

**TECHNICAL SUPPLEMENT**

# Testing Text Timing Around Canvass Attempts

---

Eric Wilson

## Summary

The primary analysis in this supplement is treatment on the treated (TOT). For Groups A, B, and C, the denominator is voters who both received a canvass attempt and actually received the text treatment. For Group D, the comparison denominator is voters who received a canvass attempt.

Across pooled treated-voter results, Group B has the strongest topline turnout result at 51.8%, compared with 50.0% for the canvass-only control. The pooled B vs D difference is +1.8 points (95% CI: -0.7 to +4.4,  $p = 0.162$ ).

The clearest heterogeneous effect appears among women. In the pooled female treated sample, Group A and Group B outperform female controls by +4.3 points ( $p = 0.020$ ) and +4.0 points ( $p = 0.032$ ), respectively.

Cost results continue to favor the simpler treatments. Using pooled exact counts, cost per incremental vote is \$9.00 for Group A, \$5.42 for Group B, and \$53.71 for Group C.

## Study design and analytic definitions

This supplement uses the campaign's final phase-level CSV outputs and exact gender subgroup files to restate the experiment with exact N sizes, vote counts, confidence intervals, and p-values. It is intended for readers who want to scrutinize the design and the size of the observed effects.

The experiment has four groups. Group A received a text before the canvass attempt. Group B received a text after the canvass attempt. Group C received a text before and after. Group D received a canvass attempt only and serves as the control.

The primary frame here is treatment on the treated. In the treatment groups, a voter is counted as treated only if the voter both received a canvass attempt and actually received the text treatment in the correct order. In the control group, the treated denominator is voters who received a canvass attempt. This framing is closest to the substantive question the experiment was trying to answer: what happens when the intended treatment sequence is actually delivered.

Intent-to-treat results for the full assigned universe are included as a sensitivity check in an appendix. Those numbers are useful as a campaign management lens because they mix treatment effectiveness with delivery and reach.

**Table 1. Assigned universe and treated-voter denominators by phase**

Period	Universe A	Treated A	Treat rate A	Treated B	Treat rate B	Treated C	Treat rate C	Universe D	Treated D	Canvass rate D
Phase 1	3691	1980	53.6%	2054	55.5%	2064	55.8%	3678	2320	63.1%
Phase 2	2537	701	27.6%	589	23.3%	555	22.0%	2534	815	32.2%
Pooled	6228	2681	43.0%	2643	42.4%	2619	42.1%	6212	3135	50.5%

## Primary turnout results (treated voters)

Table 2 through Table 4 report exact vote counts and turnout rates within the treated-voter sample. Pairwise p-values compare each treatment group to Group D using a two-sided difference-in-proportions test. Confidence intervals are reported for the difference in turnout, in percentage points.

**Table 2. Phase 1 TOT turnout results**

Group	N	Votes	Turnout	Diff vs D	95% CI (pp)	p-value
A	1980	911	46.0%	-0.5 pts	-3.4 to 2.5	0.765
B	2054	983	47.9%	+1.4 pts	-1.6 to 4.4	0.357
C	2064	963	46.7%	+0.2 pts	-2.8 to 3.1	0.899
D	2320	1078	46.5%	ref	ref	ref

**Table 3. Phase 2 TOT turnout results**

Group	N	Votes	Turnout	Diff vs D	95% CI (pp)	p-value
A	701	458	65.3%	+5.5 pts	0.6 to 10.3	0.029
B	589	386	65.5%	+5.7 pts	0.6 to 10.8	0.031
C	555	355	64.0%	+4.1 pts	-1.1 to 9.3	0.127
D	815	488	59.9%	ref	ref	ref

**Table 4. Pooled TOT turnout results**

Group	N	Votes	Turnout	Diff vs D	95% CI (pp)	p-value
A	2681	1369	51.1%	+1.1 pts	-1.5 to 3.7	0.398
B	2643	1369	51.8%	+1.8 pts	-0.7 to 4.4	0.162
C	2619	1318	50.3%	+0.4 pts	-2.2 to 3.0	0.778
D	3135	1566	50.0%	ref	ref	ref

The omnibus four-group test is not significant in Phase 1 (chi-square = 1.53,  $p = 0.675$ ) or in the pooled treated-voter sample (chi-square = 2.25,  $p = 0.522$ ). In Phase 2 the overall four-group comparison is directionally stronger but still below the 0.05 threshold (chi-square = 6.72,  $p = 0.081$ ). The pairwise

pattern is still informative. Group B is the strongest performer in the pooled data, while Group C does not improve on the simpler one-text variants.

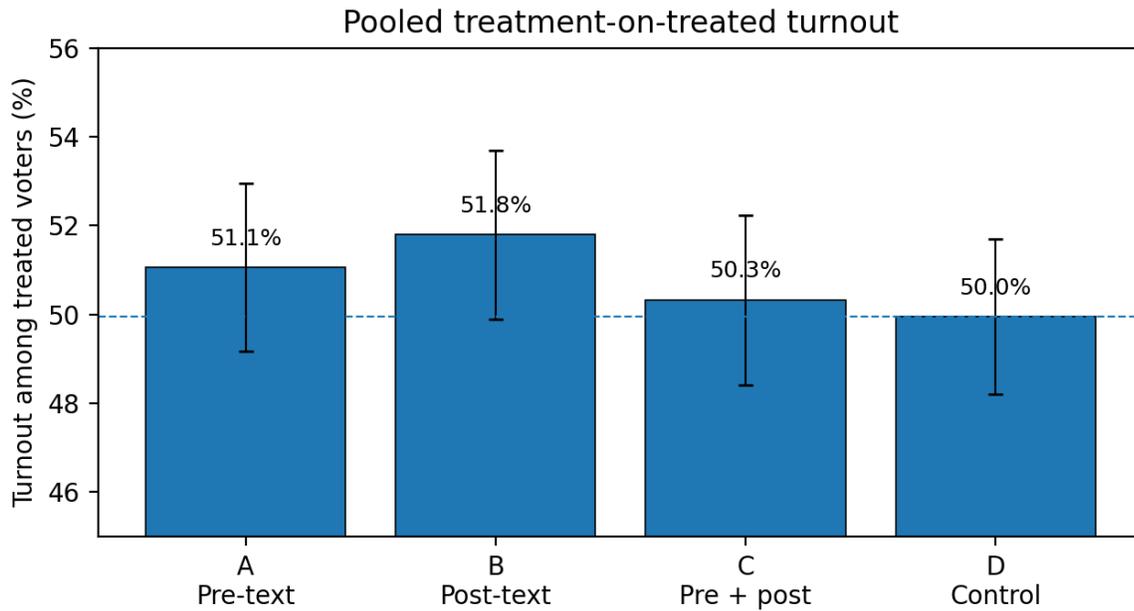


Figure 1. Pooled treatment-on-treated turnout with 95% confidence intervals.

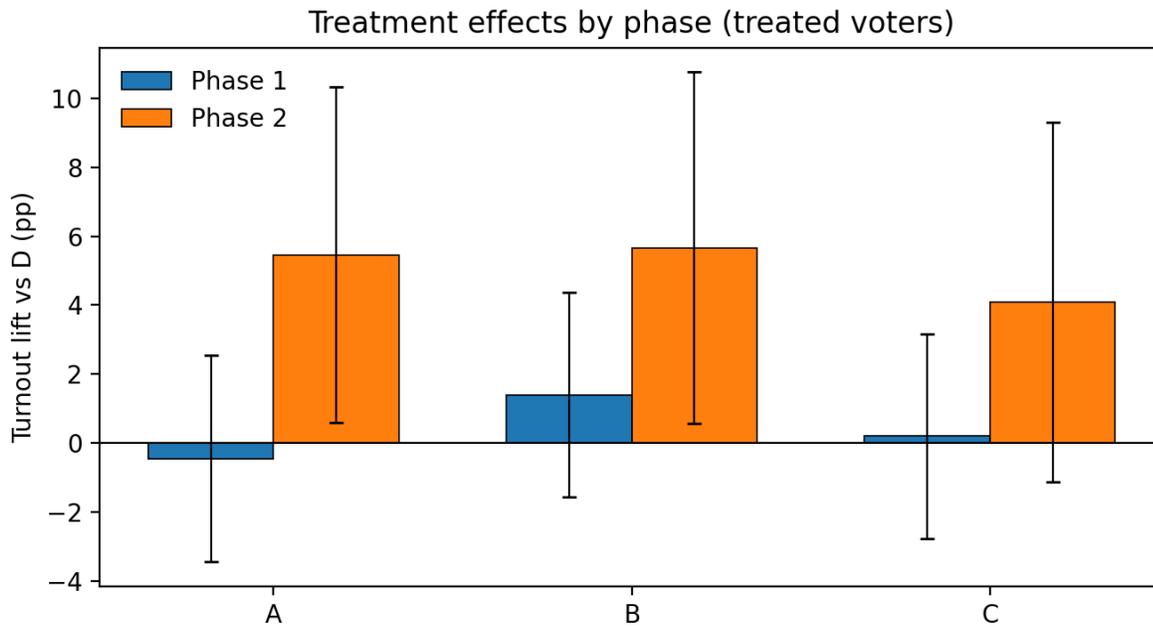


Figure 2. Turnout lift over control by phase, treatment on the treated.

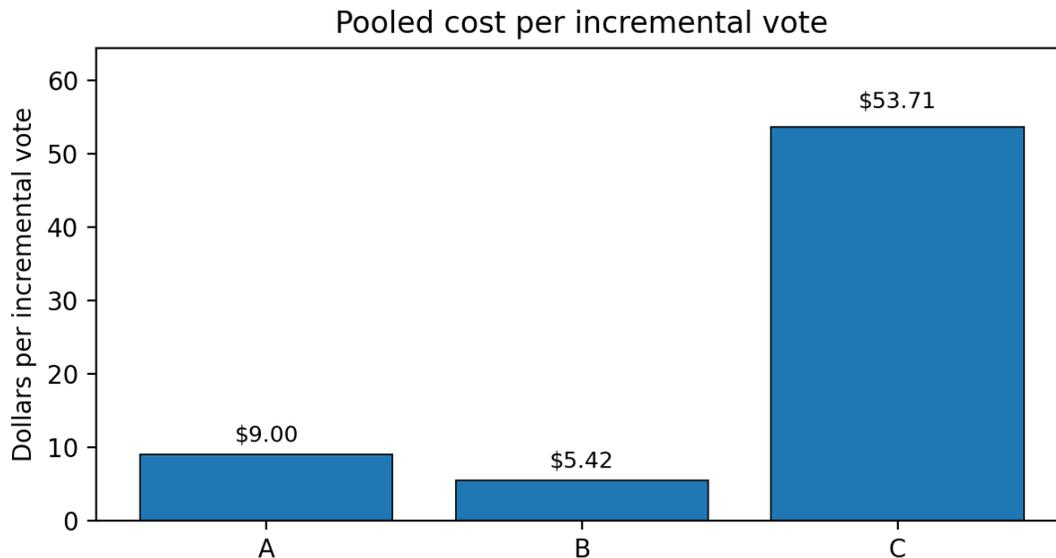
## Cost analysis

Because cost inputs are fixed at \$5.00 per canvass attempt and \$0.10 per text, the efficiency analysis is driven by observed turnout. Cost per vote summarizes average program efficiency. Cost per incremental vote isolates the marginal cost of the added text layer relative to the canvass-only control.

Using pooled exact counts, Group B has the strongest cost profile. Its treated-voter cost per vote is \$9.85, slightly below the control at \$10.01, and its estimated cost per incremental vote is \$5.42. Group A remains workable at \$9.00 per incremental vote. Group C is much less efficient because the additional text cost is paired with only a very small turnout lift in the pooled sample.

**Table 5. Pooled treated-voter cost metrics**

Group	Treated N	Votes (treated)	Treated cost	Cost per treated voter	Cost per vote (treated)	Cost per incremental vote
A	2681	1369	\$13,673.10	\$5.10	\$9.99	\$9.00
B	2643	1369	\$13,479.30	\$5.10	\$9.85	\$5.42
C	2619	1318	\$13,618.80	\$5.20	\$10.33	\$53.71
D	3135	1566	\$15,675.00	\$5.00	\$10.01	ref



*Figure 3. Pooled cost per incremental vote.*

## Gender subgroup results

The gender subgroup is the strongest source of heterogeneity in the study. The pooled female sample shows clear gains for both Group A and Group B. Male treated-voter results cluster tightly around the control and do not show a comparable pattern.

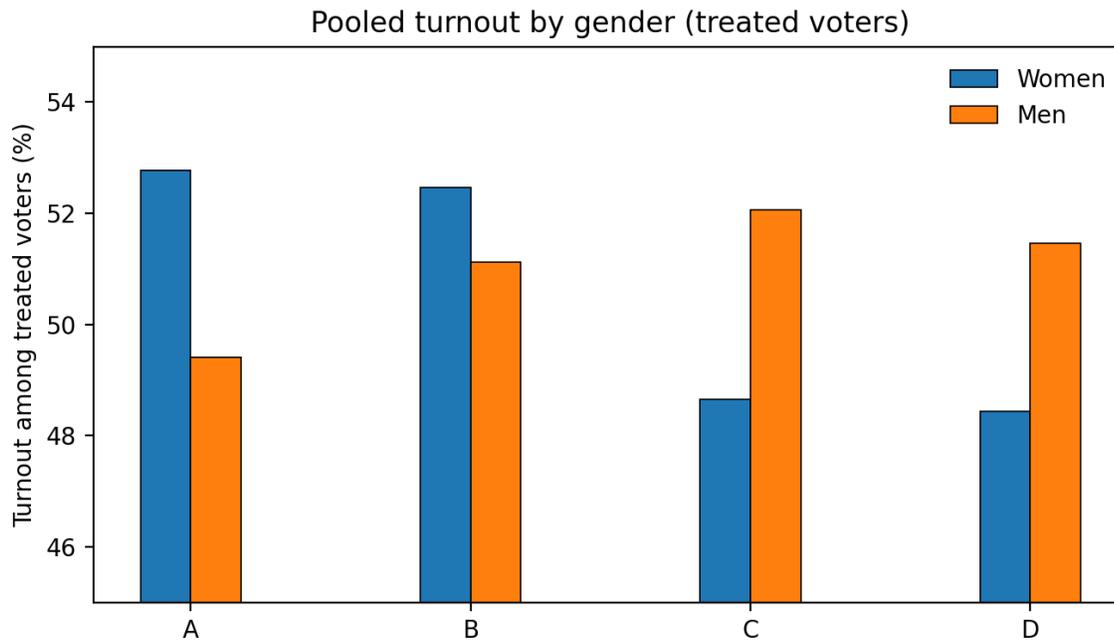
For women, Group A posts 52.8% turnout versus 48.4% in the female control group, a +4.3 point difference (95% CI: 0.7 to 8.0,  $p = 0.020$ ). Group B posts 52.5%, a +4.0 point difference (95% CI: 0.4 to 7.7,  $p = 0.032$ ). Group C is effectively flat.

For men, the corresponding differences versus control are -2.1 points for Group A ( $p = 0.268$ ), -0.3 points for Group B ( $p = 0.851$ ), and +0.6 points for Group C ( $p = 0.751$ ).

A formal interaction test reinforces that contrast. The treatment-by-gender interaction is statistically significant for Group A versus control (6.4 points, 95% CI 1.2 to 11.6,  $p = 0.015$ ). The interaction for Group B is directionally similar but weaker (4.4 points,  $p = 0.097$ ).

**Table 6. Pooled treated-voter turnout by gender**

Subgroup	Group	N	Votes	Turnout	Diff vs D	95% CI (pp)	p-value
Women	A	1332	703	52.8%	+4.3 pts	0.7 to 8.0	0.020
Women	B	1298	681	52.5%	+4.0 pts	0.4 to 7.7	0.032
Women	C	1334	649	48.7%	+0.2 pts	-3.4 to 3.9	0.907
Women	D	1563	757	48.4%	ref	ref	ref
Men	A	1348	666	49.4%	-2.1 pts	-5.7 to 1.6	0.268
Men	B	1344	687	51.1%	-0.3 pts	-4.0 to 3.3	0.851
Men	C	1285	669	52.1%	+0.6 pts	-3.1 to 4.3	0.751
Men	D	1570	808	51.5%	ref	ref	ref



*Figure 4. Pooled treated-voter turnout by gender.*

## Canvass outcome metrics and mechanism checks

The canvass disposition data help test the original priming hypothesis. If pre-canvass texting makes the door interaction itself more effective, we would expect better completion rates or lower not-home rates in the pre-text conditions.

Phase 1 provides little support for that story. Completion and refusal rates are nearly identical across groups. The only notable movement is a higher not-home rate in Group B and Group C relative to control, which does not line up with the priming theory.

Phase 2 shows more dispersion in canvass outcomes, but those shifts should be interpreted cautiously. Each voter received only one canvass attempt, and Group B's text is sent after that attempt. That means the

higher completion rate in Group B cannot be caused by the text itself at the door. The more defensible reading is that turnout effects, if present, occur after exposure rather than through better canvass contact.

**Table 7. Phase 1 treated-voter dispositions**

Metric	A	B	C	D
Complete	17.2%	17.1%	16.8%	16.9%
Not home	70.6%	74.1%	74.1%	71.5%
Refused	8.1%	8.3%	8.2%	8.4%

**Table 8. Phase 2 treated-voter dispositions**

Metric	A	B	C	D
Complete	19.8%	23.8%	18.7%	19.0%
Not home	66.2%	68.1%	71.4%	68.3%
Refused	8.4%	7.3%	9.7%	7.9%

## Intent-to-treat sensitivity analysis

The intent-to-treat appendix keeps the full assigned universe in the denominator. This is useful because it captures dilution from non-contact, non-delivery, and operational losses.

The ITT results are much smaller than the treated-voter results. In the pooled universe, Group A is 52.8%, Group B is 53.3%, Group C is 53.8%, and Group D is 53.5%. None of those differences is statistically distinguishable from the control. This is not inconsistent with the treated-voter findings. It mainly reflects how quickly small treatment effects become hard to detect once the denominator expands to the entire assigned universe.

**Table 9. Pooled intent-to-treat turnout results**

Group	N	Votes	Turnout	Diff vs D	95% CI (pp)	p-value
A	6228	3290	52.8%	-0.7 pts	-2.4 to 1.1	0.445
B	6231	3318	53.2%	-0.3 pts	-2.0 to 1.5	0.772
C	6227	3351	53.8%	+0.3 pts	-1.4 to 2.1	0.733
D	6212	3324	53.5%	ref	ref	ref

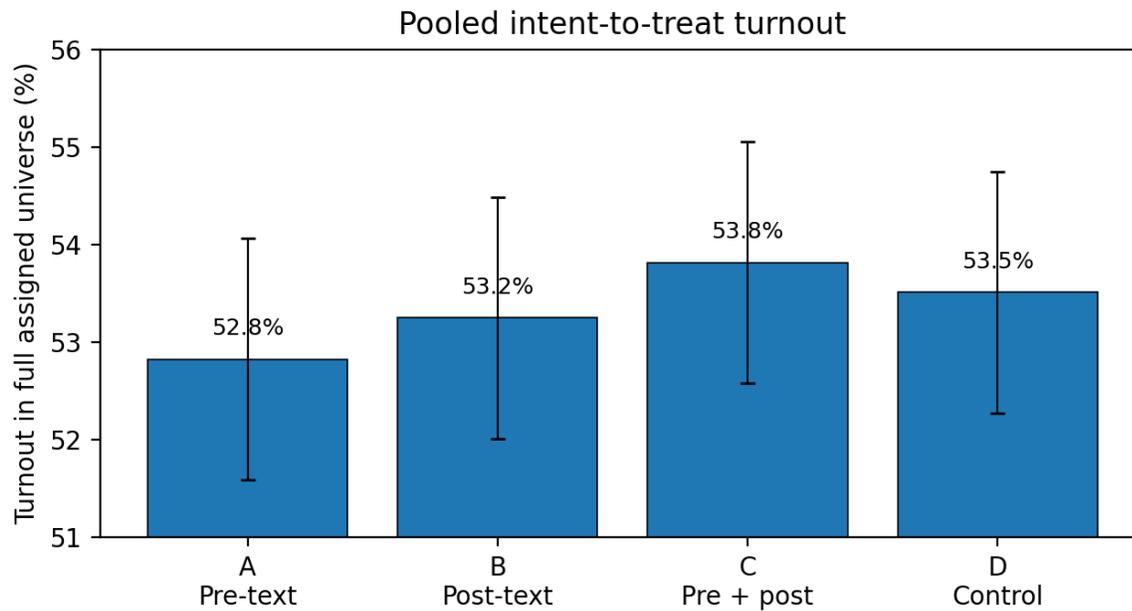


Figure 5. Pooled intent-to-treat turnout with 95% confidence intervals.

## Interpretation

For a campaign reader with a statistical background, the cleanest summary is this: the pooled treated-voter results are directionally strongest for post-canvass texting, the cost analysis also favors that variant, and the most durable subgroup effect appears among women. Those findings are strong enough to guide testing and targeting decisions.

At the same time, the topline pooled treatment effects remain modest, and the full-universe ITT estimates are substantially smaller. That combination argues for disciplined interpretation. The study points to a promising tactic and a promising audience, not to a universal rollout rule.

If a campaign wants the most evidence-consistent application of these results, it should prioritize a single post-canvass text as the default low-cost overlay, especially in female-targeted universes, and continue testing rather than adding more layers of complexity.

## Statistical notes

- All p-values reported in the main tables are from two-sided difference-in-proportions tests comparing each treatment group to the canvass-only control. Confidence intervals are 95% Wald intervals for the difference in turnout rates.
- The four-group omnibus tests use Pearson chi-square tests on the corresponding 2 x 4 turnout tables.
- The gender interaction tests compare the female treatment effect and the male treatment effect as a difference-in-differences in proportions.
- Because cost per incremental vote divides marginal cost by a small observed turnout difference, it should be interpreted as a directional planning metric rather than as a fixed market price for votes.