Introducing ROLE: Online-Only Appendix

Section 1. Comparing ROLE with Other Datasets

Table A1 compares ROLE with other recent datasets on rebel organizations. Among these, Prorok (2016) and Doctor (2020) both focus on particular aspects of individual rebel leaders' backgrounds, namely rebel leaders' war culpability (Prorok 2016) and their military and political experiences (Doctor 2020). In comparison, ROLE codes a much richer set of biographical information on rebel leaders, including birth and childhood details, family background, schooling information, adult personal and professional experiences, international experiences, language ability, mode of death, and more.

Table A1: Comparing ROLE with Recent Datasets on Rebel Groups and Rebel Leaders[1]

Original Dataset	*Content (about rebel groups)*	*Temporal coverage*	*Unit of analysis*	*Sample*	*List of civil conflicts from:*
Tiernay (2015)	Rebel leaders and leadership change	1989-2003	Leader	197 rebel leaders	Cunningham, Gleditsch & Salehyan (2009)
Prorok (2016)	Rebel leaders' war culpability	1980-2011	Leader; dyad	521 rebel leaders	NSA Dataset (Cunningham, Gleditsch & Salehyan 2013)
Rebel Leader Ascension Dataset (Cunningham & Sawyer 2019)	Rebel leaders and leader ascension method	1989-2011	Dyad-year		UCDP Armed Conflict Database
Lutmar & Terris (2019)	Rebel group leadership change	1946-2010	Rebel group-month	284 rebel groups	Regan (2002)
Foundations of Rebel Group Emergence (FORGE) (Braithwaite & Cunningham 2020)	Origins of rebel groups	1946-2011	Rebel group; dyad-year	428 rebel groups	NSA Dataset (Cunningham, Gleditsch & Salehyan 2013)
Rebel Leaders in Civil War Dataset (RLCW) (Doctor 2020)	Rebel leaders' prior military and political experiences	1989-2014	Leader; leader-year	206 rebel leaders	UCDP One-Sided Violence Dataset (Allansson, Melander & Themmer 2017)
ROLE (Authors)	Rebel leader biographies and attributes (from birth to death)	1980-2011	Leader; leader-year	488 rebel leaders	NSA Dataset via Prorok's (2016) rebel leader list

Section 2. Origins of ROLE Rebel Leader Sample

As explained in the main text, our sample of rebel leaders originates from a study conducted by Prorok (2016). As part of her research on the effects of responsibility for war initiation among state and rebel leaders, Prorok identified the top leaders of all rebel organizations in the Non-State Actors in Armed Conflict Dataset (NSA) (Cunningham, Gleditsch, and Salehyan 2013). Building on Prorok's list of leaders ensured that our database was fully compatible with the NSA dataset on

[1] Some of these data sources also contain information on state leaders. The table focuses on the portions of each one containing information on rebel groups and/or rebel leaders.

rebel organizations and the wider UCDP/PRIO family of conflict databases (Gleditsch et al. 2002) so that researchers can easily combine ROLE with a wide array of conflict data.

As is clear from Table A1 above, however, the number of cases in ROLE (488) is slightly different from the number in Prorok (521). Here, we outline how we arrived at our final number so that readers fully understand our sample. Excluding state leaders, the original Prorok database had 543 cases of rebel leadership, but 22 were not used in her study because the leaders were unidentified (leaving her with 521). We then added nine cases to that number, with the additions coming from two sources. First, we split up the power-sharing cases in her data (in which multiple leaders were combined into a single entry) so that we could collect individual leader-level data on them. There were five such cases, including one with three leaders.[2] Similarly, we split up a pair of twins who jointly led a movement into individual cases, adding another observation.[3] Overall, this added seven cases to the list. Second, we added two other cases that appear to have been absent from Prorok's dataset.[4] In sum, these changes added nine observations, yielding a total of 530. Yet, while Prorok's dataset is primarily designed to cover the period 1980-2011, it also includes a number of rebel leadership cases that end well before 1980. These cases were likely included for reasons specific to her empirical analysis, but we drop them from ROLE because they are a small and unrepresentative set of pre-1980 (or 1945-1980) leaders. After dropping these 42 observations, we arrive at our final count of 488 cases of rebel leadership between 1980 and 2011 included in the ROLE database.

---

[2] Specifically, these cases were: (1) Ekow Dennis and Edward Adjei-Ampofo, (2) Merid Negusie and Amsha Desta, (3) Jorge Soto, Rodrigo Asturias, and Rolando Moran, (4) Isak Chisi Swu and Thuingaleng Muivah, and (5) Radulan Sahiron and Yasser Igasan.

[3] These were the young twins Johnny and Luther Htoo, who jointly led the God's Army in Myanmar.

[4] These were: (1) Luis Augusto Turcias Lima, and (2) Abu Umar al-Baghdadi.

Section 3. Control Variables

We provide more detailed information on the variation and origins of the control variables included in the main results here. In particular, Table A2 shows the type and range of all of the organizational and contextual control variables used in the base models as well as the data sources from which they were drawn.

Table A2: Supporting Information on Organizational and Contextual Control Variables

Variable	*Type of Variable*	*Observed Range*	*Original Source*
Rebel strength	Ordinal variable	0–3	NSA Dataset (Cunningham, Gleditsch, & Salehyan 2013)
Rebel centralization	Ordinal variable	0–2	NSA Dataset (Cunningham, Gleditsch, & Salehyan 2013)
Foreign support	Dummy variable	0–1	Dangerous Dyads (San-Akca 2016)
Natural resource use	Dummy variable	0–1	Rustad & Binningsbo (2012)
Real GDP per capita	Continuous variable	132.82–43499.89	Gleditsch (2002)
Polity score	Continuous variable	-6–7	Vreeland (2008)
Incompatibility/ territorial dispute	Dummy variable	0–1	UCDP/PRIO Armed Conflict Dataset (Petterson & Oberg 2020)
Conflict duration	Count variable (years)	0–45	Prorok (2016) (for campaign start dates)

Section 4. Incomplete Data

As noted in the manuscript, we were able to gather complete or near-complete information on around 77% of the cases, with varying levels of information on the others. We also noted that incomplete data clustered in certain types of states, which were largely poorer and less democratic than their peers. To probe this formally, we constructed a binary indicator *COMPLETE* which takes a value of 1 if roughly 2/3 of the variables on the leader are coded, and 0 otherwise.[5] We predict this variable using logistic regression models, with a number of key contextual variables included. At the country level, we include the country's GDP, Polity score, and Composite Index of National Capabilities (CINC) score (a common measure of state strength in IR).[6] At the conflict level, we include the conflict's intensity, duration, and incompatibility type (territorial or not). We also add dummy variables for each region of the world as independent variables.

Table A3: Predictors of Incomplete Rebel Leader Data

	Incomplete Data	Incomplete Data
GDP per capita	-0.00*	-0.00*
	(0.00)	(0.00)
Polity score	-0.10**	-0.12**
	(0.04)	(0.04)
CINC score	0.49***	0.67***
	(0.14)	(0.19)
Territorial dispute	0.04	0.32
	(0.38)	(0.36)
Conflict duration	-0.13***	-0.13***
	(0.03)	(0.04)
Conflict intensity	-2.16***	-2.10***
	(0.50)	(0.48)
Africa		0.02
		(0.56)
Asia		-0.77
		(0.75)
Middle East		-0.65
		(0.56)
Constant	1.48**	1.73*
	(0.55)	(0.69)
Observations	1,776	1,691

Results from logit regression models. Robust country-clustered standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table A3 presents the results from this analysis. Poverty, autocracy, CINC score, conflict intensity, and conflict duration all significantly predict incompleteness. In particular, incomplete

[5] On the approximately three dozen core variables in the dataset, the average number of missing data points is two for the complete or near-complete cases, and 25 for the incomplete ones.
[6] In particular, CINC scores measure state strength by combining information on six different components: population, urban population, iron and steel production, energy consumption, military expenditure, and military personnel. These data are compiled and published as part of the Correlates of War project.

data are more likely in poorer, stronger, and less democratic countries as well as shorter and less intense conflicts. None of the three region dummies are significant, though the Europe dummy had to be dropped since there were no missing observations there. This suggests that there may be more or better information on the European cases than on cases from the Global South (though there are few European cases to start with).

While we cannot include missing leaders in our models, we can account for *the amount of available information on leaders more broadly (among missing and non-missing observations)* as a useful proxy. Indeed, "information availability" can be conceptualized as a continuous variable that varies among both missing and non-missing cases, and is a potential unmeasured confounder in the analysis. Specifically, if leaders with more information on them are more likely to have their experiences (e.g., combat or education) recorded, and there is more information available in cases with more terrorism (since it attracts media and scholarly attention), then information availability could be an unmeasured confounder lurking behind our results.

We address this issue in two ways. First, we replicate the base models with all five of the significant predictors of missingness identified above: GDP per capita, Polity score, CINC score, conflict intensity, and conflict duration.[7] The addition of these factors helps ensure that our results are robust to predictors of low information availability on rebel leadership. Second, we add a more direct measure of information availability to the analysis: *GSCHOLAR*, which measures the number of search hits each leader's name receives on Google Scholar. While imperfect, this is an indicator of scholarly attention to the individuals in the sample and thus proxies for their prominence in the historical record. If information availability is a strong confounder behind our results, we should see our results meaningfully change when this measure is included (since it could be systematically

[7] GDP per capita, Polity score, and duration were already in our base models, so this simply required the addition of CINC score and conflict intensity.

related to "higher" values on both leader attributes and terrorism, creating an apparent but spurious relationship). However, Table A4 shows that our core results are unchanged, speaking against the idea that information availability or scarcity is driving our results.

Table A4: Replication with Measure of Information Availability

	Terrorist Fatalities	Terrorist Fatalities	Terrorist Fatalities	Terrorist Fatalities
Leader Attributes				
Education	-0.62***	-0.53***	-0.58***	-0.47***
	(0.12)	(0.11)	(0.14)	(0.13)
Combat experience	-1.91***	-1.51***	-2.04***	-1.55***
	(0.41)	(0.35)	(0.39)	(0.39)
Age	-0.02	-0.03*	-0.02	-0.03*
	(0.01)	(0.01)	(0.02)	(0.01)
Military experience	0.75	0.47	0.92*	0.64+
	(0.48)	(0.43)	(0.36)	(0.37)
Organizational Features				
Rebel strength	-0.02	-0.03	0.15	0.18
	(0.24)	(0.23)	(0.31)	(0.32)
Rebel centralization	-0.16	-0.14	0.27	0.34
	(0.26)	(0.25)	(0.25)	(0.27)
Foreign support	-0.84*	-0.61+	-0.52	-0.28
	(0.37)	(0.35)	(0.44)	(0.44)
Natural resource use	0.52	0.62	0.67	0.68
	(0.44)	(0.44)	(0.48)	(0.46)
Contextual Factors				
GDP per capita	0.00	0.00	0.00*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Polity score	0.21***	0.21***	0.18***	0.19***
	(0.05)	(0.05)	(0.05)	(0.05)
Territorial dispute	0.96*	0.88*	1.57**	1.43**
	(0.40)	(0.38)	(0.48)	(0.44)
Conflict duration	0.01	0.01	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Information Availability				
CINC score	-0.23	-0.28		
	(0.19)	(0.18)		
Conflict intensity	1.64***	1.65***		
	(0.24)	(0.24)		
Google Scholar hits			-0.00*	-0.00
			(0.00)	(0.00)
Constant	3.99***	3.87***	5.40***	5.10***
	(1.21)	(1.19)	(1.25)	(1.25)
Observations	971	971	1,004	1,004

Results from negative binomial regression models. The first two models use TAC's high estimate of annual terrorist fatalities, while the latter two use its low one. Robust country-clustered standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Section 5. Model Specification

Table A5: Replication with Additional Covariates

	Terrorist Fatalities	Terrorist Fatalities
Leader Attributes		
Age	-0.03	-0.04*
	(0.02)	(0.02)
Education	-0.51**	-0.39*
	(0.16)	(0.15)
Military experience	1.22*	0.97+
	(0.52)	(0.53)
Combat experience	-2.16***	-1.70***
	(0.41)	(0.33)
Organizational Features		
Rebel strength	0.58+	0.56
	(0.35)	(0.35)
Rebel centralization	0.27	0.36
	(0.29)	(0.25)
Foreign support	-0.36	-0.05
	(0.42)	(0.40)
Natural resource use	0.53	0.58
	(0.67)	(0.57)
Contextual Factors		
Real GDP per capita	0.00	-0.00
	(0.00)	(0.00)
Polity score	0.19***	0.20***
	(0.04)	(0.04)
Territorial dispute	1.45***	1.37***
	(0.41)	(0.39)
Conflict duration	0.01	0.02
	(0.03)	(0.03)
Additional Covariates		
Territorial control	-0.47	-0.53
	(0.58)	(0.57)
Diaspora support	0.63	0.43
	(0.40)	(0.36)
Theocratic ideology	0.58	0.55
	(1.19)	(1.08)
Previously active	0.67	0.87
	(0.69)	(0.58)
Africa	-0.75	-0.54
	(0.69)	(0.62)
Constant	4.30**	4.22**
	(1.57)	(1.44)
Observations	958	958

Results from negative binomial regression models. The first model uses TAC's high estimate of annual terrorist fatalities, while the second one uses its low one. Robust country-clustered standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

We also conduct some model specification checks to ensure the robustness of our findings. While the base models include a wide range of covariates culled from existing studies of rebel use of terrorism, we add several others to account for additional factors. These include: a measure of rebel territorial control, which can impact incentives for indiscriminate violence (Kalyvas 2006), the existence of a sympathetic diaspora, which is linked to more hardline militant tactics (Piazza 2018), a measure of theocratic group ideology, which is linked in some research to commitment to violence (Toft and Zhukov 2015), a measure of whether a conflict has been previously active, given the strong effects of conflict on national development (Gates et al. 2012) which could shape opportunities for would-be rebel leaders to gain experiences like education, and a dummy variable for cases in Africa given the prevalence of more conventional civil wars there (Fortna 2015).

Table A5 presents the results with these additions to the primary model. As is apparent, the central conclusions are broadly similar to the main text. In particular, rebel leader education and combat experience are still associated with significantly less terrorism in both models. Meanwhile, age is associated with significantly less terrorism in one of two models, and military experience is associated with significantly more (although in the second model this effect is only significant at the 10% level). As for the organizational and contextual covariates, the results mirror those in the text, with democracy and territorial conflicts linked to more terrorism.[8] Overall, the results are thus quite similar to those in the main text and speak to the robust impact of rebel leader attributes on rebel organizations' use of terrorism.

[8] We do find in Table A4 that GDP per capita and foreign support have significant effects at the 5% level in one of the four models each. However, our view is that this should not be overinterpreted as the bulk of the evidence does not show significant effects for these variables.

Section 6. Inter-Coder Reliability (ICR) Tests

To address potential concerns about data quality and reliability, we conducted inter-coder reliability checks on ROLE. Inter-coder, or inter-rater, reliability refers to the degree of sameness in specific data entries coded by more than one individual.[9] As Gwet (2014, 6) notes: "inter-rater reliability is concerned about the reproducibility of measurements by different raters…." Data often suffer from multiple-coder bias, where various researchers code the same variable differently for different or the same observations. Salehyan (2015, 107-108) notes:

> *Whenever possible, researchers should also check the validity of the coding procedure by computing intercoder reliability statistics [e.g. Cohen's Kappa]… rarely do research projects report such statistics. Often it is unfeasible to double-code all data points and variables of interest. Yet researchers can randomly sample a percentage of cases for double-coding in order to refine and improve the coding procedure as well as aid in the training of researchers.*

The simplest way to measure agreement between coders is to calculate the percentage of agreed-upon data entries between them (Osgood 1959). Nevertheless, methodologists early on in the development of inter-coder reliability (ICR) studies realized that merely calculating percent agreement ignores the possibility that multiple coders may agree on something by chance or by a "deterministic rating procedure" (Gwet 2014, 29). A simple percentage of agreement could capture not only the reliability of multiple coders but also their agreement on chance, as well as their forced agreement through a heavy-handed codebook.

Indeed, before the full-fledged development of inter-rater reliability studies, scientists understood the necessity of controlling for chance agreement (Benini 1901). Various subsequent efforts independently converged around formulas such as:

[9] The esoteric distinction between "rater" and "coder" derives from "coders" using codebooks to aid them in the rating of a data entry (Scott 1955). Some fields refer to "judges" instead (Kassarjian 1977).

$$\frac{p_a - \frac{1}{q}}{1 - \frac{1}{q}}$$

where p_a is the percent agreement and q the number of nominal categories in the rating scale (Gwet 2014, 30). Still, various agreement coefficients emerged with Cohen's (1960) Kappa becoming the standard. To fully account for "chance," Cohen (1960) estimates "the expected percent chance agreement," as denoted by p_e.[10] Cohen puts forth his Kappa Coefficient as follows:

$$\kappa = \frac{Pa - Pe}{1 - Pe}$$

In an effort to generalize the Kappa Coefficient, Fleiss (1971) adapts the measure for multiple coders—defining the percent change agreement that any pair of coders could provide the same classification. Fleiss does not however account for missingness in one or more coders' values for a given subject. Gwet (2008) devises AC_1 to overcome some mathematical limitations of the Kappa Coefficient.[11] We present Fleiss's Kappa, Gwet's AC_1, and other relevant metrics (including straightforward percent agreement) in checks conducted on ROLE.

To build ROLE, we completed the initial multi-year data collection process by personally coding observations as well as delegating portions of the coding process to 16 research assistants from five different universities. As a follow up, we adopted Salehyan's (2015) procedural advice and randomly sampled 40 leaders (about 8% of our data) and hired two research assistants to each independently *re*-collect data on all of the variables for them. We then calculated ICR statistics for assessing the degree of agreement between the coding sets.

[10] While correcting for the potential for chance agreement, Cohen's Kappa (1960) assumes independence of coders—limiting the measure's utility in some applications. Scott's Pi (1955) is a similar to Cohen's Kappa but is more suitable for studying the frequency of use of categories by different coders. Krippendorff's Alpha (1980) is nearly identical to Scott's Pi.
[11] In particular, there are a number of instances where Cohen's Kappa yields paradoxically low coefficients (Cicchetti and Feinstein 1990).

Table A6 below shows the results of the ICR exercise. The table shows the median number of coders for each ROLE variable recoded, the average number of coders, the percent agreement between the coders, and the three leading ICR measures: Krippendorff's Alpha, Fleiss' Kappa Coefficient, and Gwet's AC_1. A standard benchmark for assessing ICR measures are as follows: 0.8000-1.000 indicates an excellent degree of agreement between coders; 0.6000-0.7999 signifies substantial agreement; 0.4000-0.5999 reflects moderate agreement; 0.2000-0.3999 suggests some agreement; and 0.000-0.1999 indicates a poor degree of agreement (Landis and Koch 1977). ROLE's coding rules generate a high degree of reliability in the data collected.

Table A6: ICR Results

ROLE Variable	***n***	**Average # of Coders**	**Percent Agreement**	**Krippendorff's Alpha**	**Fleiss' Kappa**	**Gwet's AC_1**
Western Educated	30	1.83	0.8800	0.7200	0.6800	0.8080
Military Experience	32	1.81	0.8846	0.7650	0.7494	0.7861
Combat Experience	32	1.59	0.8421	0.6783	0.6737	0.6917
Married	32	1.66	1.0000	n/a	1.0000	1.0000
Study Abroad	32	1.78	0.8800	0.7633	0.7479	0.7710
Military Abroad	32	1.75	0.9167	0.8025	0.7562	0.8734

Section 7. Checking for Multicollinearity

Finally, we also conduct Variance Inflation Factor (VIF) test to check for multicollinearity between our main explanatory variables for the full models in Table A2 in the main text. Generally, VIF values greater than 10 signify a problematic level of redundancy between variables, though some analysts use a stricter standard such as values over 5 indicating concern. By either standard, the check reveals no multicollinearity between variables in our model.

Table A7: Multicollinearity Test

	VIF	1/VIF
Leader Attributes		
Age	1.23	0.811
Education	1.25	0.803
Military experience	1.23	0.815
Combat experience	1.32	0.758
Organizational Features		
Rebel strength	1.14	0.878
Rebel centralization	1.06	0.940
Foreign support	1.07	0.934
Natural resource use	1.33	0.753
Contextual Factors		
GDP per capita	1.53	0.654
Polity score	1.47	0.682
Incompatibility	1.35	0.741
Conflict duration	1.27	0.787
Mean VIF	1.27	

Section 8. Replication with Terrorist Attacks

In addition to the above analyses, we also replicate our primary models using estimates of terrorist attacks rather than fatalities from TAC. In particular, Table A8 replicates the results from Table 2 in the text using low and high estimates of terrorist attacks initiated per organization-year, as opposed to low and high estimates of annual terrorist fatalities. As is apparent, our substantive findings are unchanged, with both education and combat experience significantly diminishing the number of terrorist attacks perpetrated in both pairs of models. The control variables also behave in a similar fashion, with significantly more terrorism in territorial conflicts and democratic states and few other significant effects. These models show the robustness of our substantive findings to alternative measures of terrorism use or activity.

Table A8: Replication with Measures of Annual Terrorist Attacks

	(M1) Terrorist Attacks	(M2) Terrorist Attacks	(M3) Terrorist Attacks	(M4) Terrorist Attacks
Leader Attributes				
Education		-0.35**		-0.29*
		(0.13)		(0.13)
Combat experience		-1.05***		-0.98***
		(0.26)		(0.26)
Age		-0.02		-0.02*
		(0.01)		(0.01)
Military experience		0.14		-0.27
		(0.36)		(0.36)
Organizational Features				
Rebel strength	-0.03	-0.16	-0.02	-0.24
	(0.34)	(0.29)	(0.33)	(0.29)
Rebel centralization	0.44*	0.28	0.48*	0.33+
	(0.22)	(0.21)	(0.19)	(0.19)
Foreign support	0.03	-0.06	0.16	0.18
	(0.30)	(0.38)	(0.30)	(0.40)
Natural resource use	0.52	0.56	0.62	0.76
	(0.41)	(0.47)	(0.39)	(0.48)
Contextual Factors				
GDP per capita	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Polity score	0.22***	0.23***	0.23***	0.22***
	(0.05)	(0.05)	(0.05)	(0.05)
Territorial dispute	1.12**	1.04**	1.24***	1.19***
	(0.37)	(0.34)	(0.33)	(0.28)
Conflict duration	0.03	-0.00	0.03	0.01
	(0.03)	(0.03)	(0.03)	(0.02)
Constant	-0.32	2.61*	0.16	3.36***
	(1.01)	(1.16)	(0.90)	(0.99)
Observations	1,374	1,004	1,374	1,004
Logged Likelihood	-2407.55	-1912.27	-3005.69	-2384.03

Results from negative binomial regression models. M1-M2 use TAC's high estimate of annual terrorist attacks, while M3-M4 use its low one. Robust country-clustered standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

References

Allansson Marie, Erik Melander & Lotta Themnér (2017) Organized Violence, 1989–2016. *Journal of Peace Research* 54(4):574-587.

Benini, R (1901) *Principii di Demongraphia: Manuali Barbera Di Scienze Giuridiche Sociali e Politiche* (No. 29). Firenze, Italy: G. Barbera.

Braithwaite, Jessica & Kathleen Gallagher Cunningham (2020) When Organizations Rebel. *International Studies Quarterly* 64(1):183-193.

Cicchetti, Domenic V. & Alvan R. Feinstein (1990) High Agreement but Low Kappa: II. Resolving the Paradoxes. *Journal of Clinical Epidemiology* 43: 551-558.

Cohen, Jacob (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20: 37-46.

Cunningham, David E., Kristian Skrede Gleditsch, & Idean Salehyan (2009) It Takes Two: A Dyadic Analysis of Civil War Duration and Outcome. *Journal of Conflict Resolution* 53(4): 570–97.

Cunningham, David E., Kristian Skrede Gleditsch, & Idean Salehyan (2013) Non-State Actors in Civil Wars: A New Dataset. *Conflict Management and Peace Science* 30 (5): 516-531.

Cunningham, Kathleen Gallagher & Katherine Sawyer (2019) Conflict Negotiations and Rebel Leader Selection. *Journal of Peace Research* 56(6):619-634.

Doctor, Austin (2020) A Motion of No Confidence: Leadership and Rebel Fragmentation. *Journal of Global Security Studies* 5(4):598-616.

Fleiss, Joseph L (1971) Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin* 76 (5): 378-382.

Fortna, Virginia Page (2015) Do Terrorists Win? *International Organization* 69(3):519-556.

Gates, Scott, Havard Hegre, Havard Mokleiv Nygard, & Havard Strand (2012) Development Consequences of Armed Conflict. *World Development*, 40(9): 1713-22.

Gleditsch, Kristian Skrede (2002) Expanded Trade and GDP Data. *Journal of Conflict Resolution* 46(5): 712–24.

Gwet, Kilem L. (2008) Computing Inter-Rater Reliability in the Presence of High Agreement. *British Journal of Mathematical and Statistical Psychology* 61(1): 29-48.

Gwet, Kilem L. (2014) *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters*, 4th ed. Gaithersburg, MD: Advanced Analytics.

Kalyvas, Stathi (2006) *The Logic of Violence in Civil War*. Cambridge, UK: Cambridge University Press.

Kassarjian, Harold H. (1977) Content Analysis in Consumer Research. *Journal of Consumer Research* 4: 8-18.

Krippendorff, Klaus (1980) *Content Analysis: An Introduction to Its Methodology*. Beverly Hills, CA: Sage.

Landis, Richard J., & Gary G. Koch (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33(1): 159-74.

Lutmar, Carmela, & Lesley Terris (2019) Introducing a New Dataset on Leadership Change in Rebel Groups, 1946–2010. *Journal of Peace Research* 56(2):306-315.

Osgood, C.E. (1959) The Representational Model and Relevant Research Methods. In *Trends in Content Analysis*, ed. Ithiel de Sola Pool. Urbana: University of Illinois Press.

Pettersson, Therese, & Magnus Öberg (2020) Organized Violence, 1989-2019. *Journal of Peace Research* 57(4): 535-47.

Piazza, James (2018) Transnational Ethnic Diasporas and the Survival of Terrorist Organizations. *Security Studies* 27(4): 607-32.

Prorok, Alyssa K. (2016) Leader Incentives and Civil War Outcomes. *American Journal of Political Science* 60(1):70-84.

Prorok, Alyssa K. (2018) Led Astray: Leaders and the Duration of Civil War. *Journal of Conflict Resolution* 62(6):1179-1204.

Regan, Patrick (2002) Third Party Interventions and the Duration of Intrastate Conflicts. *Journal of Conflict Resolution* 46(1):55-73.

Rustad, Siri A., and Helga M. Binningsbo (2012) A Price Worth Fighting For? Natural Resources and Conflict Recurrence. *Journal of Peace Research* 49(4): 531–46.

Salehyan, Idean (2015) Best Practices in the Collection of Conflict Data. *Journal of Peace Research* 52(1):105–109

San-Akca, Belgin (2016) *States in Disguise: Causes of External State Support for Rebel Groups.* New York: Oxford University Press.

Scott, William A. (1955) Reliability of Content Analysis: The Case of Nominal Scale Coding. *Public Opinion Quarterly*, 19: 321-325.

Tiernay, Michael (2015) Killing Kony. *Journal of Conflict Resolution* 59(2):175-206.

Toft, Monica, & Yuri Zhukov (2015) Islamists and Nationalists: Rebel Motivation and Counterinsurgency in Russia's North Caucasus. *American Political Science Review* 109(2): 222-38.

Vreeland, James Raymond (2008) The Effect of Political Regime on Civil War: Unpacking Anocracy. *Journal of Conflict Resolution* 52(3): 401–25.