UCDP Non-state Conflict Codebook
Version 25.1
Uppsala Conflict Data Program

Department of Peace and Conflict Research
Uppsala University

This version compiled and updated by Therese Pettersson (2025)
Replacing the earlier versions by Marie Allansson & Mihai Croicu (2017), Therése Pettersson (2014), Marcus Nilsson & Therése Pettersson (2013), Therése Pettersson (2012), Ralph Sundberg (2010), and Joakim Kreutz & Kristine Eck (2005))

Contents

1. Introduction

This document describes the Non-State Conflict Dataset, a project within the Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University. The UCDP Non-State conflict project has been developed with support from the Human Security Report Project, Simon Fraser University, in Vancouver, Canada.

In the development of the definition of non-state conflict, the input from Kristine Eck, Peter Wallensteen, Margareta Sollenberg, Lotta Themnér, Ralph Sundberg, Stina Högbladh, Therése Pettersson and Johan Brosché, have been instrumental. The UCDP non-state conflict project is also grateful for additional advice and feedback from Andrew Mack, Zoe Nielsen, Ole Magnus Thiesen, and others.

Case-specific information about the cases of non-state conflict is available at www.ucdp.uu.se. Questions regarding the definitions and the content of the dataset can be directed to ucdp@pcr.uu.se

2. Definition of Non-State conflict

2.1   Non-state conflict

A non-state conflict is defined by the Uppsala Conflict Data Program (UCDP) as *"the use of armed force between two organized armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year."*

The separate elements of the definition are operationalized as follows:

(1) *Use of armed force*: the use of arms, resulting in deaths.

    (1.1) Arms: any material means, e.g. manufactured weapons but also sticks, stones, fire, water, etc.

(2) *25 deaths*: a minimum of 25 battle-related deaths per year

    (2.1) *battle-related deaths:* deaths directly related to the use of armed force between the warring groups.

(3) *Organized groups*: consists of either

    (3.1) formally organized groups: any non-governmental group of people having announced a name for their group and using armed force against another similarly formally organized group, *or*

    (3.2) informally organized groups: any group without an announced name, but who uses armed force against another similarly organized group, where the violent activity meets the following requirement:

        (3.2.a.) there is a clear pattern of violent incidents that are connected and in which both groups use armed force against the other.

(4) *State*: a state is

    (4.1) an internationally recognized sovereign government controlling a specified territory, *or*

    (4.2) an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.

(5) *Government*: the party controlling the capital of the state.

2.2   Fatality estimates

UCDP codes three different fatality estimates – low, best and high – based on the reliability of reports and the conflicting number of deaths that can be reported for any violent event.

- *Low estimate:* The UCDP Low estimate consists of the aggregated low estimates for all battle-related incidents during a year. If different reports

provide different estimates and a higher estimate is considered more reliable, the low estimate is also reported if deemed reasonable.

- *Best estimate*: The UCDP Best estimate consist of the aggregated most reliable numbers for all battle-related incidents during a year. If different reports provide different estimates, an examination is made as to what source is most reliable. If no such distinction can be made, UCDP as a rule include the lower figure given.
- *High estimate*: The UCDP High estimate consists of the aggregated high estimates for all battle-related incidents during a year. If different reports provide different estimates and a lower estimate is considered more or equally reliable, the high estimate is also reported if deemed reasonable. If there are incidents when there is some uncertainty about what parties have been involved, these are also included in the high estimate.

It is the best estimate of battle-related deaths that determines whether a dyad will be included in the UCDP Non-state Conflict data (i.e. the best estimate needs to be 25 or higher).

2.3 External state support in non-state conflict

In some cases, external states can be involved as secondary warring parties in non-state conflicts. Examples include the US troop support to SDF in its fight against IS in Syria, and the UAE supporting the Forces of Hadi against the Yemeni al-Qaida affiliate AQAP. The 19.1 version of the dataset was the first to code troop support from an external state in non-state conflicts. Troop support to a non-state conflict can never be coded if the country in question has its own incompatibility with the group, or if the country is already a secondary warring party to a state-based conflict against the same group in the same country. Hence, state-based armed conflict trumps non-state conflict. The flowchart below elaborates on under which circumstances an external state can be included as a secondary warring party (SWP) in non-state conflict.
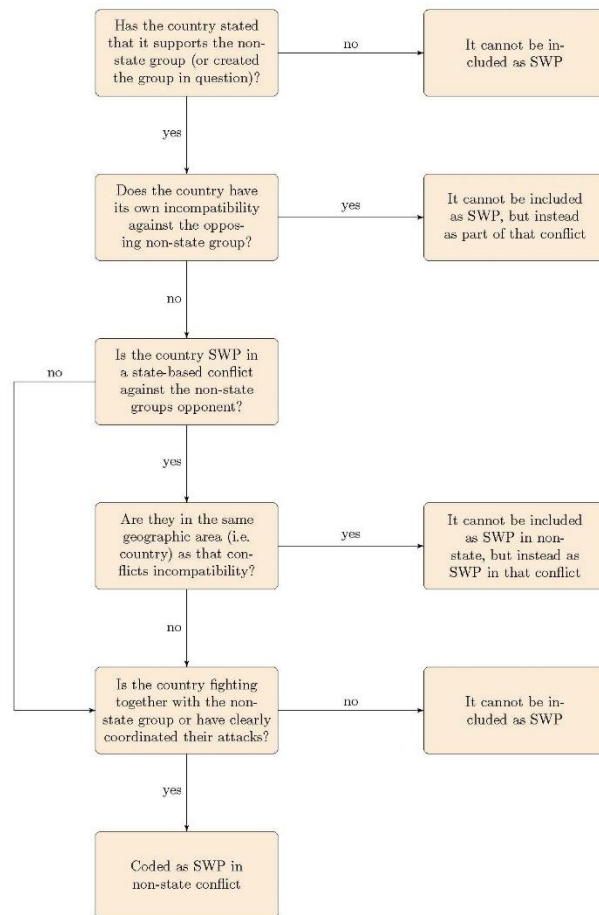
Figure 1: Secondary warring parties in non–state conflicts.

3. Variables in the Non–State conflict Dataset

Variable name	Content	Type
conflict_id	The unique identifier of the non–state conflict.	Integer
dyad_id	The unique identifier of the non–state dyad (a pair of two opposing actors). Note that one non–state conflict has, per definition, one and only one non–state dyad. The inclusion of both dyad IDs and conflict IDs in the dataset is meant to allow easier integration of this dataset with other UCDP products such as the UCDP/PRIO Armed Conflict Dataset, the UCDP Dyadic Dataset or the UCDP GED.	Integer
org	This variable indicates the organizational level of the warring sides. The level of organization is determined according to the following categories: **Organizational level 1** (formally organized groups): Rebel groups and other organized groups that have a high enough level of organization so as to be possible to include in the state–based armed conflict category. These include rebel groups with an announced name, as well as military factions (Forces of…). This level of organization captures fighting between highly organized rebel groups and fatalities are recorded according to the criteria set for battle–related deaths in the state–based conflict category. **Organizational level 2** (informally organized groups): Groups composed of supporters and affiliates to political parties and candidates. These are commonly not groups that are permanently organized for combat, but who at times use their organizational structures for such purposes. In addition to supporters of political parties and candidates, included in this category is also fighting between groups composed of supporters of other organizations such as the supporters of al–Ahly football team fighting against the supporters of al–Masry football team in Egypt 2012. Battle–related deaths are recorded according to section 3.2.a of the definition of non–state conflict. **Organizational level 3** (informally organized groups): Groups that share a common identification along ethnic, clan, religious, national or tribal lines. These are not groups that are permanently organized for combat, but who at	Integer

	times organize themselves along said lines to engage in fighting. This level of organization captures aspects of what is commonly referred to as 'communal conflicts', in that conflict stands along lines of communal identity. Battle-related deaths are recorded according to section 3.2.a of the definition of non-state conflict.	
side_a_name	The party that constitute Side A in the conflict. For each conflict the parties are listed in alphabetical order, using the latest known names of the parties involved. Comma separated if multiple.	String
side_a_name_fulltext	The full original name of the actor, in English.	String
side_a_name_mothertongue	The full original name of the actor, in mother tongue.	String
side_a_id	The unique identifier of the groups that make up Side A. For conflicts with multiple actors fighting together a temporary coalition ID has been assigned. From version 17.1 of the dataset and onwards, the ID system for conflicts, actors and dyads changed in order to make it unique across all UCDP core datasets and all UCDP types of violence. To download a conversion table containing new and old IDs, visit http://ucdp.uu.se/downloads/.	Integer
side_a_components	For conflicts with multiple actors fighting together as a joint (temporary) coalition, the components of the coalition (in the form of a string of actor IDs) are listed here. Comma separated.	String
side_a_2nd	side_a_2nd lists all states that enter a non-state conflict with troops to actively support side A in the dyad. See section 2.2 for information on under which conditions this is applicable. This variable is not part of the API version of the dataset. Comma separated if multiple.	String
gwno_a_2nd	The Gleditsch and Ward country codes of side_a_2nd. This variable is not part of the API version of the dataset. Comma separated if multiple.	String
side_b_name	The party that constitute Side B in the conflict. For each	String

	conflict the parties are listed in alphabetical order, using the latest known names of the parties involved. Comma separated if multiple.	
side_b_name_fulltext	The full original name of the actor, in English.	String
side_b_name_mothertongue	The full original name of the actor, in mother tongue.	String
side_b_id	The unique identifier of the groups that make up Side B. For conflicts with multiple actors fighting together a temporary coalition ID has been assigned. From version 17.1 of the dataset and onwards, the ID system for conflicts, actors and dyads changed in order to make it unique across all UCDP core datasets and all UCDP types of violence. To download a conversion table containing new and old IDs, visit http://ucdp.uu.se/downloads/.	Integer
side_b_components	For conflicts with multiple actors fighting together as a joint (temporary) coalition, the components of the coalition (in the form of a string of actor IDs) are listed here. Comma separated.	String
side_b_2nd	side_b_2nd lists all states that enter a non-state conflict with troops to actively support side B in the dyad. See section 2.3 for information on under which conditions this is applicable. This variable is not part of the API version of the dataset. Comma separated if multiple.	String
gwno_b_2nd	The Gleditsch and Ward country codes of side_b_2nd. This variable is not part of the API version of the dataset. Comma separated if multiple.	String
ext_troop_supp	type is a binary variable that codes whether the non-state conflict received external state support in the form of troops. If there is at least one country listed in the side_a_2nd OR side_b_2nd, this variable is coded as 1. If not, a 0 is coded. See section 2.3 for information on under which conditions this is applicable. This variable is not part of the API version of the dataset.	Integer
start_date	The first time there is a recorded event in a given dyad that results in at least one fatality. This date	Date (YYYY-MM-DD)

	is the same for all years in which the conflict has been active, regardless of whether the conflict has been active in several episodes or not. The start_date is coded as precisely as possible. For certain conflicts we can pinpoint the start of the conflict down to a single event, taking place on a specific day. For other conflicts, this is not possible, due to lack of precise information.	
start_prec	The start_prec (start precision) is coded to highlight the level of certainty for the date set in the start_date variable. 1. Day, month and year are precisely coded; there is good information on the event. 2. Day is assigned; month and year are precisely coded. This precision score is assigned if the first event which causes at least one fatality takes place within a period of 2–6 days. 3. Day is unknown; month (or a period of 30 days, not necessarily a calendar month) and year are precisely coded. The day is known to be in a given period of 30 days. The date is set to the last date of the period. 4. Month is assigned, year is precisely coded. The date is set to the last day of the assigned month. 5. Day and month are assigned, year is precisely coded. Day and month are set as precisely as possible. For example, if an event is known to have taken place between March and July, the date is set to 31 July with precision score 5. This is an automatic aggregation from the UCDP Georeferenced Event Dataset. For more complex inquiries in the temporal dimension of the conflict, you are advised to use the UCDP GED.	Integer
start_date2	start_date2 gives the date, as precise as possible, when a given episode of conflict activity reached 25 battle-related deaths.	Date (YYYY-MM-DD)
start_prec2	Precision scores calculated as per start_prec above This is an automatic aggregation from the UCDP Georeferenced Event Dataset. For more complex inquiries in the temporal dimension of the conflict, you are advised to use the UCDP GED.	Integer

ep_end	ep_end is a binary variable that codes whether the conflict is inactive the following year and an episode of the conflict thus ends. If the conflict is inactive the following year(s), this variable is coded as 1. If not, a 0 is coded. For the latest year in the dataset, it is unknown whether the conflict will be recorded as active or inactive in the following year, and the variable is always given the code 0.	Integer
ep_end_date	This variable is only coded in years where ep_end has the value 1. If a conflict year is followed by at least one year of conflict inactivity, the ep_end_date variable lists, as precise as possible, the last date of recorded combat.	Date (YYYY-MM-DD)
ep_end_prec	Precision scores calculated as per start_prec above. This is an automatic aggregation from the UCDP Georeferenced Event Dataset. For more complex inquiries in the temporal dimension of the conflict, you are advised to use the UCDP GED.	Integer
year	The year of observation (1989-2024)	Integer
best_fatality_estimate	The best fatality estimate for the given conflict-year. This is an automatic aggregation (summing) of all the Best figures for all incidents reported for the given conflict-year in the UCDP Georeferenced Event Dataset.	Integer
low_fatality_estimate	The low fatality estimate for the given conflict-year. This is an automatic aggregation (summing) of all the Low figures for all incidents reported for the given conflict-year in the UCDP Georeferenced Event Dataset.	Integer
high_fatality_estimate	The high fatality estimate for the given conflict-year. This is an automatic aggregation (summing) of all the High figures for all incidents reported for the given conflict-year in the UCDP Georeferenced Event Dataset.	Integer
location	The countries where fighting took place in the conflict-year. Comma-separated if multiple.	String

This variable should never be used for any geographical or spatial analyses of conflict as the distribution of violence as well as the relative magnitude of violence by country is not captured. In effect, a country is listed here if even one dead in the given conflict has occurred in that country.

In fact, UCDP provides much better geographic coverage of conflict (including distribution of violence for each conflict and each country) in the UCDP Georeferenced Event Dataset (GED).

Like the UCDP Non-State Conflict Dataset, GED is global and covers the same period (1989–2024).

gwno_location	The Gleditsch and Ward code for the countries where fighting took place in the conflict-year. Comma-separated if multiple. This variable should never be used for any geographical or spatial analyses of conflict as the distribution of violence as well as the relative magnitude of violence by country is not captured. In effect, a country is listed here if even one dead in the given conflict has occurred in that country. In fact, UCDP provides much better geographic coverage of conflict (including distribution of violence for each conflict and each country) in the UCDP Georeferenced Event Dataset (GED). Like the UCDP Non-State Conflict Dataset, GED is global and covers the same period (1989–2024).	String
region	The continents (regions) where violence took place: 1 = Europe (GWNo: 200–399), 2 = Middle East (GWNo: 630–699) 3 = Asia (GWNo: 700–999) 4 = Africa (GWNo: 400–626) 5 = Americas (GWNo: 2–199).	String
version	The version of the dataset: 25.1	Float

4. Data Collection Methods

This dataset is the result of:

1.   an automatic filtering and aggregation of the UCDP Georeferenced
Event Dataset from incident/event level to the conflict/dyad-year level.

2.   information gathering and coding of a number of extra variables
at the aggregate conflict or actor level (such as organization type).

The original reporting underlying the dataset is collected from three sets of
sources:

1. global newswire reporting
2. global monitoring and translation of local news performed by the BBC
3. secondary sources such as local media, NGO and IGO reports, field reports,
   social media such as Twitter and Telegram, books etc.

The process is done in a "two-pass" system, first by consulting newswire
sources for the entire globe then by consulting local/specialized sources
based on information obtained from the first pass.

A detailed description of the process including a detailed description of
the passes and the search-strings employed is provided in the UCDP
GED Codebook version 25.1, Section 4.

5. Version name convention

In 2017, the version name convention was changed, giving all UCDP datasets the
same version number across the board. This was done so that users more
easily can see which UCDP dataset corresponds with which.

This codebook corresponds to Version 25.1 of the UCDP Non-state Conflict
Dataset. For every new release, substantial changes will be documented in a
separate document. This should be helpful to researchers trying to replicate a
particular study. We recommend that whenever this dataset is used, the version
number should be cited.

The version number is a combination of a year and a number. The year refers to
when the dataset is updated with new observations. If there are changes in the
data between yearly updates, or if there are substantial changes in the
structure of the dataset, the number behind the year is incremented.

6. Format availability

The data is available in CSV (respecting the RFC 4180 specification), Excel (XLSX), Rdata (3.x version) and STATA (2010 format).

The data is available for machine-to-machine interaction through a public API. Documentation for how to use the API is available at http://ucdp.uu.se/apidocs.

7. A note on UCDP ID changes

As of version 17.1 of all UCDP datasets, the ID system for conflicts, actors and dyads was changed in order to make them unique across all UCDP core datasets and all UCDP types of violence. This allows easier aggregation and disaggregation of data as well as simplify data management for users, especially when combining multiple UCDP products together.

For example, a non-state conflict cannot have the same ID as a (different) state-based conflict or as a case of one-sided violence (which was possible before).

Further, actor/side IDs representing a government were decoupled from their corresponding Gleditsch and Ward country codes (GWNo). Gleditsch and Ward country codes for state sides/actors are provided in separate columns from their sides ID (i.e. side_a_id is different from gwno_a but represents the same entity).

This means all IDs used in the dataset from version 17.1 and onwards are no longer compatible with those in older versions of UCDP products. Further, this means that external products relying on UCDP IDs for data management tasks will have to be adapted to work with the new ID systems.

A conversion table between the new and old ID systems is available here: http://ucdp.uu.se/downloads/

8. References

Gleditsch, Kristian S. & Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." International Interactions 25: 393-413

Davies, S., Pettersson, T., Sollenberg, M., & Öberg, M. (2025). Organized violence 1989–2024, and the challenges of identifying civilian victims. Journal of Peace Research, 62(4).