Report on Results for Study 2 (aka Survey 5)

Jake Bowers

20 September 2021

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/jwbowers/Documents/PROJECTS/ThePolicyLab/COVID-VaccinesSocialNorms-private
```

```
source(here("src", "R", "000_constants_and_utils.R"))
source(here("src/R", "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	3.0783	3.1124	-0.0341
dem_rep_oth	2.0602	2.0602	0.0000
age	50.2410	50.3876	-0.1466
female	0.5482	0.5482	0.0000
race_new	1.2430	1.2430	0.0000
faminc_new_Imp	6.0891	6.1779	-0.0887
educ	3.7108	3.6847	0.0261
trust_in_govt	2.2426	2.2651	-0.0224
trust_in_science	3.9106	3.9056	0.0050
covid_subj_know	3.5522	3.4197	0.1325
relig_scale	0.5082	0.4998	0.0084

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 idential 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
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	0.940	0.000	0.000	0.000
50%	0.500	1.000	4.000	0	0	0	1.030	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	0.026	-0.022	-0.034	0	0	0	0.024	0.013	-0.002	0.004

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=0.99 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.

```r
xbres$overall
```

```
        chisquare df p.value
matched     1.863  9  0.9935
```

```r
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
ideo5	3.082	3.108	0.026	0.018	0.410	1
relig_scale	0.499	0.509	0.009	0.031	0.575	1
trust_in_science	3.902	3.915	0.013	0.013	0.577	1
educ	3.709	3.687	-0.022	-0.015	0.674	1
faminc_new.NATRUE	0.114	0.120	0.006	0.019	0.674	1
faminc_new	6.114	6.139	0.024	0.008	0.766	1
age	50.331	50.297	-0.034	-0.002	0.858	1
covid_subj_know	3.484	3.488	0.004	0.004	0.939	1
trust_in_govt	2.255	2.253	-0.002	-0.003	0.940	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 percieve 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.

```r
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.

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

	outcome	q60_3a	q60_5a	q60_1a	q60_7a	q60_ingroup	q60_2a	q60_outgroup	q60_6a	q6
	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	M
	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	1s
	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	M
	Mean :0.661	Mean :3.7	Mean :4.4	Mean :4.01	Mean :3.74	Mean :4.2	Mean :3.64	Mean :3.3	Mean :3.06	M
	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	3r
	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	M
	NA	NA	NA	NA	NA	NA's :426	NA	NA's :426	NA	N

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:

```r
lm_perc_fam
```

```
       Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
q60_1a  0.09558    0.01315   7.268 1.427e-12  0.06974   0.1214 497
```

```r
lm_perc_neigh
```

```
       Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
q60_2a  0.05282    0.01391   3.799 0.0001635   0.0255  0.08014 497
```

```r
lm_perc_city
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_3a  0.03843    0.01636   2.349   0.0192  0.00629  0.07057 497
```

```r
lm_perc_state
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_4a  0.02162    0.01804   1.198   0.2313 -0.01382  0.05706 497
```

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

```r
## 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  0.02521    0.04581  0.5504   0.5827 -0.06517   0.1156 182
```

```r
### Effect of perceptions of rep among rep
lm_perc_rep_rep
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_6a  0.09464    0.03908   2.422  0.01722  0.01712   0.1722 101
```

```r
## Effect of out-group
### Effect of perceptions of democrats among republicans
lm_perc_dem_rep
```

```
        Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_5a -0.004474    0.03184 -0.1405   0.8885 -0.06764  0.05869 101
```

```r
### Effect of perceptions of republicans among democrats
lm_perc_rep_dem
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_6a -0.01181     0.0244 -0.4842   0.6289 -0.05994  0.03632 182
```

```r
## Effects effect of perceptions of dems and reps among "other" group
lm_perc_dem_ind
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_5a  -0.0032    0.02095 -0.1528   0.8787 -0.04449  0.03809 212
```

```r
lm_perc_rep_ind
```

```
        Estimate Std. Error   t value Pr(>|t|) CI Lower CI Upper  DF
q60_6a 3.103e-18    0.02166 1.433e-16        1  -0.0427   0.0427 212
```

```r
## Effects of perceptions of independents among dem, rep, and independents
lm_perc_ind_dem
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_7a -0.01592    0.02597 -0.6128   0.5407 -0.06715  0.03532 182
```

```r
lm_perc_ind_rep
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_7a  0.07692    0.03507   2.194  0.03056 0.007357   0.1465 101
```

```r
lm_perc_ind_ind
```

```
       Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_7a  0.05157    0.02201   2.342  0.02009 0.008171  0.09496 212
```

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

```r
## These include only democrats and republicans
lm_perc_ingroup
```

```
            Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_ingroup  0.06486    0.02903   2.235  0.02622 0.007728    0.122 284
```

```r
lm_perc_outgroup
```

```
              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q60_outgroup -0.008377    0.01958 -0.4278   0.6691 -0.04692  0.03017 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..

```r
## 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)
  )
```

```r
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	0	0	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:

```r
## Fixed effects version:
lm_perc_fam
```

```
       Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
q60_1a  0.09558    0.01315   7.268 1.427e-12  0.06974   0.1214 497
```

```r
## Pair-differenced version
lm_robust(outcome ~ q60_1a, data = dat5_paired)
```

```
             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
(Intercept) -0.005922    0.02190 -0.2704 7.870e-01 -0.04895  0.03711 496
q60_1a       0.095537    0.01316  7.2592 1.514e-12  0.06968  0.12139 496
```

But now we can test for differences in effects between partisan pairs: Effect of ingroup perceptions does not differ appreciably between republicans (`pid==2`) and democrats (looking at the size of the interaction term (`q60_ingroup:I(pid == 2)TRUE`)). We do see, no surprise, that republicans are much less likely to be vaccinated (the `I(pid == 2)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.

```r
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)                  0.04839    0.03279  1.4758   0.1411 -0.01616  0.11294 281
q60_ingroup                  0.02399    0.04523  0.5305   0.5962 -0.06503  0.11301 281
I(pid == 2)TRUE             -0.09460    0.06679 -1.4164   0.1578 -0.22606  0.03687 281
q60_ingroup:I(pid == 2)TRUE  0.06351    0.06043  1.0509   0.2942 -0.05544  0.18245 281
```

```r
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)                   0.049731    0.03229  1.54012  0.12466 -0.01383  0.113292 281
q60_outgroup                 -0.012594    0.02440 -0.51617  0.60614 -0.06062  0.035435 281
I(pid == 2)TRUE              -0.141608    0.06834 -2.07202  0.03918 -0.27614 -0.007079 281
q60_outgroup:I(pid == 2)TRUE  0.001337    0.04068  0.03287  0.97380 -0.07873  0.081408 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 reduce = TRUE
-- This means we are considering biases that reduce the absolute value of the current estimate.
-- The null hypothesis deemed problematic is H0:tau = 0

Unadjusted Estimates of 'D':
  Coef. estimate: 0.096
  Standard Error: 0.013
  t-value (H0:tau = 0): 7.268

Sensitivity Statistics:
  Partial R2 of treatment with outcome: 0.096
  Robustness Value, q = 1: 0.277
  Robustness Value, q = 1, alpha = 0.05: 0.211

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 27.7% of the residual variance of both th

-- Robustness Value, q = 1, alpha = 0.05: unobserved confounders (orthogonal to the covariates) that explain more than 21.1% of the residual varia
```

```
sens_perc_fam
```

```
Sensitivity Analysis to Unobserved Confounding

Model Formula: "Data provided manually"

Null hypothesis: q = 1 and reduce = TRUE

Unadjusted Estimates of ' D ':
  Coef. estimate: 0.096
  Standard Error: 0.013
  t-value: 7.268

Sensitivity Statistics:
  Partial R2 of treatment with outcome: 0.096
  Robustness Value, q = 1 : 0.277
  Robustness Value, q = 1 alpha = 0.05 : 0.211

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[grep("rank", lms, invert = TRUE)]
robustness_values <- sapply(lms, function(lmname) {
  get_rv_qa_value(get(lmname))
})

sort(robustness_values, decreasing = TRUE)
```

```
 lm_perc_fam_rep      lm_perc_fam  lm_perc_rep_rep  lm_perc_fam_dem  lm_perc_ind_rep  lm_perc_ingroup    lm_perc_neigh
         0.22074          0.21118          0.20867          0.16351          0.15965          0.12447          0.08781
 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
         0.08391          0.04195          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  78  68  88   0
        0.5  12  24   8  32  22 140  44
        1     0   5  11  42  48 113 138
```

```
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)
})
```

```r
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    0.182   0.000    0.127     0.238     lm_perc_fam_rank
2  q60_2a_rank    0.098   0.001    0.043     0.153   lm_perc_neigh_rank
3  q60_3a_rank    0.083   0.002    0.029     0.137    lm_perc_city_rank
4  q60_6a_rank    0.208   0.003    0.073     0.344 lm_perc_rep_rep_rank
5  q60_7a_rank    0.115   0.007    0.031     0.199 lm_perc_ind_ind_rank
6  q60_7a_rank    0.169   0.014    0.035     0.303 lm_perc_ind_rep_rank
7  q60_4a_rank    0.048   0.084   -0.006     0.102   lm_perc_state_rank
8  q60_7a_rank   -0.047   0.240   -0.124     0.031 lm_perc_ind_dem_rank
9  q60_6a_rank   -0.037   0.355   -0.117     0.042 lm_perc_rep_dem_rank
10 q60_5a_rank    0.028   0.539   -0.061     0.117 lm_perc_dem_ind_rank
11 q60_5a_rank    0.000   1.000   -0.092     0.092 lm_perc_dem_dem_rank
12 q60_5a_rank    0.000   1.000   -0.130     0.130 lm_perc_dem_rep_rank
13 q60_6a_rank    0.000   1.000   -0.090     0.090 lm_perc_rep_ind_rank
```

```r
## 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] 4.837e-10
```

```r
lm_perc_fam_rank ## notice basically same
```

```
            Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
q60_1a_rank   0.1821    0.02821   6.453 2.601e-10   0.1266   0.2375 497
```

```r
## Some issues with uniroot whtn p>.05
find_gamma_lim(g = 1, y = "outcome", z = "q60_5a_rank", dat = dat5, return = "all")
```

```
$pval
[1] 0.3366

$deviate
[1] 0.4216

$statistic
[1] 2

$expectation
[1] 0

$variance
[1] 22.5
```

```r
## 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] 2.551
```

```r
## 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] 0.05099
```

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

```
[1] 0.001946
```

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

```
[1] 0.5
```

```r
## 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
      2.551       1.670       1.443       1.336       1.187       1.019          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(theg) {
  if (is.na(theg)) {
    return(NA)
  }
  message(theg)
  my_amplify(theg, lambda = seq(theg + .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        8.055       2.524       2.122       3.457
lambda       3.552       2.444       2.337       2.671
```

The p-value that we see for a given Γ (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] 0.33
```

```
## 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 = -0.73, df = 202, p-value = 0.5
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.18758  0.08645
sample estimates:
     cor
-0.05153
```

```
## 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 = -1.8, df = 364, p-value = 0.08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.19269  0.01061
sample estimates:
```

```
   cor
-0.092
```

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

```
[1] 0.3985
```

```r
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)                     0.55905    0.10945  5.1080 0.0000004456  0.34408 0.774019 566
q60_5a                         -0.01739    0.02419 -0.7192 0.4722993041 -0.06490 0.030109 566
I(dem_rep_oth == 1)TRUE        -0.32831    0.17110 -1.9188 0.0555146237 -0.66439 0.007765 566
q60_5a:I(dem_rep_oth == 1)TRUE  0.14779    0.03592  4.1141 0.0000446348  0.07723 0.218354 566
```

```r
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  0.02521    0.04581  0.5504   0.5827 -0.06517   0.1156 182
```

```r
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 -0.004474    0.03184 -0.1405   0.8885 -0.06764  0.05869 101
```

3 References