

# ucdplink: Data on battle and war-related violence

## Codebook

Corinne Bara\*

Maurice P. Schumann†

---

\*ETH Zürich. Email: [corinne.bara@sipo.gess.ethz.ch](mailto:corinne.bara@sipo.gess.ethz.ch)

†Hertie School Berlin. Email: [m.schumann@phd.hertie-school.org](mailto:m.schumann@phd.hertie-school.org)

# Background

This codebook describes the variables in the conflict-month dataset on battle and war-related violence. This dataset can be created by using the function **ucdplink** of the R package with the same name, and is described in detail in the following article (please cite when using this data):

Bara, Corinne and Maurice P. Schumann (2025). Who, what, and where? Linking violence to civil wars. *Research & Politics*.

## INSTALLATION

The R package can be installed from github: ucdplink R package.

## CITATION

When using the data, please always cite the source datasets in addition to the article by Bara & Schumann:

### **UCDP/PRIO Armed Conflict Dataset v.24.1.:**

Davies, Shawn, Garoun Engström, Therese Pettersson & Magnus Öberg (2024). Organized violence 1989-2023, and the prevalence of organized crime groups. *Journal of Peace Research* 61(4).

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).

### **UCDP Georeferenced Event Dataset (GED) Global version 24.1.:**

Sundberg, Ralph and Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4).

### **WZONE: Zones of Armed Conflicts:**

Kikuta, Kyosuke. 2022. “A New Geography of Civil War: A Machine Learning Approach to Measuring the Zones of Armed Conflicts.” *Political Science Research and Methods* 10(1): 97–115.

## Special Decisions

This section briefly summarises some decisions that were taken in the application of our linking procedure to UCDP data.

**Constant Side B:** In the original UCDP conflict data (UCDP ACD), which is annual data, the Side B can change from year to year. To give an example: In the Colombian conflict, the FARC and ELN may be both active in one year, while in the next year only the FARC is active, and in the year thereafter only the ELN, etc. In making our actor links, we have turned this information into constant Side B information. This means that every rebel group that ever was active in a conflict episode is a possible side B to which we can link violent events from the UCDP GED. This is important because otherwise we would for instance consider an attack on civilians by ELN rebels in a year in which they did not engage in armed clashes against the government as NOT war-related.

**Coalitions:** In events of non-state violence, the actor on either side can be a coalition of non-state groups. The same applies to one-sided violence, in which the perpetrator can be a coalition (mostly of rebels). Coalitions have different Actor IDs than the individual groups making up a coalition. Importantly, coalition actors are never a Side A or Side B in a UCDP ACD conflict, hence coalition violence is hard to link to conflicts. In violent GED events with coalition actors we thus split the fatalities from the event evenly between the coalition members, and then linked the event to a conflict via the individual members' actor IDs.

**Al-Qaida:** As a default and in line with common practice in conflict research, we drop the conflict with the ID 418 (United States against al-Qaida) from the dataset (though users can choose to include it in the R package). With the exception of the 9/11 attacks on US territory, most of the fighting happened in other countries (al-Qaida's transnational terrorism and the war on terror by the US and its allies). The conflict zone is accordingly large and

spans a big part of the globe, making spatial attribution unsuitable in this case.<sup>1</sup>

**Postwar violence:** Because the user can choose to include conflicts not only while they are active, but also after they have (even temporarily) ended, this data can be used to study postwar violence. However, we suggest the user do not extend the postwar period for too long, i.e., not study violence ten years after the war. The reason has to do with overlapping conflict zones and is best illustrated with a fictitious example: Rebel group A fights the government between 1991 and 1995 in conflict zone A. Ten years later, in 2005, rebel group B fights the same government over a completely different thing in conflict zone B. Conflict zones A and B partially overlap, i.e., the war in the 1990s and ten years later takes place in a similar area. If the government now kills civilians in a massacre in the overlapping part of the conflict zone in 2005, this violence would be attributed as war-related violence to both conflicts A and B. However, assuming this is postwar violence in the context of conflict A ten years later is probably unreasonable. We think conflict zone overlap is not otherwise a big problem. If different conflicts are fought around the same time in a similar area (such as in India’s Northeast), they are often part of conflict complexes where we could not - also not with case information - be sure to which conflict a violent event really “belongs”, and attributing it to multiple conflicts as war-related is therefore a sound strategy.

---

1. See also conflict description by the UCDP: <https://ucdp.uu.se/statebased/878>.

# Variables

Variable name	Description
conflict_id	The unique identifier of the conflict, taken from the UCDP ACD.
epi_id	The unique identifier of a conflict episode. New episodes start after one or more calendar years of inactivity. Consists of the conflict_id and a count of episodes since 1946.
mdate	Month of observation (written as daily date with the first of the month as standard).
year	Year of observation.
active	Indicator of a conflict's activity status.  0. Conflicts that are active (as of UCDP definition of activity)  1. Conflicts that were once active but are not currently (post-war/interwar)
postwarmonth	Counts the months since a conflict was last active. Takes on a value of 0 for currently active wars.
location	The country in which the civil war takes place, i.e., whose government is being challenged by a non-state armed group. Taken from the UCDP ACD.
type_of_conflict	Whether conflict is pure intrastate (3) or internationalized intrastate (4), meaning there is foreign involvement with troops on side A and/or side B. Taken from the UCDP ACD.

Variable name	Description
incompatibility	The main conflict issue. Taken from the UCDP ACD. <ol style="list-style-type: none"> <li>1. Incompatibility about territory</li> <li>2. Incompatibility about government</li> <li>3. Incompatibility about government AND territory</li> </ol>
territory_name	The name of the territory over which the conflict is fought. Taken from the UCDP ACD. Missing if conflict is not over territory.
side_a	The name of Side A in the conflict (the country/government). Taken from the UCDP ACD.
side_a_id	The unique identifier of the actor on side A (always the government). Taken from the UCDP ACD.
side_a_2nd	List of states that enter a conflict with troops to support side A. Taken from the UCDP ACD. Comma separated if multiple. Missing if no supporters in that year.
side_b	Name of the non-state armed group that challenges the government. Can vary annually. Comma separated if multiple. Taken from the UCDP ACD.
side_b_id	The identifier of the non-state armed group that challenges the government. Can vary annually. Comma separated if multiple. Taken from the UCDP ACD.
sideb_constant	Side B IDs of all non-state armed groups that ever challenged the government in this conflict episode (epi.id). Does not change over time within one episode. Comma separated if multiple.

Variable name	Description
side_b_2nd	List of states that enter a conflict with troops to support side B. Taken from the UCDP ACD. Comma separated if multiple. Missing if no supporters in that year.
start_date	The date (daily) of the first battle-related death in the conflict. This is when the conflict, not the current conflict episode, first started. Taken from the UCDP ACD.
start_date2	The start date (daily) of an episode, i.e., when it became active as of UCDP definition. Taken from the UCDP ACD.
gwno_a	The Gleditsch and Ward country codes of the country in which the civil war takes place, i.e., whose government is being challenged by a non-state armed group. Taken from the UCDP ACD.
gwno_a_2nd	Gleditsch and Ward country codes of states that enter a conflict with troops to support side A. Taken from the UCDP ACD. Comma separated if multiple. Missing if no supporters in that year.
gwno_b_2nd	Gleditsch and Ward country codes of states that enter a conflict with troops to support side B. Taken from the UCDP ACD. Comma separated if multiple. Missing if no supporters in that year.
region	Region of the conflict country. Taken from the UCDP ACD. <ol style="list-style-type: none"> <li>1. Europe</li> <li>2. Middle East</li> <li>3. Asia</li> <li>4. Africa</li> <li>5. Americas</li> </ol>

Variable name	Description
version	Version of the UCDP datasets that were used to create this data.
missingzone	Indicates whether a conflict zone was available for spatial linking of violence for this particular conflict. If not, violence variables that require spatial linking are missing.
battledeaths	Monthly fatalities in battle between the conflict’s main protagonists (government, rebels, and potentially supporting states). Aggregated from UCDP GED (Type 1), “best” estimate, and linked to conflict via conflict ID.
collateralciv	Monthly civilian fatalities (“collateral damage”) from battle between the conflict’s main protagonists. Aggregated from UCDP GED (Type 1), “deaths_civilians” estimate, and linked to conflict via conflict ID.
rebelosv	Monthly fatalities from one-sided violence against civilians by a conflict’s rebel group(s) (Side B). Aggregated from UCDP GED (Type 3), “best” estimate, and linked to conflict via actor ID.
interreb	Monthly fatalities from fighting between a conflict’s rebel groups (Side B), or between a conflict’s rebel group(s) and other armed groups. Aggregated from UCDP GED (Type 2, organization level 1), “best” estimate, and linked to conflict via actor ID.
interreb_collciv	Monthly civilian fatalities (“collateral damage”) from fighting between a conflict’s rebel groups (Side B), or between a conflict’s rebel group(s) and other armed groups. Aggregated from UCDP GED (Type 2, organization level 1), “deaths_civilians” estimate, and linked to conflict via actor ID.



Variable name	Description
zone_wrv	Monthly fatalities from all war-related violence perpetrated in the conflict zone, no matter by whom. Aggregated from UCDP GED (Types 2 and 3), “best” estimate, and linked to conflict spatially. Missing when conflict zone is missing.
zone_wrv <sub>civ</sub>	Monthly civilian fatalities from all war-related violence perpetrated in the conflict zone, no matter by whom. Aggregated from UCDP GED (Types 2 and 3), “deaths_civilians” estimate, and linked to conflict spatially. Missing when conflict zone is missing.
zone_osv	Monthly fatalities from all one-sided violence perpetrated in the conflict zone, no matter by whom. Aggregated from UCDP GED (Type 3), “best” estimate, and linked to conflict spatially. Missing when conflict zone is missing.
zone_nsv	Monthly fatalities from all fighting between non-state groups of any type and organization level. Aggregated from UCDP GED (Type 2), “best” estimate, and linked to conflict spatially. Missing when conflict zone is missing.
zone_nsv_collciv	Monthly civilian fatalities (“collateral damage”) from all fighting between non-state groups of any type and organization level. Aggregated from UCDP GED (Type 2), “deaths_civilians” estimate, and linked to conflict spatially. Missing when conflict zone is missing.
govosv	Monthly civilian fatalities from one-sided violence against civilians by any government. Aggregated from UCDP GED (Type 3), “best” estimate, and linked to conflict spatially. Missing when conflict zone is missing.

Variable name	Description
confgovosv	Monthly civilian fatalities from one-sided violence against civilians by the conflict government. Aggregated from UCDP GED (Type 3), “best” estimate, and linked to conflict spatially. OSV perpetrated by other governments in the conflict zone is not counted here, only OSV by the Side A in the conflict. Missing when conflict zone is missing.
combatantosv	Sum of rebelosv and confgovosv. Missing when conflict zone (and thus confgovosv) is missing.
nsag_osv	Monthly civilian fatalities from one-sided violence against civilians by any non-state armed group, whether or not they also fight the government or not. Aggregated from UCDP GED (Type 3), “best” estimate, and linked to conflict spatially. Note that rebosv is a more narrow category than nsag_osv, as the former only includes OSV by non-state armed groups that are the rebels (Side B) in the conflict. Missing when conflict zone is missing.
inter_nsag	Monthly fatalities from fighting between non-state armed groups, whether or not they also fight the government or not. Aggregated from UCDP GED (Type 2, organization level 1), “best” estimate, and linked to conflict spatially. Note that interrebel is a more narrow category than inter_nsag, as the former only includes fighting by non-state armed groups that are rebels (Side B) in the conflict. Missing when conflict zone is missing.
communal	Monthly fatalities from fighting between identity groups. Aggregated from UCDP GED (Type 2, organization level 3), “best” estimate, and linked to conflict spatially. Missing when conflict zone is missing.