
UCDP One-sided Violence Codebook

Version 23.1

Uppsala Conflict Data Program

Department of Peace and Conflict Research
Uppsala University

This version compiled and updated by Therese Pettersson (2023)

Replacing the earlier versions by Marie Allansson and Mihai Croicu (2017), Therése Pettersson (2012), Joakim Kreutz (2008) and Joakim Kreutz & Kristine Eck (2005)

When using this data, please always cite:

Eck, Kristine & Lisa Hultman (2007). One-sided violence against civilians in war: insights from new fatality data. *Journal of Peace Research* 44(2): 233-246.

Davies, Shawn, Therese Pettersson & Magnus Öberg (2023). Organized violence 1989-2022, and the return of conflict between states. *Journal of Peace Research* 60(4).

When appropriate, also cite this codebook: Pettersson, Therese (2023) UCDP One-sided Violence Codebook v 23.1 (<https://ucdp.uu.se/downloads/>).

Always include the Version number in analyses using the dataset.

Contents

1. Introduction.....	3
2. Definition of One-sided violence.....	3
2.1. One-sided violence.....	3
2.2. Fatality estimates.....	4
3. Variables in the One-sided violence dataset.....	5
4. Data Collection Methods.....	8
5. Version name convention.....	8
6. Format availability.....	9
7. A note on UCDP ID changes.....	9

1. Introduction

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

Case-specific information about the cases of one-sided violence 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 One-sided violence

2.1. One-sided violence

One-sided violence is the use of armed force by the government of a state or by a formally organized group against civilians which results in at least 25 deaths. Extrajudicial killings in custody are excluded.

The separate elements of the definition are operationalized as follows:

- (1) *Use of armed force*: use of arms in order to exert violent force, resulting in death
 - (1.1) Arms: any material means, e.g. manufactured weapons but also sticks, stones, fire, water, etc.
- (2) *25 deaths*: a minimum of 25 civilian deaths per year and per actor.
- (3) *Government*: the party controlling the capital of the state.
- (4) *Formally organized group*: any non-governmental group of people having announced a name for their group and using armed force.
- (5) *State*: a state is
 - (5.1) an internationally recognized sovereign government controlling a specified territory,
 - or
 - (5.2) an internationally unrecognized government controlling a specified territory whose sovereignty is not disputed by another internationally recognized sovereign government previously controlling the same territory.
- (6) *Civilians*: unarmed people who are not active members of the security forces of the state, or members of an organized armed militia or opposition group. Government officials, such as members of parliament, governors, and councilors, are also excluded and are instead seen as representatives of the government of a state.

Note that the UCDP definition of civilian does not refer to the definition of civilian under international law.

- (7) *Extrajudicial killings in custody*: when the government of a state kills a person in its custody.
(7.1) *custody*: when the person is located in a prison or another type of governmental facility.

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 incidents of one-sided violence 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 incidents of one-sided violence 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 incidents of one-sided violence 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 one-sided violence that determines whether or not an actor will be included in the UCDP One-sided Violence data (i.e. the best estimate needs to be 25 or higher).

3. Variables in the One-sided violence dataset

Variable name	Content	Type
conflict_id	A unique conflict identification code for each individual observation of one-sided violence in the dataset. Used only for merging purposes across the UCDP datasets.	Integer
dyad_id	A unique dyad identification code for each individual observation of one-sided violence in the dataset. Used only for merging purposes across the UCDP datasets.	Integer
actor_id	<p>This is a numerical identifier that describes each individual actor. The dataset makes use of the general actor/side ID employed by UCDP.</p> <p>Warning: Unlike the previous versions of the dataset, this variable is NO LONGER the Gleditsch and Ward identifier (GWNo) for the state actors. Use gwnoa below instead!</p> <p>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/.</p>	Integer
coalition_components	If multiple actors are perpetrating violence together as a joint (temporary) coalition, the components of the coalition (in the form of a string of actor IDs) are listed here.	String
actor_name	<p>The government of a state or the name used by a formally organized group perpetrating the violence.</p> <p>Comma separated if multiple.</p>	String
actor_name_fulltext	The full original name of the actor, in English.	String
actor_name_mothertongue	The full original name of the actor, in mother tongue.	String
year	<p>The calendar year of observation. UCDP employs a threshold for inclusion of at least 25 deaths caused by one-sided violence for each actor in a calendar year.</p> <p>There are cases when actors are responsible for one-sided violence of a scale that does not meet the threshold.</p>	Integer

These observations are coded as absence of one-sided violence.

Episodes that do not exceed the threshold for actors that have exceeded the threshold at least once during the 1989-2022 period can be traced through the UCDP GED dataset.

best_fatality_estimate	<p>The best fatality estimate for the given episode.</p> <p>This is an automatic aggregation (summing) of all the Best figures for all incidents reported for the given dyad-year in the UCDP Georeferenced Event Dataset.</p>	Integer
low_fatality_estimate	<p>The low fatality estimate for the given episode.</p> <p>This is an automatic aggregation (summing) of all the Low figures for all incidents reported for the given dyad-year in the UCDP Georeferenced Event Dataset.</p>	Integer
high_fatality_estimate	<p>The high fatality estimate for the given episode.</p> <p>This is an automatic aggregation (summing) of all the High figures for all incidents reported for the given dyad-year in the UCDP Georeferenced Event Dataset.</p>	Integer
is_government_actor	<p>For some purposes, it may be necessary to disaggregate between government and non-governmental actors in the dataset. This variable records if the actor is the government of a state and is coded as 1 for government actors and 0 for non-governmental actors.</p>	Integer
location	<p>The countries where this episode of violence took place. Comma-separated if multiple.</p> <p>WARNING: 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.</p> <p>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).</p> <p>Like the UCDP One-Sided Violence Dataset, GED is global and covers the same period (1989-2022).</p>	String

gwno_location The Gleditsch and Ward code for the countries where this episode of one-sided violence took place. Comma-separated if multiple. String

WARNING: 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 one-sided violence (including distribution of violence for each dyad and each country) in the UCDP Georeferenced Event Dataset (GED).

Like the UCDP One-Sided Violence Dataset, GED is global and covers the same period (1989-2022).

gwnoa	The Gleditsch and Ward country identifier if the actor perpetrating one-sided violence is a state. Empty otherwise.	Integer
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: 23.1	Float

4. Data Collection Methods

This dataset is the result of an automatic filtering and aggregation of the UCDP Georeferenced Event Dataset from incident/event level to the actor-year level.

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 23.1, 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 23.1 of the UCDP One-sided Violence 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/>.