

# Report on Results for Study 2 (aka Survey 5)

Jake Bowers

17 March 2022

## 0.1 The Design

We created a research design where pairs of survey respondents were created so as to be nearly identical on ideology and exactly identical on PID. These pairs were also created so that the overall research design compared favorably with a pair-randomized experiment and otherwise produced extremely homogenous comparisons in terms of age, education, gender, race, partisanship, family income, and missingness on family income.

```
library(here)

## here() starts at /Users/kwilso14/repo/brown/covid/COVID-VaccinesSocialNorms-private
source(here("src", "R", "000_constants_and_utils.R"))
source(here("src/R", "001_rmd_setup.R"))

## Loading required package: survival
## Loading required package: SparseM
##
## Attaching package: 'SparseM'
## The following object is masked from 'package:base':
##
##      backsolve
library(kableExtra)
library(tidyverse)
library(estimatr)
library(coin)

load(file = here(MATCHES_DIR, "dat_plus_matches_study2.rda"), verbose = TRUE)

Loading objects:
  dat5
  tab
  pair_diffs
  xbres_vars
  xbres
load(file = here(MATCHES_DIR, "outcome_analysis_study2.rda"))
```

Our design consists of 996 respondents placed into pairs (so we have 498 pairs). Is this a good research design for our question? Our main goal is to break the relationship between perceptions and potential confounders within pair — that is, we aim to create pairs such that the person who perceives more of their family will be vaccinated is no more or less likely to be the older member of the pair (or the democrat in the pair, or the more well-educated member of the pair, etc.). Within pair comparisons of perceptions in such a design then cannot be said to be confounded by the variables that we balanced.

We present a few pieces of information about our design here.

First, we show balance on average. The following table shows that the pairs have nearly the same mean

values on the key covariates that we aim to balance. That is, the means of the covariates are basically the same across the pairs. Not sure we want to present this as I think that within pair differences probably matter more for us.

```
kable(tab[-nrow(tab), ])
```

	Mean 1	Mean 2	Diffs
ideo5	3.0723	3.1185	-0.0462
dem_rep_oth	2.0602	2.0602	0.0000
age	50.2751	50.4337	-0.1586
female	0.5482	0.5482	0.0000
race_new	1.2430	1.2430	0.0000
faminc_new_imp	6.0772	6.1879	-0.1107
educ	3.7028	3.6847	0.0181
trust_in_govt	2.2406	2.2664	-0.0258
trust_in_science	3.9187	3.9026	0.0161
covid_subj_know	3.5462	3.4137	0.1325
relig_scale	0.5122	0.4977	0.0145

Second, what about within-pair differences? The following table tells us that across the 498 pairs, half differ by less than 1 point on the 5 point ideology scale, less than 1 point on the 6 point education scale, less than 5 year of age, and more than half of the pairs were identical in gender and race. All pairs were identical in terms of party (democrat, republican or other — which combined independents and “other” party responses). Half have identical family income. And 90% are the same on whether or not family income was missing. The worse differences we see are 2 points difference in ideology (for 10% of pairs) and less than 10% of pairs differed by as much as 10 years of age and a difference of 3 points on the 16 point family income scale. We think that these are quite small differences in substantive terms, and again, suggest that we have created a research design that limits the confounding effects of covariates on our targetted relationship.

```
pd_means <- sapply(pair_diffs, mean)
pd_quants <- sapply(pair_diffs, function(x) {
  quantile(abs(x), probs = c(0, .1, .25, .5, .75, .9, 1))
})

pd_tab <- rbind(pd_quants, pd_means)

kableExtra::kable(pd_tab[, -c(1, 9)], digits = 3)
```

	ideo	educ	age	female	race_new	dem_rep_oth	faminc	trust_in_science	trust_in_govt	covid_subj_know	r
0%	0.000	0.000	0.000	0	0	0	0.000	0.000	0.000	0.000	
10%	0.000	0.000	1.000	0	0	0	0.000	0.000	0.000	0.000	
25%	0.000	0.000	2.000	0	0	0	0.955	0.000	0.000	0.000	
50%	1.000	1.000	4.000	0	0	0	1.060	0.500	0.333	1.000	
75%	1.000	1.000	6.000	0	0	0	2.000	0.500	0.667	1.000	
90%	1.000	2.000	7.000	0	0	0	3.000	1.000	1.000	2.000	
100%	1.000	2.000	7.000	0	0	0	3.000	1.000	1.000	3.000	
pd_means	0.022	-0.022	0.118	0	0	0	-0.002	0.022	0.000	-0.024	

The average of the within pair differences is shown on the bottom row of the above table. Again, we see very small differences within pair after creating this design.

Third, we compare our design to an established standard: a pair-randomized experiment. If we were to randomize which member of a pair perceived more versus less vaccination in their social network, we would expect no systematic relationships between perceptions and these covariates within pairs — across many covariates, we would see some relationships, but if we were really to randomize, we would know exactly the distribution that these differences would take. This kind of firm knowledge of a standard for an unconfounded design suggests that we compare our non-randomized design to the equivalent randomized design in the process of arguing in favor of the design. @hansenbowers2008 provide a single test to summarize the relationship between many covariates and a randomized intervention — thus allowing us to avoid the multiple comparisons problem of looking at many different covariate tests. That result for this design provides us with  $p=0.98$  for the test of the hypothesis that the covariate-to-perception relationship is what we would observe in an experiment: that is, we have no strong argument against that null in this case. The subsequent table summarizes the variable-by-variable results. Again, we see that the person who is ranked as perceiving more positive health behaviors is no more or less likely to be older, etc.. That is, those covariates cannot strongly confound any comparisons we make within these pairs. The final column in the table of variable-by-variable results shows the p-values after adjusting for multiple testing using the Holm FWER adjustment.

```
xbres$overall
```

```

      chisquare df p.value
matched 2.568 9 0.979
xbres_vars <- data.frame(xbres$results[, c("rankperc=0", "rankperc=1", "adj.diff", "std.diff", "p"), "matched"])
xbres_vars$p_adjusted <- p.adjust(xbres_vars$p, method = "holm")
kable(arrange(xbres_vars, p), digits = 3)

```

	rankperc.0	rankperc.1	adj.diff	std.diff	p	p_adjusted
trust_in_science	3.900	3.922	0.022	0.022	0.345	1
ideo5	3.084	3.106	0.022	0.016	0.487	1
age	50.295	50.414	0.118	0.007	0.535	1
faminc_new.NATrue	0.114	0.122	0.008	0.025	0.579	1
covid_subj_know	3.492	3.468	-0.024	-0.026	0.653	1
educ	3.705	3.683	-0.022	-0.015	0.669	1
relig_scale	0.503	0.507	0.004	0.013	0.817	1
faminc_new	6.127	6.124	-0.002	-0.001	0.980	1
trust_in_govt	2.253	2.254	0.000	0.001	0.988	1

The above table shows that the average age of the person who perceives more vaccination intentions is, on average, about .4 of a year younger than the person who perceives fewer vaccination intentions — but this difference is not substantively or statistically meaningful when we consider the whole research design. Similarly, we see about 12.7% of those who perceive fewer vaccination intentions to have not reported their family income whereas 11.6% reported family income among those who perceive more vaccination intention. Again, these differences are not distinguishable from those that we would see in a truly randomized experiment. This is not such an experiment, but, in regards an observational study design, it hews closely to the randomized standard.

## 0.2 Outcome Analysis

Our analysis of outcomes is quite simple: we regress reports of respondents' own intentions to be vaccinated (or reported vaccination) on reports of those respondents' perception of such intentions in their social networks, conditional on pair. This amounts to comparing the own-intentions of the higher-perceiver within pair to the own-intentions of the lower-perceiver within pair and taking the average. We report HC2 standard errors.

I am just printing out the results here in raw form. First the scales of the variables to make interpretation easier (the q60 variables were rescaled to have a 0 point for their lowest value). The outcome is 0=not vaccinated and not definitely planning to be vaccinated, 1=vaccinated or definitely planning on it.

```

allvars_used <- unique(unlist(lapply(ls(patt = "~lm_perc"), function(x) {
  all.vars(formula(get(x)))
})))

```

The q60\_ingroup and q60\_outgroup variables exclude people who reported neither democrat nor republican partisanship.

```
kableExtra::kable(summary(dat5[, allvars_used]))
```

	outcome	q60_3a	q60_5a	q60_1a	q60_7a	q60_ingroup	q60_2a	q60_outgroup	q60_6a	q60_4a
Min.	:0.000	Min. :0.0	Min. :0.0	Min. :0.00	Min. :0.00	Min. :0.0	Min. :0.00	Min. :0.0	Min. :0.00	Min. :0.00
1st Qu.	:0.000	1st Qu.:3.0	1st Qu.:4.0	1st Qu.:3.00	1st Qu.:3.00	1st Qu.:3.0	1st Qu.:3.00	1st Qu.:2.0	1st Qu.:2.00	1st Qu.:2.00
Median	:1.000	Median :4.0	Median :5.0	Median :5.00	Median :4.00	Median :5.0	Median :4.00	Median :3.0	Median :3.00	Median :3.00
Mean	:0.663	Mean :3.7	Mean :4.4	Mean :4.01	Mean :3.73	Mean :4.2	Mean :3.63	Mean :3.3	Mean :3.05	Mean :3.05
3rd Qu.	:1.000	3rd Qu.:5.0	3rd Qu.:5.0	3rd Qu.:5.00	3rd Qu.:5.00	3rd Qu.:5.0	3rd Qu.:5.00	3rd Qu.:5.0	3rd Qu.:4.00	3rd Qu.:4.00
Max.	:1.000	Max. :6.0	Max. :6.0	Max. :6.00	Max. :6.00	Max. :6.0	Max. :6.00	Max. :6.0	Max. :6.00	Max. :6.00
NA		NA	NA	NA	NA	NA's :426	NA	NA's :426	NA	NA

Perceptions of vaccination intentions of others strongly predicts own intentions — recall that we are very strongly holding constant ideology here and holding pid exactly constant:

```
lm_perc_fam
```

```

      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
q60_1a 0.08913    0.01312   6.796 3.091e-11 0.06336  0.1149 497
lm_perc_neigh

```

```

      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
q60_2a 0.04962    0.01374   3.61 0.0003375 0.02261  0.07662 497
lm_perc_city

```

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_3a	0.03601	0.01658	2.171	0.03038	0.003425	0.06859	497
lm_perc_state							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_4a	0.02488	0.01823	1.365	0.1729	-0.01094	0.0607	497

We also asked for perceptions of vaccination intentions by Democrats, Republicans and Independents:

```
## Effect of in-group
### Effect of perceptions of democrats among democrats
lm_perc_dem_dem
```

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_5a	0.02429	0.04411	0.5507	0.5825	-0.06274	0.1113	182
lm_perc_rep_rep							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_6a	0.07767	0.0382	2.033	0.04465	0.001891	0.1534	101
lm_perc_dem_rep							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_5a	0.002222	0.03064	0.07252	0.9423	-0.05857	0.06301	101
lm_perc_rep_dem							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_6a	-0.003906	0.02407	-0.1623	0.8712	-0.05139	0.04358	182
lm_perc_dem_ind							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_5a	-0.006441	0.0221	-0.2914	0.771	-0.05001	0.03713	212
lm_perc_rep_ind							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_6a	-0.01347	0.02163	-0.6228	0.5341	-0.05611	0.02917	212
lm_perc_ind_dem							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_7a	-0.01527	0.0244	-0.6258	0.5322	-0.0634	0.03287	182
lm_perc_ind_rep							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_7a	0.06897	0.03469	1.988	0.04949	0.0001568	0.1378	101
lm_perc_ind_ind							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_7a	0.0438	0.02309	1.897	0.05924	-0.001724	0.08932	212

We could also look at the average effect of the ingroup versus outgroup, but I think the preceding results are more clear:

```
## These include only democrats and republicans
lm_percingroup
```

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60ingroup	0.05396	0.02828	1.908	0.05738	-0.001702	0.1096	284
lm_perc_outgroup							

	Estimate	Std. Error	t value	Pr(> t )	CI Lower	CI Upper	DF
q60_outgroup	-0.00104	0.01905	-0.05457	0.9565	-0.03854	0.03646	284

I don't think that the main point of our short paper involves talking about differences in effects across different partisan pairs, but here is a little analysis. First, reduce the dataset to pair differences. For example, below, in bm=1, person 1 has outcome=1, and person 2 has outcome=0 (i.e. outcome=1-0=1), both are republicans, person 1 perceives more people in their family vaccinated, etc.. (i.e. q60\_1a=2 2 is a positive number), etc..

## To look at whether democrat pairs differ from republican pairs, we could reduce the data to pair level:

```
dat5_paired <- dat5 %>%
  group_by(bm) %>%
  summarize(
    outcome = diff(outcome),
    q60_1a = diff(q60_1a),
    q60_2a = diff(q60_2a),
    q60_3a = diff(q60_3a),
    q60_4a = diff(q60_4a),
    q60_5a = diff(q60_5a),
    q60_6a = diff(q60_6a),
    q60_7a = diff(q60_7a),
    q60_ingroup = diff(q60_ingroup),
    q60_outgroup = diff(q60_outgroup),
    pid = unique(dem_rep_oth)
  )

kableExtra::kable(head(dat5_paired))
```

bm	outcome	q60_1a	q60_2a	q60_3a	q60_4a	q60_5a	q60_6a	q60_7a	q60_ingroup	q60_outgroup	pid
1	0	1	0	1	1	4	1	3	1	4	2
2	-1	1	-1	-1	-1	0	2	-1	NA	NA	3
3	1	4	2	3	0	0	0	3	0	0	1
4	0	-2	-1	-2	0	-1	-1	-2	-1	-1	1
5	0	1	0	1	1	0	0	2	0	0	1
6	-1	-3	-2	-1	1	0	0	-1	0	0	2

This way of looking at the data is basically the same as we did above with fixed effects for pairs. Notice the close similarity in this analysis:

## Fixed effects version:  
lm\_perc\_fam

```
Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
q60_1a 0.08913 0.01312 6.796 3.091e-11 0.06336 0.1149 497
```

## Pair-differenced version

```
lm_robust(outcome ~ q60_1a, data = dat5_paired)
```

```
Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
(Intercept) 0.007061 0.02223 0.3177 7.509e-01 -0.03661 0.05073 496
q60_1a 0.089201 0.01312 6.7994 3.029e-11 0.06343 0.11498 496
```

But now we can test for differences in effects between partisan pairs: Effect of ingroup perceptions does not differ appreciably between republicans ( $\text{pid}==2$ ) and democrats (looking at the size of the interaction term ( $\text{q60\_ingroup:I(pid} == 2)\text{TRUE}$ )). We do see, no surprise, that republicans are much less likely to be vaccinated (the  $\text{I(pid} == 2)\text{TRUE}$  term), but this is not central to our paper, so I'm not pursuing the task of statistical tests for group-differences in effects here.

```
lm_robust(outcome ~ q60_ingroup * I(pid == 2), data = dat5_paired, subset = pid != 3)
```

```
Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
(Intercept) 0.04959 0.03219 1.5403 0.1246 -0.01378 0.11296 281
q60_ingroup 0.02489 0.04314 0.5770 0.5644 -0.06003 0.10982 281
I(pid == 2)TRUE -0.07436 0.06565 -1.1326 0.2584 -0.20360 0.05488 281
q60_ingroup:I(pid == 2)TRUE 0.04901 0.05843 0.8388 0.4023 -0.06600 0.16402 281
```

```
lm_robust(outcome ~ q60_outgroup * I(pid == 2), data = dat5_paired, subset = pid != 3)
```

```
Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
(Intercept) 0.049225 0.03235 1.52174 0.1292 -0.01445 0.11290 281
q60_outgroup -0.004099 0.02409 -0.17016 0.8650 -0.05151 0.04331 281
I(pid == 2)TRUE -0.108558 0.06606 -1.64332 0.1014 -0.23859 0.02148 281
q60_outgroup:I(pid == 2)TRUE 0.002365 0.03930 0.06018 0.9521 -0.07500 0.07973 281
```

## 1 Sensitivity Analysis

This study was not randomized, but we are making statistical inferences quantities as if it were. We know that we have some biases here, but are limited in regards our data. Although we have adjusted for some of the largest drivers of perceptions, we have not managed to control them all. How large of an unobserved effect on perceptions have to be in order to over turn our results? We present two modes of sensitivity analysis here.

## 1.1 Hazlett and Cinelli

Our first approach builds on the approach in Cinelli and Hazlett in which they posit a linear regression model and imagine unobserved confounders. They propose a “Robustness Value” or a value of the influence of an unobserved confounder on **both** the outcome and treatment above which our substantive interpretation of our result would change. For example, we print out the results for the effect of perceptions of family below. The key piece for us is the Robustness Value,  $q=1, \alpha=.05$ : this is the value above which we could no longer claim to detect effects (the other values are more liberal — values that would be required to make our relationships exactly zero in magnitude).

```
library(sensemakr)

sens_perc_fam <- sensemakr(
  estimate = lm_perc_fam$coef["q60_1a"],
  se = lm_perc_fam$std.error["q60_1a"], dof = lm_perc_fam$df
)
summary(sens_perc_fam)
```

Sensitivity Analysis to Unobserved Confounding

Model Formula: "Data provided manually"

Null hypothesis:  $q = 1$  and  $\text{reduce} = \text{TRUE}$

-- This means we are considering biases that reduce the absolute value of the current estimate.  
-- The null hypothesis deemed problematic is  $H_0: \tau = 0$

Unadjusted Estimates of 'D':

Coef. estimate: 0.089  
Standard Error: 0.013  
t-value ( $H_0: \tau = 0$ ): 6.796

Sensitivity Statistics:

Partial R2 of treatment with outcome: 0.085  
Robustness Value,  $q = 1$ : 0.262  
Robustness Value,  $q = 1, \alpha = 0.05$ : 0.194

Verbal interpretation of sensitivity statistics:

-- Partial R2 of the treatment with the outcome: an extreme confounder (orthogonal to the covariates) that explains 100% of the residual variance  
-- Robustness Value,  $q = 1$ : unobserved confounders (orthogonal to the covariates) that explain more than 26.2% of the residual variance of both the  
-- Robustness Value,  $q = 1, \alpha = 0.05$ : unobserved confounders (orthogonal to the covariates) that explain more than 19.4% of the residual variance  
sens\_perc\_fam

Sensitivity Analysis to Unobserved Confounding

Model Formula: "Data provided manually"

Null hypothesis:  $q = 1$  and  $\text{reduce} = \text{TRUE}$

Unadjusted Estimates of 'D':

Coef. estimate: 0.089  
Standard Error: 0.013  
t-value: 6.796

Sensitivity Statistics:

Partial R2 of treatment with outcome: 0.085  
Robustness Value,  $q = 1$ : 0.262  
Robustness Value,  $q = 1, \alpha = 0.05$ : 0.194

For more information, check summary.

So, we would need some covariates that are unrelated the variables included in the pairing that would explain more than 20.7 % of the variation in **both** the outcome **and** perceptions in order for our statistical tests to not-reject the null of no effects.

The other Robustness Values are displayed below (notice that the values of 0 are for the models where we could not reject the null of no relationship (for example, the effects of perceptions of democrats among republicans).

All of the values are pretty large: what can we imagine that could predict perceptions and vaccination intention independently of all of the variables in the pairing as strongly as would be required below?

```

get_rv_qa_value <- function(model) {
  res <- sensemakr(
    estimate = model$coef[1],
    se = lm_perc_fam$std.error[1], dof = lm_perc_fam$df
  )
  return(res$sensitivity_stats$rv_qa)
}

lms <- ls(patt = "^lm_perc")
lms <- lms[grepl("rank", lms, invert = TRUE)]
robustness_values <- sapply(lms, function(lmname) {
  get_rv_qa_value(get(lmname))
})

sort(robustness_values, decreasing = TRUE)

```

lm_perc_fam	lm_perc_fam_rep	lm_perc_rep_rep	lm_perc_fam_dem	lm_perc_ind_rep	lm_perc_ingroup	lm_perc_neigh
0.19442	0.18382	0.16238	0.16146	0.13715	0.09179	0.07822
lm_perc_ind_ind	lm_perc_city	lm_perc_dem_dem	lm_perc_dem_ind	lm_perc_dem_rep	lm_perc_ind_dem	lm_perc_outgroup
0.05970	0.03432	0.00000	0.00000	0.00000	0.00000	0.00000
lm_perc_rep_dem	lm_perc_rep_ind	lm_perc_state				
0.00000	0.00000	0.00000				

## 1.2 Rosenbaum Style

Paul Rosenbaum developed an approach to sensitivity analysis that Hazlett and Cinelli build on. It involves positing an unobserved factor that, within pair, changes the probability of “selection into treatment” away from uniform. Our “treatment” variable here has 7 categories, so we change the analysis a bit here: the question becomes whether the person who perceives more within a pair (rank of perceptions = 1) also tends to be the person with more positive vaccination intentions. In this formulation, if two people do not differ in perceptions, then this pair adds nothing to the analysis and they are dropped.

```

dat5 <- dat5 %>%
  group_by(bm) %>%
  mutate(
    q60_1a_rank = rank(q60_1a) - 1,
    q60_2a_rank = rank(q60_2a) - 1,
    q60_3a_rank = rank(q60_3a) - 1,
    q60_4a_rank = rank(q60_4a) - 1,
    q60_5a_rank = rank(q60_5a) - 1,
    q60_6a_rank = rank(q60_6a) - 1,
    q60_7a_rank = rank(q60_7a) - 1
  ) %>%
  ungroup()

## Notice that sometimes the higher perceiver has a 1 and sometimes a 6 (mostly higher numbers of course).
with(dat5, table(q60_1a_rank, q60_1a, exclude = c()))

```

	q60_1a						
q60_1a_rank	0	1	2	3	4	5	6
0	30	67	26	82	70	85	0
0.5	12	22	8	30	22	140	42
1	0	5	11	41	49	115	139

```

lm_perc_fam_rank <- lm_robust(outcome ~ q60_1a_rank, fixed_effects = ~bm, data = dat5)
lm_perc_neigh_rank <- lm_robust(outcome ~ q60_2a_rank, fixed_effects = ~bm, data = dat5)
lm_perc_city_rank <- lm_robust(outcome ~ q60_3a_rank, fixed_effects = ~bm, data = dat5)
lm_perc_state_rank <- lm_robust(outcome ~ q60_4a_rank, fixed_effects = ~bm, data = dat5)
## Do perceptions of Democrats or Republicans matter differently for Democrats/Republicans?
## dem_rep_oth: 1=dem, 2=rep, 3=other (indep, dk, other)
lm_perc_dem_dem_rank <- lm_robust(outcome ~ q60_5a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 1)
lm_perc_dem_rep_rank <- lm_robust(outcome ~ q60_5a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 2)
lm_perc_dem_ind_rank <- lm_robust(outcome ~ q60_5a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 3)
lm_perc_rep_dem_rank <- lm_robust(outcome ~ q60_6a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 1)
lm_perc_rep_rep_rank <- lm_robust(outcome ~ q60_6a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 2)
lm_perc_rep_ind_rank <- lm_robust(outcome ~ q60_6a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 3)
lm_perc_ind_dem_rank <- lm_robust(outcome ~ q60_7a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 1)
lm_perc_ind_rep_rank <- lm_robust(outcome ~ q60_7a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 2)
lm_perc_ind_ind_rank <- lm_robust(outcome ~ q60_7a_rank, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 3)

## These are just to help us understand the results below:
## the pattern of estimates and statistical tests should be the same as when we use the full scale
lm_rank_res_lst <- lapply(ls(patt = "_rank$"), function(lmnm) {
  res <- tidy(get(lmnm))
  res$model <- lmnm
  return(res)
})

```

```

lm_rank_res <- bind_rows(lm_rank_res_lst) %>%
  select(term, estimate, p.value, conf.low, conf.high, model) %>%
  mutate(across(where(is.numeric), round, 3)) %>%
  arrange(p.value)
lm_rank_res

  term estimate p.value conf.low conf.high model
1 q60_1a_rank  0.164  0.000   0.108   0.220 lm_perc_fam_rank
2 q60_2a_rank  0.085  0.002   0.030   0.140 lm_perc_neigh_rank
3 q60_6a_rank  0.167  0.013   0.037   0.297 lm_perc_rep_rep_rank
4 q60_3a_rank  0.065  0.018   0.011   0.119 lm_perc_city_rank
5 q60_7a_rank  0.143  0.033   0.012   0.274 lm_perc_ind_rep_rank
6 q60_7a_rank  0.089  0.047   0.001   0.177 lm_perc_ind_ind_rank
7 q60_4a_rank  0.048  0.094  -0.008   0.104 lm_perc_state_rank
8 q60_7a_rank -0.056  0.162  -0.134   0.022 lm_perc_ind_dem_rank
9 q60_6a_rank -0.038  0.416  -0.131   0.054 lm_perc_rep_ind_rank
10 q60_6a_rank -0.023  0.579  -0.103   0.058 lm_perc_rep_dem_rank
11 q60_5a_rank  0.020  0.670  -0.073   0.113 lm_perc_dem_ind_rank
12 q60_5a_rank  0.000  1.000  -0.090   0.090 lm_perc_dem_dem_rank
13 q60_5a_rank  0.000  1.000  -0.126   0.126 lm_perc_dem_rep_rank

## here is the p-value with no bias
find_gamma_lim(g = 1, y = "outcome", z = "q60_1a_rank", dat = dat5, return = "all")$pval

[1] 1.88e-08
lm_perc_fam_rank ## notice basically same

      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
q60_1a_rank  0.1639    0.02855  5.741 1.642e-08  0.1078    0.22  497
## Some issues with uniroot whtn p>.05
find_gamma_lim(g = 1, y = "outcome", z = "q60_5a_rank", dat = dat5, return = "all")

$pval
[1] 0.3791

$deviate
[1] 0.3078

$statistic
[1] 1.5

$expectation
[1] 0

$variance
[1] 23.75
## Now find the Gamma (or size of the selection effect) required to drive p>=.05
sens_perc_fam_G <- uniroot(f = find_gamma_lim, z = "q60_1a_rank", y = "outcome", dat = dat5, lower = 1, upper = 20)
sens_perc_fam_G$root

[1] 2.182
## Here is the Gamma value: the covariate has to cause the person to be 4 times more likely to be the higher perceiver within the same pair
## Maximum treatment odds ration between higher and lower person within set.
## Showing the the pvalue is .0501 in this case
find_gamma_lim(g = sens_perc_fam_G$root, z = "q60_1a_rank", y = "outcome", dat = dat5, return = "all")$pval

[1] 0.05098
find_gamma_lim(g = 1, z = "q60_6a_rank", y = "outcome", dat = filter(dat5, dem_rep_oth == 2), return = "all")$pval

[1] 0.007153
find_gamma_lim(g = 1, z = "q60_5a_rank", y = "outcome", dat = filter(dat5, dem_rep_oth == 1), return = "all")$pval

[1] 0.5
## Now for the other variables
gamma_vals <- sapply(grep("^q60_[0-9]a.*_rank", names(dat5), value = TRUE), function(nm) {
  message(nm)
  if (nm == "q60_5a_rank") { ## Perceptions of Democrats only potentially non zero among Dems
    res_G <- try(uniroot(f = find_gamma_lim, z = nm, y = "outcome", dat = filter(dat5, dem_rep_oth == 1), lower = 1, upper = 20), silent = TRUE)
  }
  if (nm == "q60_6a_rank") { ## Perceptions of Reps only non zero among Reps
    res_G <- try(uniroot(f = find_gamma_lim, z = nm, y = "outcome", dat = filter(dat5, dem_rep_oth == 2), lower = 1, upper = 20), silent = TRUE)
  }
  if (!(nm %in% c("q60_5a_rank", "q60_6a_rank"))) {
    res_G <- try(uniroot(f = find_gamma_lim, z = nm, y = "outcome", dat = dat5, lower = 1, upper = 20), silent = TRUE)
  }
  if (inherits(x = res_G, "try-error")) {
    res_G$root <- NA
  }
})

```



```

}
return(res_G$root)
})

```

```

Warning in res_G$root <- NA: Coercing LHS to a list
sort(gamma_vals, decreasing = TRUE, na.last = TRUE)

```

```

q60_1a_rank q60_6a_rank q60_2a_rank q60_3a_rank q60_7a_rank q60_4a_rank q60_5a_rank
2.182      1.413      1.318      1.162      1.030      1.008      NA

```

Notice that the stronger the relationship, the larger the bias required to drive the p-value above .05. Strongest effect is family, for example.

Helping to interpret these gamma values

```

amps <- lapply(gamma_vals, function(the_g) {
  if (is.na(the_g)) {
    return(NA)
  }
  message(the_g)
  my_amplify(the_g, lambda = seq(the_g + .001, 10))
})
names(amps) <- names(gamma_vals)

amps_min <- sapply(amps[!is.na(amps)], function(mat) {
  mat[2, ]
})
## We are not looking at Independents because they are to heterogeneous
## So only consider q60_1 to 6 (not 7).
## The 4a result arises from using ranks (p=.08). So excluding that one too.
amps_min[, c("q60_1a_rank", "q60_2a_rank", "q60_3a_rank", "q60_6a_rank")]

```

```

      q60_1a_rank q60_2a_rank q60_3a_rank q60_6a_rank
delta      5.939      2.054      1.511      2.408
lambda     3.183      2.319      2.163      2.414

```

The p-value that we see for a given  $\Gamma$  (like `r gamma_vals[[1]]`) can be produced from an unobserved covariate that increases the treatment odds (within all pairs) by  $\lambda$  and increases the odds of a positive pair difference in outcomes by  $\delta$ . So, if we had such a covariate that increases the odds of being the higher perceiver by about 9 it would also have to increase the odds of own vaccination intention by about 4.4 in order to yield a p-value of .051.

## 2 Relationships with Correlations

```

## among dems, perceptions of dems
with(dat5[dat5$dem_rep_oth == 1, ], cor(outcome, q60_5a))

```

```

[1] 0.3285

```

```

## among reps, perceptions of dems
with(dat5[dat5$dem_rep_oth == 2, ], cor.test(outcome, q60_5a))

```

Pearson's product-moment correlation

```

data: outcome and q60_5a
t = -0.52, df = 202, p-value = 0.6
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.1731 0.1013
sample estimates:
cor
-0.03656

```

```

## among dems, perceptions of reps
with(dat5[dat5$dem_rep_oth == 1, ], cor.test(outcome, q60_6a))

```

Pearson's product-moment correlation

```

data: outcome and q60_6a
t = -1.4, df = 364, p-value = 0.2
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.17457 0.02936
sample estimates:

```

```

cor
-0.07337
## among reps, perceptions of reps
with(dat5[dat5$dem_rep_oth == 2, ], cor(outcome, q60_6a))

[1] 0.3919
lm_robust(outcome ~ q60_5a * I(dem_rep_oth == 1), data = dat5, subset = dem_rep_oth != 3)

              Estimate Std. Error t value    Pr(>|t|) CI Lower CI Upper  DF
(Intercept)      0.54239    0.10860   4.9944 0.0000007875  0.32908  0.75570 566
q60_5a           -0.01232    0.02407  -0.5119 0.6089420740 -0.05961  0.03496 566
I(dem_rep_oth == 1)TRUE -0.30811    0.17032  -1.8090 0.0709820144 -0.64264  0.02643 566
q60_5a:I(dem_rep_oth == 1)TRUE 0.14205    0.03581   3.9664 0.0000823232  0.07171  0.21239 566
lm_robust(outcome ~ q60_5a, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 1)

              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_5a  0.02429    0.04411  0.5507  0.5825 -0.06274  0.1113 182
lm_robust(outcome ~ q60_5a, fixed_effects = ~bm, data = dat5, subset = dem_rep_oth == 2)

              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_5a  0.002222    0.03064  0.07252  0.9423 -0.05857  0.06301 101

```

### 3 References