The impact of a relational contact

Gabriel Zucker, Director of Research
July 17, 2020

Abstract

Estimates of the value of a relational contact in the progressive turnout space have begun to settle around 1-2pp in 2020. We contend that this figure is too low, due to under-probed discrepancies between ITT and TOT estimates. In this paper, we present, to our knowledge, the first systematic meta-analysis of the TOT impact of a friend-to-friend voter contact, based on a range of assumptions about the underlying contact rates of existing studies. The primary models produce estimated effects ranging from 1.2 to 5.0pp. Our preferred specification (which we consider fairly conservative) shows an effect of 2.8pp (SE=.8); we believe other specifications in the 4pp range to be optimistic but still far from implausible. Such estimates, furthermore, are more consistent with what we know about the impact of other GOTV techniques. While analysts are generally justified in being conservative under such high uncertainty, we believe that undue caution would lead the progressive movement to systematically undervalue relational programs, whose impact would be measured at the bottom tail of a wide distribution, while non-relational programs are measured with much greater precision. We believe 2.5-3pp is a conservative but reasonable working estimate for the value of a relational contact in 2020.

1. Background

To date, most relational RCTs have followed the following paradigm: via some method, volunteers make a list of their friends known to the campaign/researcher. The researcher then provides back a random subsample of that list to the volunteer, encouraging her to contact all of them. The ITT effect is defined as the difference in turnout between friends on the treatment list and friends held back in a control list.

In most cases, contact rates are meaningfully less than 100%. In this sense, relational contacts are no different than other GOTV tactics; canvassing, for example, regularly has a contact rate around 30%. Notably, though, the variance in contact rates across studies is quite high. The 2018 VoteWithMe test, for example, estimated contact rates of around

1%; while the 2016 For Our Future WI test reported a contact rate of 86%. Moreover, the contact rate is generally not measured with much precision: Because relational programs put control in the hands of supporters rather than the campaign, relational contacts generally cannot be tracked with the specificity of canvass or phonebank attempts. With such high variance and uncertainty surrounding the contact rate, most meta-analyses to date have focused on ITT estimates of relational programs, and the impact of relational programs has frequently come to be reported this way.

But, frequently, what is of interest is the TOT estimate, the actual value of a friend-to-friend contact: the turnout boost accrued to a voter when they actually are contacted by a friend.[1] Because contact-rates are generally so low, this number is many times larger than the ITT figure. And, to our knowledge, no reliable meta-analysis estimate of this number currently exists. This paper intends to provide such an estimate.

The balance of this paper is structured as follows. Section 2 discusses the overall methodology. Section 3 presents the component studies and the assumed TOT effects, as well as studies not included in the analysis. Section 4 shows the results of the meta-analysis. Section 5 concludes.

2. Methodology

2.1 TOT/SE Construction

We reviewed the results of 15 relational studies. The majority of these, as discussed above, do not report contact rates. Based on common-sense heuristics and available information from similar implementations, we provide three models of possible contact rates for each study. (More precisely, we provide models of the compliance rate: the difference in contact level between treatment and control groups.) The three models are:

- **TOT Medium:** reasonably conservative (contact rates err on the *large* side, thus underestimating the TOT).
- **TOT Low** (high contact / low TOT): *extremely* conservative; it is not plausible that effects could be *lower* than these models.
- **TOT High** (low contact / high TOT): optimistic, but still generally plausible.

The point estimates and standard errors of the TOT effects are then constructed using the Bloom adjustment.

2.2 Scaling to 2020

The Analyst Institute generally recommends scaling effects downwards in presidential election years, with presidential effect sizes 57% of mid-term effects. However, we contend that this scaling factor may be too aggressive for relational programs. The

[1] Note, for purposes of this analysis, we will generally make the aggressively simplifying assumption that the value of a friend-to-friend contact is fixed, regardless of the medium, the messenger, the recipient, or other variables. In practice, this is likely not to be the case: contacts are probably more valuable by phone or in person than via text or email; contacts are probably more valuable from low-propensity to high-propensity voters than vice versa, etc.

scaling is well-established for traditional GOTV programs, and this makes intuitive sense: Traditional interventions are roughly equally effective in on- and off-year cycles, in the abstract, but their marginal impacts are notably smaller in higher-activity elections simply because there is so much background noise. But relational programs may be more effective in higher-salience cycles, when a larger portion of the population is able to speak more forcefully to their friends about the import of the election. Vote Tripling's research suggests that few voters are willing to remind their friends to vote in elections in which they will not vote themselves. The fact that turnout and supporter engagement is higher overall in presidential cycles thus opens the door to more friend-to-friend contacts, including contacts from friends with lower propensity to vote, who may be seen as more powerful messengers. The established 57% scaling factor does not account for this positive effect on relational contacts' effectiveness, whose magnitude is unknown.

As a result, we suggest that, at a minimum, the true ITT is likely somewhere between the scaled and unscaled estimates. In the meta-analysis, we report results from both scaled and unscaled estimates.

2.3 Meta-Analysis

We combine the scaled TOTs and SEs in a DerSimonian-Laird random-effects model, implemented using the Stata package -metaan-. Study weights are calculated by the DL model, determined primarily by the inverse of the study's variance, which is to say studies are of course weighted in inverse proportion to their uncertainty. Note that the uncertainty surrounding the contact rate varies by study, and in principle should contribute to the weights as well: studies with greater uncertainty around the contact rate have effectively higher standard errors for any single TOT estimate than studies that report contact rates more reliably.

Modeling this uncertainty would weight still lower those studies with low and uncertain contact rates. For simplicity, we skip this step, and do not model the uncertainty due to the contact rate.

3. Existing Research

We reviewed all studies tagged as "relational" in the Analyst Institute database, as well as additional studies flagged by our advisors. The studies are discussed here in three categories: (1) studies using RO platforms, which sync a volunteer's contacts to the voter file and encourage users to message matched voters, (2) phone-based programs, in which organizations encourage supporters to call specified contacts from lists volunteers provide, and (3) other designs, with unique methodological issues.

3.1 Category 1: Platforms

Table 1 shows results from all platform RO studies.

The majority of these programs used Empower, formerly known as MyRVPList, in the 2018 midterms, with various organizations employing the app to implement relational programs. The intensity of volunteer engagement varied from organization to organization, ranging from For Our Future, which issued several calls to action culminating in a one-on-one follow-up, to Faith in Action and Planned Parenthood, which do not report having issued reminders beyond the initial call to action. Row 7 shows results from all Empower users, including the studies in 1-6, among others. Write-ups of all Empower studies reported two effect sizes: an overall effect and a (usually much) larger effect for friends who were initially not matched to the voter file. Relying on this second effect seems to run the risk of cherry-picking positive data, and so this study relies only on the much more modest overall effects.

Planned Parenthood was the only one of these organizations to report plausible contact rates; their study reported treatment contact of 50% and control contact of 10%, with significant uncertainty around both of these estimates. For Our Future reported contact rates of 12% in the treatment group and 4% in the control group, which the authors considered likely to be highly incomplete; several other Empower studies also noted that reported contact rates were in the single digits, and dismissed these as unrealistically low. It is notable that both Planned Parenthood and For Our Future reported high contact rates in the control group relative to treatment; it seems likely that effective compliance rates would only be around 75-80% even in the optimistic scenario that all treatment voters were contacted.

But of course it is unrealistically optimistic to think that all treatment voters were contacted: Anecdotally, based on conversations with Analyst Institute, it is generous to assume that contact rates in treatment were much above 50% in any given implementation of these studies. Given the control contacts rate, a treatment contact rate of 50% would translate to a compliance rate of 40%.

The contact rates in Table 1 assume that the median implementation would have a compliance rate around 40%. Organizations with more reminders are assumed to have slightly higher compliance; organizations with large numbers of contacts per volunteer are assumed to have lower compliance, as it is more likely that volunteers picked only some of their targets to contact. Win Justice, for example, had 15 targets per volunteer, and even if 60% of volunteers made some contacts, and they contacted an average of 7.5 of their contacts, this would still translate only to a treatment contact rate of 30%, even before accounting for likely leakage into the control group. ROC, on the other hand, with 4.4 contacts per volunteer would be assumed to have better contact rates, despite having a similar reminder regime.

It seems likely that contact rates in the pooled study (Row 7) are even lower than for any individual organizational implementation, since organizations in the study likely made extra efforts to recruit and organize volunteers; but this is subject to still more uncertainty, and high variance across volunteers. Given the uncertainty — and given that this study is partially duplicative of the results in Rows 1-6 — we drop the Empower (pooled) study in many of our specifications.

Table 1: Platform RO studies

	Platform/Study		Year	N (trtmnt)	Mean targets / vol	Messaging to volunteers	ITT	Contact rate reported?	Compliance rate *assumed*		
									H	M	L
1	Empower	For Our Future[2]	2018	1118	4.7[3]	Several calls to action, and a one-on-one follow-up	3.7	Very incomplete: 12% T, 4% C	65%	45%	25%
2		ROC[4]	2018	12378	4.4[5]	Several calls to action	1.2	No	60%	40%	20%
3		Win Justice[6]	2018	13551	15[7]	Several calls to action	0.3	No	50%	30%	10%
4		Voces de la frontera[8]	2018	2890	15[9]	Two calls to action	0.6	No	50%	30%	10%
5		Faith in Action[10]	2018	490	4.6[11]	Encouraged to take action, but no ongoing communication	3.8	No	55%	40%	25%
6		Planned Parenthood[12]	2018	1505	6.5[13]	Encouraged to take action, but no report of	0.9	Incomplete: 50% T, 10% C	55%	40%	25%

[2] https://members.analystinstitute.org/research/for-our-future-results-memo-11003
[3] 1118 targets and 240 volunteers.
[4] https://members.analystinstitute.org/research/roc-action-results-memo-10980
[5] 12378 targets and 2912 volunteers.
[6] https://members.analystinstitute.org/research/win-justice-relational-voter-contact-results-memo-11026
[7] 13551 targets and 903 volunteers.
[8] https://members.analystinstitute.org/research/voces-de-la-frontera-action-relational-voter-turnout-results-memo-11097
[9] 2890 targets and 193 volunteers.
[10] https://members.analystinstitute.org/research/faith-in-action-relational-persuasion-results-memo-11247
[11] 490 targets and 106 volunteers.
[12] https://members.analystinstitute.org/research/planned-parenthood-votes-2018-rvt-test-11193
[13] 1505 targets and 233 volunteers.

					ongoing communication					
7	Empower (pooled)[14]	2018	34738	7.1[15]	Varied across implementing partners	0.5	No	40%	20%	7.5%
8	OutreachCircle[16]	2018	17930	155[17]	Nudged to email after initial match	1.3	Auto-email only: 60% were sent emails; 19% opened emails	60%	40%	20%
9	Vote With Me[18]	2018	11149120	70[19]	Nudged to message after initial match but no follow-up	-0.1	In-app contacts only: 1% contact rate	3%	1%	.5%

All effect sizes in pp. H = high / M = medium / L = low.

[14] https://members.analystinstitute.org/research/empower-relational-voter-turnout-test-results-memo-11167
[15] 34738 targets and 4886 volunteers.
[16] https://members.analystinstitute.org/research/analysis-of-outreachcircle-mobilization-experiment-2018-11014
[17] 17930 targets and 116 volunteers.
[18] https://members.analystinstitute.org/research/votewithme-relational-voter-turnout-results-memo-11275
[19] 11,149,120 targets and 159,423 volunteers.

Arguably, Studies 1-6 should be incorporated into the meta-analysis with diminished weights, given that they were all partially run as a single experiment, and thus their errors may be correlated. For simplicity, we skip this clustering, though it should perhaps be included in further research.

The OutreachCircle and Vote With Me studies differed from the Empower tests in a few key ways. First, friends were matched automatically from phone contacts, in a more opaque and less iterative matching process that may have seemed to users less reliable than that in Empower; and, second, long lists of 90% of all matching phone contacts were returned to users, meaning it is unlikely that even the most committed volunteers would have independently messaged a significant portion of their treatment group.

OutreachCircle solved for the long friend lists (averaging 155 per volunteer) by allowing volunteers to send automated emails through the platform. But, being automated emails from an unknown platform, these probably lacked some of the punch of a true friend-to-friend contact. Volunteers were encouraged to follow up with additional, more organic, contacts, but these additional contacts were not tracked, there appeared to be no follow-up to further motivate them, and there is fundamentally no way of knowing how frequently they occurred, or if they occurred much at all. **Given that the vast majority of the study may have consisted in auto-emails that are not properly friend-to-friend contacts, we drop the OutreachCircle study in some of our specifications.**

The Vote With Me study, meanwhile, did not have an auto-email feature, and had barely any follow-ups to volunteers encouraging them to reach out to friends. The study estimates contact rates of no more than 1%, which would leave the study entirely underpowered for most reasonable effect sizes. The reported -.1pp effect estimate is almost surely just noise (p =.7; different specifications have different signs), and the likely statistical fluke of the estimate falling just below zero has a large impact on the overall model after scaling, in some specifications.

3.2 Category 2: Match and Call Friends

Table 2: Match and call studies

	Study	Year	N (trtmnt)	ITT	Contact rate reported?	Compliance rate *assumed*		
						H	M	L
10	For Our Future WI[20]	2016	2941	1.7	86% T, 2% C	84%	84%	84%
11	COLOR / LIPS[21]	2008	42	22.6	High rate suggested	100%	92%	85%
12	PICO[22]	2016	448	4.2	No	80%	50%	20%
13	Battleground TX[23]	2016	3314	0.6	16.4% (phone call only)	45%	30%	16%

All effect sizes in pp. H = high / M = medium / L = low.

[20] https://members.analystinstitute.org/research/future-wisconsin-relational-voting-program-evaluation-9217
[21] https://members.analystinstitute.org/research/results-from-the-color-2008-general-election-experiment-375
[22] https://members.analystinstitute.org/research/pico-relational-organizing-test-8968
[23] https://members.analystinstitute.org/research/battleground-texas-social-network-gotv-test-8972

A second set of studies did not rely on apps; volunteers simply provided a list of contacts and were encouraged to call a randomized portion of them. The fact that these studies took place off of apps meant that volunteers were generally under more supervision from campaigns and therefore were more likely to record their contacts, more likely to follow through at all, and more likely to reach their friends via phone rather than text message or email, which is plausibly a higher-return medium. Of these studies, all but one recorded contact rates in the study.

For Our Future WI is the only study in this meta-analysis to report a high contact rate with certainty; as such, in some model specifications, it represents the outright majority of the modeled weight, which provides sufficient reason to probe it more deeply. Executed in Wisconsin in 2016, the intervention required significant buy-in from volunteers and so, unsurprisingly, produced a network of very high-propensity voter contacts: the average turnout score among the sample was extremely high, at 88. As expected, subgroup analysis showed much larger effects among lower-propensity voters; but the prevalence of high-propensity voters means that the average effect remains quite low overall, as there simply is not much room to boost the turnout of a group who are all already expected to vote. As a robustness check, given the inordinate impact of this study on the overall effect sizes, we drop it in several specifications.

The COLOR / LIPS study, the only study from before 2016, reported astronomical effect sizes, which may seem unlikely to generalize in the face of all the other evidence. The small sample size, however, means that this study rarely receives much weight in the meta-analysis. Contact rates are not reported but are implied to be high.

The Battleground TX program contained both a phone call and a postcard; the contact rate (16.4%) is reported for the phone call only. The contact rate by postcard is implied to be near 100%, which means the effective contact rate is somewhat higher, in proportion to the portion of the effect ascribed to the postcard. The M and H compliance rates are essentially functions of the portion of the impact ascribed to the postcard.

The study also reports a larger effect size of 2.9pp for regions with higher contact rates (where rates are in the 30% range). While this implies a significantly larger TOT, using this estimate runs the risk of cherry picking positive data and is disregarded.

PICO reports no contact rate information, and no information to guess at a likely rate. In the absence of data, the assumed rates here vary widely.

3.3 Category 3: Other

The remaining studies have additional measurement issues.

The VA Faith in Public Life study encouraged volunteers to contact friends via multiple media: phone call, text, and in-person conversation. None of these appear to have been randomized; but a pledge card was randomized. The impact of the pledge card (2.8pp) thus appears to be a lower bound estimate of the impact of the entire program. The various ITT estimates here consider if the rest of the program had been similarly randomized.

The Vote Tripling 2018 study has two sources of ambiguity. First, there are two layers of uncertainty in the contact rate; the experiment did not reliably measure who made relational contacts, or even who intended to do so. The contact rate thus could have easily been anywhere from 10% to 40%.[24] Second, the experiment did not collect information on the recipients of the intended relational contacts; the impact was measured only on cohabitants. Thus the ITT, which measures only the impact on cohabitants, is a very low lower bound. First name data from pledged triplers suggests that around 13% of the actual impact accrued to cohabitants,[25] so the true ITT could have easily been up to four times larger.

In both studies, standard errors are scaled up for the larger ITT effect sizes, to maintain a constant Z. For both studies, the three TOT models are calculated respectively as ITT-L/CR-H = TOT-L; ITT-M/CR M=TOT-M; ITT-H/CR-L=TOT-H.

Table 3: Other

	Study	Year	N (trtmnt)	ITT			Compliance rate *assumed*		
				L	M	H	H	M	L
14	VA Faith in Public Life[24]	2018	798	2.8	3.5	4.2	100 %	85 %	70 %
15	Vote Tripling TX	2018	10692	1.1	2	4	40%	20 %	10%

All effect sizes in pp. H = high / M = medium / L = low.

3.4 Studies Not Included

Three arguably relational studies are not included in the above review of existing research:

1. LCV (2014), ITT=2.5pp[25]— A mail test in which treatment subjects received postcards with a picture of a friend, while control subjects received generic make-a-plan postcards. We do not consider this a study of the impact of true friend-to-friend contact, as the mail was not sent by the friend, and the experiment did not encourage any organic relational contact.

2. Next Gen (2017), multiple reported ITTs[26] — A Facebook test in which contacts tagged their friends or sent them automatically generated DMs. We do not consider this a study of the impact of a true friend-to-friend contact, since the contacts were not direct (tagging), or not organic (auto-DMs).

[24] https://members.analystinstitute.org/research/faith-in-public-life-relational-mail-11041
[25] https://members.analystinstitute.org/research/league-of-conservation-voters-social-organizing-test-5811
[26] https://members.analystinstitute.org/research/networked-mobilization-the-role-of-visibility-in-peer-to-peer-voter-mobilization-on-facebook-10509

3. Sister District (2018), multiple reported ITTs[27] — A Facebook test in which contacts tagged their friends in status updates. As above, we do not consider this a study of the impact of a true friend-to friend contact.

[27] https://members.analystinstitute.org/research/relational-voter-turnout-and-facebook-tagging-11011

3.5 Summary of TOTs and SEs

The included studies with their implied TOT rates and standard errors are shown in Table 4. Columns to the right show studies adjusted for 2020 effects using the .57 factor.

Table 4: TOTs and SEs for all studies

	Study	Year	Notes	TOT			2020-Adjusted TOT		
				L	M	H	L	M	H
1	For Our Future	2018	1-6 are partially correlated	5.7 (3.2)	8.2 (4.7)	14.8 (6.0)	3.2 (1.8)	4.7 (2.7)	8.4 (4.8)
2	ROC	2018	1-6 are partially correlated	2.0 (1.8)	3.0 (2.8)	6.0 (5.5)	1.1 (1.0)	1.7 (1.6)	3.4 (3.1)
3	Win Justice	2018	1-6 are partially correlated	.6 (1.4)	1.0 (2.3)	3.0 (7.0)	.3 (.8)	.6 (1.3)	1.7 (4.0)
4	Voces de la frontera	2018	1-6 are partially correlated	1.2 (3.4)	2.0 (5.7)	6.0 (17.0)	.7 (1.9)	1.1 (3.2)	3.4 (9.7)
5	Faith in Action	2018	1-6 are partially correlated	6.9 (4.9)	9.5 (6.8)	15.2 (10.8)	3.9 (2.8)	5.4 (3.8)	8.7 (6.1)
6	Planned Parenthood	2018	1-6 are partially correlated	1.6 (2.9)	2.3 (4.0)	3.6 (6.4)	.93 (1.7)	1.3 (2.3)	2.1 (3.6)
7	Empower (pooled)	2018	Likely duplicative of 1-6 — dropped in some models	1.3 (1.5)	2.5 (3.0)	6.7 (8.0)	.7 (.9)	1.4 (1.7)	3.8 (4.6)
8	OutreachCircle	2018	Not entirely true F2F — dropped in some models	2.2 (1.7)	3.3 (2.5)	6.5 (5.0)	1.2 (1.0)	1.9 (1.4)	3.7 (2.3)
9	Vote With Me	2018	Sign is ambiguous	-3.3 (3.3)	-10.0 (10.0)	-20.0 (20.0)	-1.9 (1.9)	-5.7 (5.7)	-11.4 (11.4)
10	For Our Future WI	2016	Accounts for much of weight; dropped in some for robustness	2.0 (1.2)	2.0 (1.2)	2.0 (1.2)	2.0 (1.2)	2.0 (1.2)	2.0 (1.2)
11	COLOR / LIPS	2008		22.6 (10.8)	24.6 (11.7)	26.6 (12.7)	22.6 (10.8)	24.6 (11.7)	26.6 (12.7)
12	PICO	2016		5.3 (3.1)	8.4 (5.0)	21.0 (12.5)	5.3 (3.1)	8.4 (5.0)	21.0 (12.5)
13	Battleground TX	2016		1.3 (2.7)	2.0 (4.0)	3.8 (7.5)	1.3 (2.7)	2.0 (4.0)	3.8 (7.5)

14	VA Faith in Public Life	2018		2.8 (2.0)	4.1 (3.0)	6.0 (4.3)	1.6 (1.1)	2.3 (1.7)	3.4 (2.4)
15	Vote Tripling TX	2018		2.8 (1.7)	10.0 (6.0)	40.0 (24.0)	1.6 (.9)	5.7 (3.4)	22.8 (13.7)

TOT models from each study, with scaled standard errors in parentheses.

Note that standard errors are not reported in studies 10, 12, and 14. For 10 and 14, we reverse engineered the approximate standard error from the reported p-value. Study #12 provides no information on certainty, and so we assume a standard error commensurate to other studies of the same size.

4. Meta-Analysis

For each of the six models of TOT estimates and associated standard errors, we use a DerSimonian-Laird random-effects model to estimate an overall effect size. Table 5 shows the results; Row 1 shows the top line results including all studies. Without 2020 adjustment: the conservative model estimates 2pp TOT effects, the moderate model around 3pp effects, and the low-contact model around 4pp effects. Using 2020 adjustment, these are all about 1pp lower. The robustness checks in Rows 2 and 3 essentially do not change the results at all: despite the concerns raised about the Empower pooled and Outreach Circle studies, they do not have a meaningful impact on the overall effect sizes. Forest plots of these results are shown in the Appendix.

As discussed in Section 3.2, the For Our Future study, being the only somewhat large high-contact-rate study in the sample, commands as much as 61% of the total weight in some models, which ascribes a perhaps implausible portion of the weight to one study. As a robustness check, Rows 4 and 5 duplicate Rows 1 and 3 without For Our Future included. These models produce somewhat higher estimates especially in the H model, with effects now estimated around 7pp with unadjusted or 4pp with adjusted figures. The results of this check suggest that the overall effects may be slightly understated due to this one study; the bulk of the evidence points towards slightly higher impacts.

Table 5: Meta-analysis

	TOT *not* adjusted			TOT adjusted		
	L	M	H	L	M	H
(1) All studies	2.0 (.5)	2.8 (.8)	4.1 (1.3)	1.2 (.3)	2.0 (.5)	3.0 (.8)
(2) Without Empower pooled (#7)	2.1 (.6)	2.9 (.8)	4.7 (1.5)	1.3 (.4)	2.0 (.6)	3.0 (.9)
(3) Without Empower pooled (#7), OutreachCircle (#8)	2.1 (.6)	2.8 (.8)	5.0 (1.8)	1.3 (.4)	2.0 (.6)	3.0 (1.0)
(4) Without For Our Future WI (#10)	2.0 (.6)	3.4 (1.0)	7.2 (2.0)	1.1 (.3)	1.9 (.6)	4.1 (1.2)
(5) Without #7, #8, #10	2.1 (.7)	3.6 (1.2)	7.4 (2.3)	1.2 (.4)	2.1 (.7)	4.2 (1.4)

Average effects estimated with DerSimonian-Laird random-effects model. Standard errors in parentheses.

5. Discussion

Predictably, the vast uncertainty around contact rates in the existing relational research yields a wide variety of estimates for the TOT effect of a friend-to-friend contact in 2020, ranging from 1.2pp to 5.0pp in primary specifications, and up to 7.4pp in some robustness checks. However, several conclusions seem justified:

- Only the most incredibly conservative assumptions yield effects in the 1pp range — specifications using our implausibly low “L” model, and additionally applying

aggressive 2020 adjustments that we argue are, at best, too aggressive in the case of relational programs.

- Our "M" model, which we contend is a middle-of-the-road yet still fairly conservative estimate, yields projections in the 3pp range without adjustments, and in the 2pp range with adjustments. • In our view, the single most plausible — but still cautious — specification is that of the "M" model without 2020 adjustment. This model shows a 2.8pp effect (SE=.8).

- Meanwhile, the "H" specifications span effect sizes from 3.0 to 5.0pp among primary specifications, and even then are hampered by implausibly high weights assigned to the For Our Future WI study. Recall that the "H" model is aggressive in its TOT estimates, though in our view not unrealistically so.

In our view, the weight of this imperfect evidence supports a working estimate of 2.5-3pp for the TOT of a relational contact in 2020. Anything less requires highly conservative assumptions at multiple stages of the analysis.

And, while in certain contexts making such deeply cautious assumptions is admirable, we contend this approach would unfairly bias GOTV research against relational work. Less relational approaches, being implemented directly by campaigns, are simply subject to far less uncertainty, and assumptions wind up playing a much less influential role. Analysts will systematically underestimate the relative importance of relational programs if their impact is defined by the worst-case scenario of a wide distribution, while non-relational programs are measured with precision. It is not a valid comparison.

Another way entirely to consider this analysis is by comparing relational programs to other, better understood, techniques. Canvassing is generally estimated to have a TOT impact of around 5-6pp. To a first order of approximation, we intuitively expect that a friend-to-friend contact should have an impact at least comparable to that of a canvassing interaction. Given everything we know about psychology and human behavior, **it does not seem plausible that a single canvass interaction of average quality would be worth several friend-to-friend reminders.**

More careful measurement of contact rates — perhaps via incentivized endline surveys — will be necessary to measure TOT impacts of relational contact more rigorously; and, even then, the large standard errors on the ITT estimates in these generally small-N studies would limit precision. However, the evidence certainly suggests that prevailing estimates in the 1-2pp range are too low.

Appendix

Figures 1 and 2 (Table 5, Row 1)

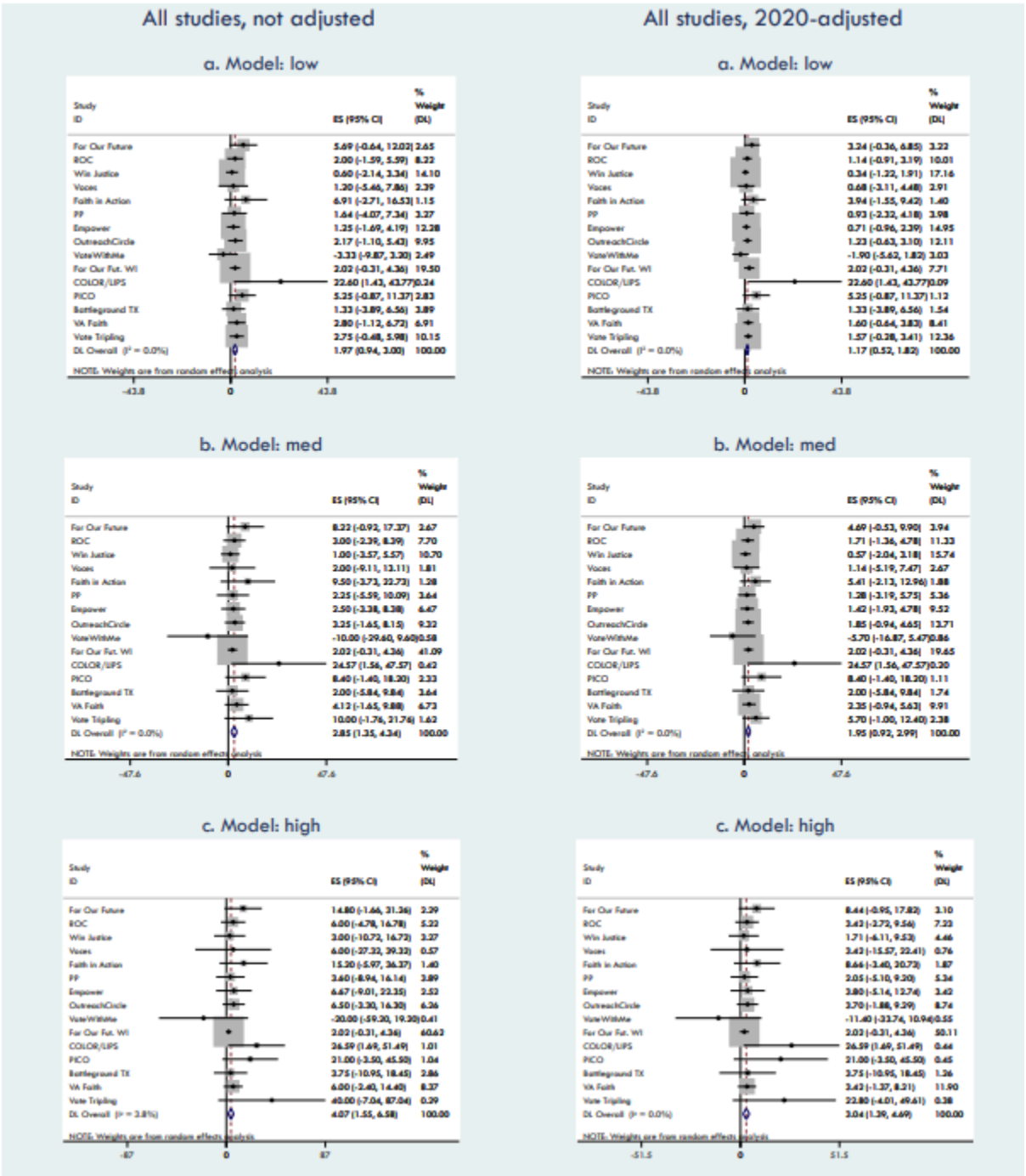
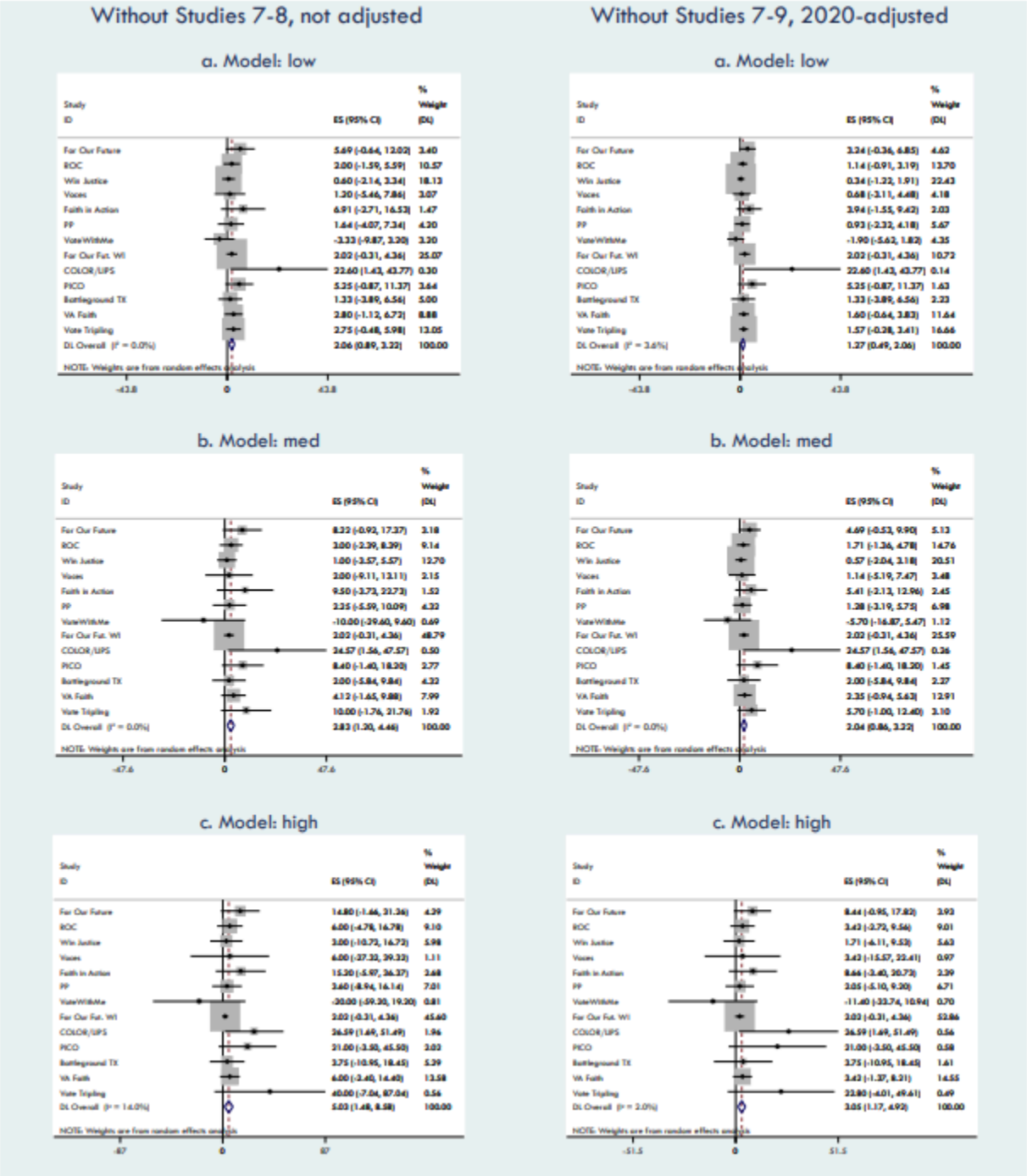
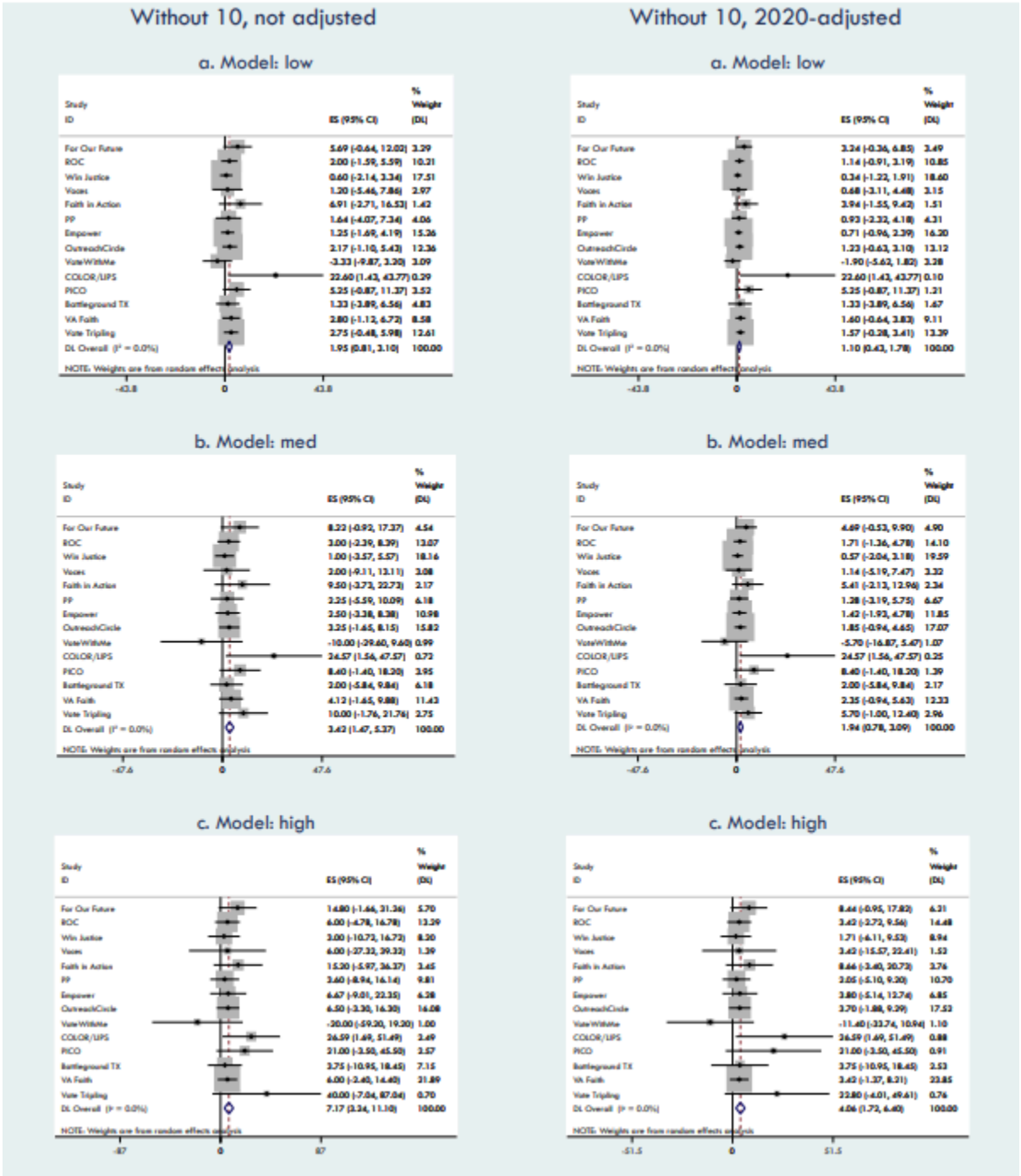


Figure 3 and 4 (Table 5, Row 3)



Figures 5 and 6 (Table 5, Row 4)



Figures 7 and 8 (Table 5, Row 5)