TAMANI Project DD

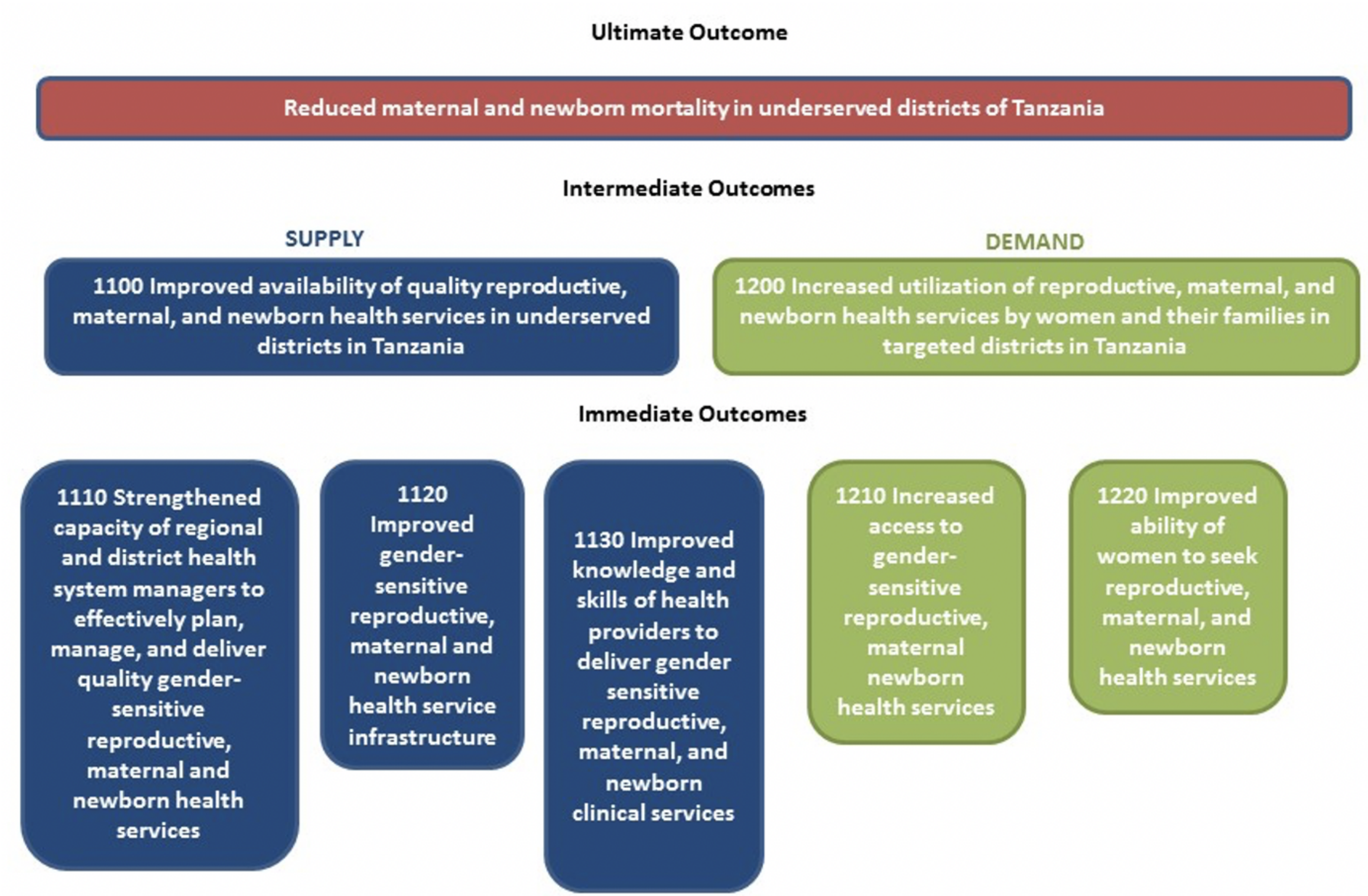
Sam Harper

joint with Erin H., Ari, Ilona, Holly

(all errors are mine)

2021-11-23

Tabora Maternal Newborn Health Initiative (TAMANI)



Tabora Maternal Newborn Health Initiative (TAMANI)

Interventions

- Training health care providers in basic and comprehensive Emergency Obstetric and Newborn Care and family planning (EmONC)
- Community Health Worker training
- Community-based dialogues focused on gender and women's empowerment

Main outcomes

- Unmet needs for family planning
- Skilled birth attendance
- 4+ antenatal visits
- Contraceptive prevalence
- Respectful maternity care

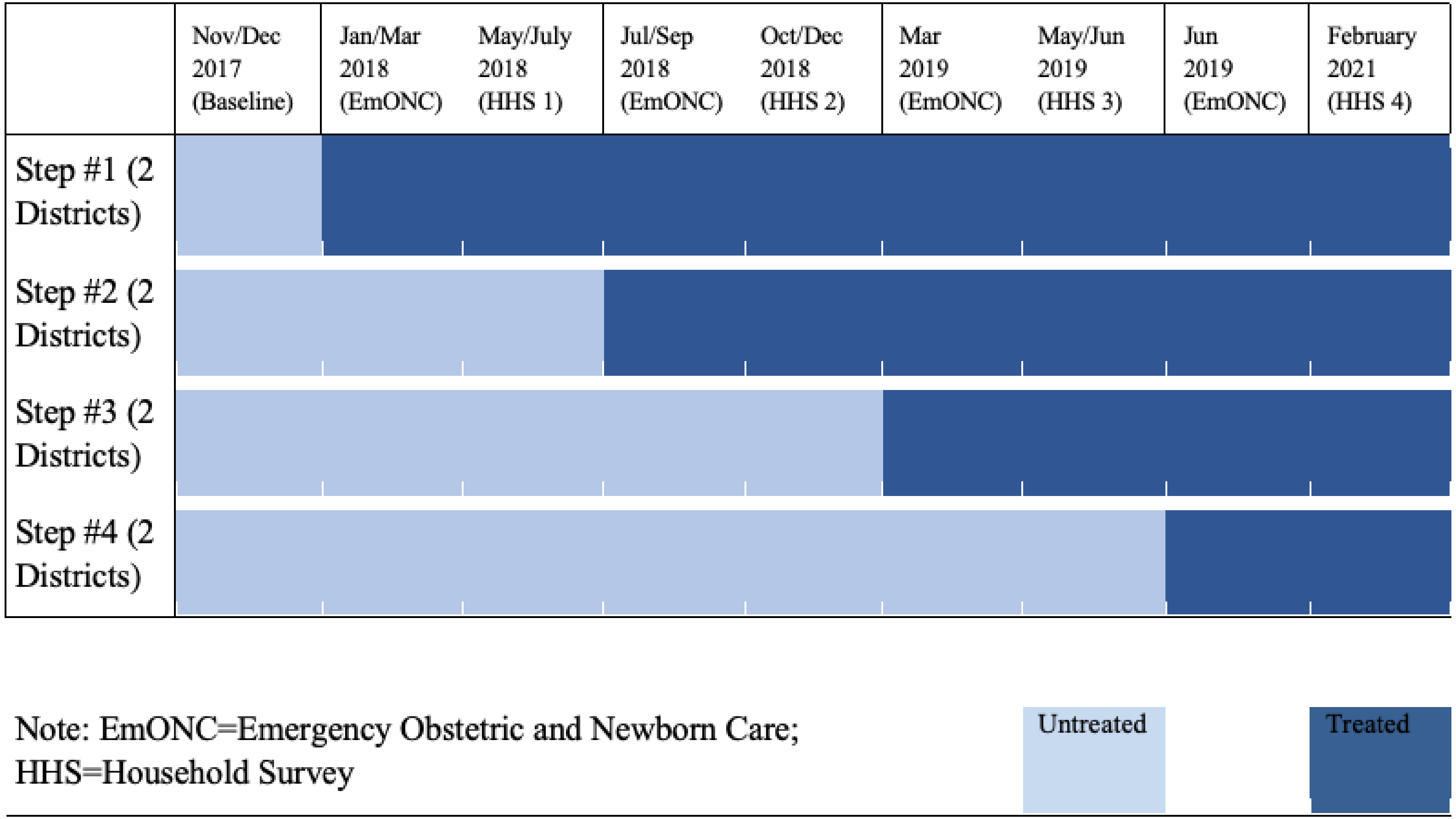
Tabora Maternal Newborn Health Initiative (TAMANI)

- Implemented by CARE Canada/Tanzania in 8 districts in Tabora region, Tanzania

Source: https://commons.wikimedia.org/w/index.php?curid=47130439, Google Maps

Initial Design: Stepped Wedge

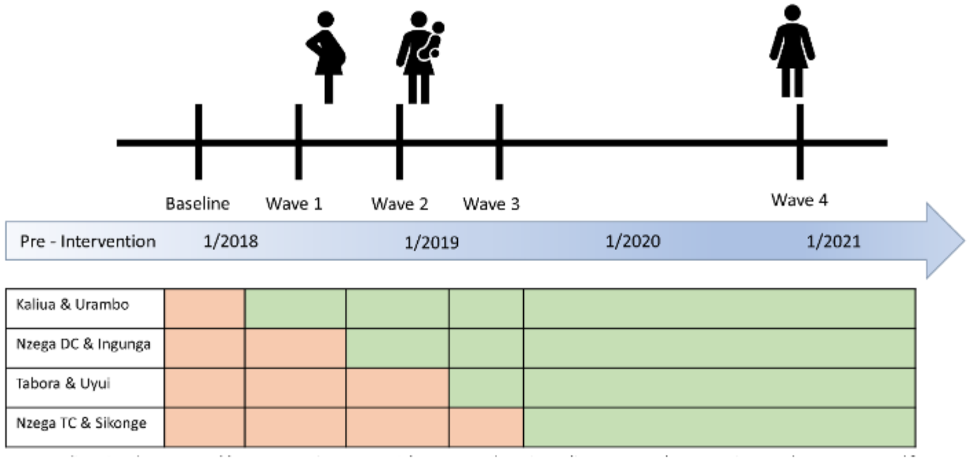
- Political constraints led to breaking random timing.
- Switching of order of two districts.
- Analyze as DD?



A note about defining 'treated'

- *Treated* at wave 4 For immediate outcomes (contraceptive prevalence, unmet need for family planning).
- Same woman reports on a child that was born in January 2019 (before the intervention was implemented).
- *Untreated* at wave 4 for for Jan '19 delivery outcomes (skilled birth attendance and respectful care).
- Pregnancy began in the spring of 2018, so pregnancy outcomes (antenatal care visits) are considered “untreated”.

Respondent at wave 4 (January 2021).



Basic setup for DD with variable timing

We have different districts that are exposed to our intervention at different times. We often use OLS (or LPM) to fit:

$$y_{it} = \alpha_i + \tau_t + \beta^{DD} D_{it} + \epsilon_{it}$$

where

- y_{it} is the outcome for unit i at time t.
- α_i are unit-specific fixed effects.
- τ_t are fixed effects for each time period.
- D_{it} is a time-varying treatment indicator.
- β^{DD} is the difference-in-differences estimate.

Key points from Goodman-Bacon (2019)

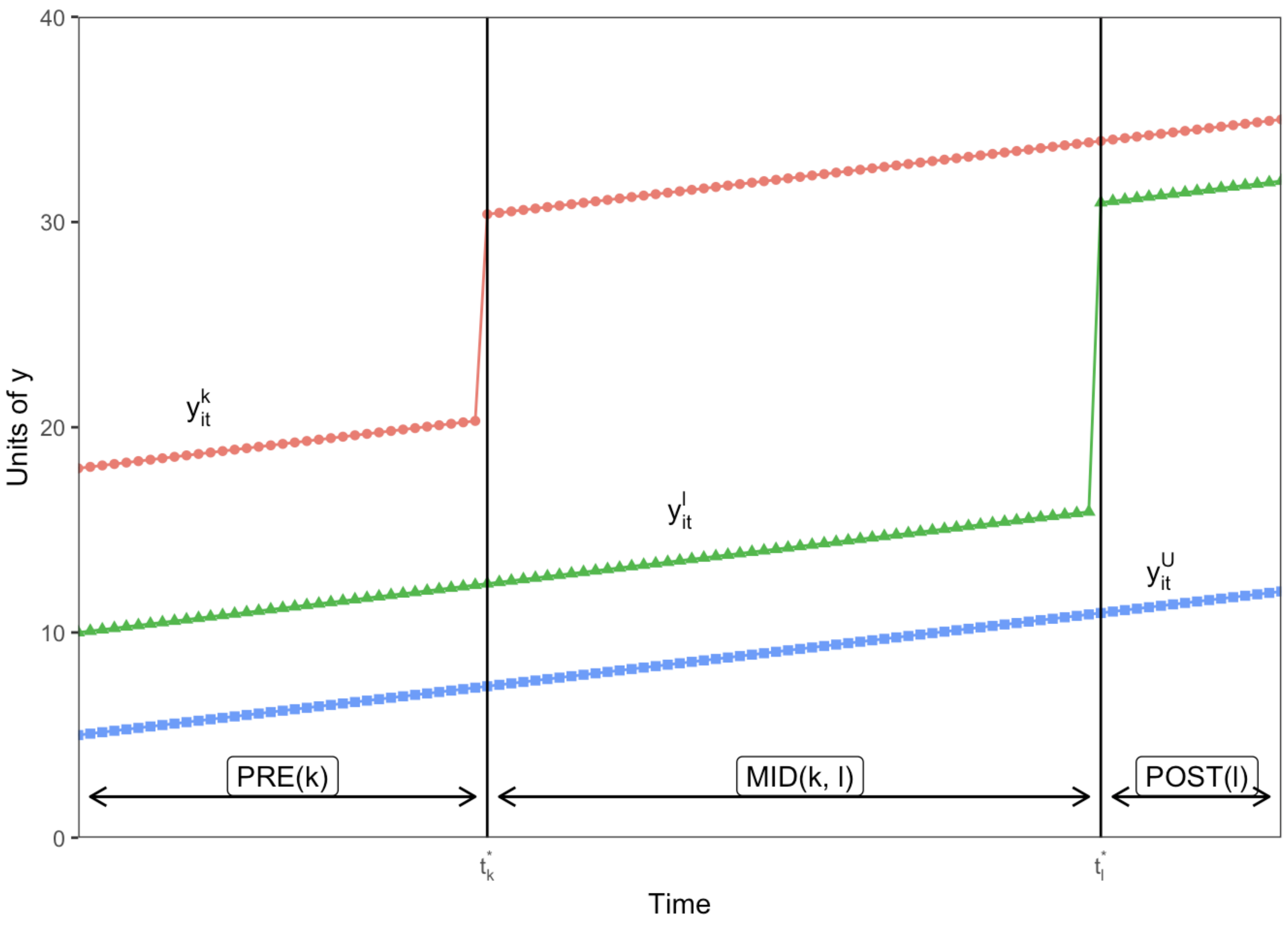
- With OLS, DD with treatment timing is a variance-weighted average of many 2x2 ATTs.
- Weights are a function of both group sizes *and* variances.
- Can lead to β^{DD} that is a poor summary of group-specific effects.

1. Early-adopters (*k*) vs. never treated (*U*)

2. Later-adopters (*l*) vs. never treated (*U*).

3. Early (*k*) vs. later (*l*) adopters.

4. **Later (l) vs. earlier (k) adopters.**



Graph from https://andrewcbaker.netlify.app/2019/09/25/difference-in-differences-methodology/

Data structure: Individual-level

- sba_birth = SBA present
- txdel = treated
- time = survey wave
- pid = person ID
- group = time when group first treated

district	sba_birth	txdel	time	dist_id	p_id	group
Kaliua DC	0	0	1	1	237	2
Kaliua DC	1	1	5	1	2168	2
Kaliua DC	0	1	5	1	6723	2
Kaliua DC	0	0	1	1	2269	2
Kaliua DC	1	1	5	1	5084	2
Kaliua DC	1	1	5	1	2058	2

Data structure: Pooled by district

- tsba = total SBA births
- tpop = total pop (births)
- txdel = treated
- time = survey wave
- group = time when group first treated

district	dist_id	time	tsba	tpop	psba	txdel	group
Kaliua DC	1	1	409	705	0.580	0	2
Kaliua DC	1	2	49	73	0.671	1	2
Kaliua DC	1	3	63	77	0.818	1	2
Kaliua DC	1	4	9	10	0.900	1	2
Kaliua DC	1	5	94	114	0.825	1	2
Nzega DC	2	1	465	634	0.733	0	3
Nzega DC	2	2	60	76	0.789	0	3
Nzega DC	2	3	43	57	0.754	1	3
Nzega DC	2	4	9	10	0.900	1	3

TWFE models (OLS)

- Individual and aggregate basically identical.
- Clustered SEs approximately the same.
- Intervention increased the Pr(SBA) by 6 pp (95% CI -1.3 to 13.3).

$$y_{sba} = \alpha + \beta * txdel + \gamma_{district} + \delta_{time} + \epsilon$$

	Individual	**Aggregate**
txdel	0.060	0.060
	(0.037)	(0.040)
	[-0.013, 0.133]	[-0.019, 0.138]
Num.Obs.	5555	40
Std. Errors	Clustered (district)	Clustered (district)
FE: district	X	X
FE: time	X	X

Callaway-Sant'Anna Approach: Group-Time cohorts

- The CS approach starts with defining Group-Time cohorts.
- Groups defined by when they were *first* treated.

	P(SBA)					**Total Pop**				
Group	**Time1**	**Time2**	**Time3**	**Time4**	**Time5**	**Time1**	**Time2**	**Time3**	**Time4**	**Time5**
2	0.631	0.711	0.814	0.846	0.826	898	90	102	13	144
3	0.678	0.770	0.664	0.800	0.784	1455	165	125	30	232
4	0.778	0.807	0.673	0.750	0.868	1128	140	101	16	167
5	0.854	0.939	0.854	0.800	0.855	556	66	41	10	76

- Different aggregation schemes for ATTs are possible.
- Can allow for covariates via regressions adjustments, IPW and DR.

Callawy-Sant'Anna implementation

- Includes options for different structure, SE calculation, weights, etc.

```
# Use not-yet-treated as comparison group
did::att_gt(yname = "sba_birth", # name of the LHS variable
  tname = "time", # name of the time variable
  idname = "p_id", # name of the id variable
  gname = "group", # name of the first treatment period
  data = d_ind, # dataset
  xformla = NULL, # conditional parallel-trends
  weightsname = NULL, # can add weights
  est_method = "reg", # estimation method
  control_group = "notyettreated", # set the control group
  bstrap = TRUE, # compute bootstrapped SE
  biters = 1000, # bootstrap iterations
  print_details = FALSE, # if TRUE, print detailed results
  panel = FALSE, # panel or repeated cross-sectional
  clustervars = NULL) # cluster ID
```

See https://bcallaway11.github.io/did/articles/multi-period-did.html for R, https://econpapers.repec.org/software/bocbocode/S458976.htm for Stata

Estimates from CS approach

- Note there is no *overall* estimate, lots of heterogeneity.
- Each treatment group has an ATT at each time period.
- Can be combined to produce different aggregate ATTs.

term	group	time	estimate	std.error	conf.low	conf.high
ATT(2,2)	2	2	0.011	0.053	-0.094	0.116
ATT(2,3)	2	3	0.260	0.057	0.149	0.371
ATT(2,4)	2	4	0.269	0.174	-0.072	0.610
ATT(3,2)	3	2	0.045	0.043	-0.039	0.129
ATT(3,3)	3	3	0.018	0.070	-0.118	0.155
ATT(3,4)	3	4	0.170	0.160	-0.145	0.484
ATT(4,2)	4	2	-0.062	0.043	-0.147	0.022
ATT(4,3)	4	3	-0.048	0.082	-0.208	0.112
ATT(4,4)	4	4	0.130	0.186	-0.234	0.495

Note:

P-value for pre-test of parallel trends assumption: 0.31

Re-creating the Group-Time ATTs

- ATT(2,2) means estimating ATT *at time 2* for the group *first treated at time 2*
- For ATT(2,2) we are comparing Pr(SBA) between:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:
Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g22	time1	time2	Long diff	ATT_2_2
0	0.745	0.814	0.069	**NA**
1	0.631	0.711	0.080	**0.011**

This estimate says that intervention increased the probability of an SBA birth by 0.01 for Group 2 at Time 2.

- For ATT(2,3) the groups being compared are:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- ATT(2,3) means estimating the ATT *at time 3* for the group *first treated at time 2*
- Need to exclude any group treated at time 3 to avoid bias.
- The group first treated at time 3 is excluded.
- Provides an estimate of the lagged impact of intervention.

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:
Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g23	time1	time3	Long diff	ATT_2_3
0	0.803	0.725	-0.078	**NA**
1	0.631	0.814	0.182	**0.26**

Etc., etc., etc...

- For ATT(4,2) the groups being compared are:

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- ATT(4,2) means estimating the ATT *at time 2* for the group *first treated at time 4*
- Need to exclude any group treated at time 2 to avoid bias.
- The group first treated at time 2 is excluded.
- Provides an estimate of the lead effects or non-parallel trends.

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g42	time1	time2	Long diff	ATT_4_2
0	0.727	0.818	0.091	**NA**
1	0.778	0.807	0.029	**-0.062**

Our 'hand-calculated' ATT(4,2) of -0.062 is the same as the regression-based CS estimate:

term	group	time	estimate	std.error	conf.low	conf.high
ATT(2,2)	2	2	0.011	0.055	-0.098	0.120
ATT(2,3)	2	3	0.260	0.057	0.149	0.372
ATT(2,4)	2	4	0.269	0.161	-0.046	0.585
ATT(3,2)	3	2	0.045	0.043	-0.039	0.130
ATT(3,3)	3	3	0.018	0.064	-0.107	0.144
ATT(3,4)	3	4	0.170	0.156	-0.136	0.476
ATT(4,2)	4	2	-0.062	0.046	-0.152	0.027
ATT(4,3)	4	3	-0.048	0.083	-0.211	0.115
ATT(4,4)	4	4	0.130	0.189	-0.241	0.502

"Final" ATT(4,4) compares only 2 groups (4 clusters):

group	time1	time2	time3	time4	time5
2	0.631	0.711	0.814	0.846	0.826
3	0.678	0.770	0.664	0.800	0.784
4	0.778	0.807	0.673	0.750	0.868
5	0.854	0.939	0.854	0.800	0.855

Note:

Red = treated, Gray = untreated

- The 2x2 (weighted averages):

g44	time3	time4	Long diff	ATT_4_4
0	0.854	0.80	-0.054	**NA**
1	0.673	0.75	0.077	**0.13**

Questions / Help

Aggregating

What is the right way to conceptualize aggregating the ATTs?

Clustering

How to think about the 'right' standard error?

Dynamic Effects

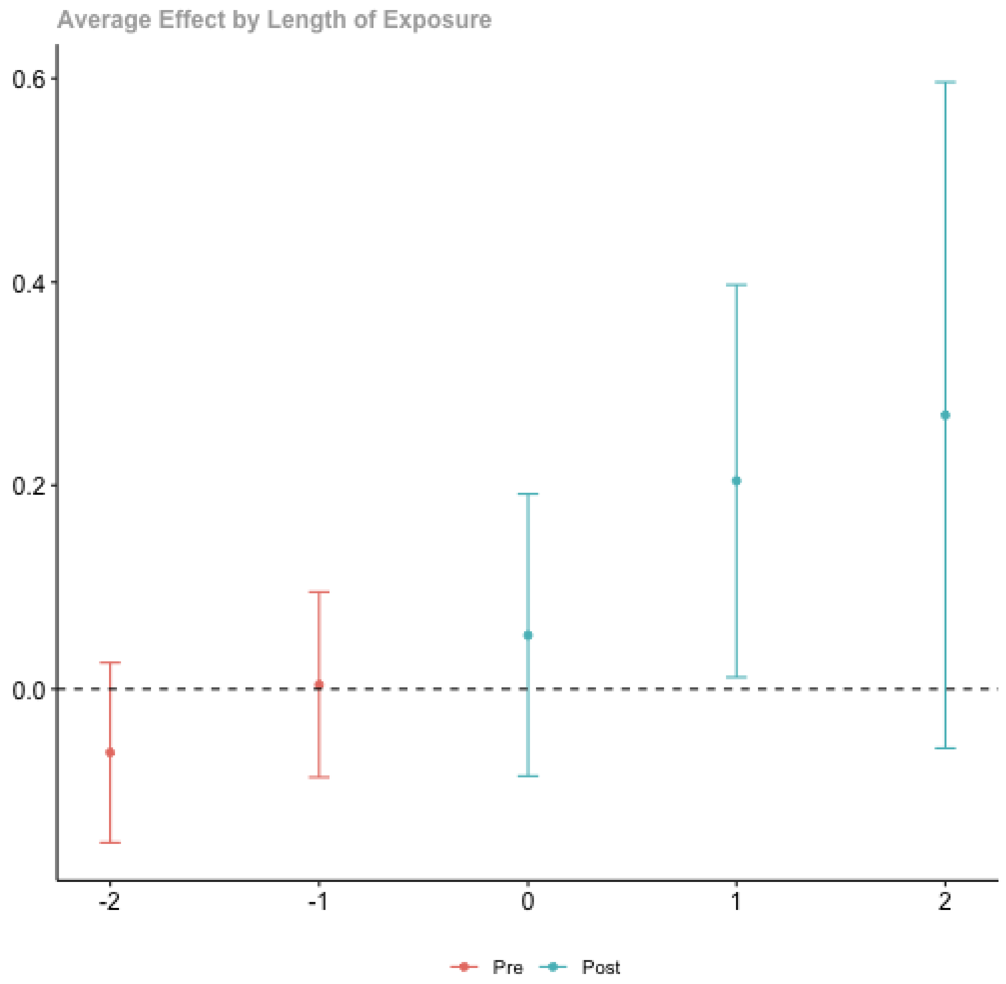
- Estimates by length of exposure
- Overall summary:

$$\overline{ATT} = 0.176(-0.02, 0.37)$$

event.time	estimate	std.error	conf.low	conf.high
-2	-0.062	0.045	-0.151	0.026
-1	0.004	0.047	-0.087	0.096
0	0.053	0.071	-0.086	0.192
1	0.204	0.098	0.012	0.397
2	0.269	0.167	-0.058	0.596

- TWFE estimate was

$$\overline{ATT} = 0.06(-0.013, 0.133)$$



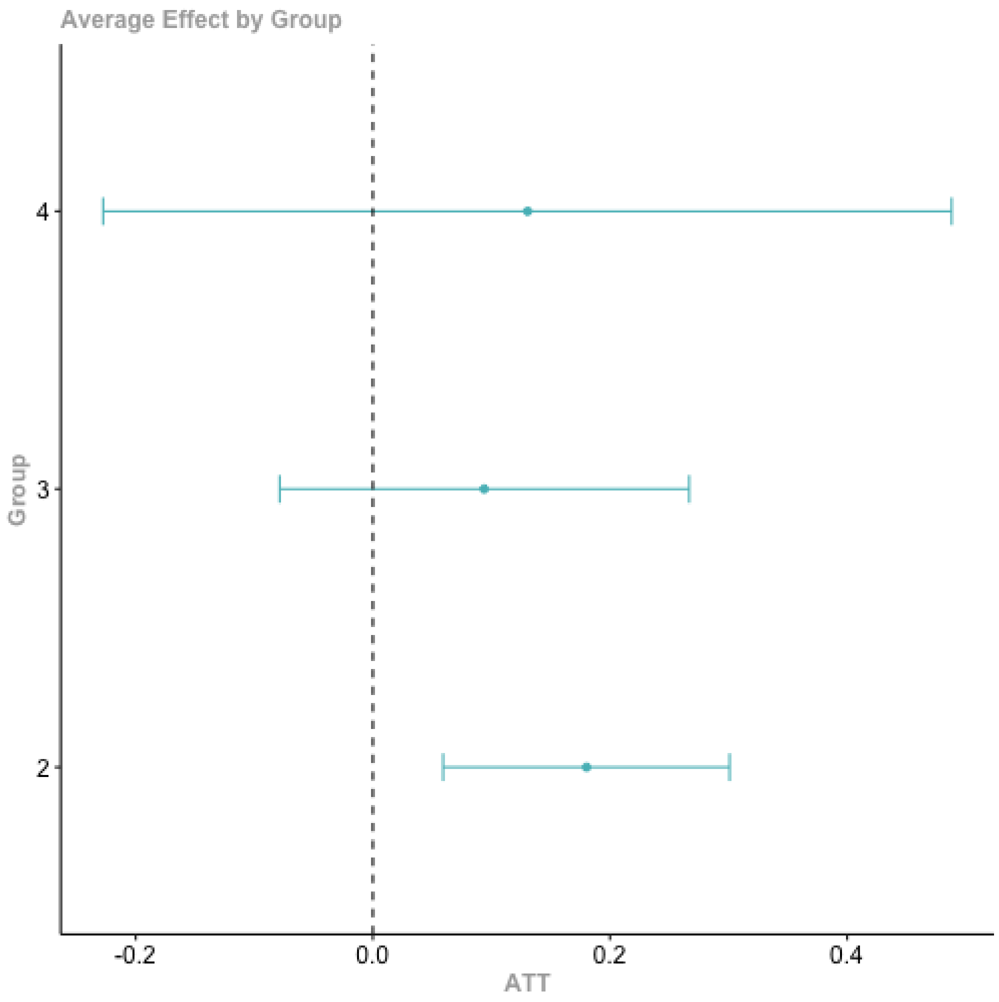
Group-specific ATTs

- Group-specific estimates
- Weighted avg of each group-time ATT for group g

group	estimate	std.error	conf.low	conf.high
Average	0.128	0.097	-0.061	0.318
2	0.180	0.062	0.059	0.301
3	0.094	0.088	-0.079	0.267
4	0.130	0.183	-0.228	0.488

- TWFE estimate was

$$\overline{ATT} = 0.06(-0.013, 0.133)$$



Calendar Time ATTs

- ATT in time period t for groups that have participated in the treatment by time period t.

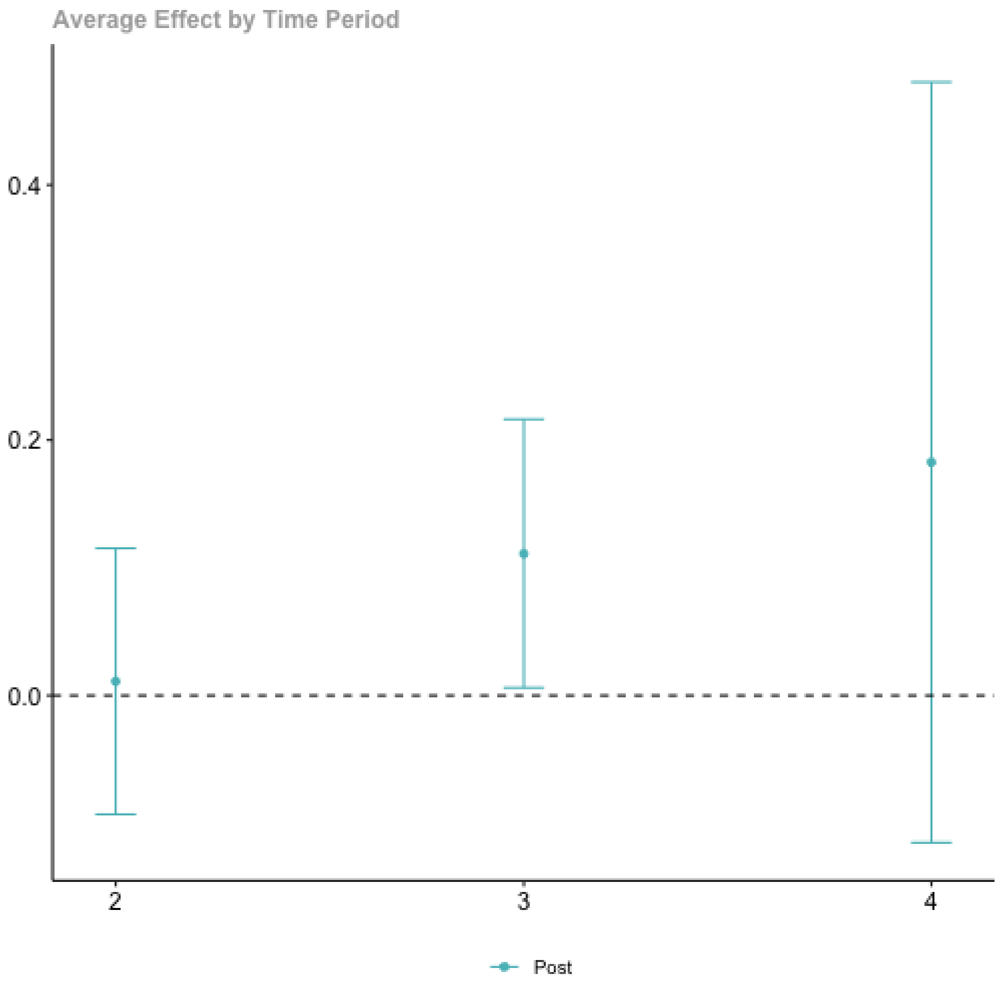
- Overall summary:

$$\overline{ATT} = 0.102(-0.02, 0.22)$$

time	estimate	std.error	conf.low	conf.high
2	0.011	0.053	-0.093	0.115
3	0.111	0.054	0.006	0.216
4	0.183	0.152	-0.115	0.480

- TWFE estimate was

$$\overline{ATT} = 0.06(-0.013, 0.133)$$



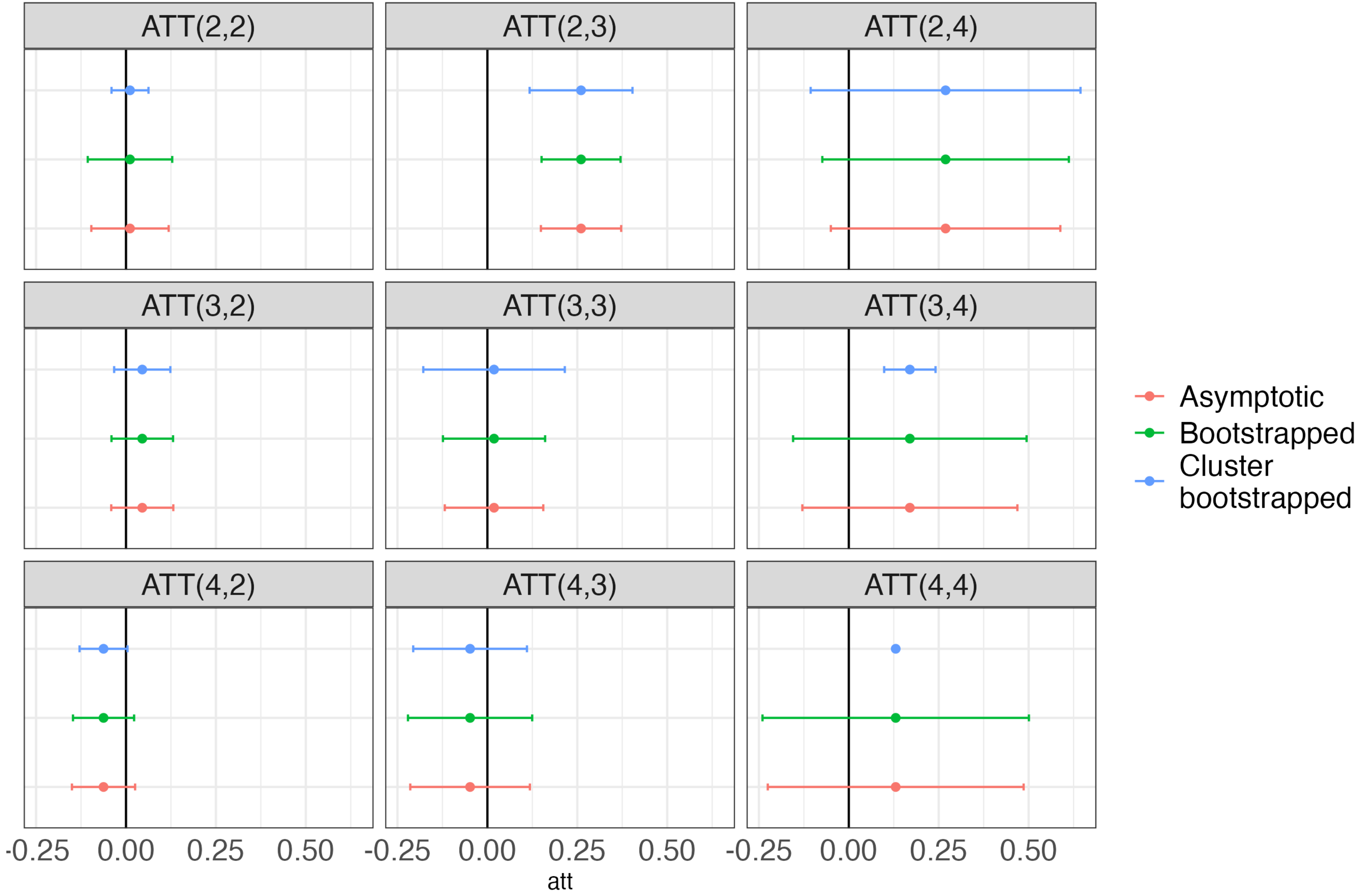
Options for Standard errors

- Asymptotic
- Bootstrap
- Cluster bootstrap

Comment from the Callaway-Sant'Anna paper:

Remark 10. In DiD applications, it is common to use "cluster-robust" inference procedures; see, e.g., Wooldridge (2003) and Bertrand et al. (2004). However, we note that the choice of whether to cluster or not is usually not obvious, and depends on the kind of uncertainty one is trying to reflect; see, e.g., Abadie et al. (2017) for a discussion in a cross-sectional setup.[14] In the case that one wishes to account for clustering to reflect "cluster-based" sampling uncertainty, we note that this can be done in a straightforward manner using a small modification of the multiplier bootstrap described above, provided that the number of cluster is "large." More precisely, instead of drawing observation-specific V's, one simply needs to draw cluster-specific V's; see, e.g., Sherman and Le Cessie (2007), Kline and Santos (2012), Cheng et al. (2013), and MacKinnon and Webb (2018, 2020). If the number of clusters is "small," however, the application of the aforementioned bootstrap procedure is not warranted.[15]

- Some SEs hard to understand.
- Cluster bootstrapped smaller in most cases.
- Absurdly small for ATT(4,4) that has small sample and only 4 clusters.

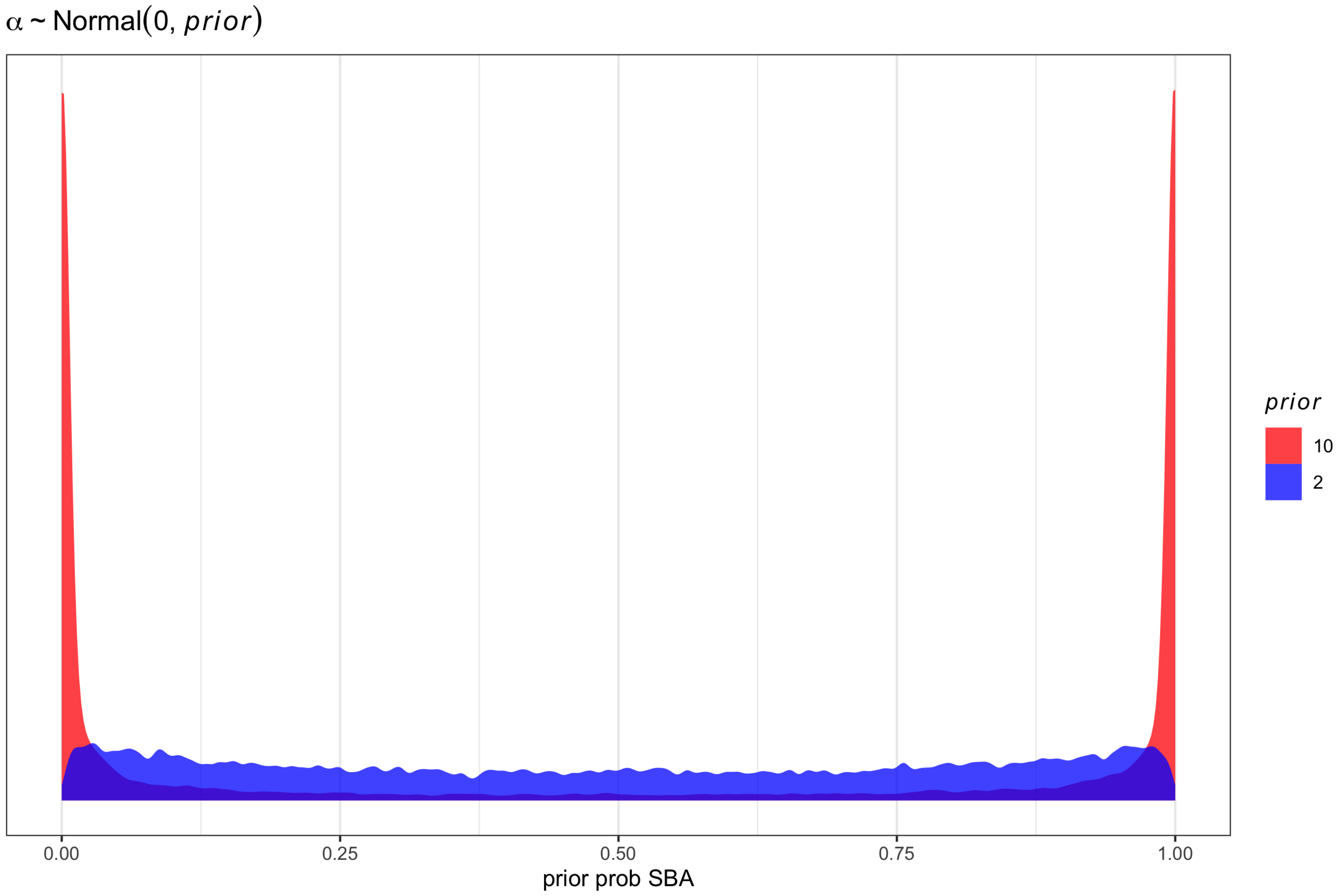


Other attempts to estimate ATT(4,4)

- Cluster-adjustment always excessively narrow.
- Ignoring clustering also seems wrong.
- Bayesian estimation more conservative.

Method	ATT(4,4)	SE	95% CI
Diff in proportions	0.130	0.182	-0.228, 0.489
Diff in proportions (svy clustered)	0.130	0.002	0.123, 0.138
OLS, robust SE	0.130	0.184	-0.232, 0.493
OLS, cluster robust SE	0.130	0.002	0.123, 0.138
Logit ME, robust	0.123	0.150	-0.171, 0.418
Logit ME, cluster robust	0.123	0.015	0.093, 0.154
Bayesian logit, flat priors	0.120	0.130	-0.230, 0.330
Bayesian logit, regularizing priors	0.060	0.100	-0.160, 0.230

Bayesian priors for intercept



Bayesian priors for ATT