Non-State Actor Mass Atrocity Dataset

v. 1
October 2017

Cyanne E. Loyle, PhD[1]
Indiana University and Peace Research Institute (Oslo)

Introduction

The purpose of this dataset is to assess the structural and strategic conditions under which non-state actors commit mass atrocities. In other words, it is designed to answer questions such as when do non-state actors choose to deliberately, and on large scale, target civilians?

Unit of Analysis

The unit of analysis for this dataset is the non-state actor group. The data relies on a definition of non-state actors which includes rebel groups and pro-government militias (PGMs). Data for these groups was taken from the Non-State Actor Dataset (Cunningham et al. 2013) and the Pro-Government Militia Database (Carey et al. 2013).

This definition (and subsequent coding) does not include armed criminal gangs, ethnic groups engaging in violence against other ethnic or identify based groups (i.e. groups not engaged in violence against the government or as an actor in a civil or interstate war), or terrorist organizations not engaged in a specific incompatibility against the government.[2]

Based on data collection and limitations, the inclusion criteria for the dataset is as follows:

1. *I include groups that have been formed between 1 January 1989 and 31 December 2007.* This coding decision was made because of existing data restrictions. The two datasets from which my actor list is derived are the NSA dataset and the PGMD v.1.1. The NSA dataset spans from 1945 to 2011 and the PGMD v.1.1 spans from 1987 to 2007. Further, I built on events data in the UCDP GED v.5 (Sundberg and Melander 2013) that covers the years 1989 through 2015. In order to utilize all three datasets, the overlap period (1989-2007) was selected.

2. *I analyze only those group that have been involved in an armed conflict.* For rebels this meant that they had a UCDP conflict ID (which they did by definition since the NSA dataset builds off of the UCDP data). For PGMs this meant that they were assigned a UCDP conflict ID in the SVAC 1.0 dataset.[3] This coding decision was made to ensure a comparable structural environment for all groups and to hone in on the contexts in which mass atrocities are more likely.

Sources

Data for this dataset was compiled from a variety of existing sources, primarily, the Non-State Actor Dataset (Cunningham et al. 2013), the Pro-Government Militia Database (Carey et al 2013), the NAGs dataset (San-Akca 2015), and the UCDP Georeferenced Events Dataset v5 (2016).

Variables not included in existing data sources were coded by graduate and undergraduate research assistants at Indiana University. To code this information, RAs consulted media reports, NGO reports,

---

[2] Of note, rebel groups and PGMs which commit acts of terrorist violence are included in this analysis.

[3] The SVAC dataset v.1.0 includes all active armed conflicts in the period 1989-2009, as defined by the UCDP/PRIO Armed Conflict database (Gleditsch et al. 2002) and the UCDP Dyadic Dataset (Harbom, Melander & Wallensteen 2009; Harbom & Wallensteen 2010). It includes conflicts that have either been active in one or more of the years 1989-2009 (the study period) or were active in one or more of the 5 years preceding the study period. An armed conflict is defined as: "a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths" (Gleditsch et al. 2002).

online databases, academic books and articles, and other online sources. Media reports were obtained through Google and Lexis Nexis searches. Examples include BBC, the Telegraph, the New York Times, the Economist, and the Guardian. Some local media sources were also included where applicable. NGO reports were obtained through Google searches. By far the most cited NGO was Human Rights Watch. Reports produced by UN agencies and tribunals, such as the UN Refugee Agency (UNHCR) and the International Criminal Tribunal for the Former Yugoslavia were also cited. All databases consulted are publically available online. The two most cited databases were the UCDP Encyclopedia and the Pro-Government Militia Database (PGMD V1.1), but other databases such as the Mapping Militant Organizations database, and data based on the Country Studies Series by the Federal Research Division of Congress were also included. Other online sources were consulted when information was difficult to find. These sources were obtained through Google searches.

Identification Variables

Actor ID (atrocity_actorlist_id)
A unique ID was created for each group in the dataset

UCDP Conflict ID (ucdpid)
UCDP ID for the conflict in which a group was involved.
- Rebel: Coded from "ucdpid" variable in Non-State Actor (NSA) dataset v. 3.3 (Cunningham, Gleditsch, and Salehyan 2012)
- PGM: Coded from "ucdpid" variable in Sexual Violence in Armed Conflict (SVAC) Dataset v.1.0-2013 (Cohen and Nordås, 2013)

Group Name (group_name)
Non-state actor group name
- Rebel: Coded from "side b" in Non-State Actor (NSA) dataset v. 3.3 (Cunningham, Gleditsch, and Salehyan 2012)
- PGM: Coded from "ActorName" in Sexual Violence in Armed Conflict (SVAC) Dataset v.1.0-2013 (Cohen and Nordås, 2013)

Country (country)
Country name in which the armed conflict is taking place

Source: Coded from "Location" variable in UCDP ACD (Gleditsch et al. 2002)

Regime Type (regime_type)
Regime type for the country in which the armed conflict is taking place, averaged over the groups lifespan

0= Democracy
1= Anocracy
2= Autocracy

Source: Polity IV (Marshall and Jaggers, 2003)

GDP (gdp)

Per capita GDP for the country in which the armed conflict is taking place, averaged over the groups lifespan

Source: Penn World Tables v.9 (Feenstra, Inklaar, and Timmer 2015)

Region (region)

Within which region does conflict occur?

       1 = Europe

       2 = Middle East

       3 = Asia

       4 = Africa

       5 = Americas

Source: Coded from "region" variable in UCDP ACD (Gleditsch et al. 2002)

Atrocity measure

In coding mass atrocity for this project, I relied on a measure of civilian deaths from the UCDP Georeferenced Events Dataset v5 (2016) and new atrocity data collection. This measure captures the intentional and deliberate targeting of civilians by rebel group and other non-state actors. Information is recorded in event form, including the number of people killed in an individual event.

For rebel groups, all atrocity variables are based on the UCDP GED v.5 (Sundberg and Melander 2013). If a rebel group does not appear in the UCDP GED, external sources were consulted to determine whether the group had perpetrated violence against civilians. If those sources consulted indicated that the group did not target civilians, all atrocity variables were coded as 0. Only one rebel group, TNSM in Pakistan, attacked civilians, but does not appear in the UCDP GED v.5. Sources indicated that the group primarily targeted security forces and carried out minimal violence against civilians. Therefore, TNSM is coded as 0 for all atrocity variables except for *deaths_civilians* and *event.* The latter two variables are coded as missing.

For PGMs, original event data was collected. First, a list was made of all PGMs for which *target* in the PGMD v.1.1 includes "civilians" or "unarmed political opposition." Second, a Lexis Nexis search of major world publications was conducted for each of these groups for the years 1989 through 2007. Any description of an event in which the group killed at least one civilian was recorded. The resulting event data was used to construct all atrocity variables.

Total Civilian Deaths (deaths_civilians)

Estimated number of civilian deaths attributed to NSA group

- For Rebel Group: Derived from "deaths_civilians" variable in UCDP GED (Sundberg and Melander 2013). First, all events occurring outside of 1989 - 2007 time frame were dropped. Second, the collapse(sum) command in Stata was used to sum all civilian_deaths for each group.

- For Pro-Government Militia: Derived from "civ_deaths" variable in PGM_Events Dataset (self-collected, see above). First, events occurring over multiple years were dropped. Second, the collapse(sum) command in Stata was used to sum all civilian_deaths for each group. ***Note: for all PGMs that appear in UCDP GED (Sundberg and Melander 2013) (e.g. Janjaweed), GED data is used to construct atrocity variables. See above for information about the construction of atrocity measures for these groups.

Atrocity Event

Number of events in which NSA group killed a certain threshold of civilians

- For Rebel Groups: Derived from "deaths_civilians" variable in UCDP GED (Sundberg and Melander 2013). Measure is a count of the total number of events perpetrated by each group between 1989 and 2007 resulting in civilian casualties.

- For Pro-Government Militia: Derived from "civ_deaths" variable in PGMD-Events Dataset. Measure is a count of the total number of events perpetrated by each group between 1989 and 2007 that result in civilian casualties. Events coded as occurring over multiple years and those for which "civ_deaths" was coded as missing (i.e. no_number is coded as 1) were not included in the total.

These variables are constructed with different thresholds. There is a 1000+ civilian deaths (atrocity_event_1000) variable, a 500+ civilian deaths event (atrocity_event_500) variable, etc.

Atrocity Year

Number of years in which NSA group killed a certain threshold of civilians

- For Rebel Groups: Derived from "deaths_civilians" variable in UCDP GED (Sundberg and Melander 2013). Measure is a count of the total number of years between 1989 and 2007 in which the group killed civilians at a certain threshold. Measure created by using the collapse (sum) command in STATA to collapse "deaths_civilins" by group name ("side_a") and "year" variables. A count variable was then created which summed the number of years between 1989 and 2007 in which the group killed civilians at a certain threshold.

- For Pro-Government Militia: Derived from "civ_deaths" variable in PGMD-Events Dataset. Measure is a count of total number of events perpetrated by each group between 1989 and 2007 that result at a certain threshold of civilian casualties. Events which are coded as occurring over multiple years were dropped. The collapse (sum) command was used to collapse "civ_deaths" by group ("side_b") and "year." A variable that counted the number of years in which the group killed civilians at a certain threshold was created.

These variables are constructed with different thresholds. There is a 1000+ civilian deaths per year (atrocity_year_1000) variable, a 500+ civilian deaths per year (atrocity_year_500) variable, etc.

Group Characteristics

PGM (pgm)
Codes whether or not the group is a pro-government militia.
> 0= No
> 1= Yes

Source: Pro-Government Militias Database (PGMD) v. 1.1-2013 (Carey, Mitchell, and Lowe 2013)

Government Relationship (gov_rel)
For PGMs, codes the relationship of the PGM with the government
> 0= Informal
> 1= Semioffical

Source: Pro-Government Militias Database (PGMD) v. 1.1-2013 (Carey, Mitchell, and Lowe 2013)

Rebel (rebel)
Codes whether or not the group is a rebel group.
> 0= No
> 1= Yes

Source: Non-State Actor (NSA) dataset v. 3.3 (Cunningham, Gleditsch, and Salehyan 2012)

Group Duration (group_duration)
Records how many years between 1989 and 2007 a group is active

- For Rebel Group: Derived from "startdate" and "enddate" variables in NSA (Cunningham, Gleditsch, and Salehyan 2012). If "startdate "is before 1989, "startdate" is coded as 1989. If "enddate" is after 2007, "endddate" is coded as 2007. A variable is created which counts the number of years between "startdate" and "enddate".

- For Pro-Government Militia: Derived from "date_formed" and "date_dissolved" variables in PGMD (Carey, Mitchell, and Lowe 2013). If "date_formed" is before 1989,"date_formed" is coded as 1989. If "date_dissolved" is after 2007, "date_dissolved" is coded as 2007. A variable is created which counts number of years between "date_formed" and "date_dissolved."

Islamist (islamist)
Codes whether or not a group expresses an islamist ideology
> 0= No
> 1= Yes

- For Rebel Groups: The coding for this variable comes from Wood and Thomas (2017). Wood and Thomas (2017) define Islamist rebels as those "that seek to organize society and government in accordance with a strict (fundamentalist) interpretation of Islamic law" (35). Groups coded as "religious" in either the Terrorist Organization Profiles (TOPs) or National Consortium for the Study of Terrorism and Responses to Terrorism (START, 2015) databases are

examined more closely to determine whether existing studies or reports refer to the group as Islamic, the group publicly advocates the implementation of a system of government based on Sharia law, or the group advocates the establishment of a theocratic state based on Islamic principles among its primary political objectives.

- For pro-government militias: For PGMs, existing studies and reports were examined to determine whether the group was an Islamist group.[4] We followed the same criteria as Wood and Thomas (2017) in deciding whether a group was islamist. For example, a BBC report describes Turkish Hizbullah as "…an extreme fundamentalist Islamist terrorist organization [that] released a manifesto…[calling] for Kurds to be governed by Islamic law." The report continues, "Hizbullah is regarded as a violent extreme Islamic fundamentalist organization in Turkey seeking to install - through the use of force, if necessary - an Islamic regime based on the Islamic law in Muslim Turkey, a country defined as secular by its Constitution." According to the report the group publicly advocates the implementation of a system of government based on Sharia law and it is therefore coded as Islamic.

Note- Just because a group's membership is Muslim does not mean the group is Islamist. For example, discussing the Mohajir Quami Movement a report in The Guardian describes members as "Mohajirs, Urdu-speaking refugees" and goes on to explain that they"… entered Sind … when a young firebrand leader, Altaf Hussein, founded the Mohajir Quami Movement to demand official recognition of a fifth nationality in Pakistan." The report indicates that while the group's membership is Muslim, they do not seek the political outcomes associated with Islamism. Instead, the group is demanding that Urdu-speaking refugees be recognized as a distinct nationality in Pakistan. The group is therefore coded as being ethno-nationalist, but not as being Islamist.

Leftist (leftist)
Codes whether or not a group expresses a leftist ideology
> 0= No
> 1= Yes

- For Rebel Group: For most rebel groups, the coding for this variable comes from Wood and Thomas (2017). Wood and Thomas (2017) define leftist rebels as "…typically draw[ing] on Marxist philosophies of class struggle and promote revolution as a strategy through which to liberate peasants, workers, and other oppressed class groups." They continue, "…rebellions adopting these ideologies typically propose a fundamental reshaping of existing social hierarchies in order to 'liberate' the population" (34). Groups coded as "leftist" in either the Terrorist Organization Profiles (TOPs) or National Consortium for the Study of Terrorism and Responses to Terrorism (START, 2015) databases are coded as being *leftist*. Those groups that do not appear in these datasets are examined more closely to determine whether existing studies or reports indicate that "groups adopt a Marxist-inspired ideology (e.g. socialist, communist, Maoist, or Marxist-Leninist)…" (39).

- For Pro-Government Militias: For PGMs, that do not appear in the Wood and Thomas (2017) data, existing studies and reports were examined to determine whether the group was a leftist group. We followed the same criteria as Wood and Thomas (2017) in deciding whether a group

---

[4] Note that data was also collected for one rebel group, Movement for Peace in Tajikistan, which did not appear in the Wood and Thomas (2017) data.

was leftist. For example, a report by the Mapping Militants project at Stanford University indicates that "Of the five major Tamil militant groups in the 1980s, including the EPRLF, TELO, EROS, PLOTE, and LTTE, PLOTE was the group that adhered most to Marxist-Leninist ideology." Because the report indicates that the People's Liberation Organization of Tamil Eelam (PLOTE) adheres to a Marxist-Leninist ideology, PLOTE is coded as leftist.

Ethnic Membership (ethnic_membership)
Records whether or not the group recruits ethnically
> 0= No
> 1= Yes

- For Rebel Groups: Derived from "ethnic recruitment" variable in ACD2EPR data.

- For Pro-Government Militias: Derived from "membership" variable in PGMD v1.1 (Carey, Mitchell, and Lowe 2013). Coded as 1 for groups coded as "ethnic". All other PGMs are coded as 0.

Separatist (separatist)
Does the rebel group have separatist aims or is the PGM fighting in a conflict with a rebel group with separatist aims
> 0= No
> 1= Yes

Source: Coded from "Incomp" variable in UCDP ACD (Gleditsch et al. 2002). Coded as 1 if UCDP recorded "Incomp" as an "incompatibility concerning territory".

Central Command (central_command)
Codes whether or not the group has a central command structure
> 0= No
> 1= Yes

- For Rebel Groups: Coded from binary "centcontrol" variable in Non-State Actor (NSA) dataset v. 3.3 (Cunningham, Gleditsch, and Salehyan 2012).

- For Pro-Government Militias: Information for this variable was self-coded from news articles, human rights reports, etc. (see source information above). PGM command structure was self-coded. Groups were coded as having a central command when they had an identifiable leader, received structured military training, and/or when it was clear they were receiving and following orders from an established governing party (such as the military, a political party, or a country's executive). For example, the CDR in Sierra Leone was described as "an organized armed faction" and "formed by the government". The Kalagala Action Plan in Uganda was described as "well organized", "government trained" and "wearing green uniforms".

Strength of Central Command (strengthcent)
Ordinal measure for the strength of the central command
> 1= NA (no central command)

2= Low
3= Moderate
4= High

- For Rebel Group: Coded from ordinal "strengthcent" variable in Non-State Actor (NSA) dataset v. 3.3 (Cunningham, Gleditsch, and Salehyan 2012).

- For Pro-Government Militias: Information for this variable was self-coded from news articles, human rights reports, etc. (see source information above). Material was read to establish the strength of the command structure of the PGM. Command strength is coded as "high" when there was evidence that the group was well organized, disciplined, and cohesive, such as the example of the Kalagala Action Plan wearing uniforms, as mentioned above. "High" command structure groups also demonstrate evidence of being organized into units with sub-commanders or levels of command. Command strength is coded as "moderate" when there was evidence of a group leader or organization structure, but where there was no evidence that that the group was disciplined or highly organized. The default coding for this variable was "moderate" central command. "Low" command structure groups demonstrated little cohesion.

Political Party Link (pol_party_link)
Codes whether or not the group has a link to an existing political party
　　0= Group not linked to a political party
　　1= Group is linked to a political party

- For Rebel Group: Derived from "rebpolwing" variable in Non-State Actor (NSA) dataset v. 3.3 (Cunningham, Gleditsch, and Salehyan 2012). Groups coded as having an "explicit," "acknowledged," or "alleged" link are coded as 1. Those coded as having "no" link are coded 0.

- For Pro-Government Militia: Derived from "government-link" variable in Pro-Government Militias Database (PGMD) v. 1.1-2013 (Carey, Mitchell, and Lowe 2013). Group coded as 1 if PGMD coded as linked to a "political party" and 0 otherwise.

Conflict Characteristics

Conflict Duration (conflict_duration)
How many years, within the 1989-2007 time frame, is the conflict active

Source: Derived from "year" variable in UCDP ACD (Gleditsch et al. 2002). Variable is a count of the number of years between 1989 and 2007 in which each conflict was active.

Incompatibility (incompatibility)
What is the primary incompatibility of the conflict
　　1 = Incompatibility concerning territory
　　2 = Incompatibility concerning government

Source: Coded from "Incomp" variable in the UCDP ACD (Gleditsch et al. 2002)

Battle Deaths (battle_deaths_total)
How many total battle deaths occurred in the conflict between 1989 and 2007 (continuous)

Source: Derived from "bdhigh" variable in UCDP Battle deaths dataset v.5. Years before 1989 and after 2007 were dropped. The collapse (sum) command in STATA was used to collapse "bdhigh" by "conflictid" to create a measure of the total number of battle deaths for each conflict.

State Civilian Deaths (state_civ_deaths)
Number of civilians killed by the government

Source: UCDP GED v. 5 (Sundberg and Melander 2013)

State Atrocity Binary (state_atrocity_binary)
Has the government ever carried out a mass atrocity?
>   0= No
>   1= Yes

Source: Early Warning Project's Episodes of State Mass Killing Dataset

State Atrocity during Group's life Binary (state_atrocity_grouplife)
Has the government ever carried out a mass atrocity while the group was in existence
>   0= No
>   1= Yes

Source: Early Warning Project's Episodes of State Mass Killing Dataset

Rebels in Conflict (rebels_in_conflict)
Total number of rebel groups in the armed conflict in which a NSA group is active

PGMs in Conflict (pgm_in_conflict)
Total number of PGMs in the armed conflict in which a NSA group is active

NSAs in Conflict (nsa_in_conflict)
Total number of rebel groups and PGMs in the armed conflict in which a NSA group is active

Works Cited

Carey, Sabine C., Neil J. Mitchell and Will Lowe. 2013. "States, the Security Sector, and the Monopoly of Violence: A New Database on ProGovernment Militias". Journal of Peace Research 50(2): 249-258.

Cohen, Dara Kay, and Ragnhild Nordås. "Sexual violence in armed conflict: Introducing the SVAC dataset, 1989–2009." Journal of Peace Research 51.3 (2014): 418-428.

Croicu, Mihai and Ralph Sundberg, 2016, "UCDP GED Codebook version 5.0", Department of Peace and Conflict Research, Uppsala University

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.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand (2002) Armed Conflict 1946-2001: A New Dataset. Journal of Peace Research 39(5).

Marshall, Monty G. and Keith Jaggers. 2003. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2002.

Melander, Erik and Therése Pettersson & Lotta Themnér (2016) Organized violence, 1989-2015. Journal of Peace Research 53(5).

NPWJ. Conflict Mapping in Sierra Leone: Violations of International Humanitarian Law 1991 to 2002. 2004: 189. http://www.npwj.org/ICC/Conflict-Mapping-Sierra-Leone-Violations-International-Humanitarian-Law-1991-2002.html

OHCHR. Report of the Mapping Exercise documenting the most serious violations of human rights and international humanitarian law committed within the territory of the Democratic Republic of the Congo between March 1993 and June 2003. June 2010 p.82-83 http://www.ohchr.org/Documents/Countries/CD/DRC_MAPPING_REPORT_FINAL_EN.pdf

The Ottawa Citizen. 'The Butcher of Kosovo': The name Arkan strikes fear in his enemies. He takes credit for the death of thousands of civilians during the war in Croatia and Bosnia. He's back in business in Kosovo, Bruce Culp reports. March 29, 1999. (accessed in Lexis Nexis)

San-Akca, Belgin. "Dangerous Companions: Cooperation between States and Nonstate Armed Groups (NAGs)" v.04/2015. nonstatearmedgroups.ku.edu.tr.

Sundberg, Ralph, and Erik Melander, 2013, "Introducing the UCDP Georeferenced Event Dataset", Journal of Peace Research, vol.50, no.4, 523-532

Wood, Reed M., and Jakana L. Thomas. "Women on the frontline: Rebel group ideology and women's participation in violent rebellion." *Journal of Peace Research* 54.1 (2017): 31-46.