Online Appendix

Fighting in Cyberspace:

Internet access and the substitutability of cyber and military operations

Nadiya Kostyuk*and Erik Gartzke†

February 9, 2023

Contents

1 Summary Statistics and Correlation Plots **2**

2 Empirical Strategy **2**

3 Main Results **4**

4 Robustness Checks **6**
4.1 Under-reporting of Cyber Operations 8
4.2 Alternative Model Specification 16
4.3 Alternative Time Dimension 18
4.4 Alternative Measures of Covariates 18
4.5 Alternative Behavior: Major Powers 18
4.6 Alternative Lags 18
4.7 Types of Cyber Operations 18
4.8 Additional Controls 21

*Assistant Professor, School of Public Policy and School of Cybersecurity and Privacy, Georgia Institute of Technology, nkostyuk3@gatech.edu

†Professor of Political Science and Director of the Center for Peace and Security Studies (cPASS), University of California, San Diego, egartzke@ucsd.edu.

1 Summary Statistics and Correlation Plots

Figure 1 depicts the correlation plot and Table 1 shows the summary statistics for the main variables of interest.

Figure 1: Correlation Plot

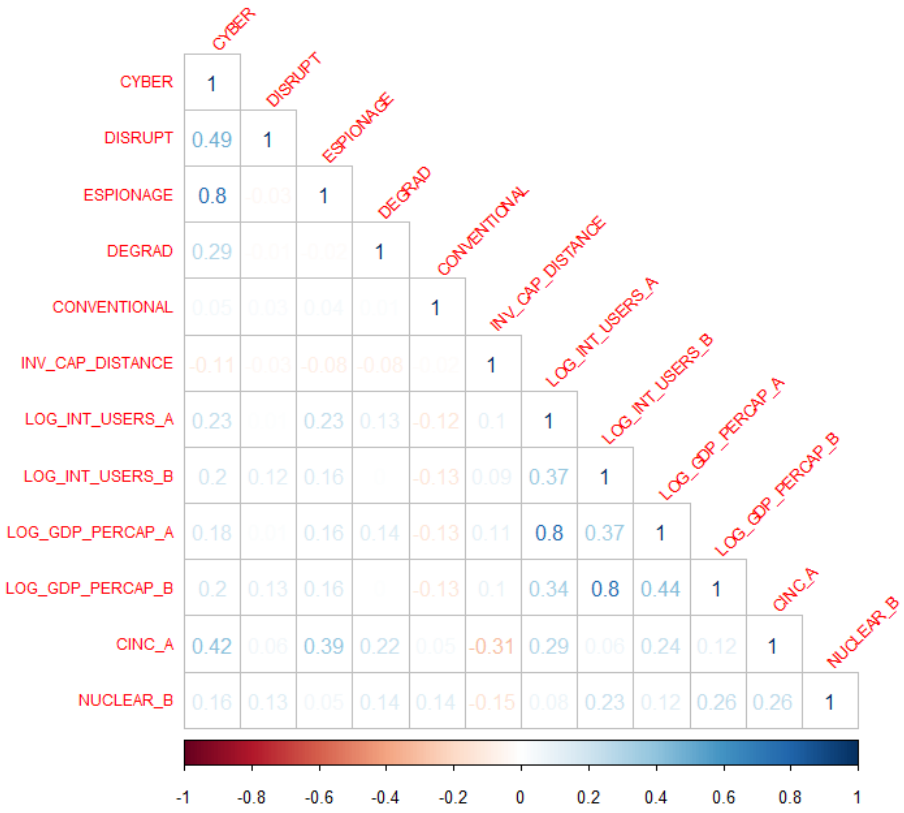


Table 1: Summary Statistics

	Cyber	Degradation	Disruption	Espionage	Convent.	Nuclear_B
Minimum	0.000	0.000	0.000	0.000	0.000	0.000
Median	0.000	0.000	0.000	0.000	0.000	0.000
Mean	0.0798	0.007	0.021	0.052	0.264	0.2087
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
	Int_Users_A (sc, log)	Int_Users_B (sc, log)	GDP_perCapita_A (sc, log)	GDP_perCapita_B (sc, log)	CINC_A (sc)	Distance (sc)
Minimum	-1.510	-1.529	-2.069	-2.047	-0.213	-0.955
Median	0.066	0.062	-0.129	-0.133	-0.419	-0.359
Mean	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	1.770	1.713	2.203	2.1505	3.822	3.713

`sc`: standardized around mean; `log`: logarithmized

2 Empirical Strategy

We model concurrent substitution and complementarity of cyber and military operations using logistic regression with the following probability conditional on the type of operations conducted

(cyber versus military):

$$logit(P(Y_{a,b;t}|Type, Military_{a,b;t-1}, Cyber_{a,b;t-1}, X_{d(a,b);t}, Z_{a;t}, N_{b;t})) = \beta_0 + \beta_1 Type + \beta_2 Military_{a,b;t-1} + \beta_3 Type * Military_{a,b;t-1} + \beta_4 Cyber_{a,b;t-1} + \beta_5 Type * Cyber_{a,b;t-1} + GX_{d(a,b);t} + FZ_{a;t} + MN_{b;t} + u_{a,b} + s_{a,b} * Type + w_{a,b;t} + v_{a,b;t} * Type, \quad (1)$$

In Equation 1, $Y_{a,b;t}|Type, Military_{a,b;t-1}, Cyber_{a,b;t-1}, X_{d(a,b);t}, Z_{a;t}, N_{b;t}$ is a random variable identifying whether an attacker a uses cyber or military operations against a target b in period t, conditional on operation type, the full history of lags, and covariates; β_0 represents the log odds of the operation rate irrespective of operation type for the baseline case; β_1 represents twice the difference between the log odds of the two operation types for the baseline case; `Type` is a binary variable coded as '+1' if an attacker a uses a CO against target b and '-1' if a uses conventional military operations against b; $Military_{a,b;t-1}$ is a dummy variable that gets encoded as "1" if attacker a uses a military operation against target b in period $t-1$ and "0" otherwise; β_2 represents the increase in log odds of the operation of either type, given the presence of a military operation in the previous year; β_3 represents twice the difference in the increase in the log odds between types, given a previous year military operation; $Cyber_{a,b;t-1}$ is a dummy variable that gets encoded as "1" if attacker a uses a CO against b in period $t-1$ and "0" otherwise; β_4 represents the increase in log odds of the operation of either type, given a CO in the previous year; and β_5 represents twice the difference in the increase in the log odds between types, given a previous year CO.

We also include dyad-, attacker-, and target-specific covariates. Specifically, $X_{d(a,b);t} = \left[x_{1d(a,b);t}, \ldots, x_{kd(a,b);t}\right]'$ is a matrix of k dyad-specific exogenous variables, and G is a one-dimensional vector of coefficients.[1] Our dyad-specific exogenous variable is the inverse distance between the capitals of two countries. $Z_{a;t} = [z_{1;t}, \ldots, z_{ka;t}]'$ is a matrix of k attacker-specific exogenous variables, and F is a two-dimensional vector of coefficients. Our attacker-specific exogenous variables include an attacker's Internet users per capita and its material capability, using Singer, Bremer and Stuckey (1972)'s Composite Index of National Capability score (v5.0). Similarly, $N_{b;t} = [n_{1b;t}, \ldots, z_{nb;t}]'$ is a matrix of k target-specific exogenous variables, and M is a two-dimensional vector of coefficients. Target-specific exogenous variables include Internet users per capita and a dummy variable for nuclear status. We use random effects (REs): i.e., random intercepts for dyad ($u_{a,b}$) and for dyad-time ($w_{a,b;t}$), and random slopes for type within dyad ($s_{a,b}$) and dyad-time ($v_{a,b;t}$).

The regression coefficients in Equation 1 estimate the sequential effects of cyber and conventional operations. To estimate concurrent effects ($\texttt{Cyber}_{a,b;t} \leftrightarrow \texttt{Military}_{a,b;t}$), we need to allow dependence between them. No dependence or correlation between $\texttt{Cyber}_{a,b;t}$ and $\texttt{Military}_{a,b;t}$ means that the two variables are independent. Positive dependence between cyber and military operations in the same dyad-time implies that an attacker uses cyber and military operations as complements. Negative dependence between types of operations in the same dyad-time implies that types are substitutes.

A generalized linear model with zero REs has no correlation within dyad or within dyad-time between the two types of operations. As a result, it does not allow for dependence between our variables of interest. A generalized linear model with only dyad-time RE ($w_{a,b;t}$) includes a random intercept for dyad-time. This allows only for positive dependence within dyad-time between the

[1] Note that a subscript a, b identifies a directional relationship of a on b. Thus $a, b \neq b, a$. Conversely, a subscript $d(a, b)$ identifies a symmetric relationship between a and b (e.g., distance between between two nations), so that $d(a, b) = d(b, a)$.

two types of operations, allowing only for the possibility that cyber and conventional operations are concurrent complements. A generalized linear model with a random slope for dyad-time for each type of operation ($v_{a,b;t} * Type$) allows only negative dependence within dyad-time between the two types of operations, allowing only for concurrent substitutes. To account for the possibility that states use cyber and military operations as either concurrent complements or substitutes, we allow both positive and negative dependence within a dyad-time between the two types of operations and include both $w_{a,b;t}$ and $v_{a,b;t} * Type$ in Equation 1.[2]

Let us see why both a random intercept for dyad-time ($w_{a,b;t}$) and a random slope for dyad-time for each type of operation ($v_{a,b;t} * Type$) allow positive and negative dependence within a dyad-time between types of operations. For simplicity, we estimate correlation between cyber and conventional conflict using the following regressions:

$$Y_0 = a_0 + \theta + \epsilon_0,\ \theta \sim N(0, \sigma_\theta^2),\ \text{and} \tag{2}$$

$$Y_1 = a_1 + \theta + \epsilon_1,\ \theta \sim N(0, \sigma_\theta^2), \tag{3}$$

where Y_0 is a dummy variable for conventional conflict; Y_1 is a dummy variable for cyber conflict; a_0/a_1 represents an average operation rate for each type of conflict; θ incorporates dyad REs; and ϵ_0/ϵ_1 represents an error term. To capture concurrent effects between operations, we estimate correlation between cyber and conventional conflict:

$$Cov(Y_0, Y_1) = E[Y_0 * Y_1] - E[Y_0] * E[Y_1] = E[\theta^2] > 0 \tag{4}$$

Equation 4 shows that a random intercept θ only allows for a positive correlation between our key variables, since the variance of a random intercept is positive, $E[\theta^2] > 0$.

Adding a slope for operation type τ to the model allows for negative dependence. Introducing τ to Equations 2 and 3 where $\tau = -1$ for military and $\tau = 1$ for cyber: $Y_0 = a_0 + \theta - \tau + \epsilon_0;\ \theta, \tau \sim N(0, \Sigma)$, and $Y_1 = a_1 + \theta - \tau + \epsilon_1;\ \theta, \tau \sim N(0, \Sigma)$. Estimating correlation between conflict types, and if Σ is diagonal $Cov(\theta, \tau) = 0$; we get: $Cov(Y_0, Y_1) = E[\theta^2] - E[\tau^2] < 0$, if $E[\theta^2] < E[\tau^2]$, and $Cov(Y_0, Y_1) = E[\theta^2] - E[\tau^2] > 0$, if $E[\theta^2] > E[\tau^2]$. Adding both a random intercept and a random slope for a type allows us to investigate whether cyber and conventional conflicts are concurrent complements or substitutes.

3 Main Results

Model fit. We start by comparing the model suggested in Equation 1 with simpler nested models with some random effects omitted. We use the Akaike Information Criterion (AIC) to determine how well each model is supported by these data—the smaller the AIC, the better the model fit. Table 2 displays the results.

[2]The resulting model is similar to vector autoregression (VAR) that accommodates for mixed effects (ME-VAR). While VAR models are familiar in political science, existing ME-VAR models do not accommodate our research design. Moreover, it is worth noting that this model uses a simultaneous equations set-up. Unlike existing methodological approaches which use this set-up and allow only for sequential correlations—accounting only for the sequential use of cyber and military operations—our model also allows simultaneous correlations, accounting for the concurrent use of cyber and military operations.

Table 2: *The likelihood of using cyber and conventional, military operations, by rivalries (Models w/o covariates)*

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
			Fixed effects			
			Dependent Variable: Probability of cyber conflict in period t			
Cyber conflict in period t-1	74.71***(41.05;135.96)	52.77***(26.76;104.04)	34.34***(16.27;72.50)	46.25***(17.69;120.94)	34.34***(16.27;72.49)	46.23***(17.68;120.87)
Conventional conflict in period t-1	0.92(0.50;1.66)	0.32**(0.16;0.64)	0.48^(0.22;1.03)	0.53(0.23;1.21)	0.48^(0.22;1.03)	0.53(0.23;1.21)
			Dependent Variable: Probability of conventional conflict in period t			
Cyber conflict in period t-1	1.47(0.82;2.64)	0.70(0.36;1.36)	1.06(0.50;2.23)	1.24(0.53;2.91)	1.06(0.51;2.23)	1.24(0.53;2.91)
Conventional conflict in period t-1	14.84***(10.81;20.37)	7.01***(4.67;10.52)	5.89***(3.87;8.99)	7.51***(3.92;14.40)	5.89***(3.87;8.99)	7.51***(3.92;14.39)
			Random Effects			
Dyad (intercept)		✓	✓	✓	✓	✓
Variance		1.37	1.52	1.55	1.52	1.55
Dyad (slope for `Type`)			✓	✓	✓	✓
Variance			0.53	0.69	0.53	0.69
Dyad-Year (intercept)				✓		✓
Variance				0.51		0.51
Dyad-Year (slope for `Type`)					✓	✓
Variance					2.1e-12	8.1e-12
Akaike Inf. Crit.	1,445.1	1,411.0	1,404.2	1,404.7	1,406.2	1,406.7

Note: Results are from a logistic regression model. The reported values are the odds ratios and confidence intervals. Odds ratio larger than 1 identify positive correlation and those smaller than 1 identify negative correlation. There are 2,300 observations. All results are based on two-tailed tests. The lowest AIC suggests that Model 3 provides the best model fit. In addition to using the AIC criterion, we also compare each model to one another using the 'anova()' for a (marginal) likelihood ratio test. None of the models that are more complex than Model 3 provide a statistically significant improvement in fit based on the log likelihood test ($p > .2$). ^p<0.1; *p<0.05; **p<0.01; ***p<0.001

Model 1 in Table 2, which provides the results of a non-mixed GLM, has the largest AIC ($AIC_1 = 1445.1$), suggesting that the model with no dependence could be further improved. To do that, we next fit a model with dyad REs to allow dependence within a dyad; this model, however, does not account for conflict type. Model 2's AIC ($AIC_2 = 1411$) is much lower than Model 1's AIC ($AIC_1 = 1445.1$), confirming that Model 2, which allows dependence within a dyad, provides a better fit. Next, we fit a model that allows dependence within a dyad to account for conflict type. Model 3 further improves fit with an AIC at a low to this point in the analysis of 1404.2.

Having accounted for dyad REs, we add dyad-time REs to understand the concurrent use of cyber and conventional operations. We start with allowing only positive dependence and add a random intercept for dyad-time ($w_{a,b;t}$ in Equation 1). This model allows for the possibility that an attacker uses cyber and conventional operations as complements against the same target in the same year. The resulting model 4 is similar to Model 3 ($AIC_3 = 1404.2$ and $AIC_4 = 1404.7$), though it is slightly larger (i.e., worse).

Next we fit the model that allows negative dependence by adding a random slope for dyad-time for each conflict type ($v_{a,b;t} * Type$ in Equation 1). This model accounts for the possibility that an attacker uses cyber and conventional operations against the same target as substitutes within the same year. Model 5 displays the results. Model 5's fit ($AIC_5 = 1406.2$) is again worse than Model 3's fit ($AIC_3 = 1404.2$), or even Model 4's fit ($AIC_4 = 1404.7$). Lastly, Model 6 allows both positive and negative correlation within dyad-year, accounting for the possibility that an attacker uses cyber and conventional operations against the same target as either complements or substitutes within the same year. Model 6's fit ($AIC_6 = 1406.7$) is worse than Model 3's fit ($AIC_3 = 1404.2$), Model 4's fit ($AIC_4 = 1404.7$), and Model 5's fit ($AIC_5 = 1406.2$). In addition to using the AIC criterion, we also compare each model to all others using the 'anova()' for a (marginal) likelihood ratio test. None of the models that are more complex than Model 3 provide a statistically significant improvement in fit based on the log likelihood test ($p > .2$).

Model 3 provides the best fit; rivals do not appear to be using cyber and conventional operations as concurrent complements or substitutes, but instead use them independently. We next utilize Model 3 to assess sequential cyber/conventional interactions.

Results: Table 3 which displays the results includes a number of models, starting with the base model with no covariates to the model that accounts for the effect of each covariate on the likelihood of different types of conflict—cyber, conventional, or any (cyber and conventional). Our main findings are that a rival's Internet dependency explains its propensity towards cyber versus conventional conflict.

4 Robustness Checks

In addition to the analyses already presented, we conduct the following sets of robustness checks:

1. accounting for under-reporting of COs (Section 4.1).
2. alternative model specification (Section 4.2);
3. alternative time dimension (Section 4.3);
4. alternative measures of covariates (Section 4.4);
5. alternative behavior, examining major powers (Section 4.5);

Table 3: *The likelihood of using cyber and conventional, military operations, by rivalries (odd ratios and confidence intervals)*

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
			Fixed effects		
			Dependent Variable: Probability of cyber conflict in period t		
Cyber conflict in period t-1	28.99***(13.68;61.44)	15.30***(7.09;33.03)	14.53***(6.62;31.88)	14.58***(6.64;32.02)	14.57***(6.63;32.01)
Conventional conflict in period t-1	0.57(0.26, 1.25)	0.69(0.31;1.54)	0.55(0.24;1.23)	0.54(0.24;1.23)	0.54(0.24;1.22)
Attacker's internet users per capita (log, sc)	——	2.15**(1.33;3.46)	1.79*(1.12;2.87)	1.78*(1.11;2.86)	1.79*(1.11;2.86)
Target's internet users per capita (log, sc)	——	2.00**(1.23;3.24)	2.26**(1.37;3.72)	2.27**(1.37;3.75)	2.27**(1.37;3.75)
Attacker's CINC score (sc)	——	——	2.77***(1.88;4.09)	2.83***(1.88;4.25)	2.83***(1.88;4.25)
Distance between two states (sc)	——	——	——	1.12(0.71;1.79)	1.13(0.71;1.80)
Nuclear-armed target (dummy)	——	——	——	——	2.81*(1.07;7.41)
			Dependent Variable: Probability of conventional conflict in period t		
Cyber conflict in period t-1	1.01(0.48, 2.13)	1.25(0.58;2.67)	1.42(0.67;3.02)	1.42(0.67;3.02)	1.41(0.66;3.01)
Conventional conflict in period t-1	5.96***(3.92, 9.07)	5.62***(3.70;8.52)	5.79***(3.82;8.79)	5.79***(3.82;8.79)	5.80***(3.82;8.80)
Attacker's internet users per capita (log, sc)	——	0.68*(0.50;0.93)	0.77^(0.57;1.05)	0.78(0.57;1.05)	0.78(0.57;1.05)
Target's internet users per capita (log, sc)	——	0.73*(0.54;0.98)	0.71*(0.53;0.96)	0.71*(0.53;0.96)	0.72*(0.53;0.97)
Attacker's CINC score (sc)	——	——	1.10(0.79;1.53)	1.09(0.78;1.53)	1.10(0.78;1.53)
Distance between two states (sc)	——	——	——	1.04(0.72;1.49)	1.04(0.72;1.49)
Nuclear-armed target (dummy)	——	——	——	——	2.62*(1.20;5.71)
			Dependent Variable: Probability of any conflict in period t		
Attacker's internet uses per capita (log, sc)	0.91(0.70, 1.18)	1.21(0.89;1.63)	1.18(0.87;1.60)	1.18(0.87;1.60)	1.18(0.87;1.60)
Target's internet uses per capita (log, sc)	0.93(0.72, 1.19)	1.20(0.89;1.62)	1.27(0.93;1.73)	1.27(0.93;1.74)	1.27(0.93;1.74)
Attacker's CINC score (sc)	1.65***(1.26, 2.16)	1.67***(1.25;2.22)	1.75***(1.30;2.36)	1.76***(1.30;2.38)	1.76***(1.30;2.38)
Distance between two states (sc)	1.04(0.77, 1.40)	1.05(0.77;1.44)	1.07(0.77;1.47)	1.08(0.78;1.51)	1.08(0.78;1.51)
Nuclear-armed target (dummy)	2.19*(1.15, 4.15)	2.47**(1.25;4.88)	2.67**(1.32;5.39)	2.68**(1.33;5.42)	2.71**(1.32;5.59)
			Random Effects		
Dyad (intercept)	✓	✓	✓	✓	✓
Variance	1.02	1.22	1.36	1.36	1.36
Dyad (slope for `Type`)	✓	✓	✓	✓	✓
Variance	0.30	0.61	0.31	0.31	0.30
Akaike Inf. Crit.	1389.2	1348.6	1328.5	1330.5	1332.4

Note: Results are from a logistic regression model. The reported values are the odds ratios and confidence intervals. Odds ratio larger than 1 identify positive correlation and those smaller than 1 identify negative correlation. There are 2,300 observations. All results are based on two-tailed tests. Variables: `log`—logarithmized; `sc`—standardized. ^p<0.1; *p<0.05; **p<0.01; ***p<0.001

6. alternative measure of lags (Section 4.6);
7. accounting for different types of cyber operations (Section 4.7); and
8. accounting for additional covariates (Section 4.8).

4.1 Under-reporting of Cyber Operations

Variations in reporting can be a serious problem for conflict event data (Weidmann, 2016; Dietrich and Eck, 2020), especially for COs, due to their novelty, secrecy surrounding their execution, and the difficulty of attributing origin. In the main manuscript we briefly explained why bias may not present as serious of an issue here as one might assume. In this section we first consider the two types of biases that can result in latent events—events that happened but are not recorded in the dataset. First is non-systematic underreporting that occurs if variables in the model are not related to the probability that a CO is reported (i.e., COs are under-reported at random), conditional on the cyber conflict occurring. Second is systematic underreporting that occurs if variables in the model are related to the probability that a CO is reported (i.e., due to their transparent nature, democracies tend to report more COs that they suffered than autocracies).

Figures 2–9 present the obtained results and compare them to the true estimates. Specifically, the black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis. We use contrasts in the model to present odds ratios to describe associations between different types of conflict and covariates. Odds ratios larger than 1 identify positive correlations and odds ratios between 0 and 1 identify negative correlations. Our variables include three components: (1) odds ratios (`OR`); (2) a DV (`conflict` stands for any conflict); and (3) an explanatory variable. For instance, `OR_military_cyb(lag)` displays the odds ratios of military conflict in the period after cyber conflict takes place; `OR_cyber_internet(a)` displays the odds ratios of cyber conflict given an attacker's (log) internet per capita. Overall, the results presented in Figures 2–9 confirm that our main finding—indirect substitutability of a country's Internet dependency and its likelihood of experiencing cyber and conventional conflict—is robust to the situations that account for different types of reporting biases typical for cyber-operations.

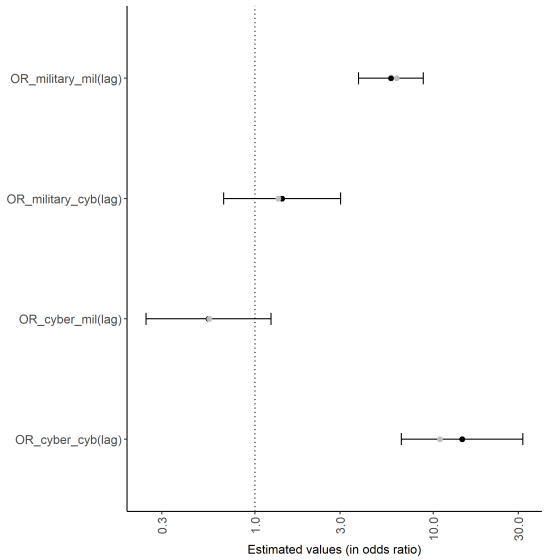
Non-Systematic Under-reporting. As earlier defined, non-systematic underreporting occurs if variables in the model are not related to the probability that a CO is reported, conditional on the cyber conflict occurring. For simplicity, we consider two possible scenarios. First, unobserved events are completely random. Specifically, probability of a cyber operation taking place between any two pairs of dyads are equally likely. While this is not a realistic assumption, it is a good starting point as it allows us to understand how under-reported, randomly distributed COs change the obtained results. We assume that 10% of non-cyber-events actually were events; this means that we added 105 event, on average, in each simulation; this means that the observed events represent only 49% of actual events, on average.[3] We run sensitivity analysis re-estimating our model 200 times using the simulated data that has 10% more of non-cyber-events randomly added as events. Since we add 10% more of non-cyber-events as events at random, we expect predictors of COs to be attenuated towards zero in the simulated results—specifically for the effect of a country's Internet dependency

[3]We also run robustness checks when the observed events represent only 90%, 80%, or 70% of actual events, on average. The results remain robust to this alternative specification.

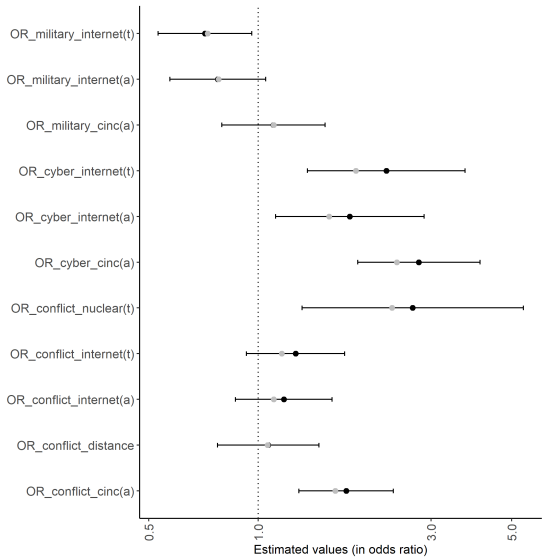
on cyber and/or conventional conflict. The effect of adding COs will also introduce the noise in the lags, so we expect the effects of lags attenuated as well (Carroll et al., 2006).

Figure 2: *10% more of non-cyber-events added as events at random (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations



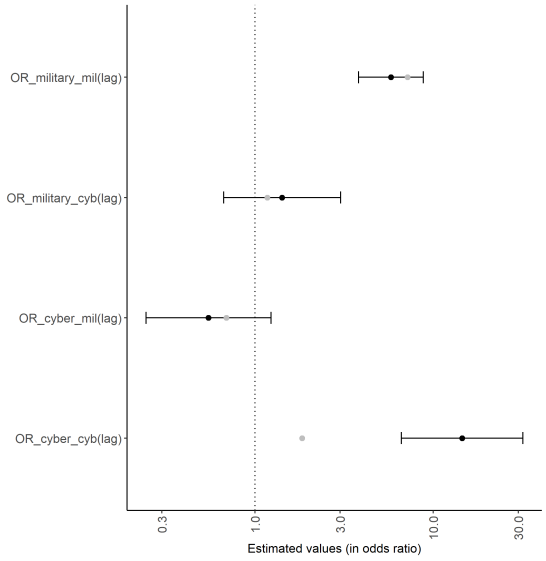
Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times using the simulated data that has 10% more of non-cyber-events added as events at random.

Figure 2 presents the results. Figure 2a shows the results for associations between cyber and military operations. For `OR_military_mil(lag)` (i.e., association between military conflict in t-1 and military conflict in t), our results from the sensitivity analysis are biased upwards by the small amount within our original confidence intervals. For `OR_cyber_mil(lag)` (i.e., association between military conflict in t-1 and cyber conflict in t) and `OR_military_cyb(lag)` (i.e., association between cyber conflict in t-1 and military conflict in t), the means of our main model almost perfectly overlap with the means of our simulated results. For `OR_cyber_cyb(lag)` (i.e., association between cyber conflict in t-1 and cyber conflict in t), our results from the sensitivity analysis are biased downwards by the small amount, attenuating towards one; but, they still lie within our original confidence intervals.

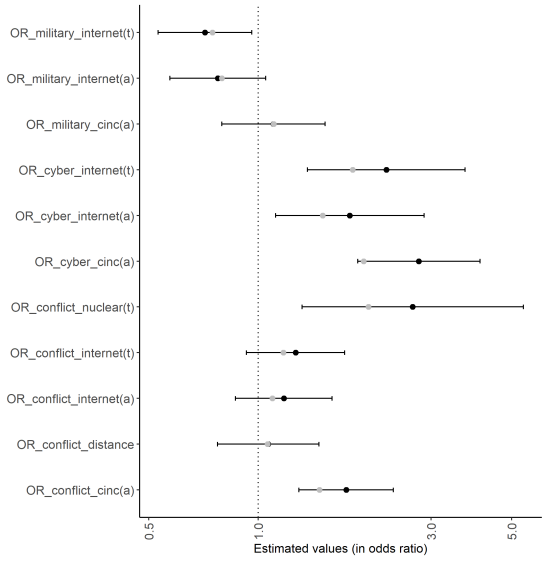
Figure 2b shows the results for associations between covariates and cyber and military operations. For `OR_military_internet(t)`, `OR_military_internet(a)`, `OR_military_cinc(a)`, and `OR_conflict_distance`, the means of our main model almost perfectly overlap with the means of our simulated results. For `OR_cyber_internet(t)`, `OR_cyber_internet(a)`, `OR_cyber_cinc(a)`, `OR_conflict_nuclear(t)`, `OR_conflict-_internet(t)`, `OR_conflict_internet(a)`, `OR_conflict_cinc(a)`, our results from the sensitivity analysis are biased downwards by the small amount within our original confidence intervals.

Figure 3: *10% more of non-cyber-events added as events from the model (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations



Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

Second, instead of randomly adding 10% of non-COs as COs, we sample hypothetical under-reported events according to the model predictions, rescaled so that the mean probability that a non-cyber-event is an under-reported event is 0.1. Specifically, zero events are sampled as events with probability of p, where

$$p = \left(\text{logistic}(\text{logit}(\hat{p}) - \frac{1}{N} \sum_{i=1}^{N} logit(\hat{p}_i)) \right) * \theta / 0.5. \tag{5}$$

In Equation 5, $\text{logit}(p) = \log \frac{p}{1-p}$; $\text{logistic}(p) = logit^{-1}(p) = \frac{e^p}{1+e^p}$; θ - the rate at which zeros are flipped to 1s (under-reporting of COs), $\hat{p}$ - model probability; i - indexes observations. Events that we added occur, on average, with the probability of θ.

We apply the approach defined in Equation 5 with $\theta = 0.1$. Specifically, we sample hypothetical under-reported events according to our main model predictions, rescaled so that the mean probability that a non-cyber-event is an under-reported event is 0.1. Figure 3 presents the obtained results and compares them to the true estimates. Figure 3a shows the results for associations between cyber and military operations. For `OR_military_mil(lag)` and `OR_cyber_mil(lag)`, our results from the sensitivity analysis are biased upwards by the small amount within our original confidence intervals. For `OR_military_cyb(lag)`, our results from the sensitivity analysis are biased downwards by the small amount within our original confidence intervals. For `OR_cyber_cyb(lag)`, our results from

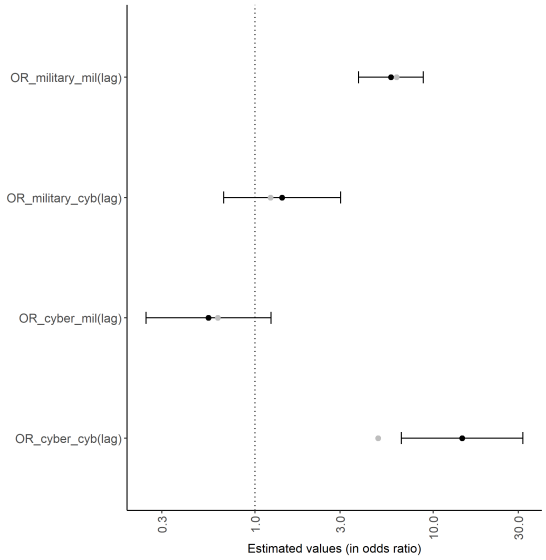
the sensitivity analysis are biased downwards, attenuating towards one.

Figure 3b shows the results for associations between covariates and cyber and military operations. For `OR_military_internet(t)` and `OR_military_internet(a)`, our results from the sensitivity analysis are biased upwards by the small amount within our original confidence intervals. For `OR_military_cinc(a)` and `OR_conflict_distance`, the means of our main model almost perfectly overlap with the means of our simulated results. For `OR_cyber_internet(t)`, `OR_cyber_internet(a)`, `OR_cyber_cinc(a)`, `OR_conflict_nuclear(t)`, `OR_conflict_internet(t)`, `OR_conflict_internet(a)`, and `OR_conflict_cinc(a)`, our results from the sensitivity analysis are biased downwards by the small amount within our original confidence intervals.

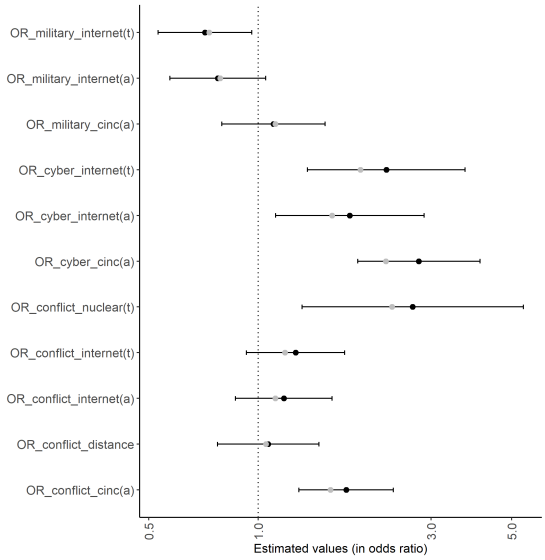
Overall, this analysis that considers systematic biases in cyber conflict suggest that our main results that there is an indirect subtitutionary effect of the Internet dependency on the conflict type do not change. This evidence confirms that despite DCID data limitation, it is suitable for analysis.

Figure 4: *10% more of non-cyber-events added as events if they occurred during peacetime, simulated using the model predictions (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations



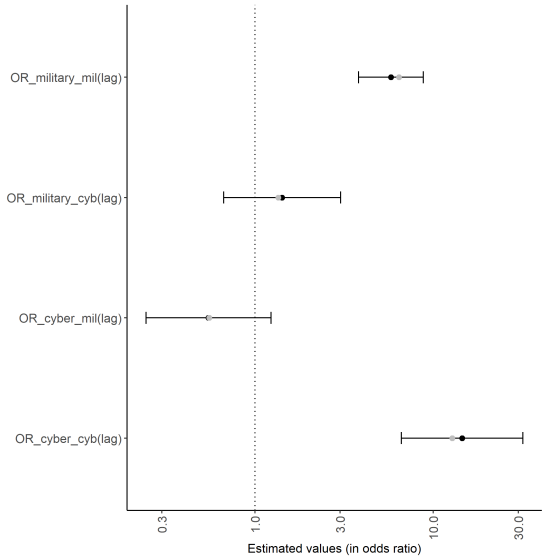
Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

Systematic Underreporting. Systematic underreporting occurs if variables in the model are related to the probability that a CO is reported. We consider the following sources of systematic biases outlined in the existing literature that might be relevant to our study. First, given that events are more likely to be reported during a military conflict than during peacetime (Weidmann, 2016),

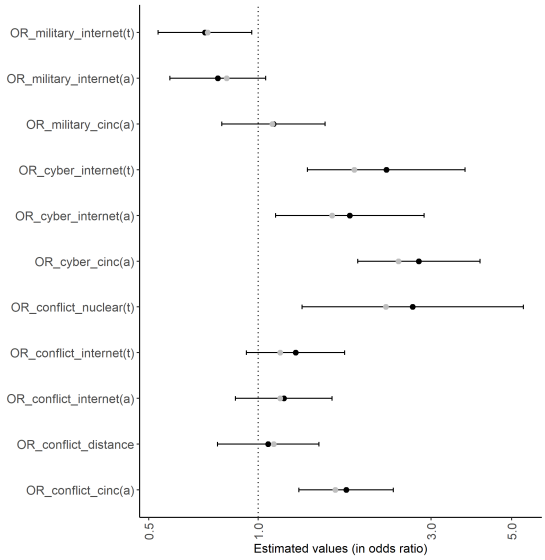
COs that occur in tandem with conflict might also be more reported than those that occur during peacetime. To consider this possibility, we add 10% more of non-COs as COs if they occurred during peacetime, using the model predictions. Figure 4 which presents the obtained results demonstrates that our results are robust to this alternative source of bias.

Figure 5: *10% more of non-cyber-events added as events if the target of COs is an autocracy, simulated using the model predictions (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations



Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

Second, autocracies might be less likely to report COs they suffered given a lower degree of government transparency in such regimes compared to democracies. To consider this possibility, we add 10% more of non-COs as COs, if the victim of COs is a autocracy, simulated using the model predictions. Figure 5 which presents the obtained results demonstrates that our results are robust to this alternative source of bias.

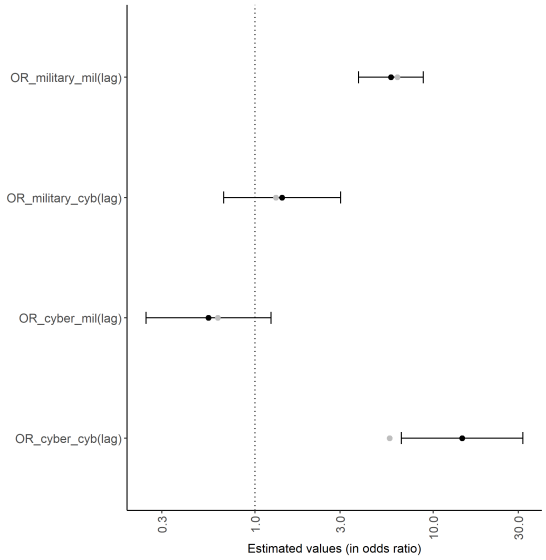
Third, given the higher reporting rates of violence in cellphone-covered areas (Weidmann, 2016, 207), we might expect countries with high Internet reliance to report more COs if they a target of COs than those countries with low Internet reliance. To consider this possibility, we add 10% more of non-COs as COs for the countries with low Internet connectivity (less than 80% of its population using Internet), simulated using the model predictions.[4] Figure 6 which presents the

[4]We create three different cut-off points to identify countries with high Internet reliance: (1) those nations who have more than 50% of their population using the Internet; (2) those nations who have more than 70% of their population using the Internet; and (3) those nations who have more than 80% of their population using the Internet. The results remain robust to these alternative cut-off points.

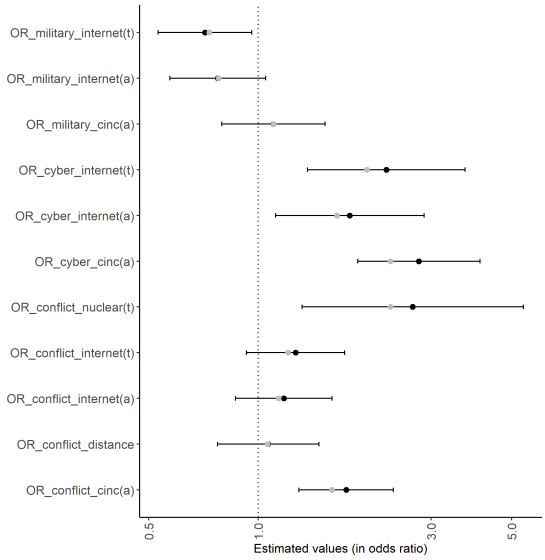
obtained results demonstrates that our results are robust to this alternative source of bias.

Figure 6: *10% more of non-cyber-events added as events if a country has low Internet connectivity (less than 80% of its population using Internet), simulated using the model predictions (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations

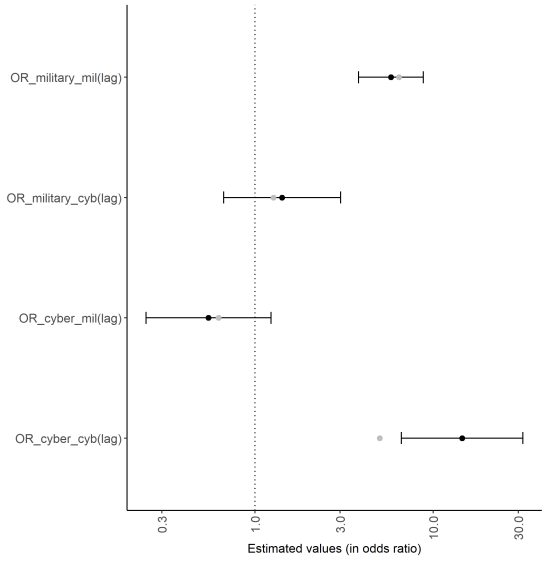


Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

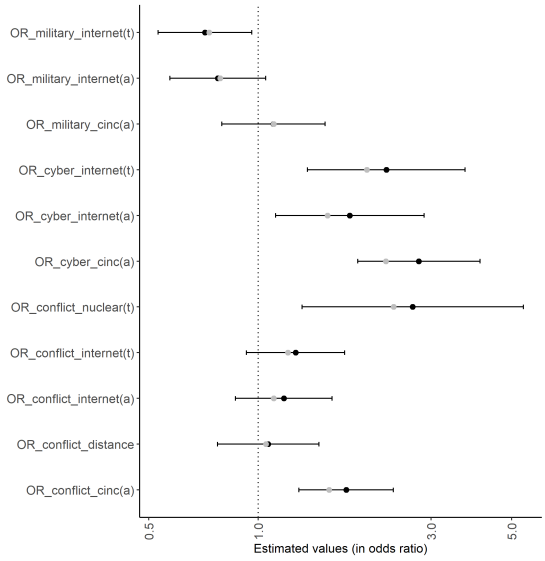
Fourth, due to resource and language constraints, DCID might have an inherent bias toward over-reporting incidents affecting countries where English is widely spoken, or there is English-language media. To consider this possibility, we add 10% more of non-COs as COs if the target of COs is a non-English-speaking country, simulated using the model predictions. Figure 7 which presents the obtained results demonstrates that our results are robust to this alternative source of bias.

Figure 7: *10% more of non-cyber-events added as events if the target of a CO is a non-English-speaking country, simulated using the model predictions (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations

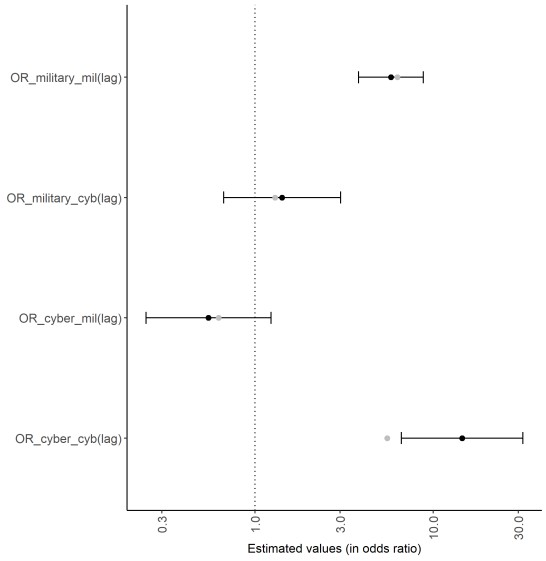


Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

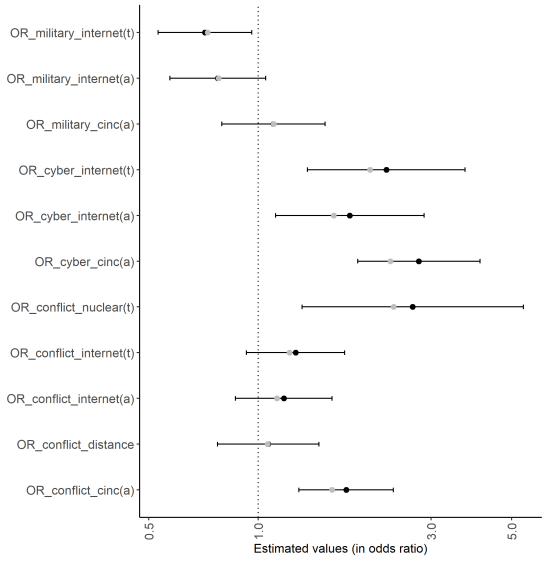
Last, reporting biases might be country-specific. For instance, COs against the United States might be more likely reported, investigated, and attributed than against any other nation, given sophisticated resources and expertise that the United States has. To consider this possibility, we add 10% more of non-COs as COs if the United States is not a target, simulated using the model predictions. Figure 8 which presents the obtained results demonstrates that our results are robust to this alternative source of bias. Additionally, the operations executed by the United States might be less likely reported and attributed given the stealth of its COs. To consider this possibility, we add 10% more of non-COs as COs if the United States is an attacker, simulated using the model predictions. Figure 8 which presents the obtained results demonstrates that our results are robust to this alternative source of bias.

Figure 8: *10% more of non-cyber-events added as events if the United States is not a target, simulated using the model predictions (simulated and true estimates)*

(a) Associations between cyber and military operations



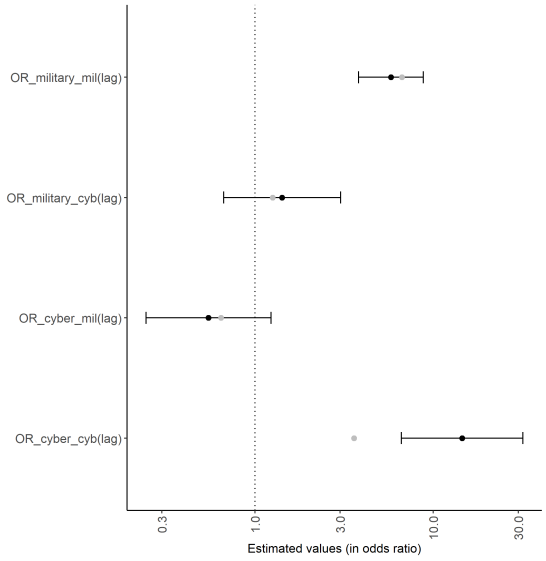
(b) Associations between covariates and cyber and military operations



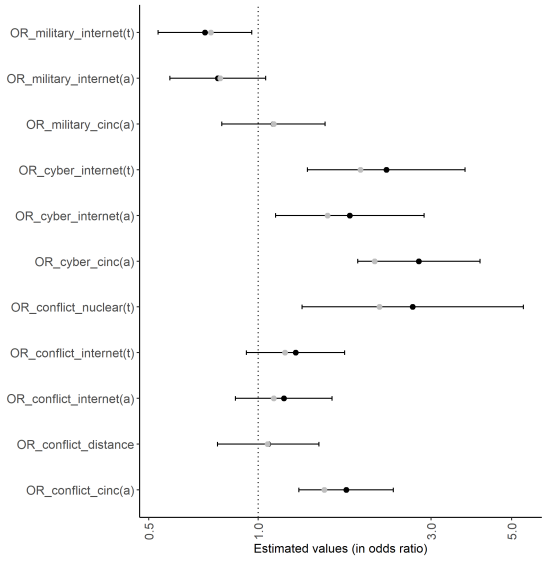
Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

Figure 9: *10% more of non-events added as events if the United States is an attacker, simulated using the model predictions (simulated and true estimates)*

(a) Associations between cyber and military operations



(b) Associations between covariates and cyber and military operations



Note: The figure presents the true and simulates estimates. The black dots display the estimates from Model 1 in Table 3. The black error bars display the confidence interval of the estimates from Model 1 in Table 3. The gray dots display the mean estimates from our sensitivity analysis, during which we re-estimated our model 200 times. We add 10% more of non-cyber-events as events, with probabilities proportional to our model.

4.2 Alternative Model Specification

We use an alternative model specification—instead of dyad REs, we also use dyad-year REs. Table 4 that presents the obtained results demonstrates that our earlier obtained results remain robust, further confirming that the model with only dyad REs (Model 1) provides the best fit.

Table 4: *Robustness Checks: Alternative Model Specification*

	Model 1	*Model 2*	*Model 3*	*Model 4*
			Fixed effects	
	Dependent Variable: Probability of cyber conflict in period t			
Cyber conflict in period t-1	14.53***(6.62;31.88)	17.84***(7.06;45.09)	15.23***(6.33;36.67)	14.47***(6.60;31.74)
Conventional conflict in period t-1	0.55(0.24;1.23)	0.59(0.25;1.39)	0.54(0.23;1.24)	0.55(0.24;1.23)
Attacker's internet uses per capita (log, sc)	1.79*(1.12;2.87)	1.83*(1.12;3.00)	1.79*(1.12;2.88)	1.79*(1.12;2.87)
Target's internet uses per capita (log, sc)	2.26**(1.37;3.72)	2.34**(1.38;3.98)	2.27**(1.37;3.76)	2.26**(1.37;3.72)
Attacker's CINC score (sc)	2.77***(1.88;4.09)	2.91***(1.92;4.40)	2.80***(1.88;4.16)	2.75***(1.87;4.06)
	Dependent Variable: Probability of conventional conflict in period t			
Cyber conflict in period t-1	1.42(0.67;3.02)	1.59(0.69;3.66)	1.41(0.66;3.04)	1.42(0.67;3.02)
Conventional conflict in period t-1	5.79***(3.82;8.79)	7.06***(3.82;13.05)	6.01***(3.58;10.07)	5.78***(3.81;8.78)
Attacker's internet uses per capita (log, sc)	0.77^(0.57;1.05)	0.76^(0.55;1.05)	0.77(0.57;1.05)	0.78(0.57;1.05)
Target's internet uses per capita (log, sc)	0.71*(0.53;0.96)	0.69*(0.51;0.95)	0.71*(0.53;0.96)	0.71*(0.53;0.96)
Attacker's CINC score (sc)	1.10(0.79;1.53)	1.10(0.78;1.55)	1.10(0.79;1.54)	1.10(0.79;1.52)
	Dependent Variable: Probability of any conflict in period t			
Attacker's internet uses per capita (log, sc)	1.18(0.87;1.60)	1.18(0.86;1.63)	1.18(0.86;1.60)	1.18(0.87;1.60)
Target's internet uses per capita (log, sc)	1.27(0.93;1.73)	1.28(0.92;1.77)	1.27(0.93;1.74)	1.27(0.93;1.73)
Attacker's CINC score (sc)	1.75***(1.30;2.36)	1.79***(1.31;2.45)	1.76***(1.29;2.39)	1.74***(1.29;2.34)
Distance between two states (sc)	1.07(0.77;1.47)	1.07(0.77;1.50)	1.07(0.77;1.48)	1.05(0.76;1.45)
Nuclear-Armed Target (dummy)	2.67**(1.32;5.39)	2.80**(1.34;5.83)	2.70**(1.32;5.53)	2.67**(1.32;5.40)
			Random Effects	
Dyad (intercept)	✓	✓	✓	✓
Variance	1.17	1.18	1.18	1.17
Dyad (slope for `Type`)	✓	✓	✓	✓
Variance	0.56	0.62	0.55	0.55
Dyad-Year (intercept)		✓		✓
Variance		0.66		1.85e-06
Dyad-Year (slope for `Type`)			✓	✓
Variance			0.30	8.37e-17
Akaike Inf. Crit.	1328.5	1329.4	1330.5	1332.6

Note: Results are from a logistic regression model. The reported values are the odds ratios and confidence intervals. Odds ratio larger than 1 identify positive correlation and those smaller than 1 identify negative correlation. There are 2,300 observations. All results are based on two-tailed tests. Variables: `log`—logarithmized; `sc`—standardized. The lowest AIC suggests that Model 1 provides the best model fit. In addition to using the AIC criterion, we also compare each model to one another using the 'anova()' for a (marginal) likelihood ratio test. None of the models that are more complex than Model 1 provide a statistically significant improvement in fit based on the log likelihood test ($p > .2$). ^p<0.1; *p<0.05; **p<0.01; ***p<0.001

4.3 Alternative Time Dimension

In this section, we use monthly data for our analysis. Model 1 in Table 5 which presents our results confirms our earlier findings. It is worth noting that since conflicts are likely to last for a number of consecutive months,[5] it is not surprising that we observe the large estimates and confidence intervals for the effects of the same types of operations in period t-1 on those operations in period t.

4.4 Alternative Measures of Covariates

We use a different measure for *distance*. Instead of an inverse distance between the countries' capitals, we use a dummy indicating whether states share a land border or are separated by less than 150 miles of water (Stinnett et al., 2002). Model 2 in Table 5 which displays the results confirms that our results are robust to this alternative measure of distance.

4.5 Alternative Behavior: Major Powers

Given that many major cyber campaigns are either attributed to major powers or target these powers, does cyber mostly explain the behavior of these major powers? To address this possibility, we create interactions between conflict variables—cyber and conventional— and either an attacker or a target being a major power. To identify a major power, we use the Correlates of War project's *State System Membership List* (v2016), according to which China, Japan, Russia, the United Kingdom, and the United States were major powers during the studied period. Model 3 in Table 5 which displays the results for an attacker shows that it is not the case. Moreover, it confirms that our results are robust to this alternative behavior. We also ran the same model, using an interaction between conflict variables—cyber and conventional—and a target as a major power. The results remain robust to this alternative model specification.

4.6 Alternative Lags

We also account for an alternative measure of lags. Instead of using period t-1, we also use t-2.[6] Model 4 in Table 5 which displays the results confirms that our results remain robust to this alternative measure of lags.

4.7 Types of Cyber Operations

We also account how different types of cyber operations can affect our results. Table 6 presents the obtained results. Model 1 in Table 6 presents results for cyber espionage campaigns. It confirms our theoretical expectations (*Hypotheses 5A* and *5B*). Model 2 in Table 6 presents results for both cyber disruption and degradation campaigns. They show that there are no real evidence of association between these attacks and the Internet users (OR = 0.89; CI (05.1, 1.53)).[7]

[5] For instance, an average duration of cyber conflict is 134 days; an average duration of conventional conflict is 132 days.

[6] Given that our data has only ten years, we do not run robustness checks using a larger measure of lags than two periods.

[7] We do not run a separate regression analysis for degradation operations because there are not enough data points (i.e., 18 events out of 2,300 sample).

Table 5: *Additional Robustness Checks*

	Model 1 `Monthly data`	*Model 2* `Alternative Distance measure`	*Model 3* `Attacker as major power`	*Model 4* `Alternative lags`
		Fixed effects		
	Dependent Variable: Probability of cyber conflict in period t			
Cyber conflict in period t-1$^{\diamond}$	599.81***(299;1200)	14.28***(6.50;31.38)	6.73***(2.20;20.62)	6.32***(2.49;16.06)
Conventional conflict in period t-1$^{\diamond}$	0.70(0.29;1.67)	0.54(0.24;1.21)	0.73(0.25;2.16)	0.29**(0.12;0.70)
Attacker's internet uses per capita (log, sc)	2.38***(1.43;3.96)	1.87**(1.16;3.01)	.58∧(0.98;2.56)	3.15**(1.55;6.40)
Target's internet uses per capita (log, sc)	1.89*(1.12;3.19)	2.33**(1.40;3.85)	2.13**(1.30;3.51)	3.79***(1.80;7.99)
Attacker's CINC score (sc)	2.57***(1.73;3.82)	2.73***(1.88;3.98)	2.11**(1.21;3.66)	4.01***(2.31;6.98)
Attacker as major power (dummy)	——	——	2.65(0.55;12.69)	——
Attacker as major power*Cyber in t-1	——	——	4.94∧(0.86;28.23)	——
Attacker as major power*Conventional in t-1	——	——	2.61(0.38;17.82)	——
	Dependent Variable: Probability of conventional conflict in period t			
Cyber conflict in period t-1$^{\diamond}$	0.94(0.50;1.76)	1.40(0.66;2.97)	1.21(0.36;4.04)	2.80*(1.18;6.65)
Conventional conflict in period t-1$^{\diamond}$	25.72***(101;156)	5.72***(3.77;8.68)	6.73***(4.22;10.75)	1.62*(1.03;2.54)
Attacker's internet uses per capita (log, sc)	0.79∧(0.62;1.00)	0.79(0.59;1.08)	0.77∧(0.56;1.04)	0.71(0.47;1.09)
Target's internet uses per capita (log, sc)	0.76*(0.60;0.96)	0.73*(0.54;0.99)	0.71*(0.52;0.95)	0.58**(0.38;0.87)
Attacker's CINC score (sc)	1.07(0.76;1.52)	1.09(0.79;1.49)	1.08(0.65;1.78)	1.03(0.64;1.67)
Attacker as major power (dummy)	——	——	1.40(0.33;5.96)	——
Attacker as major power*Cyber in t-1	——	——	2.60(0.41;16.41)	——
Attacker as major power*Conventional in t-1	——	——	1.38(0.28;6.69)	——
	Dependent Variable: Probability of any conflict in period t			
Attacker's internet uses per capita (log, sc)	1.37*(1.01;1.85)	1.22(0.90;1.66)	1.10(0.81;1.50)	1.50∧(0.96;2.34)
Target's internet uses per capita (log, sc)	1.20(0.88;1.63)	1.30(0.95;1.79)	1.23(0.90;1.68)	1.48∧(0.95;2.32)
Attacker's CINC score (sc)	1.66**(1.21;2.29)	1.73***(1.30;2.30)	1.51∧(0.97;2.33)	2.04**(1.33;3.12)
Distance between two states (sc)	1.05(0.74;1.51)	0.40(0.10;1.68)	1.10(0.79;1.52)	1.08(0.67;1.72)
Nuclear-Armed Target (dummy)	2.86*(1.27;6.42)	2.54*(1.25;5.15)	2.82**(1.39;5.76)	4.75**(1.67;13.50)
		Random Effects		
Dyad (intercept)	✓	✓	✓	✓
Variance	1.36	1.39	1.16	3.41
Dyad (slope for `Type`)	✓	✓	✓	✓
Variance	0.31	0.30	0.56	1.15
Akaike Inf. Crit.	3617.9	1327.1	1331.6	1249.3
Number of obs-ns	26,200	2,300	2,300	2,070

Note: Results are from a logistic regression model. The reported values are the odds ratios and confidence intervals. Odds ratio larger than 1 identify positive correlation and those smaller than 1 identify negative correlation. There are 2,300 observations. All results are based on two-tailed tests. Variables: `log`—logarithmized; `sc`—standardized; $^{\diamond}$ identifies period t-2 for Model 4 (Alternative lags). ∧p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table 6: *Additional Robustness Checks*

	Model 1 `Espionage`	*Model 2* `Disruption & Degradation`	*Model 3* `w GDP per capita`	*Model 4* `Espionage w GDP per capita`
		Fixed effects		
		Dependent Variable: Probability of cyber conflict in period t		
Cyber conflict in period t-1$^{\diamond}$	187.1***(45.1, 776.1)	3.28*(1.12, 9.63)	14.64***(6.66;32.17)	190.46***(45.78;792.41)
Conventional conflict in period t-1	0.56(0.16, 1.90)	0.66(0.25, 1.72)	0.54(0.24;1.22)	0.55(0.16;1.88)
Attacker's internet uses per capita (log, sc)	2.49**(1.25, 4.93)	0.89(0.51, 1.53)	2.06*(1.15;3.68)	3.08**(1.40;6.76)
Target's internet uses per capita (log, sc)	2.49**(1.25, 4.93)	0.89(0.51, 1.53)	2.39**(1.33;4.28)	1.9∧(1.00;3.83)
Attacker's CINC score (sc)	2.56***(1.68, 3.88)	1.88**(1.22, 2.88)	2.79***(1.89;4.11)	2.57***(1.69;3.91)
		Dependent Variable: Probability of conventional conflict in period t		
Cyber conflict in period t-1$^{\diamond}$	1.01(0.39, 2.63)	1.84(0.65, 5.25)	1.41(0.66;3.00)	0.99(0.38;2.57)
Conventional conflict in period t-1	6.31***(4.10, 9.70)	5.77***(3.80, 8.78)	5.74***(3.78;8.72)	6.22***(4.05;9.57)
Attacker's internet uses per capita (log, sc)	0.79(0.59, 1.06)	0.77∧(0.57, 1.04)	0.88(0.58;1.36)	0.95(0.62;1.45)
Target's internet uses per capita (log, sc)	0.79(0.59, 1.06)	0.77∧(0.57, 1.04)	0.76(0.50;1.16)	0.80(0.53;1.21)
Attacker's CINC score (sc)	1.16(0.85, 1.58)	1.08(0.78, 1.50)	1.11(0.80;1.55)	1.17(0.86;1.60)
		Dependent Variable: Probability of any conflict in period t		
Attacker's internet uses per capita (log, sc)	1.40∧(0.95, 2.07)	0.82(0.59, 1.15)	1.35(0.87;2.10)	1.71*(1.02;2.87)
Target's internet uses per capita (log, sc)	1.40∧(0.95, 2.07)	0.82(0.59, 1.15)	1.35(0.87;2.08)	1.25(0.79;1.99)
Attacker's GDP per capita (log, sc)	——	——	0.83(0.52;1.30)	0.77(0.49;1.20)
Target's GDP per capita (log, sc)	——	——	0.93(0.60;1.46)	0.91(0.58;1.41)
Attacker's CINC score (sc)	1.72***(1.27, 2.33)	1.42*(1.05, 1.93)	1.76***(1.30;2.38)	1.73***(1.28;2.36)
Distance between two states (sc)	1.08(0.79, 1.48)	1.02(0.73, 1.42)	1.08(0.78;1.49)	1.09(0.80;1.50)
Nuclear-Armed Target (dummy)	2.12*(1.06, 4.22)	3.28***(1.62, 6.65)	2.77**(1.36;5.64)	2.24*(1.11;4.53)
		Random Effects		
Dyad (intercept)	✓	✓	✓	✓
Variance	1.16	1.09	1.36	1.16
Dyad (slope for `Type`)	✓	✓	✓	✓
Variance	0.21	0.73	0.30	0.20
Akaike Inf. Crit.	1165.2	1213.4	1331.6	1167.3
Number of obs-ns	2,300	2,300	2,300	2,300

Note: Results are from a logistic regression model. The reported values are the odds ratios and confidence intervals. Odds ratio larger than 1 identify positive correlation and those smaller than 1 identify negative correlation. There are 2,300 observations. All results are based on two-tailed tests. Variables: `log`—logarithmized; `sc`—standardized; $^{\diamond}$ identifies similar types of operations as models DVs (e.g., in the model where cyber espionage is a DV (Model 1), the independent variable is lagged cyber espionage). ∧p<0.1; *p<0.05; **p<0.01; ***p<0.001

4.8 Additional Controls

It is plausible that states with more Internet access would also tend to be wealthier and therefore more attractive targets for cyber-attacks, commercial espionage and others forms of spying, and potentially less attractive to fight because of factors such as interdependence. We run additional models with GDP per capita, taken from the World Bank, as a confounding factor. Models 3 and 4 in Table 6 which present the obtained results demonstrate that the earlier obtained results hold. High correlation between GDP per capita and Internet users per capita (80% as Figure 1 shows) is why we did not include GDP per capita in our main analysis. Moreover, higher AICs in the models with GDP per capita (Models 3 and 4 in Table 6) than in the models without GDP per capita (Model 3 in Table 3 and Model 1 in Table 6) show the worse fit of the models with GDP per capita.

References

Carroll, Raymond J, David Ruppert, Leonard A Stefanski and Ciprian M Crainiceanu. 2006. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC.

Dietrich, Nick and Kristine Eck. 2020. "Known unknowns: media bias in the reporting of political violence." *International Interactions* pp. 1–18.

Singer, J David, Stuart Bremer and John Stuckey. 1972. "Capability distribution, uncertainty, and major power war, 1820-1965." *Peace, war, and numbers* 19:48.

Stinnett, Douglas M, Jaroslav Tir, Paul F Diehl, Philip Schafer and Charles Gochman. 2002. "The correlates of war (cow) project direct contiguity data, version 3.0." *Conflict Management and Peace Science* 19(2):59–67.

Weidmann, Nils B. 2016. "A closer look at reporting bias in conflict event data." *American Journal of Political Science* 60(1):206–218.