilizer_vote_g2018_agesyes	-0.001	0.004	-0.119	0.90		
income_imputed	0.000	0.000	0.462	0.64		
bachelors_imputed	0.113	0.012	9.667	0.00		
num_matched_triplees_in_cluster	0.000	0.002	-0.222	0.82		

Appendix: Age effects

For completeness, we include visualizations of the interaction effects between age and treatment, as described in the section <u>Secondary Analysis: Differential effectiveness</u>. The graphs

below show the total interaction effect, combining the interaction effects of age and age^2, but do not include the main effect of treatment. These interactions did not approach significance and we do not attempt to interpret them.

| Interaction term between **friend** age and **SMS** reminder | Interaction term between **mobilizer** age and **SMS** reminder |

| Interaction term between **friend** age and **letter** reminder | Interaction term between **mobilizer** age and **letter** reminder |
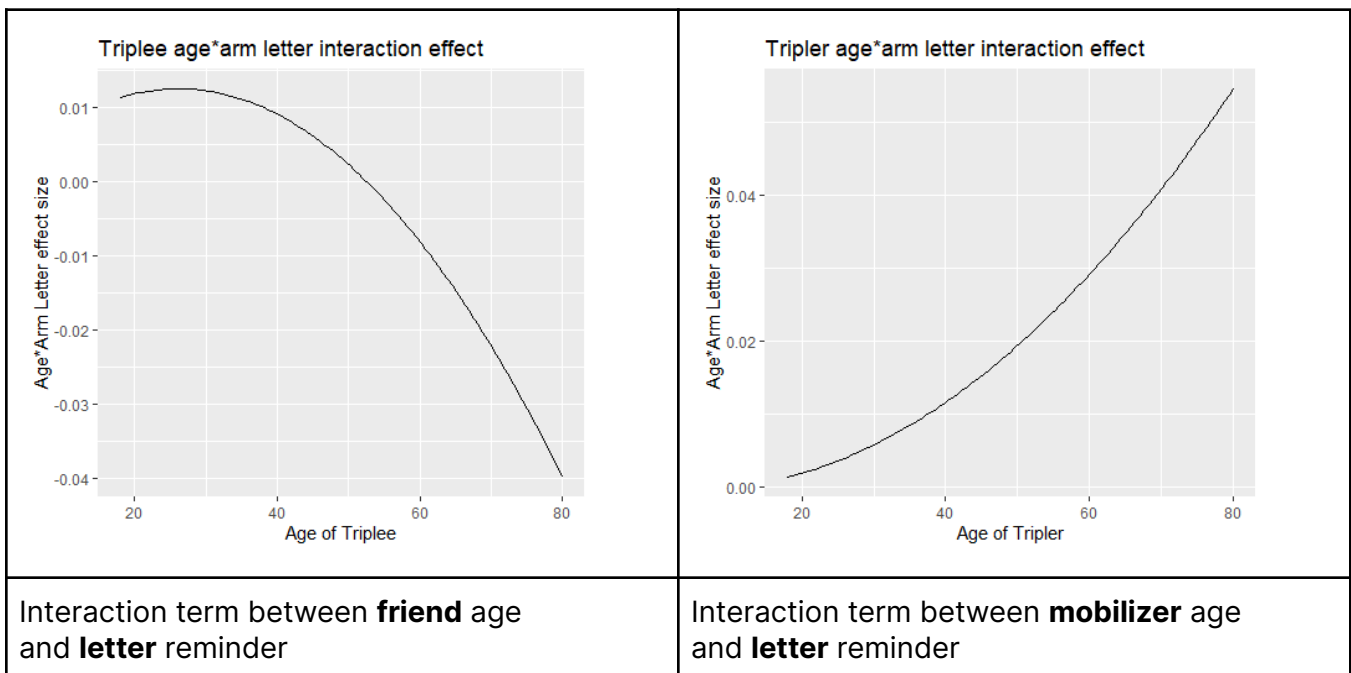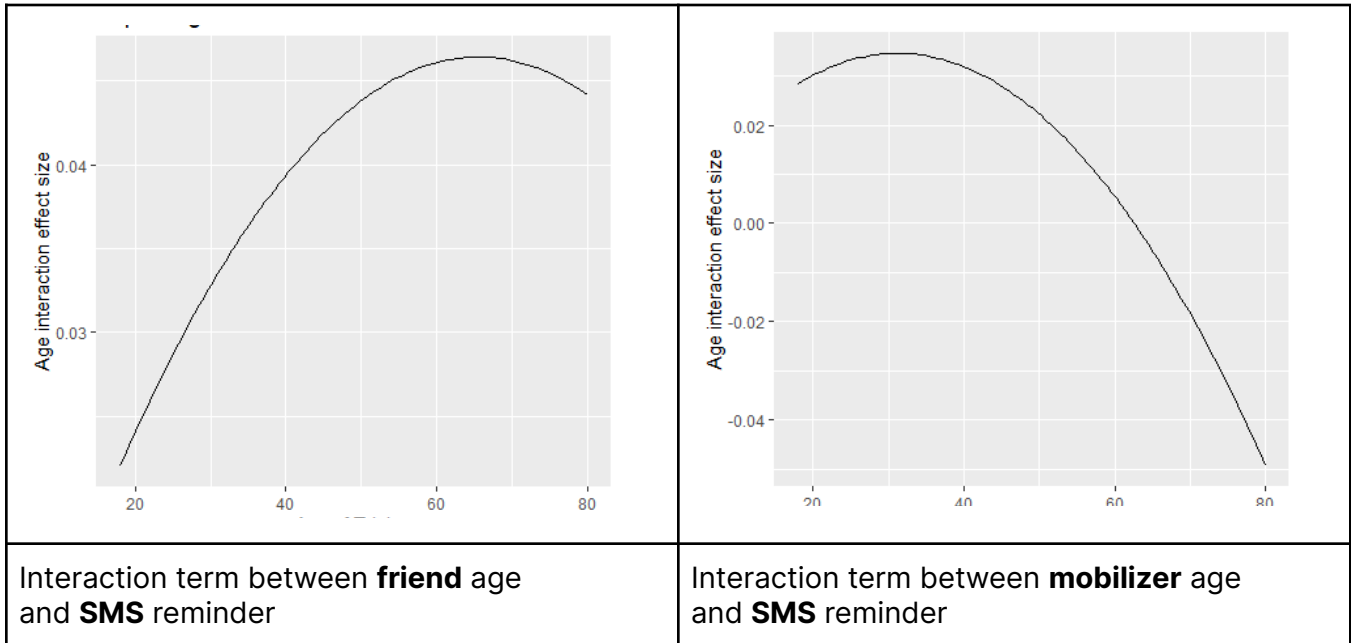
Appendix: Vote Forward letter

The template below was filled out by volunteers writing to mobilizers who lived in Florida. The other three states used identical templates except for the state-specific voting information website.

VOTE FORWARD

Dear

Thank you for chatting with a volunteer recently and pledging to remind **Test1, Test2, and Test3** to vote in the upcoming election on Tuesday, November 8! Will you get in touch with them right now?

Reminders make a big difference, especially when they come from someone we trust — and it only takes a minute. I encourage all my friends to vote because

You can learn more about this important election including who's on the ballot and how, when, and where to vote at the official Florida voter information website:

www.dos.myflorida.com/elections/for-voters

Thank you for being a voter, and for encouraging **Test1, Test2, and Test3** to join you at the polls!

Sincerely,

A volunteer wrote and sent this letter through **Vote Forward**, a nonprofit organization that serves volunteers who want to help fellow Americans vote, in collaboration with **Vote Rev Action Fund**, a voter engagement nonprofit.

If you have questions about voting, you can call 866-OUR-VOTE (866-687-8683) for help.

Use code X4ZN37 at votefwd.org/vote to let me know you plan to vote or to opt out of future mailings.

For: Dakota Henderson, 6864 Laurel Drive, Cutler Bay, MS 85975

Appendix: Text message content

Placeholders

These messages contain placeholders that were filled in with the relevant variable information. eg, if we sent the message "Thank you %FirstName%!" to Robert Reynolds, he would receive a message saying "Thank you, Robert!" Placeholders used in these messages include:

- FirstName and LastName
- collection-day: a string referencing when the initial canvassing interaction took place. Typically this referenced the day of the week ("on Monday") but in cases where sending was delayed it used a more appropriate descriptor such as "last week".
- friendstring: A list of friends the mobilizer pledged to remind. Properly formatted based on the number of entries (examples: one person = "Michael", two people = "Michael & Evie", three = "Michael, Andrea, & Evie")
- OrgName: Vote Rev
- OptOut: Instructions for opting out of further messages. Typically "Stop2Quit" but this may have varied.
- PollTime: State-specific poll closing time for election day

Messages

Welcome message

Hi %FirstName%! So great to meet you %collection-day%. Thank you for agreeing to remind %friendstring% to vote in the November election! - %OrgName% %OptOut%

Early voting

Florida:
Hi again %FirstName%! Right now, can you remind %friendstring% to vote? Florida early voting ends tomorrow, Saturday 11/5. (Some counties, including Hillsborough and Pinellas, end Sunday).

Michigan:
Hi again %FirstName%! Right now, can you remind %friendstring% to vote? There's still time to vote early in Michigan! You can send your friends to vote.org/polling-place-locator to find locations.

Arizona:
Hi again %FirstName%! AZ recommends mailing ballots BY TOMORROW to make sure they get there on time. [STOP2quit] Right now, can you remind %friendstring% to mail back their ballots?

North Carolina:

Hi again %FirstName%! Right now, can you remind %friendstring% to vote? North Carolina early voting ends TOMORROW, Saturday 11/5, at 3pm. You can send your friends to vote.org/polling-place-locator to find locations.

Day before election day

In case %friendstring% didn't vote early, will you remind them right now to vote tomorrow? You can tell them polls are open until %PollTime%, locations at vote.org/polling-place-locator. Thank you for helping everyone vote!!! [STOP2quit]

Election day

Just checking in, did you get a chance to remind %friendstring% to vote? Send us a 👍 if you have!