
UCDP Battle-related Deaths Dataset Codebook

Version 21.1

Uppsala Conflict Data Program

Department of Peace and Conflict Research
Uppsala University

This version compiled and updated by Therese Pettersson (2021)

When using this data, please always cite:

Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).

When appropriate, also cite this codebook: Pettersson, Therese (2021) UCDP Battle-related Deaths Dataset Codebook v 21.1 (<https://ucdp.uu.se/downloads/>).

Always include the Version number in analyses using the dataset.

Contents

Introduction.....	2
Definition of Armed Conflict and Battle-related Deaths	3
Battle-related Deaths.....	3
Best, Low and High Estimate of Battle-related Deaths.....	4
Variables in the Battle-related Deaths Dataset	4
Data Collection Methods	11
Version name convention.....	11
Format availability	12
A note on UCDP ID changes	12
References.....	12

Introduction

This codebook presents the variables in the UCDP Battle-related Deaths Dataset, clarifies the definitions and methods, and provides explanations concerning the treatment of the sources consulted. The UCDP Battle-related Deaths Dataset project was developed with support from the Human Security Report Project, Simon Fraser University, 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.

There are two versions of the UCDP Battle-related Deaths Dataset (BRD), depending on the unit of analysis (resolution) employed. As such, there is:

1. a dyadic version, where each individual conflict dyad (pair of two fighting actors) has its own individual entry in each year the conflict dyad was active;

Thus, each entry contains a single pair of opposing sides (one side A and one side B) in a given year.

Multiple dyads may form a single conflict, as a conflict is defined by its incompatibilities.

2. a conflict version, where each individual conflict has its individual entry.

Thus, each entry contains one side A (the contested party) and one or more contesters in a given year (each entry being one contested incompatibility).

As such, for each row in the BRD Conflict version, there are one or more rows in the BRD Dyadic version.

If you are not certain which version of the dataset to choose, it is safer to use the dyadic version; the conflict version is an automatic aggregation of the dyadic version that automatically aggregates and drops information.

The definition of conflict in this dataset follows the UCDP definition of armed conflict.

Definition of Armed Conflict and Battle-related Deaths

The Uppsala Conflict Data Program (UCDP) defines an armed conflict as a contested incompatibility that concerns government and/or territory over which the use of armed force between two parties, of which at least one is the government of a state, has resulted in at least 25 battle-related deaths in one calendar year.

The separate elements of this definition are defined as follows.

- (a) *Incompatibility that concerns government and/or territory.* The incompatibility must concern government and/or territory and it refers to the stated generally incompatible positions of the parties. An *incompatibility that concerns government* refers to the type of political system, the replacement of the central government or the change of the composition of the current government. An *incompatibility that concerns territory* refers to the status of a territory, for example, the change of the state in control of a certain territory (interstate conflict), secession or autonomy (intrastate conflict).
- (b) *Use of armed force.* This refers to the use of arms by the military forces of the parties resulting in at least 25 deaths in a year. Arms are defined as any material means of combat, for example, manufactured weapons as well as sticks, stones, fire, water, and so on.
- (c) *Party.* This refers to the government of a state or an opposition organization or alliance of opposition organizations. *The government of a state* is the party controlling the capital of a state. An *opposition organization* is any non-governmental group which has announced a name for the group and its political goals and has used armed force to achieve them.
- (d) *State.* A state is an internationally recognized sovereign government controlling a specified territory or an internationally non-recognized government controlling a specific territory which is not disputed by an internationally recognized sovereign government, which previously controlled the same territory.

Battle-related Deaths

Battle-related deaths refer to those deaths caused by the warring parties that can be directly related to combat¹. This includes battlefield fighting, guerrilla activities (e.g. hit-and-run attacks/ambushes) and all kinds of bombardments of military bases, cities and villages etc. The target for the attacks is either the military forces or representatives for the parties, though there is often substantial collateral damage in the form of civilians being killed in the crossfire, indiscriminate bombings, etc. All fatalities – military as well as civilian – incurred in such situations are counted as battle-related deaths.

¹ Battle-related deaths, which concern direct deaths, are not the same as *war-related deaths*, which includes both direct as well as indirect deaths due to disease and starvation, criminality, or attacks deliberately directed against civilians only (one-sided violence).

Best, Low and High Estimate of Battle-related Deaths

The general rule for UCDP's estimation of battle-related deaths is moderation. All battle-related deaths have to be verified in one way or another, and all estimates reported are based on UCDP expertise of each particular conflict. As a general rule, all figures are disaggregated as far as possible and any figures that are not trustworthy are disregarded in the coding process. Due to the great uncertainty of reports from conflict areas, the project provides three estimates concerning battle-related deaths for each year.

(a) *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.

(b) *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.

(c) *High estimate.* The UCDP High estimate consists of the aggregated high estimates for all battle-related incidents during a year as listed in UCDP GED. 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 important to emphasize that the fatality estimates given by UCDP are based on publicly accessible sources. Due to the lack of available information, it is possible that there are more fatalities than the UCDP high estimate, but it is very unlikely that there is fewer than the UCDP best estimate. One important caveat is that the precision of the numbers belies the uncertainty of the estimates; while UCDP estimates are based on aggregated totals of all incidents recorded in our sources, we acknowledge that numerous events are not reported in the media. Figures are revised retroactively each year as new information becomes available.

Variables in the Battle-related Deaths Dataset

The Dyadic version of the dataset contains the following fields:

Variable name	Content	Type
<code>conflict_id</code>	The unique identifier of the conflict to which the dyad corresponds, as given in the UCDP/PRIO Armed Conflict Dataset version 21.1 and the UCDP Dyadic Dataset version 21.1	Integer
<code>dyad_id</code>	The unique identifier of the dyad, as given in the UCDP Dyadic Dataset version 21.1	Integer

location_inc	The name of the country/countries whose government(s) has a primary claim to the incompatibility. If multiple countries are listed, this is comma separated.	String
side_a	The name of the country/countries of Side A in a conflict. Always the government side in intrastate conflicts, as given in the UCDP/PRIO Armed Conflict Dataset.	String
side_a_id	The unique identifier of the actor on side A in the dyad. Comma separated if multiple.	String
side_a_2nd	<p>side_a_2nd lists all states that enter a conflict dyad with troops to actively support side A in the dyad. A secondary warring party on side A shares the position in the incompatibility with Side A in the conflict. Side_a_2nd does not need to meet the 25 battle-related deaths criterion to be included in the dataset; an active troop participation is enough.</p> <p>If multiple countries are listed, this is comma separated.</p>	String
side_b	The name of the country or opposition organization in the dyad, as given in the UCDP Dyadic Dataset. In the dyadic version of the dataset, only one side_b exists per entry.	String
side_b_id	<p>The unique identifier of the actor on side B in the dyad.</p> <p>Note that in contrast with older versions of UCDP datasets, this variable is NO LONGER the Gleditsch and Ward state identifier (GWcode or GWNo) if the conflict is interstate and Side B represents a country. Use the gwno_b variable instead.</p> <p>In the dyadic version of the dataset, only one side_b_id exists per entry.</p>	Integer
side_b_2nd	<p>side_b_2nd lists all states that enter a conflict dyad with troops to actively support side B in the dyad. A secondary warring party on side B shares the position in the incompatibility with Side B in the conflict. Side_b_2nd does not need to meet the 25 battle-related deaths criterion to be included in the dataset; an active troop participation is enough.</p> <p>If multiple countries are listed, this is comma separated.</p>	String
incompatibility	<p>The main conflict issue identified per the UCDP definitions, as applied to the dyad:</p> <p>1= Incompatibility about territory</p>	Integer

2= Incompatibility about government
 3= Incompatibility about government AND territory

territory_name	The name of the territory over which the conflict is fought, provided that the incompatibility is over territory, as given in the UCDP/PRIO Armed Conflict Dataset.	String
year	The year of observation (1989-2020).	Integer
bd_best	The UCDP Best estimate for battle-related deaths in the dyad in the given year.	Integer
bd_low	The UCDP Low estimate for battle-related deaths in the dyad in the given year.	Integer
bd_high	The UCDP High estimate for battle-related deaths in the dyad in the given year.	Integer
type_of_conflict	<p>One of the following four types of conflict that the dyad is active in:</p> <p>1 = extrasystemic (between a state and a non-state group outside its own territory, where the government side is fighting to retain control of a territory outside the state system)</p> <p>2 = interstate (both sides are states in the Gleditsch and Ward membership system).</p> <p>3 = intrastate (side A is always a government; side B is always one or more rebel groups; there is no involvement of foreign governments with troops, i.e. there is no side_a_2nd or side_b_2nd coded).</p> <p>4 = internationalized intrastate (side A is always a government; side B is always one or more rebel groups; there is involvement of foreign governments with troops, i.e. there is at least ONE side_a_2nd or side_b_2nd coded).</p>	Integer
battle_location	<p>The name of the country/countries in which battle-related deaths have been recorded in this dyad-year: battle_location is a string variable, where the different countries are separated by a comma</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>	String

	Like UCDP BRD, GED is global and covers the same period (1989-2020).	
gwno_a	The Gleditsch and Ward country codes of side_a. Comma separated if multiple.	String
gwno_a_2nd	The Gleditsch and Ward country codes of side_a_2nd. Comma separated if multiple.	String
gwno_b	The Gleditsch and Ward country codes of side_b. Comma separated if multiple.	String
gwno_b_2nd	The Gleditsch and Ward country codes of side_b_2nd. Comma separated if multiple.	String
gwno_loc	The Gleditsch and Ward country codes of the incompatibility. Comma separated if multiple.	String
gwno_battle	<p>The Gleditsch and Ward country codes of battle_location. 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>	String
region	<p>The region of the incompatibility (NOT of the Battle Location):</p> <p>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).</p>	String
version	The version of the dataset: 21.1	Float

The Conflict version of the dataset contains the following fields:

Variable name	Content	Type
conflict_id	The unique identifier of the conflict, as given in the UCDP/PRIO Armed Conflict Dataset version 21.1 and the UCDP Dyadic Dataset version 21.1	Integer
dyad_id	A string variable listing all the dyads active in the current conflict-year separated by commas (',').	String
year	The year of observation (1989-2020).	Integer
location_inc	The name of the country/countries whose government(s) has a primary claim to the incompatibility. If multiple countries are listed, this is comma separated.	String
side_a	The name of the country/countries of Side A in a conflict. Always the government side in intrastate conflicts, as given in the UCDP/PRIO Armed Conflict Dataset. Comma separated if multiple.	String
side_a_2nd	side_a_2nd lists all states that enter a conflict with troops to actively support side A in the dyad. A secondary warring party on side A shares the position in the incompatibility with Side A in the conflict. Side_a_2nd does not need to meet the 25 battle-related deaths criterion to be included in the dataset; an active troop participation is enough. Comma separated if multiple.	String
side_a_id	The unique identifier of the actor on side A in the dyad.	Integer
side_b	The name of the country or opposition organization(s) of side B in a conflict, as given in the UCDP/PRIO Armed Conflict Dataset. Comma separated if multiple.	String
side_b_id	The identifier of each of the actors on side B in the conflict. Note that in contrast with older versions of UCDP datasets, this variable is NO LONGER the Gleditsch and Ward state identifier (GWcode or GWNo) if the conflict is interstate and Side B represents a country. Use the gwno_b variable instead. If more than one opposition organization or state is involved in a conflict, this is a comma-separated list of values.	String

side_b_2nd side_b_2nd lists all states that enter a conflict dyad with troops to actively support side B in the dyad. A secondary warring party on side B shares the position in the incompatibility with Side B in the conflict. Side_b_2nd does not need to meet the 25 battle-related deaths criterion to be included in the dataset; an active troop participation is enough.

Comma separated if multiple.

incompatibility The main conflict issue identified per the UCDP definitions: Integer

1= Incompatibility about government
 2= Incompatibility about territory
 3= Incompatibility about government AND territory

territory_name The name of the territory over which the conflict is fought, provided that the incompatibility is over territory, as given in the UCDP/PRIO Armed Conflict Dataset. String

bd_best The UCDP Best estimate for battle-related deaths in the conflict in the given year. Integer

bd_low The UCDP Low estimate for battle-related deaths in the conflict in the given year. Integer

bd_high The UCDP High estimate for battle-related deaths in the conflict in the given year. Integer

type_of_conflict One of the following four types of conflict over which the dyad is fought: Integer

1 = extrasystemic (between a state and a non-state group outside its own territory, where the government side is fighting to retain control of a territory outside the state system).
 2 = interstate (both sides are states in the Gleditsch and Ward membership system).
 3 = intrastate (side A is always a government; side B is always one or more rebel groups; there is no involvement of foreign governments with troops, i.e. there is no side_a_2nd or side_b_2nd coded).
 4 = internationalized intrastate (side A is always a government; side B is always one or more rebel groups; there is involvement of foreign governments with troops, i.e. there is at least ONE side_a_2nd or side_b_2nd coded).

battle_location The name of the country/countries in which battle-related deaths have been recorded in this conflict-year: battle_location is a string variable, where the different countries are separated by a comma (',').

	<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. 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 UCDP BRD, GED is global and covers the same period (1989-2020).</p>	
gwno_a	The Gleditsch and Ward country code of side_a. Comma separated if multiple.	String
gwno_a_2nd	The Gleditsch and Ward country codes of side_a_2nd. Comma separated if multiple.	String
gwno_b	The Gleditsch and Ward country codes of side_b. Comma separated if multiple.	String
gwno_b_2nd	The Gleditsch and Ward country codes of side_b_2nd. Comma separated if multiple.	String
gwno_loc	The Gleditsch and Ward country codes of the incompatibility. Comma separated if multiple.	String
gwno_battle	<p>The Gleditsch and Ward country codes of battle_location.</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>	String
region	<p>The region of the incompatibility (NOT of the Battle Location):</p> <p>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).</p>	String
version	The version of the dataset: 21.1	Float

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. an automatic join of that aggregation with dyad and conflict-year data (such as secondary warring party support or conflict typology) from the UCDP/PRIO Armed Conflict Dataset and the UCDP Dyadic Dataset.

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, 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 21.1, Section 4.

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 21.1 of the UCDP Battle-related Deaths 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.

Format availability

The data is available in Excel (XLSX), CSV (respecting the RFC 4180 specification), 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>.

A note on UCDP ID changes

In 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/>.

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

Pettersson, Therese, Shawn Davis, Amber Deniz, Garoun Engström, Nanar Hawach, Stina Högladh, Margareta Sollenberg & Magnus Öberg (2021). Organized violence 1989-2020, with a special emphasis on Syria. *Journal of Peace Research* 58(4).