Online Appendix: Supporting Information

Technical

Proof of Unbiasedness of Trajectory Balancing Estimator

The proof in this section follows very closely from Hazlett and Xu 2018, and is meant to illustrate the key findings in their paper for the purpose of our study.

First, we restate the three assumptions in the main text:

Assumption 1. Conditional ignorability $Y_{it}(0) \perp \text{Cohort}_i | \mathbf{Y}_{i,pre} \quad \forall t > T_0$

Assumption 2. Linearity in Prior Outcomes (LPO) $E[Y_{it}(0)|\mathbf{Y}_{i,pre}] = (1, Y_{i,pre})^\top \theta_t + \eta_{it}$ for $E[\eta_{it}|\mathbf{Y}_{i,pre}] = 0$

Assumption 3. Weight feasibility: there exists a set of weights w_i that are non-negative and sum to 1 such that:

$$\frac{1}{N_{\text{Cohort}_i \neq \infty}} \sum_{\text{Cohort}_i \neq \infty} y_{it} = \sum_{\text{Court}_i = \infty} w_i y_{it} \text{ for } t \leq \text{Year Before Treatment}$$

The argument we seek to prove is:

$$E[\widehat{ATT}|\mathbf{Y}_{pre}] = \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[Y_{it}|Cohort \neq \infty, \mathbf{Y}_{pre}] - \sum_{Cohort = \infty} E[w_i Y_{it}|Cohort = \infty, \mathbf{Y}_{pre}]$$

We can rewrite Y_{it} as the sum of three components: the treatment effect of courts $\tau_{it} = Y_{it}(1) - Y_{it}(0)$, $Y_{it}(0)$, and an error term η_{it}. For simplicity, let $Cohort_{cohort}, Y_{pre}] =..$ Then we can write the next line as

$$= \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[\tau_{it} + Y_{it}(0) + \eta_{it}|.] - \sum_{Cohort = \infty} E[w_i Y_{it}(0) + w_i \eta_) it|.]$$

which is a simple substitution of the term Y_{it}.

Next, we can rearrange terms and break up the expectations:

$$= \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[\tau_{it}|.] + \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[Y_{it}(0)|.] + \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[\eta_{it}|.] - \sum_{Cohort = \infty} E[w_i Y_{it}(0)|.] + \sum_{Cohort = \infty} E[w_i \eta_{it}|.]$$

Under Assumption 1, we can drop the Cohort subscripts:

$$= ATT_t + \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[Y_{it}(0)|\mathbf{Y}_{pre}] + \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[\eta_{it}|\mathbf{Y}_{pre}] - \sum_{Cohort=\infty} E[w_i Y_{it}(0)|\mathbf{Y}_{pre}] + \sum_{Cohort=\infty} E[w_i \eta_{it}|\mathbf{Y}_{pre}]$$

Then under Assumption 2:

$$= ATT_t + \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} (1, Y_{i,pre})^\top \theta_t + \frac{1}{N_{tr}} \sum_{Cohort \neq \infty} E[\eta_{it}|\mathbf{Y}_{pre}] - \sum_{Cohort=\infty} w_i (1, Y_{i,pre})^\top \theta_t + \sum_{Cohort=\infty} E[w_i \eta_{it}|\mathbf{Y}_{pre}]$$

Under Assumption 3:

$$\sum_{Cohort \neq \infty} (1, Y_{i,pre})^\top \theta_t = \sum_{Cohort=\infty} w_i (1, Y_{i,pre})^\top \theta_t$$

which allows us to write:

$$\underbrace{ATT_t}_{\text{Treatment Effect}} + \underbrace{E\Big[\frac{1}{N_{\text{Court}_i<\infty}} \sum_{\text{Court}_i<\infty} \eta_{it} - \sum_{\text{Court}_i=\infty} w_i \eta_{it}|\mathbf{Y}_{pre}\Big]}_{\text{Bias}}$$

Under our assumptions, $E[\eta_{it}|\mathbf{Y}_{pre}] = 0$, meaning shocks are mean zero conditional on the history of the outcome trajectory, so bias is zero. We test this assumption by estimating a version of equation (1) where ATT_t is zero by construction, meaning any difference detected between groups on average can be attributed to the bias term. However, we find that there is no effect in this subsample, which provides support for our assumption. Then, we have

$$E[\widehat{ATT}_t|\mathbf{Y}_{pre}] = \underbrace{ATT_t}_{\text{Treatment Effect}}$$

Under our assumptions, we have proven that

$$E[\widehat{ATT}_t|Y_{pre}] = ATT_t$$

Illustration that Uncertainty over Start Dates Biases Against a Result

There is a possibility that some Taliban courts measured in 2011 were established some time earlier. This may be concerning, since our research design depends on the timing and location of Taliban court introduction. However, the bias resulting from any measuring error will attenuate the size of treatment effects - meaning they will make it harder to find evidence that Taliban courts have an impact. We illustrate a proof below.

Recall that the balancing constraint is:

$$\frac{1}{N_{tr}} \sum_{Cohort=2011} y_{it} = \sum_{Cohort=\infty} w_i y_{it}$$

If it was the case that courts existed prior to 2011, y_{it} for the treatment group can be defined as the sum of the outcome under control $Y_{it}(0)$ and the treatment effect τ_{it}. Then we can rewrite the balancing equation as:

$$\frac{1}{N_{tr}} \sum_{Cohort=2011} Y_{it}(0) + \tau_{it} = \sum_{Cohort=\infty} w_i Y_{it}(0) \quad \text{for} \quad t \leq t_0$$

If this is the case, then the estimator for the ATT can be written as before:

$$\frac{1}{N_{tr}} \sum_{Cohort=2011} y_{it} - \sum_{Cohort=\infty} w_i y_{it} \quad \text{for} \quad t > t_0$$

However, note that w_i has been solved for in a way that absorbs τ_{it}. If τ_{it} is constant overtime, then the observed effect will be precisely zero, since the control group was reweighted to equal the treated group after treatment. If τ_{it} grows stronger overtime, the treatment effect will be nonzero, but will be smaller in magnitude than the true effect, because control units that have trajectories similar to treated units after treatment will be given higher weight.

Descriptive

Survey Questions from ANQAR

Sampling Design of ANQAR

The ANQAR survey is stratified by province, covering all of the 34 first-level administrative units in Afghanistan. The administrative district is the primary sampling unit.[75] Sampled districts are selected via a probability proportional to size

75. One exception is Kabul district (the capital), which is subdivided into additional survey units due to the size of the city, which accounts for roughly 13% of the country's population. For consistency, we utilize the administrative boundary designation for Kabul rather than these subdivisions.

approach. After districts have been selected for sampling, secondary sampling units composed of villages and settlements are randomly selected.[76] After the sampling set has been identified and before fielding a survey wave, ACSOR engages with local elders to secure permission for enumerators to enter sample villages. Once enumerators arrive at a village, a random walk method is used to identify target households. Once a household is selected, a Kish grid is used to randomize the respondent within each selected household. Prior work by condra2019civilians indicates that response rates in ANQAR are on par with other large surveys in more peaceful contexts.

Government Strength Questions

TABLE A1. *ANQAR Survey Questions*

Question	Coverage	Concept
How well does the [Government of Afghanistan, Province Governor, District Governor] of Afghanistan do its job? Is it good, fair, or bad? (Overall, Security, Economy, Development, Corruption, Essential Services)	Wave 1 - Wave32	Performance Index (15 Total Questions)
Between the two, the Anti- Government Elements (Mukhalafeen-e dawlat) and the Government, who has more influence in your mantaqa now?	Wave 1 - Wave 32	Control (1 Total Question)

Combat Data

ANSO data was collected from over 100 enumerators, who collected event information through "humanitarian and development NGOs, triangulated SMS and phone-in reports, and local news media"[77]. The effort was funded by a variety of nongovernmental organizations for the purpose of providing security incident information to keep aid workers safe.

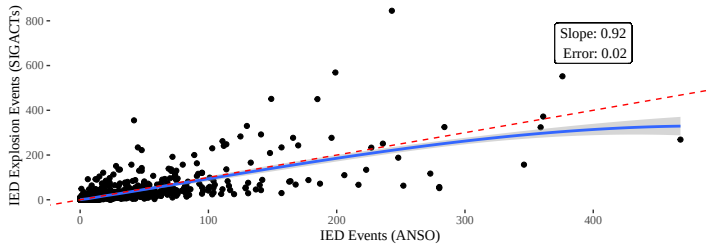
We validate our measures by measuring the correlation between the same combat outcome - improvised explosive device explosion incidents (IEDs) - across datasets. We report the scatter plot and regression in Figure A1. The slope coefficient is .9, and

76. ACSOR maintains a list of villages and settlements, which is used as the sampling frame. It is the most comprehensive list currently available.

77. Sexton 2016.

the linear and nonlinear fits of the data show a strong positive relationship between datasets. The fact that two different data collection methods are picking up on the same patterns is suggestive that different data collection methods are tracing the same pattern, increasing our confidence that ANSO is reliable.

FIGURE A1. *Database Comparison: SIGACTs IED Events Against ANSO IED Events*



Scatter plot of the IED Explosions against IED events from ANSO. Dashed line is 45 degree line (slope of 1, perfect correlation). Thick line is the LOESS fit which is highly linear. Linear regression coefficient reported in the top right corner with slope of 0.92.

Measuring Pre-Existing Taliban Influence (Control)

We rely on three measures of Taliban influence in the pre-2011 period to adjust our estimates.

Forward Operating Bases

We use the average number of forward operating bases in a district from 2008-2010.

Direct Survey Questions

ANQAR surveys directly ask respondents who has more control over the district: the government or the Taliban. We use the average responses to this question per district from the 2008-2010 period.

Background on ACSOR Control Data

Afghan Center for Socio-economic and Opinion Research (ACSOR) is a survey firm in Afghanistan responsible for fielding a wide variety of surveys on numerous topics across Afghanistan. They have a pool of over 1,000 interviewers, who come from diverse ethnic backgrounds and are both male and female.[78].

During the process of sampling villages from districts for surveys, ACSOR logged if a place was accessible, and if it was not, what the reason was for the accessibility. One of the reasons for inaccessibility was denial by the Taliban. ACSOR collected this data monthly at the district level for their internal purposes.

To measure control, we consider a district under Taliban control if the Taliban was able to deny access to a village in a district. The ability to make a place illegible to the state control par excellence in a coutnerinsurgnecy war, as it demonstrates the ability of an armed actor to deny external actors access to information.

To avoid posttreatment bias, we use the data from 2010 to measure control by the Taliban.

Relationship between ACSOR Control Data and ANQAR Survey Data

A natural concern from using survey data in a conflict zone is as follows: does the process of warfighting disrupt survey enumeration in a way that may confound one's results? We note any study that attempts to leverage public opinion data during wartime may fall prey to potential bias from inaccessibility due to the survey process.

Since our study has data on places that where access was disrupted due to Taliban influence, we can directly adjust our estimates for confounding from the data collection process, which is a feature unique to our study.

78. For more information on the survey teams, quality control, and general topics about the organization visit: ACSOR's website

Importantly, in cases where survey teams are unable to visit a village, they do not simply give up on trying to assess opinion in a place. Instead, the teams use intercept surveys to collect information, which allows them to solicit information about the district even when the Taliban has influence. Again, if one is concerned that this difference in the collection process may influence our results since it could be correlated with courts, we note that since we adjust for Taliban control through a direct measure of enumeration access, we account for this source of confounding.

Empirical

Combat Per Capita

TABLE A2. *Combat Per Capita (Logged)*

Outcome	(1) AOG	(2) IED	(3) DF	(4) IED Explosions	(5) Casualty Events
Taliban Courts	0.18**	0.24***	0.22**	0.23***	0.31***
	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)
Dataset	ANSO	ANSO	SIGACT	SIGACT	SIGACT
N. Districts	339	339	339	339	339
N. Years	6	6	6	6	6
Standard Deviation DV	1.86	1.68	1.96	1.65	1.57
Mean DV (Control)	2.14	1.69	1.8	1.26	1.13

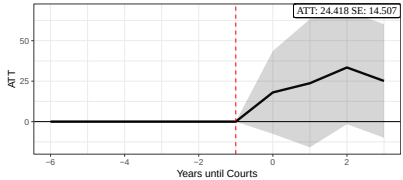
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Note: Combat outcomes transformed by per capita (attacks per 100,000 people) transformed with the natural log (plus one).

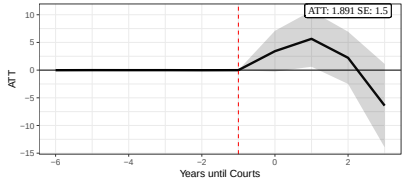
Combat Over Longer Time Period (2006-2014)

FIGURE A2. *Armed Conflict Trends: Trajectory Balanced 2006-2014*

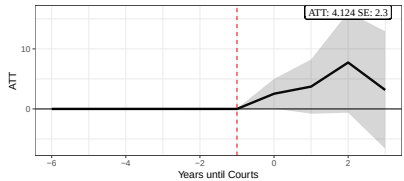
(a) *Direct Fire (SIGACT)*



(b) *IED Explosions (SIGACT)*



(c) *Casualty Events (SIGACT)*



Note: Mean balancing results using the levels of SIGACTs for the period of 2006-2014. ATT by period plotted. X-axis is the years until courts, the y-axis is the ATT, shading is the jackknifed standard error. Pooled ATT and standard error reported in the upper right corner.

Relaxing Conditional Ignorability

TABLE A3. *Civilian Attitude Results: Covariates*

Outcome	(1) State Court	(2) Taliban Approval	(3) Gov. Influence	(4) Gov. Index
Taliban Courts	-0.07^{***}	0.04^{**}	-0.07^{***}	-0.72^{***}
	(0.02)	(0.01)	(0.02)	(0.12)
N Districts	170	170	194	187
N. Years	6	6	7	7
SD DV	0.18	0.18	0.21	1.74
Mean DV	0.46	0.21	0.71	-0.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

TABLE A4. *Combat Results: Covariates*

Outcome	(1) AOG	(2) IED	(3) DF	(4) IED Explosions	(5) Casualty Events
Taliban Courts	11.6^{**}	5.86^{*}	32.46^{**}	7.68^{*}	9.82^{***}
	(3.67)	(2.92)	(11.37)	(3.26)	(2.88)
Dataset	ANSO	ANSO	SIGACT	SIGACT	SIGACT
N. Districts	339	339	339	339	339
N. Years	6	6	6	6	6
Standard Deviation DV	44.42	33.34	126.42	36.51	28.33
Mean DV (Control)	16.39	11.61	14.41	6.25	4.4

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

LPO: Kernel Balancing

TABLE A5. *Civilian Attitude Results: Kernel Balancing*

	(1) State Court	(2) Taliban Approval	(3) Gov. Influence	(4) Gov. Index
Taliban Courts	-0.0699***	0.0313	-0.0807***	-0.8431†
	(0.018)	(0.02)	(0.02)	(0.45)
N. Districts	170	170	194	187
N Years	6	6	7	7
SD DV	0.18	0.18	0.21	1.74
Mean DV	0.46	0.21	0.71	-0.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Kernel balancing results with covariates. The outcomes are civilian attitudes measured by ANQAR.

TABLE A6. *Combat Results: Kernel Balancing*

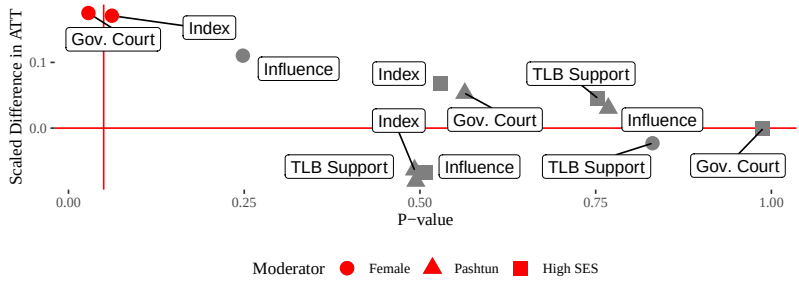
Outcome	(1) AOG	(2) IED	(3) DF	(4) IED Explosions	(5) Casualty Events
Taliban Courts	5.12	4.14†	35.29*	5.38	6.71*
	(3.3)	(2.41)	(14.03)	(3.47)	(3.1)
Dataset	ANSO	ANSO	SIGACT	SIGACT	SIGACT
N. Districts	338	338	338	338	338
N. Years	6	6	6	6	6
Standard Deviation DV	42.9	31.34	92.23	29.23	22.26
Mean DV (Control)	16.39	11.61	14.41	6.25	4.4

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Kernel balancing results with covariates. The outcomes are combat measured by ANSO and SIGACTs.

Attitude Heterogeneous Effects

FIGURE A3. *Difference in ATT by Subgroup*



Each shape is a different moderator (gender, ethnicity, or SES). The vertical axis is the scaled difference in ATT between groups - for instance, the circle shape labeled female is the difference in ATT between female and male respondents. The horizontal axis is the p-value calculated from a t-test for difference in coefficients. Each label shows what the outcome is.

Social Desirability Bias

TABLE A7. *Civilian Attitude Results: Missingness Outcome*

	State Court	Taliban Approval	Gov. Influence	Gov. Index
Taliban Courts	-0.0041	0.0049	-0.0026	-0.012
	0.0026	0.0065	0.0022	0.0068
N. Districts	167	167	192	187
N. Years	6	6	6	7

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Outcome in each column is the share of total responses to the question that are missing. For instance, the state courts outcome is the number of respondents who did not respond to the question about state courts divided by the total number of responses, aggregated to the district year. All estimates are close to zero, and we do not reject the null that any estimate is distinct from zero.

TABLE A8. *Civilian Attitude Results*

	State Court	Taliban Approval	Gov. Influence	Gov. Index
Taliban Courts	-0.0804***	0.0572**	-0.0899***	-0.6042***
	(0.0156)	(0.0181)	(0.0205)	(0.1442)
N. Districts	159	159	178	172
N Years	6	6	7	7
Mean DV	0.2	0.18	0.22	1.73
SD DV	0.46	0.19	0.73	-0.17

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$

Outcome in each column is the response aggregated to the district year. For this analysis, we filter to only respondents who report they are comfortable with the interview. When restricting our analysis to only those respondents who are unlikely to be susceptible to coercion, we document the same pattern as in the main analysis, suggesting that respondent feelings about the interview process are unlikely to be driving our findings.

Sequencing: Withdrawn Courts

TABLE A9. *Civilian Attitude Results: Withdrawn Courts*

Outcome	(1) State Court	(2) Taliban Approval	(3) Gov. Influence	(4) Gov. Index
Taliban Courts	−0.1**	0.07*	−0.11*	-0.09
	(0.04)	(0.03)	(0.05)	(0.25)
N. Districts	137	137	151	157
N. Years	3	3	4	4

Note: Outcomes are civilian attitudes measured by ANQAR. Taliban courts are using courts that are withdrawn in 2012 and 2013. We filter the data to 2011 and before to capture the effect of withdrawn courts before they are withdrawn.

TABLE A10. *Combat Results: Withdrawn Courts*

Outcome	(1) AOG	(2) IED	(3) DF	(4) IED Explosions	(5) Casualty Events
Taliban Courts	2.95	7.81*	21.8	9.56	2.96
	(3.71)	(3.31)	(18.51)	(6.5)	(3.04)
Dataset	ANSO	ANSO	SIGACT	SIGACT	SIGACT
N. Districts	258	258	258	258	258
N. Years	4	4	4	4	4
Standard Deviation DV	33.08	32.12	74.42	26.92	15.95
Mean DV (Control)	14.29	10.42	11.03	5.57	3.16

Note: Outcomes are combat measured by ANSO and SIGACT. Taliban courts are using courts that are withdrawn in 2012 and 2013. We filter the data to 2011 and before to capture the effect of withdrawn courts before they are withdrawn.

Recode: Withdrawn Courts

TABLE A11. *Civilian Attitude Results: Recoding Withdrawn Courts*

Outcome	(1) State Court	(2) Taliban Approval	(3) Gov. Influence	(4) Gov. Index
Taliban Courts	−0.07***	0.05***	−0.06***	−0.53***
	(0.01)	(0.01)	(0.01)	(0.1)
N. Districts	198	199	228	221
N. Years	6	6	7	7

Note: Outcomes are combat measured by ANSO and SIGACT. Taliban courts are all courts, coding those as withdrawn as being exposed in 2011 and exposed continuously.

TABLE A12. *Combat Results: Recoding Withdrawn Courts*

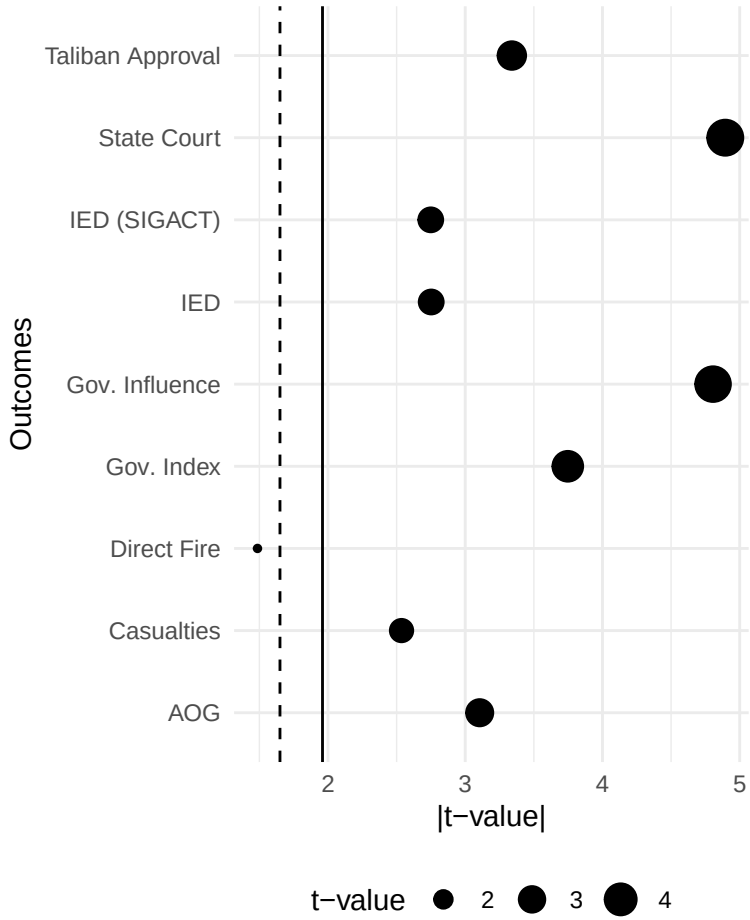
Outcome	(1) AOG	(2) IED	(3) DF	(4) IED Explosions	(5) Casualty Events
Taliban Courts	8.44***	6.74**	33.51***	8.98***	8.44***
	(2.48)	(2.17)	(10.08)	(2.5)	(2.05)
Dataset	ANSO	ANSO	SIGACT	SIGACT	SIGACT
N. Districts	398	398	398	398	398
N. Years	6	6	6	6	6
Standard Deviation DV	45.21	33.89	177.63	42.12	31.3
Mean DV (Control)	16.39	11.61	14.41	6.25	4.4

Note: Outcomes are combat measured by ANSO and SIGACT. Taliban courts are all courts, coding those as withdrawn as being exposed in 2011 and exposed continuously.

Standard Errors

Block Bootstrap

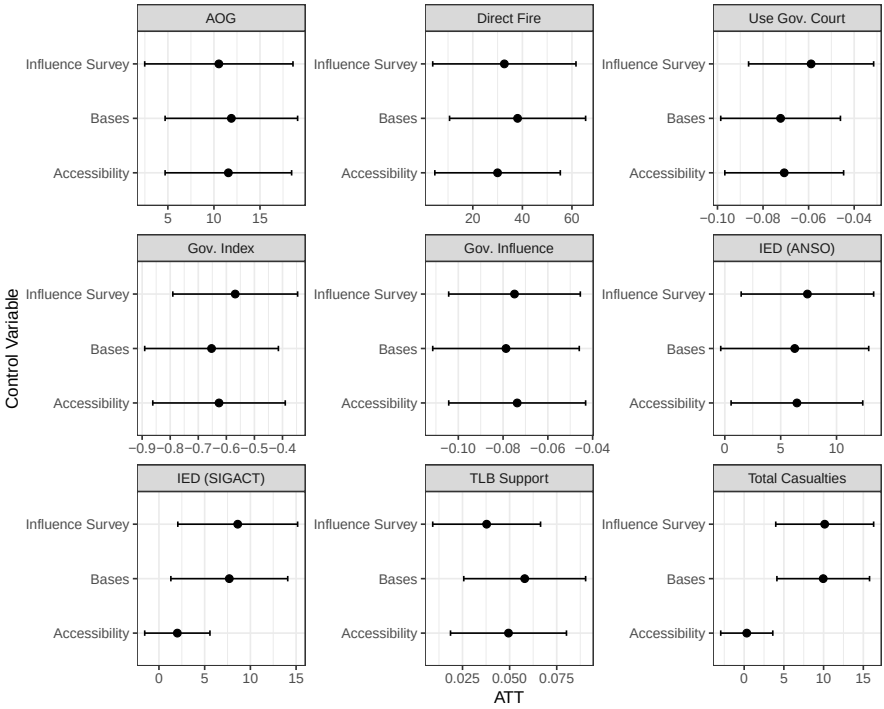
FIGURE A4. *Block Bootstrapped Standard Errors: Outcomes Against T-Statistics*



Note: horizontal axis is the absolute value of the t-statistic (t-value) and the vertical axis is the outcomes. Points are sized proportional to the t-stat. The dashed line is the t-value for $p < .10$ and the solid line is the t-value for $p < .05$.

Adjusting for Control

FIGURE A5. *Adjusting for Control*



Note: Each subfigure is a different outcome. Every estimate shows the ATT and confidence interval using a different pre-treatment measure of control.