UCDP Onset Dataset Codebook
Version 25.1

Uppsala Conflict Data Program

Department of Peace and Conflict Research
Uppsala University

This version compiled and updated by Magnus Öberg, James Dale, and Mert Can Yilmaz (2024)

When using this data, please always cite:

Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Håvard Strand (2002) Armed Conflict 1946–2001: A New Dataset. *Journal of Peace Research* 39(5): 615–637.

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).

Always include the version number in analyses using the dataset.

Please consult the codebook for the *UCDP/PRIO Armed Conflict Dataset version 25.1, 1946-2024* at: https://ucdp.uu.se/downloads/index.html#armedconflict for information on definitions not provided here.

Introduction

UCDP provides onset datasets for interstate and intrastate conflict, coding onsets in three distinct ways for each type of conflict, resulting in six distinct datasets. All datasets contain variables recording onsets of new armed conflict, and re-starts of previously active armed conflict after 1, 2, 3, 5, 10, and 20-years of inactivity. Three onset datasets are provided for each category of conflict (intrastate and interstate), differing with regard to the unit of analysis and on what basis they determine if there is new conflict or if conflict has restarted. The variants are described below and in Table 1 we use a real-world example to illustrate how they differ in the intrastate case. A flow chart in Appendix 1 provides a guide to the correct dataset depending on conflict type, unit of analysis and desired structure for the onsets.

All versions of the onset dataset contain annual observations of all states in the international system, as defined by Gleditsch and Ward (1999) between 1946 and 2021. The dataset files are plain CSV files, fields separated by ',' (comma).

Intrastate conflict

The intrastate datasets contain onsets for all intrastate (type 3) and internationalized intrastate (type 4) conflicts in the UCDP/PRIO Armed Conflict Dataset.

Version 1: UCDP Intrastate Conflict Level Onset Dataset

The *UCDP Intrastate Conflict Level Onset Dataset* uses the unique conflict id as the basis for determining whether it is a new conflict, and whether it is re-starting after 1, 2, 3, 5, 10 or 20-years. Thus, when determining whether an onset has occurred the conflict (identified by conflict id) is only compared to its own history. This means that there can be multiple onsets of conflicts with different "conflict id"s in the same country, even in the same year, and there can be onsets of conflicts while conflicts with other "conflict id"s are ongoing in the same country. Each onset is a separate row in the data, and there may thus be multiple entries for one country year.

Version 2: UCDP Intrastate Country Level Onset Dataset

The *UCDP Intrastate Country Level Onset Dataset* uses the unique country id (gwno_a) as the basis for determining whether it is a new onset of an intrastate armed conflict in a country, and whether it is re-starting after 1, 2, 3, 5, 10 or 20-years. Thus, when determining whether an intrastate conflict onset has occurred the conflict (identified by conflict id) is compared to the history of all intrastate conflict within the state (identified by gwno_a), and not just its own history. This means that there *cannot* be multiple onsets of conflicts the same country-year, and there can be *no* onsets of intrastate conflicts while intrastate conflicts with the same or other "conflict id"s are ongoing in the same country. Thus, in this dataset countries with ongoing intrastate conflict(s) *are not considered at risk* of new onsets of intrastate conflict. Each country-year can only appear once in the data.

Version 3: UCDP Intrastate Country Level Multiple Onset Dataset

The *UCDP Intrastate Country Level Multiple Onset Dataset* uses the unique conflict id as the basis for determining whether there is an onset of a new intrastate conflict, but uses country id (gwno_a) to determine whether intrastate conflict in the country is re-starting after 1, 2, 3, 5, 10 or 20-years. Thus, when determining whether a new onset has occurred the conflict (identified by conflict id) is compared only to its own history; but when determining whether an intrastate conflict restarted the conflict (identified by conflict id) is compared to the history of all intrastate conflict within the state (identified by gwno_a), and not just its own history. This allows for the onset of new intrastate conflicts (based on conflict id) in countries with ongoing intrastate conflict. Thus, in this dataset countries with ongoing intrastate conflict(s) *are considered at risk* of new onsets of intrastate conflict.

Table 1. Coding Example Illustrating Differences Between Versions 1-3 of the Intrastate Onset Dataset

Dataset	name	year	gwno_a	newconf	onset1	onset2	onset3	onset5	onset10	onset20	conflict_ids	year_prev
Conflict Level (conflict_id)												
	Bosnia-Herzegovina	1992	346	1	1	1	1	1	1	1	389	1815
	Bosnia-Herzegovina	1993	346	1	1	1	1	1	1	1	397	1815
	Bosnia-Herzegovina	1993	346	1	1	1	1	1	1	1	398	1815
	Bosnia-Herzegovina	1994	346	0	0	0	0	0	0	0		
Country Level (gwno_a)												
	Bosnia-Herzegovina	1992	346	1	1	1	1	1	1	1	389	1815
	Bosnia-Herzegovina	1993	346	0	0	0	0	0	0	0	389 397 398	
	Bosnia-Herzegovina	1994	346	0	0	0	0	0	0	0	389 397 398	
Country Level Multiple Onsets (conflict_id & gwno_a)												
	Bosnia-Herzegovina	1992	346	1	1	1	1	1	1	1	389	1815
	Bosnia-Herzegovina	1993	346	1	0	0	0	0	0	0	397	1992
	Bosnia-Herzegovina	1993	346	1	0	0	0	0	0	0	398	1992
	Bosnia-Herzegovina	1994	346	0	0	0	0	0	0	0		

Interstate conflict

The interstate datasets contain onsets for all interstate (type 2) conflicts in the UCDP/PRIO Armed Conflict Dataset.

Version 1: UCDP Interstate Conflict Level Onset Dataset

The *UCDP Interstate Conflict Level Onset Dataset* uses the unique conflict id as the basis for determining whether it is a new conflict, and whether it is re-starting after 1, 2, 3, 5, 10 or 20-years. Thus, when determining whether an onset has occurred the conflict (identified by conflict id) is only compared to its own history. This means that there can be multiple onsets of conflicts with different "conflict id"s in the same country, even in the same year, and there can be onsets of conflicts while conflicts with other "conflict id"s are ongoing in the same country. Each onset is a separate row in the data, and there may thus be multiple entries for one country year.

Version 2: UCDP Interstate Country Level Onset Dataset

The *UCDP Interstate Country Level Onset Dataset* uses the unique country id (gwno_a) as the basis for determining whether it is a new onset of an interstate armed conflict in a country, and whether it is re-starting after 1, 2, 3, 5, 10 or 20-years. Thus, when determining whether an interstate conflict onset has occurred the conflict (identified by conflict id) is compared to the history of all interstate conflict involving that state (identified by gwno_a). This means that there *cannot* be multiple onsets of conflicts the same country-year, and there can be *no* onsets of interstate conflicts while other interstate conflicts with the same or other "conflict id"s are ongoing in the same country. Thus, in this dataset countries with ongoing interstate conflict(s) *are not considered at risk* of new onsets of interstate conflict. Each country-year can only appear once in the data.

Version 3: UCDP Interstate Country Level Multiple Onset Dataset

The *UCDP Interstate Country Level Multiple Onset Dataset* uses the unique conflict id as the basis for determining whether there is an onset of a new interstate conflict, but uses country id (gwno_a) to determine whether interstate conflict in the country is re-starting after 1, 2, 3, 5, 10 or 20-years. Thus, when determining whether a new onset has occurred the conflict (identified by conflict id) is compared only to its own history; but when determining whether an interstate conflict restarted the conflict (identified by conflict id) is compared to the history of all intrastate conflict within the state (identified by

gwno_a), and not just its own history. This allows for the onset of new interstate conflicts (based on conflict id) in countries with ongoing interstate conflict. Thus, in this dataset countries with ongoing interstate conflict(s) *are considered at risk* of new onsets of intrastate conflict.

Variables in the Onset Datasets

Variable name	Content	Type
abc	Country abbreviation	String
name	Name of location according to the government side in the conflict.	String
year	Year of observation	Integer
gwno_a	The Gleditsch and Ward country code	Integer
newconf	Coded 1 if the country-year contains a new conflict in a country-year (version 1, 2, and 3)	Integer
onset1	Onset of an active armed conflict (>25 battle deaths) (version 1 and 2) or a conflict episode in a country-year (version 3) Coded as 1 if: 1) A new conflict emerges in a country-year (version 1 and 2) 2) A conflict episode starts for the first time in a country (for version 3) 3) There is <u>more than one year</u> since - the last observation of a specific conflict in a country-year (version 1) - the last observation of any conflict in a country-year (version 2) - the last observation of the location being at war in a country-year (version 3)	Integer
onset2	Onset of an active armed conflict (>25 battle deaths) (version 1 and 2) or a conflict episode in a country-year (version 3) Coded as 1 if: 1) A new conflict emerges in a country-year (version 1 and 2) 2) A conflict episode starts for the first time in a country (for version 3) 3) There is <u>more than two years</u> since - the last observation of a specific conflict in a country-year (version 1)	Integer

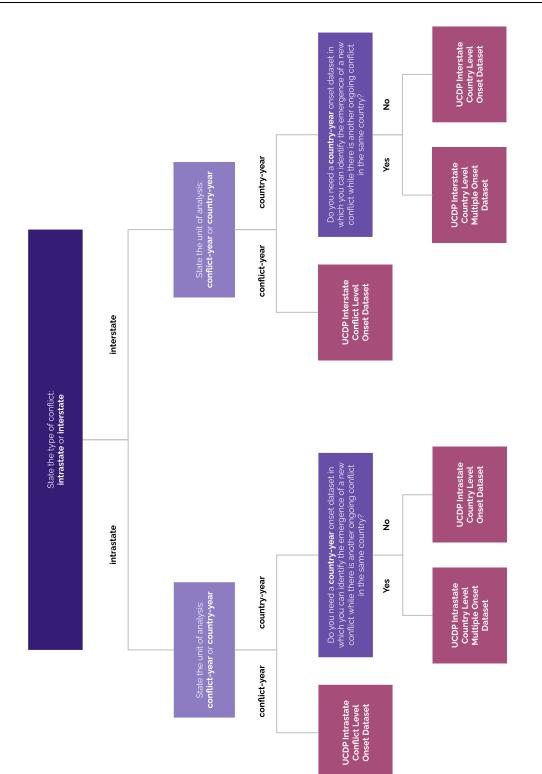|		

onset3	Onset of an active armed conflict (>25 battle deaths) (version 1 and 2) or a conflict episode in a country-year (version 3)
	
	Coded as 1 if:
	
	1) A new conflict emerges in a country-year (version 1 and 2)
	2) A conflict episode starts for the first time in a country (for version 3)
	3) There is <u>more than three years</u> since
	   - the last observation of a specific conflict in a country-year (version 1)
	   - the last observation of any conflict in a country-year (version 2)
	   - the last observation of the location being at war in a country-year (version 3)
		Integer

onset5	Onset of an active armed conflict (>25 battle deaths) (version 1 and 2) or a conflict episode in a country-year (version 3)
	
	Coded as 1 if:
	
	1) A new conflict emerges in a country-year (version 1 and 2)
	2) A conflict episode starts for the first time in a country (for version 3)
	3) There is <u>more than five years</u> since
	   - the last observation of a specific conflict in a country-year (version 1)
	   - the last observation of any conflict in a country-year (version 2)
	   - the last observation of the location being at war in a country-year (version 3)
		Integer

onset10	Onset of an active armed conflict (>25 battle deaths) (version 1 and 2) or a conflict episode in a country-year (version 3) Coded as 1 if: 1) A new conflict emerges in a country-year (version 1 and 2) 2) A conflict episode starts for the first time in a country (for version 3) 3) There is <u>more than ten years</u> since - the last observation of a specific conflict in a country-year (version 1)	Integer

	- the last observation of any conflict in a country-year (version 2) - the last observation of the location being at war in a country-year (version 3)	

onset20	Onset of an active armed conflict (>25 battle deaths) (version 1 and 2) or a conflict episode in a country-year (version 3)
	
	Coded as 1 if:
	
	1) A new conflict emerges in a country-year (version 1 and 2)
	2) A conflict episode starts for the first time in a country (for version 3)
	3) There is <u>more than twenty years</u> since
	   - the last observation of a specific conflict in a country-year (version 1)
	   - the last observation of any conflict in a country-year (version 2)
	   - the last observation of the location being at war in a country-year (version 3)
		Integer

conflict_ids	UCDP/PRIO Armed Conflict Dataset's conflict_id observations used for computing the onset gaps (version 1, 2, and 3)	Integer
year_prev	The previous onset recorded year. 1) The year of the previous observation of a specific conflict in a country-year (version 1) 2) The year of the previous observation of any conflict in a country-year (version 2) 3) The year of the previous observation of the location being at war in a country-year (version 3) If there is no previous observation, then the variable takes the value of 1815.	Integer

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

Appendix 1

State the type of conflict: **intrastate** or **interstate**

intrastate

interstate

intrastate branch

State the unit of analysis: **conflict-year** or **country-year**

conflict-year

country-year

UCDP Intrastate Conflict Level Onset Dataset

Do you need a **country-year** onset dataset in which you can identify the emergence of a new conflict while there is another ongoing conflict in the same country?

Yes

No

UCDP Intrastate Country Level Multiple Onset Dataset

UCDP Intrastate Country Level Onset Dataset

interstate branch

State the unit of analysis: **conflict-year** or **country-year**

conflict-year

country-year

UCDP Interstate Conflict Level Onset Dataset

Do you need a **country-year** onset dataset in which you can identify the emergence of a new conflict while there is another ongoing conflict in the same country?

Yes

No

UCDP Interstate Country Level Multiple Onset Dataset

UCDP Interstate Country Level Onset Dataset

Appendix 2

Conflict Level Onset Dataset

Time →

Conflict A
Conflict B
Conflict C

Step 1: for each conflict event following a gap, find most recent event belonging to **the same conflict**

Step 2: identify separate onsets for each conflict, relative to **only to the same conflict** (numbers denote length of conflict-free gap preceding onset)

		2					4							5		
			3				3					6				
												16				

Country Level Onset Dataset

Time →

Conflict A
Conflict B
Conflict C

Step 1: conflict identities discarded, all times coded as 'conflict' or 'no conflict' and gaps identified accordingly

Step 2: identify single set of onsets not associated with any conflict (numbers denote length of conflict-free gap preceding onset)

| | 2 | | | | | 3 | | | | | | 2 | | 1 | | |

Country Level Multiple Onset Dataset

Time →

Conflict A
Conflict B
Conflict C

Step 1: for each conflict event following a gap, find most recent event belonging to **any conflict**

Step 2: identify separate onsets for each conflict, relative to **all other conflicts** (numbers denote length of conflict-free gap preceding onset)

| | | 2 | | | | | 3 | | | | | | | 1 | | |
| | | | | | | | 3 | | | | | 2 | | | | |