

The Ethnic One-Sided Violence Dataset

Codebook

Livia Isabella Schubiger, ETH Zurich & Hanne Fjelde, Uppsala University

May 27, 2019

PROJECT TEAM:

LARS-ERIK CEDERMAN & LIVIA SCHUBIGER, ETH ZURICH

SIMON HUG & MARCELO OLARREAGA, UNIVERSITÉ DE GENÈVE

HANNE FJELDE, LISA HULTMAN & MARGARETA SOLLENBERG, UPPSALA UNIVERSITY

When using the data, please cite the following publication (and this codebook, if appropriate):
H Fjelde, L Schubiger, L Hultman, LE Cederman, S Hug, M Sollenberg: Introducing the ethnic one-sided violence dataset, *Conflict Management and Peace Science*, forthcoming.

RESEARCH ASSISTANTS:

MOHAMMAD AGHDAM, BETTINA GREUTER, ELISE CLERC, MARTINA DOMINKOVIĆ,
LORENZO FIORITO, JOHAN VIGNE, UNIVERSITÉ DE GENÈVE
DANIELA ARAUZ, PIIA BRAENFORS, FABIAN MORGENTHALER, RALUCA STROE, FRAN
VILLAMIL, MARJENN WITT, MATTHEW ZELINA, ETH ZURICH
HENRIK PERSSON, PAULINA CRUZ VELASQUEZ, SARA SMITH, TOM RENVALL, GUDLAUG
OLAFSDOTTIR & JACOB LEWIS, UPPSALA UNIVERSITY

Contents

- 1 Summary** **1**

- 2 Data Structure** **1**

- 3 Coding Criteria** **3**
 - 3.1 Explicit vs. Implicit Information on Ethnicity 4
 - 3.2 Intention for Ethnic Targeting 5
 - 3.3 Data Sources 9

- 4 Variables** **10**

- References** **15**

List of Tables

- 1 One-sided Violence 1989-2013 2

1 Summary

Confrontations between state forces and armed opposition movements often cause considerable harm to civilian populations, much of which results from deliberate attacks against noncombatants (e.g., Valentino, Huth and Balch-Lindsay, 2004; Kalyvas and Kocher, 2007). While violence against civilians is not always based on ethnic targeting, ethnicity and associated ‘markers’ are among the primary criteria based on which civilians are collectively victimized in many armed conflicts (e.g., Fjelde and Hultman, 2014). And yet, whereas data on the perpetrators of one-sided violence is available on a global scale (Eck and Hultman, 2007), we lack comparable information about the collective identity of the victims of civilian targeting. The Ethnic One-Sided Violence Dataset (EOSV) is the first dataset with global coverage providing information on the ethnic identity of the victims of deliberate violence against noncombatants by armed groups. Specifically, the EOSV dataset disaggregates the civilian victim category in the Uppsala Conflict Data Program One-Sided Violence dataset (Eck and Hultman, 2007) by specifying the victims’ ethnic affiliation. Each ethnic victim group is then coded using the identifiers in the Ethnic Power Relations dataset on politically relevant ethnic groups worldwide (Cederman, Wimmer and Min, 2010; Vogt et al., 2014). These data will allow researchers to study, for example, *whose* civilian constituency is targeted by which armed actor. Moreover, the dataset allows researchers to connect the information on targeted ethnic groups to other relevant information, in particular the *Family of Ethnic Power Relations Datasets* (e.g., Vogt et al., 2014).

The data collection was coordinated between different research teams in Switzerland (ETH Zurich and University of Geneva) and Sweden (Uppsala University).¹

2 Data Structure

The basic unit of analysis is the perpetrator-year as in the UCDP One-Sided Violence (OSV) Dataset (Eck and Hultman, 2007). The OSV dataset provides data on the estimated number of

¹In Switzerland, the data collection is funded by the Swiss Network of International Studies (SNIS). This codebook is partially based on the project proposal ‘Civilian Victimization and Conflict Escalation’ submitted to SNIS in May 2014 on behalf of Lars-Erik Cederman, Simon Hug, Marcelo Olarreaga, and Livia Schubiger. In Sweden, the data collection is funded by the Swedish Research Council, through the project ‘Threatening Ties: Understanding wartime civilian targeting along ethnic lines’ (E0136501) led by Hanne Fjelde (together with Lisa Hultman and Margareta Sollenberg) that also entails a collaboration with Paul Huth and Jacob Aronson at the University of Maryland.

fatalities caused by *direct*² and *intentional*³ violence against civilians⁴ exerted by either government forces or non-state armed groups. For every perpetrator responsible for killing at least 25 civilians in one calendar year, the OSV dataset records estimates of the number of civilian victims based on the cross-validation of news reports. Table 1 lists the number of actor-years for the period 1989-2013, based on Eck and Hultman (2007, data version 1.4-2014). The table shows that there were 795 cases of at least 25 civilian killings in terms of perpetrator-years, the perpetrators being either state actors or non-state armed groups, such as rebel groups and militias. Based on this dataset, we code ethnic victim groups and link campaigns of one-sided violence to data on politically relevant ethnic groups.

Region	Actor-Years	Percent
Europe	33	4.15
Middle East	72	9.06
Asia	244	30.69
Africa	389	48.93
Americas	57	7.17
Total	795	100.00

Table 1: One-sided Violence 1989-2013

Specifically, we disaggregate the civilian victims into ethnic victim groups using ethnic group identifiers from the EPR core dataset, henceforth EPR (Vogt et al., 2014), which records all politically relevant ethnic groups as well as ethnic groups’ access to political power (Cederman, Wimmer and Min, 2010; Vogt et al., 2014). We largely follow the data structure proposed by Rügger (2013) and Rügger and Bohnet (2014) for the ethnic composition of refugee movements. As in the case of refugee movements, several ethnic groups can be affected by campaigns of civilian killings. This is taken into account by coding several ethnic groups per perpetrator and year where one-sided violence affects more than one group.

²‘Direct’ here refers to the criterion that only deaths directly caused by armed actors are included when referring to victims of one-sided violence, while deaths resulting indirectly from armed conflict such as through disease and starvation are excluded (Eck and Hultman, 2007). Note that this definition of ‘direct’ violence differs from the one proposed by Balcells (2010). Balcells (2010) distinguishes between violence against civilians in a context of face-to-face contact between the perpetrator and the victim and instances of violence during which no such direct contact occurs.

³The criterion of *intentionality* leads to the exclusion of civilian fatalities caused by ‘cross-fire’ and similar situations: “Often conflict parties attack each other with disregard to the civilian population, and while this may violate international law, it does not constitute one-sided violence” (Eck and Hultman, 2007, 235). Also excluded are extrajudicial executions in custody.

⁴We follow UCDP in defining civilians as any person unarmed at the time of being targeted, excluding lightly armed noncombatants, policemen, or military personnel (even if unarmed).

3 Coding Criteria

While we define ethnicity as the shared belief in a common descent and culture (Weber, 1996[1978]),⁵ we follow Cederman, Wimmer and Min (2010) in focusing mainly on *politically relevant ethnic groups*, that is, on ethnic groups whose interests are claimed to be represented by at least one political actor (for example, a political party) in the national political arena, or whose members are systematically excluded in certain domains of public politics. These politically relevant ethnic groups are included in the EPR core dataset (Vogt et al., 2014). We additionally include information on some ethnic groups not listed in EPR (as outlined below).

We try to identify, for each perpetrator and year, if the civilians killed belong to one or several ethnic groups. Whenever members of a given ethnic group are clearly identifiable as victims of one-sided violence, we assign the respective fatalities to the categories ‘Ethnic Group 1,’ ‘Ethnic Group 2,’ ‘Ethnic Group 3,’ etc, along with the EPR identifier (Vogt et al., 2014) if available. We do not require ethnic groups to be disproportionately targeted, nor do we require a stated intent of the perpetrators to target members of specific ethnic groups in order to record the ethnic identity of civilian victims. However, where information on the alleged intention of perpetrators is available (for example through reports documenting the screening of civilians for ethnic markers), we code this information separately (see section 3.2). We provide binary information (ethnic group affected yes/no), rather than group-specific fatality estimates, in the published dataset. If an ethnic group is listed for a given perpetrator and year, this implies that at least some victims were directly or indirectly (i.e., through explicit or implicit information) identified as belonging to that ethnic group. The minimal threshold for inclusion is, in principle, one fatality per ethnic group. However, as the level of precision in the information we have to determine the number of victims belonging to a certain group varies greatly across cases, and as information on violence against civilians as well as on victims’ identity tends to be biased, we prefer to not release fatality counts, and to interpret this variable in a binary way. See also Fjelde et al. (2019).

⁵Markers indicating a shared descent and culture can for example be a common language, phenotypical features, or religious beliefs (Weber, 1996[1978]; Cederman, Wimmer and Min, 2010).

3.1 Explicit vs. Implicit Information on Ethnicity

As a general rule, we only assign ethnic group identifiers if the ethnic identity of civilian victims is explicitly mentioned in at least one of our sources. This implies, for example, that the geographically concentrated occurrence of one-sided violence in areas primarily inhabited by a specific ethnic group is not sufficient to assign fatalities to specific ethnic groups. Instead, we corroborate the ethnic character of violence through sources that *explicitly* refer to the ethnic identity of civilian victims (be this through quantitative or qualitative data).⁶

As an exception to this rule, we infer the ethnic identity based on non-explicit information in special cases where the context in which one-sided violence occurs strongly implies that ethnicity is very likely to play an important role even though the ethnic identity of civilian victims is not specified in any of our sources. While there is no fixed set of situations in which such a coding based on ‘implicit information’ is allowed, codings based on implicit information have to be flagged, and coders have to carefully justify why the ethnic identity of civilians can be inferred from the context in which the targeting took place. One example of a situation where we would code the ethnic identity of civilian victims even if not explicitly mentioned is civilian killings that occur geographically concentrated in ethnically homogeneous areas (these areas being populated by the ethnic group opposed to the perpetrator) in an ongoing ethnic conflict with a recent history of ethnic targeting.⁷ Another example where we would code the ethnic identity of the victims even if it is not explicitly mentioned is where the civilians killed are members of an ethno-nationalist party that we know has a homogenous ethnic base.

For every perpetrator, year, and ethnic victim group, we report whether the ethnic identity assignment was based on explicit or implicit information.⁸ Since we aggregate information to years, our codings may be based on both types of information. Where information on the ethnic identity of civilian victims is available but ethnicity seems irrelevant for patterns of targeting for a given perpetrator and year, information on the ethnicity of civilian victims is nevertheless reported (as we code the targeting intention in a separate variable, see section 3.2). Whenever the ethnic

⁶Note that this approach differs from Fjelde and Hultman (2014), who rely on the GeoEPR-ETH (Wucherpfennig et al., 2011) and UCDP GED data (Sundberg, Lindgren and Padsokocimaite, 2010; Sundberg, 2013) to infer the character of one-sided violence at the level of grid cells.

⁷In other words, while geographic information alone is not sufficient to infer the ethnic identity of civilian victims, geographically concentrated targeting in ethnically homogeneous areas *and* in the context of an ongoing ethnic conflict with a history of ethnic targeting is.

⁸Justifications of the codings are available from the authors upon request.

identity of civilian victims is not reported or otherwise clearly identifiable (i.e., there is neither explicit nor implicit information), these deaths are not assigned to any ethnic group.

Sometimes, information on religion is available only, without the specific religious group being recorded as such in EPR. In these cases, we proceed as follows: If religiously defined victim groups can be linked to specific EPR groups, they are coded as ethnic victim groups, otherwise they are not. The first case is given, for example, if specific EPR groups practice the religion in question, while other EPR groups of the country do *not*. The second case applies, for example, if the religion of the victims is practiced by *all* EPR groups in the country. If ethnicity and religion strongly overlap in a country, but only information on religion is available, then we code these in the same way as ‘ethnic umbrella groups’ (i.e., assign the fatalities to all ethnic groups) unless a more precise coding based on ‘implicit information’ (such as geography and context) is warranted. If the relationship between ethnic and religious cleavages is unclear, we consult the EPR coding team, which has in-depth case knowledge on collective identity dimensions for each country.

3.2 Intention for Ethnic Targeting

As outlined above, we do *not* rely on a stated intent of perpetrators in order to record the ethnic identity of victims. In other words, it is not necessary that a given perpetrator explicitly states an intention to target members of a particular ethnic group in order for us to record the ethnic identity of civilian victims. Such an explicit statement is rarely given, even in cases of extreme levels of strategic ethnic targeting. Moreover, a motive to perpetrate violence along ethnic lines is sometimes not even necessary for violence to hit certain ethnic groups disproportionately. Artillery shelling, for example, may affect primarily the members of specific ethnic groups residing in targeted areas, with the profiling scheme of the perpetrator possibly being geography (or other criteria) instead of ethnicity. However, to the extent that information on the alleged intention of perpetrators is available, we include this information as well.

We assess the extent to which civilian targeting may have been based on intentional ethnic profiling by evaluating, first, the evidence that one-sided violence was based on selective, individualized targeting, collective targeting (i.e., profiling based on potential victims’ alleged membership in particular groups), or targeting that was completely arbitrary in the selection of victims

(Kalyvas, 2006; Steele, 2009; Wood, 2012; Gutiérrez Sanín and Wood, 2017). Targeting that is individualized and purely based on behavior does not meet the criteria of ethnic profiling. Campaigns of violence that are completely arbitrary in the selection of victims are not based on collective targeting either. Where collective targeting occurs, it can be based on identity-related criteria other than ethnicity. Moreover collective targeting is sometimes based on behavior *and* identity (Gutiérrez Sanín and Wood, 2017). For example, members of a given ethnic group may be collectively targeted, but only if they engage in some sort of behavior (e.g., protests), while members of other ethnic groups may engage in the same type of behavior without being victimized.

In order to assess the intention for ethnic targeting, we first of all evaluate for each perpetrator, ethnic victim group, and year whether the available evidence points to people being targeted at least partially based on their ethnic identity (i.e., their alleged ethnic affiliation), based on other collective identities (e.g., political or sexual orientations), or based on their behavior. Specifically, we evaluate our sources with regard to evidence of civilians being screened for ethnic ‘markers,’ of explicit or alleged strategies of armed groups (or armed group leaders) to target members of specific ethnic groups, of other indicators of ethnic targeting – or none or any combination of the above.

Clear statements by the perpetrators of ethnic profiling in relation to civilian targeting are rare, but clearly important to inform our coding decisions. In cases where such public announcements exist, we use them as a basis for coding ethnic targeting. More commonly, a perpetrator may deny systematic ethnic targeting, even where several other, independent sources (such as Amnesty International and Human Rights Watch) report evidence to support that ethnic targeting did occur. In these cases, the statements by the perpetrators denying systematic ethnic one-sided violence do not affect the coding decisions for the intention variable.

Based on this assessment, we code a variable (*int*) that indicates whether for a given perpetrator, year, and ethnic victim group,

- the targeting of the majority of killed civilians belonging to that ethnic group was purely behavioral, indiscriminate, or based on non-ethnic collective targeting (intention = 0; no ethnic targeting), or
- if there are indications of repeated ethnic screening in relation to civilian targeting and/or

a substantial proportion of civilians were killed in incidents with strategic ethnic targeting (intention = 1; some level of ethnic targeting). In order to code ethnic targeting (intention = 1), there should be indications that at least half (i.e., 50% or more) of the civilian victims of ethnic violence were killed (per perpetrator, ethnic group, and year) in events with ethnic screening, or events that were part of campaigns of strategic identity-based targeting.⁹ The level of confidence/uncertainty based on which the type of targeting can be coded is reflected in a separate variable (*intconf*).¹⁰

A list of hypothetical examples and how we categorize them to inform our coding decisions is provided below. Note that in these examples, ‘identity-based screening’ exclusively refers to screening based on ethnicity, while ‘collective targeting’ refers to violence based on profiling that focuses on non-ethnic collective identities. Note also that while the membership in a political organization certainly has a behavioral element, we consider targeting based on such membership as ‘collective targeting’ (not behavioral targeting), as there are many types of behavior that can be implied by the membership in a given organization. An example of collective targeting that could be identity-based or not is the victimization of members of ethno-nationalist parties. We adopt a conservative approach and do not code this as ethnic targeting unless there is direct or indirect evidence that members of similar organizations representing other ethnic groups are not being targeted; the same applies to any type of collective action (see below).

- Ex. Behavioral Criteria (No Ethnic Targeting; Intention = 0):
 - Members of a certain ethnic group are targeted only while engaging in a certain type of behavior such as participating in anti-government protests and demonstrations.
- Ex. Behavioral Criteria and Identity-based Screening (Ethnic Targeting; Intention = 1):
 - Members of a given ethnic group are targeted if engaging in a certain type of behavior while members of other ethnic groups engaging in the same type of behavior are not.
 - State forces attack villagers and claim to search for suspected protestors after protest marches demanding autonomy for ethnic group X. Sources indicate that during these attacks many members of ethnic group X were killed regardless of their protest activities, while members of other ethnic groups were spared.
- Ex. Behavioral and Collective Targeting (No Ethnic Targeting; Intention = 0):

⁹Strategic ethnic targeting implies ethnic targeting as an organizational policy (?).

¹⁰Case-specific comments on the type of targeting as well as the level and source of uncertainty are available from the authors upon request.

- State forces search villages and kill suspected protestors after protest marches that had no ethnic dimension in terms of who participated or the issue of contention. While it is unclear whether only protestors were being targeted, the targeting was reportedly based on political criteria (e.g., party membership) rather than ethnic ones.
- Ex. Strategy (Ethnic Targeting; Intention = 1):
 - Explicit statements by armed group leaders or government representatives making reference to plans or orders to target members of particular ethnic groups.
 - Credible (i.e., independent) sources such as AI or HRW indicate that a certain armed organization engages in systematic ethnic targeting.
- Ex. Identity-based Screening (Ethnic Targeting; Intention = 1):
 - Campaigns of violence against civilians during which some civilians are being killed and others spared depending on whether they are judged to belong to a given ethnic group.
- Ex. Identity-based Screening and Collective Targeting (Ethnic Targeting; Intention = 1):
 - Civilians are targeted based on their alleged membership in a given ethno-political organization, while members of similar organizations representing other ethnic groups are not.
 - Civilians are targeted based on their alleged membership in a given ethno-political organization, but also in other situations that indicate identity-based screening.
- Ex. Identity-based screening, Strategies, and Behavioral Criteria (Ethnic Targeting; Intention = 1):
 - Civilians are targeted in many individual episodes or events which fit into all of the above-mentioned categories; the majority of civilians are killed in a context that suggests identity-based screening.

Note that whenever enough information is available, we determine the type of targeting on the event level, i.e., for every single event, before aggregating to the perpetrator-victim group-year level. Only when the information is not precise enough or there are too many small-scale events to code each instance of one-sided killings separately do we rely on more general information (e.g., yearly reports documenting military campaigns against a particular ethnic group).

If the context strongly implies that ethnic targeting took place, but our sources are not specific enough to determine this with a very high level of confidence, or if explicit information indicates ethnic targeting, but this information cannot conclusively be linked to particular events of one-sided violence recorded in the UCDP dataset, this is reflected in the uncertainty variable (Intconf1, Intconf2 etc). Note, however, that if information on ethnic targeting can be conclusively linked to particular events in which together a majority of the civilian victims per perpetrator, victim group,

and year were subject to a particular type of targeting, and only the remaining events cannot be identified based on the information given, the uncertainty variable ('intconf') can still be coded with the highest certainty. In other words, the majority rule (at least 50% of the victims per perpetrator, victim group, and year) also applies to the coding of the uncertainty variable.¹¹

3.3 Data Sources

We rely on different sources to identify the ethnic identity of civilian victims by perpetrator and year, as well as targeting patterns. The qualitative case narratives provided in the UCDP conflict encyclopedia (<http://www.pcr.uu.se/research/ucdp/database>) are a good starting point, but are typically insufficient to disaggregate the victim groups by perpetrator and year. We rely in principle on the same source pool underlying the original UCDP one-sided violence codings, but use additional and/or alternative sources of information where this is helpful or necessary.¹² Our sources include daily reports from news agencies or newspapers and reports of human rights organizations or truth commissions, as well as secondary sources, such as qualitative case studies.¹³

Regardless of the type of information used, we try to corroborate our coding decisions based on several sources and assess the level of uncertainty of our codings based on whether different sources agree.

¹¹Whenever possible, we work with the actual UCDP GED data (i.e., event data) that underlie the UCDP actor-year data. However, where these data are not available (not all countries have been event coded at the time the dataset described in this codebook was created), or where there are too many individual events to code all of them individually, we work with more general information. We also rely on additional information where necessary (see next section). Regardless of the type of sources used, the same principle applies: If the type of targeting for more than 50% of civilian victims per victim group-perpetrator-year can be clearly identified, we code a high level of certainty. However, the highest certainty level can only be coded if the type of targeting is determined based on events that clearly underlie the UCDP actor-year fatality data (whether we have actual event data provided by UCDP or rely on other sources). In other words, while not all events have to be clearly identified in terms of the type of targeting and the overlap with UCDP in order to code the highest level of certainty, the events based on which coders determine the type of targeting (0/1) do. Otherwise a lower level of certainty is coded.

¹²The reliance on alternative or additional sources can be necessary for example because UCDP records many, at times thousands of individual events for certain actors, in which case we start with summary reports provided by other organizations (such as Human Rights Watch or Amnesty International). In other cases the UCDP sources are not informative when it comes to the ethnicity of civilian victims.

¹³Moreover, for some conflicts quantitative datasets are available that document the character of violence in a detailed way. Information on the geographic concentration of one-sided violence may also be helpful to infer the character of violence. However, as outlined above, while such geographic indicators can be a useful starting point, we refrain from relying exclusively on geographic information.

4 Variables

The first nine variables come directly from the UCDP One-Sided Violence Dataset (Eck and Hultman, 2007):

Actor ID (Integer)

Identifier of perpetrator of one-sided violence as indicated in the UCDP One-Sided Violence Dataset.

Actor Name (String)

Name of armed group perpetrating one-sided violence.

Year (Integer)

Calendar year to which the number of civilian victims per perpetrator refers.

Best Fatality Estimate: Number of Civilian Victims (Integer)

Number of civilian victims per perpetrator-year as indicated in the UCDP One-Sided Violence Dataset (best estimate).

High Fatality Estimate: Number of Civilian Victims (Integer)

Number of civilian victims per perpetrator-year as indicated in the UCDP One-Sided Violence Dataset (high estimate).

Is Government Actor (Integer)

This variable records if the actor is the government of a state and is coded as 1 for government actors and 0 for nongovernmental actors.

Location (String)

Name of the countries where one-sided violence took place in the perpetrator-year. Comma-separated if multiple.

GWNO Location (Integer)

The Gleditsch and Ward code for the countries where one-sided violence took place in the perpetrator-year (Gleditsch and Ward, 1999). Comma-separated if multiple.

Region (Integer)

The continents (regions) where one-sided violence took place:

1. Europe (GWNo: 200-399),
2. Middle East (GWNo: 630-650 and 652-699)
3. Asia (GWNo: 700-999)
4. Africa (GWNo: 400-626 and 651)
5. Americas (GWNo: 2-199).

The following are the new EOSV variables:

EthnGr1: Ethnic Group 1 (String)

Name of ethnic group exposed to one-sided violence perpetrated by a given armed group in the respective year; names as in EPR if available (Vogt et al., 2014). If an ethnic group is identified which does not fulfill the criteria for inclusion in EPR (e.g. because of size or lack of national relevance), a commonly used name of that group is used.

EthnGrID1: Ethnic Group 1 ID (Integer)

Identifier of ethnic group exposed to one-sided violence perpetrated by a given armed group in the respective year; identifiers as in EPR if available (Vogt et al., 2014). For ethnic groups that do not fulfill the criteria for inclusion in EPR (e.g. because of size or lack of national relevance), the identifier is coded as missing.

Typ1: Type (Integer)

Variable stating whether

1. the ethnic group assignment is based on explicit information,
2. the ethnic group assignment is based on implicit information,
3. the ethnic group assignment is based on explicit and implicit information,

Note:

- We assign ethnic group identifiers even if ethnicity is judged irrelevant for patterns of targeting.
- If neither explicit nor implicit information on the ethnic identity of civilian victims is available, these deaths are not assigned to any ethnic group.

Int1: Intention 1 (Integer)

Variable indicating whether one-sided violence (by this perpetrator and in this particular year) against members of ethnic group 1 was purely behavioral, indiscriminate, or based on non-ethnic collective targeting in the majority of killings (intention = 0; no ethnic targeting), or there are indications of ethnic screening or strategic identity-based targeting in the majority of killings (intention = 1; some ethnic targeting).

- 0: No ethnic targeting.
- 1: Some level of ethnic targeting.

Intconf1: Intention Confidence 1 (Integer)

Variable describing level of uncertainty when it comes to variable 'Int1.'

1. Low level of uncertainty; e.g., sources agree on intention.
2. Some level of uncertainty; e.g., sources agree on intention, but remain vague; most but not all sources agree on intention; most sources agree on intention, patterns of ethnic targeting cannot be conclusively linked to the events underlying the OSV data; ethnic targeting can be inferred from the context but there is minimal explicit information.
3. High level of uncertainty; e.g., sources disagree on intention; scant information on intention available.

Dats1: Data Sources 1 (String)

List of data sources used per perpetrator, ethnic victim group, and year, if other than UCDP original source.

Repeating all EOSV variables for each additional ethnic victim group: EthnGr2, EthnGr3, EthnGr4, EthnGr5, EthnGr6, EthnGr7.

References

- Balcells, Laia. 2010. "Rivalry and Revenge. Violence against Civilians in Conventional Civil Wars." *International Studies Quarterly* 54:2.
- Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. "Why do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(1):87–119.
- Eck, Kristine and Lisa Hultman. 2007. "Violence Against Civilians in War." *Journal of Peace Research* 44(2):233–246.
- Fjelde, Hanne and Lisa Hultman. 2014. "Weakening the Enemy: A Disaggregated Study of Violence against Civilians in Africa." *Journal of Conflict Resolution* 58(7):1230–1257.

- Fjelde, Hanne, Lisa Hultman, Livia Isabella Schubiger, Lars-Erik Cederman, Simon Hug and Margareta Sollenberg. 2019. "Introducing the ethnic one-sided violence dataset." *Conflict Management and Peace Science* XX(X):Forthcoming.
- Gleditsch, Kristian Skrede and Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* p. 393413.
- Gutiérrez Sanín, Francisco and Elisabeth Jean Wood. 2017. "What Should We Mean by 'Pattern of Political Violence'? Repertoire, Targeting, Frequency, and Technique." *Perspectives on Politics* 15(1):20–41.
- Kalyvas, Stathis N. 2006. *The Logic of Violence in Civil War*. Cambridge: Cambridge University Press.
- Kalyvas, Stathis N. and Matthew Kocher Kocher. 2007. "How 'Free' Is Free Riding in Civil Wars? Violence, Insurgency, and the Collective Action Problem." *World Politics* 59(2):177–216.
- Rüegger, Seraina. 2013. "Conflict Actors in Motion: Refugees, Rebels and Ethnic Conflict." PhD Thesis. ETH Zürich.
- Rüegger, Seraina and Heidrun Bohnet. 2014. "Quo Vadis? The Direction of Refugee Movements." ETH Zürich, Unpublished Manuscript.
- Steele, Abbey. 2009. "Seeking Safety: Avoiding Displacement and Choosing Destinations in Civil Wars." *Journal of Peace Research* 46:419–430.
- Sundberg, Ralph, Mathilda Lindgren and Ausra Pads kocimaite. 2010. "UCDP GED Codebook version 1.0-2011." Department of Peace and Conflict Research, Uppsala University.
- Sundberg, Ralph Melander, Erik. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50(4):523–532.
- Valentino, Benjamin A., Paul Huth and Dylan Balch-Lindsay. 2004. "Draining the Sea: Mass Killing and Guerrilla Warfare." *International Organization* 58(2):375–407.
- Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker and Luc Girardin. 2014. "The Family of Ethnic Power Relations Datasets: Integrating Data

on Ethnicity, Geography and Conflict.” Draft presented at the ENCoRe Conference in Uppsala, October 2014.

Weber, Max. 1996[1978]. The Origins of Ethnic Groups. In *Ethnicity.*, ed. John Hutchinson and Anthony D. Smith. Oxford, New York: Oxford University Press pp. 35–40.

Wood, Elisabeth Jean. 2012. Rape during War Is Not Inevitable: Variation in Wartime Sexual Violence. In *Understanding and Proving International Sex Crimes*, ed. Morten Bergsmo, Alf B. Skre and Elisabeth Jean Wood. Torkel Opsahl Academic EPublisher.

Wucherpfennig, Julian, Nils Benedikt Weidmann, Luc Girardin, Lars-Erik Cederman and Andreas Wimmer. 2011. “Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset.” *Conflict Management and Peace Science* 28(5):423–437.