Codebook for:
One for All? State Violence and Insurgent Cohesion

Livia I. Schubiger

International Organization

Table 1 describes the variables used to generate the results presented in the paper and appendix. In the cross-sectional data, the unit of analysis is the conflict dyad; in the time-series data, it is the dyad-year. While based on the UCDP Dyadic Conflict Dataset (Harbom, Melander and Wallensteen, 2008, v 18.1), the time-series data include inactive years (see below). *MKactive* is an indicator of whether an armed group's alleged civilian constituency was affected by state-led collective targeting in the context of state-orchestrated mass killings at any point during the armed group activity after conflict onset (cross-sectional dataset) or in a particular year (time-series dataset). The coding procedure here starts with the dataset and coding notes of Ulfelder and Valentino (2008),[1] which identify episodes of state-led mass killings and the target constituencies. It also draws on additional sources on a case-by-case basis. This variable is coded 1 if one of the following links between a state-sponsored mass killing and an armed group applies: (i) the link between the mass killing and a particular armed group is explicitly mentioned in the mass killing data coding notes; additionally, this seems to be the primary target group; (ii) the link between the state-led mass killing and a particular insurgent group is explicitly mentioned in the mass killing coding notes, yet there are other targeted groups; (iii) the link between the state-led mass killing and a particular armed group and year is not explicitly mentioned in the mass killing coding notes, but there is otherwise an indication that this group's alleged constituency was targeted in this episode. The variable is coded zero if none of these links apply in the relevant time period.

Frag is an indicator for whether the armed group of a dyad was affected by splintering during a given year (time-series dataset). This variable is coded 1 if a faction broke away and subsequently ended up being involved in any form of organized violence recorded by UCDP. In the cross-sectional dataset, frag_postMK1 and frag_postMK1alt indicate whether fragmentation occurred after the initial conflict period or exposure to state violence (main and alternative definition), while frag_preMK1 and frag_preMK1alt are indicators for prior insurgent fragmentation. The main sources for the fragmentation variables are the Foundations of Rebel Group Emergence (FORGE) Dataset (Braithwaite and Cunningham, 2019, v. 1.0), as well as the UCDP Actor Dataset (Pettersson, 2015, v. 2.2.-2015). Additional sources were consulted on a case-by-case basis, for example to determine the year of a split or to determine if splintering occurred as opposed to distinct dynamics of group formation and disintegration.

The time-series dataset includes years of conflict inactivity if the armed group in a given year was clearly or potentially active (despite the conflict not reaching the conventional threshold of 25 battle-related deaths). *Conflictactive* is an indicator of whether an armed conflict was active during a given year according the the UCDP Dyadic Conflict Dataset (Harbom, Melander and Wallensteen, 2008); inactive years are included and coded zero if the armed group is clearly or potentially still active as an armed opposition group (various sources).[2] *Dyadsfirstyear* indicates the number of active dyads in a conflict during the first year, also based on the UCDP Dyadic Conflict Dataset. *Excluded* indicates whether the rebel group of a conflict dyad recruits its members from an ethnic group excluded from

[1] Dataset and coding notes as of 2017.

[2] If a conflict does not resume activity, the maximum number of years a dyad can be included and coded 0 is 5.

state power. This variable is based on the ACD2EPR Dataset (Wucherpfennig et al., 2012, v. 2018) and the Ethnic Power-Relations Dataset (Vogt et al., 2015, v. 2018). If ethnicity is politicized in a country, recruitment from ethnic groups classified as politically "irrelevant" is also coded 1. *Excluded population* indicates the fraction of the population that is excluded from access to political power and adopted from Vogt et al. (2015). *Irregular2* is a dummy variable identifying irregular civil wars. This variable is adopted from Balcells and Kalyvas (2014) but expanded to code missing values where possible. The variables *prevactive*, *rebsupnum*, *fightcapnum*, and *terrcontnum* are based on the Non-State Actor Dataset (Cunningham, Gleditsch and Salehyan, 2009, v. 3.4-1). The variable *e_v2x_neopat* is based on the Varieties of Democracy Dataset (Sigman and Lindberg, 2017; Varieties of Democracy, 2018). More information on included cases, variables, coding procedures, and data sources is provided in the article.

Variable	**Description**	**Sources**
	Unit & Time Identifiers and Time at Risk	
dyadid	dyad identifier	ucdp dyadic dataset (v. 18.1).
ccode**	country identifier	
year*	calendar year	
counterdyad*	number of years since conflict onset	
	State Violence	
MKactive	armed group constituency affected by state-led collective targeting in context of state-sponsored mass killings (1: yes, 0: no)	based on Ulfelder & Valentino 2008 and associated coding notes, as well as additional case-specific sources.
MKactive_alt*	corresponds to MKactive, 1st mass killing episode only	
MKprevious*	state-led mass killings affected country prior to this conflict and/or year (1: yes, 0: no)	
	Fragmentation	
frag*	insurgent fragmentation (1: yes; 0: no)	main sources: FORGE dataset (v. 1.0) & UCDP actor dataset (v. 2.2.-2015); additional sources by case.
frag_postMK1**	insurgent fragmentation (1: yes; 0: no)	
frag_preMK1**	prior insurgent fragmentation (1: yes; 0: no)	
frag_postMK1alt**	insurgent fragmentation, alternative definition (1: yes; 0: no)	
frag_preMK1alt**	prior insurgent fragmentation, alternative definition (1: yes; 0: no)	
mktosplit1*	years from 1st mass killing onset to 1st split for "treated" cases	
fracount*	number of splits up to given year	
	Covariates	
intensityfirstyear**	intensity of conflict during first year (1: low intensity, 2: high intensity)	ucdp dyadic dataset (v. 18.1).
dyadsfirstyear	number of dyads per conflict, first year	
conflictactive*	intensity of conflict (1: conflict active, 0: conflict inactive, but armed group clearly or potentially active)	conflict activity: ucdp dyadic dataset; armed group activity: additional sources by case.
irregular2	irregular war (1: yes; 0: no)	main source: Kalyvas & Balcells 2014; additional sources by case
excluded	recruitment from excluded ethnic groups	acd2epr dataset & epr dataset (v. 2018)
exclpopn	fraction of excluded population	epr dataset (v. 2018)
prevactive	conflict previously active (1: yes; 0: no)	non-state actor dataset (v. 3.4-1).
rebsupnum	external rebel support (1: explicit or alleged; 0: no)	
fightcapnum	relative fighting capacity (1: high or moderate; 0: low)	
terrcontnum	territorial control (1: yes; 0: no)	
e_v2x_neopat	neopatrimonial rule index	varieties of democracy dataset

* time-series dataset; ** cross-sectional dataset

Table 1: Variables and Sources

References

Balcells, Laia and Stathis N. Kalyvas. 2014. "Does Warfare Matter? Severity, Duration, and Outcomes of Civil Wars." *Journal of Conflict Resolution* 58(8):1390–1418.

Braithwaite, Jessica Maves and Kathleen Gallagher Cunningham. 2019. "When Organizations Rebel: Introducing the Foundations of Rebel Group Emergence (FORGE) Dataset." *International Studies Quarterly* 64(1):183–193.

Cunningham, David E., Kristian S. Gleditsch and Idean Salehyan. 2009. "It Takes Two: A Dyadic Analysis of Civil War Duration and Outcome." *Journal of Conflict Resolution* 53(4):570–597.

Harbom, Lotta, Erik Melander and Peter Wallensteen. 2008. "Dyadic Dimensions of Armed Conflict, 1946-2007." *Journal of Peace Research* 45(5):697–710.

Pettersson, Therese. 2015. UCDP Actor Dataset Version 2.2. Technical report Uppsala Conflict Data Program.

Sigman, Rachel and Staffan I. Lindberg. 2017. "Neopatrimonialism and Democracy: An Empirical Investigation of Africa's Political Regimes." The Varieties of Democracy Institute, Working Paper Series 2017: 56.

Ulfelder, Jay and Benjamin Valentino. 2008. "Assessing Risks of State-Sponsored Mass Killing." `SSRN: https://ssrn.com/abstract=1703426orhttp://dx.doi.org/10.2139/ssrn.1703426`.

Varieties of Democracy. 2018. "Codebook V 8 - April 2018." University of Gothenburg, V-Dem Institute.

Vogt, Manuel, Nils-Christian Bormann, Seraina Regger, Lars-Erik Cederman, Philipp Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family." *Journal of Conflict Resolution* 59(7):1327–1342.

Wucherpfennig, J., Nils Metternich, Lars-Erik Cederman and Kristian Skrede Gleditsch. 2012. "Ethnicity, the State and the Duration of Civil War." *World Politics* 64(1):79–115.