

# The Militarized Interstate Participant (MIP) Codebook

---

This document provides user information on the coding procedures for the Militarized Interstate Participants (MIP) data, 1816-2014. We include a discussion of data formats and variable information for the participant-level data. We also discuss the addition of new variables for the data, including `fatalmin` and `fatalmax`. Any questions or errors should be reported to the authors.

---

July 13, 2023

## Citation

We ask that users of this dataset cite the following article and note the version number of the data they are using in their study:

- D. M. Gibler and S. V. Miller, "The Militarized Interstate Confrontation (MIC) Dataset, 1816-2014." *Journal of Conflict Resolution*, 2023. doi:[10.1177/00220027221104704](https://doi.org/10.1177/00220027221104704)

## Rationale for Militarized Interstate Confrontations (MIC) and Militarized Interstate Participants (MIP) Datasets

Our data collection efforts began with the original Militarized Interstate Dispute (MID) dataset from the Correlates of War Project (CoWMID) (?). We found an overwhelming number of errors in the original data and corrected those data, providing all of our suggested changes to the CoWMID Project dataset hosts. However, after more than nine years of communications with CoWMID and several exchanges in *International Studies Quarterly*,<sup>1</sup> it is apparent that CoWMID admits their pre-2002 data is incredibly error-prone but that they have neither the resources nor inclination to fix those thousands of admitted errors. Therefore, we are introducing the Militarized Interstate Confrontation (MIC) data, which are based on proper use of MID coding rules and include numerous advancements in both information and presentation.<sup>2</sup>

We use this document to discuss how we coded the data, using original CoWMID coding documents, and we provide annotations on specific variables most often used (and, when appropriate, misused) by conflict scholars. We also discuss important new variables we have added to the dataset and several variables we remove from the data.

## Data Files Associated with the Militarized Interstate Participant Data

We provide the following files for users of the MIP data set. We advertise the data of interest in the list below first as the more accessible comma-separated values file (.csv), but we also have the same data available in a Stata data file format (.dta) or an R serialized data frame (.rds). Please note that we are constantly revising these data; users should report the version number of the dataset used in any research. Future releases of the data will come with a text file summarizing changes to the data.

- mic-part-[version].csv: This is the participant-level MIC data with one case per confrontation-participant episode. Multilateral confrontations can have multiple entries for a participant on one or both sides of a confrontation. We also include Stata and R versions of the data.

---

<sup>1</sup>See ? and ? for examples, but also our final response which is posted on this website in ?. Note that our final response demonstrates that the CoWMID hosts (1) were ignorant of basic coding rules that led to the creation of the dataset—coding rules that should still be in use today— and (2) admitted there are thousands of errors in their released data.

<sup>2</sup>We previously released these suggested changes to CoWMID and the public in 2013 and have amended these data over time. CoWMID has referred to our suggestions as the GML data based on the author listing of our first *ISQ* piece, but this is the first time we are releasing the cases as a coherent dataset, and this is our first release of a codebook.

## Two Levels of Data and the Variables in Each Dataset

We keep the confrontation-level (MIC) and participant-level (MIP) presentation of the data and faithfully code values for each variable according to the rules established in ? and amended by ? and ?. Our separate change, drop, merge, and could-not-find reports provide details on conflict cases in which we disagree with the CoWMID team's application of their own coding rules as established in prior publications. Here, we describe the variables we include and exclude from the data and also describe several changes we made in order to better facilitate studies of conflict.

### *Militarized Interstate Confrontation (MIC) Participant-Level Data*

The following variables are included in the MIC participant-level data, with cautionary notes about each variable when appropriate:

- **micnum:** The confrontation number for each case, as originally established by CoWMID for each case. Note that we have also added numerous confrontation cases that are not in the CoWMID data as disputes. These cases begin with **confrontnum** values in the 9000's to demonstrate completely new cases.
- **ccode:** This is the CoW country code for that particular participant in the confrontation. Note that states can enter, exit, and re-enter a confrontation a theoretically infinite number of times; states can also change sides in the confrontation. For these cases the country code will remain the same, but the **sidea** variable can change values and the date variables will change to reflect the dates of each participation in the confrontation.
- **stmon:** The start month of that particular participant in the confrontation, with values ranging from 1 to 12. There are no missing month values; if the month of a militarized action could not be determined, the action was not coded.
- **stday:** The start day of that particular participant in the confrontation, with values ranging from 1 to 31 and missing days reported as values of -9. (Note that we have done our best to appeal to the historical record to eliminate missing days in the data, but there are numerous cases where only the month of the militarized action is known.)
- **styear:** The start year of that particular participant in the confrontation, with values ranging from 1816 to 2014. There are no missing year values; if the year of a militarized action could not be determined, the action was not coded.
- **endmon:** The end month of that particular participant in the confrontation, with values ranging from 1 to 12. Note that the end date of a confrontation can be coded due to the last incident, a set period of time following certain incidents (e.g. a blockade that last six months), or the implementation of a treaty (see especially, ?). For the last militarized event in a confrontation, please see our MIE data, 1946-2014. The MICEndings dataset can also be used to determine whether the end date corresponds to a militarized event or settlement. MICEndings also includes data on whether a treaty was signed, provisions of the treaty or agreement, etc.
- **endday:** The end day of that particular participant in the confrontation, with values ranging from 1 to 31 and missing days reported as values of -9. (Again, we have done our best to appeal to the historical record to eliminate missing days in the data, but there are numerous cases where only the end month of the militarized action is known.)

- `endyear`: The end year of that particular participant in the confrontation, with values ranging from 1816 to 2014.
- `sidea`: Side A is defined by whichever state initiated the first militarized event in the confrontation. All states coordinating with the original Side A state are also considered members of Side A. The variable is dichotomous, with a “1” indicating Side A and “0” indicating Side B. Note that Side A is the lower-number ccode for confrontations that began with clashes (confrontations in which the initiator could not be determined).
- `fatalmin`: This is an estimate of the minimum number of military battle-deaths for that particular participant in the confrontation. Missing values are often reported for fatality because `fatalpre` is unknown but near one of the battle-death thresholds. For example, if military battle-deaths were likely, between 20 and 30, but not definitively 25 or less or 26 or more, a missing fatality would be coded because it was unclear which value of the scale should be entered. This variable also allows researchers to better determine confrontations with military battle-deaths (or fatal MIDs). Minimum fatality levels also better allow integration of war data (see the section on the integrated war data provided below).
- `fatalmax`: This is an estimate of the maximum number of military battle-deaths for that particular participant in the confrontation. Used in conjunction with `fatalmin`, the most likely range of military battle-deaths in the confrontation can be determined.
- `hiact`: The “highest” level of militarized action for that particular participant in the confrontation. Note that this scale is actually not ordinal, however. Possible values [with `hostlev` in brackets] include the following: 0 No militarized action [1], 1 Threat to use force [2], 2 Threat to blockade [2], 3 Threat to occupy territory [2], 4 Threat to declare war [2], 5 Threat to use CBR weapons [2], 6 Threat to join war, 7 Show of force [3], 8 Alert [3], 9 Nuclear alert [3], 10 Mobilization [3], 11 Fortify border [3], 12 Border violation [3], 13 Blockade [4], 14 Occupation of territory [4], 15 Seizure [4], 16 Attack [4], 17 Clash [4], 18 Declaration of war [4], 19 Use of CBR weapons [4], and 22 War battle [5].
- `hostlev`: The “highest” recorded hostility level for that particular participant in the confrontation. Again, this scale should not be treated as ordinal. Possible values include the following: 1 No militarized action, 2 Threat to use force, 3 Display of force, 4 Use of force, and 5 War.
- `orig`: This variable is coded as “1” for states that participated in the confrontation from the beginning and “0” otherwise. Note that originators cannot always be determined by start dates since multiple states can enter a confrontation with different missing days in the same confrontation—i.e., originators begin the confrontation early in the month and joiners enter later in the month, but exact days could not be determined. Researchers should consult our MIE data and event numbers for these cases.
- `version`: The MIC participant-level data version number. We are constantly reviewing and updating our datasets, and we ask users of our data to always report which version number of the data they are using in their research.

#### *Variables Not Included in the Militarized Interstate Participant (MIP) Data*

We omit the following variables, with explanations following each variable definition.

- `dispnun4`: The MID<sub>4</sub> release of the CoWMID data provided a new dispute number. All values are missing for pre-1993 cases. Our focus is on the entire domain of conflict data, 1816-2014, so we see no need to include this variable in our data.
- `stabb`: We do not include the state abbreviation because merging with the state system data containing full state names is relatively easy, the abbreviation often provides little information, and we suspect the abbreviation can cause coder confusion. As we note in our MID change notes, there are cases in all CoWMID versions in which similarly-abbreviated states were confused—e.g., Nigeria and Niger in MID<sub>4</sub> and Indonesia and India in MID<sub>3</sub>. Further, `stabb` is not consistently applied throughout the data as, for examples, RUS becomes USR and then RUS again, or YUG becomes SRB, etc.
- `fatality`: We drop the seven-category CoWMID measure of fatalities. Researchers can still calculate their own ordinal fatality categories based on the information available in `fatalmin` and `fatalmax`. However, we believe studies will be better served by continuous measures of fatalities (that are perhaps logged).
- `fatalpre`: CoW's fatality estimates, with precision, are routinely unknown if they are not zero. We instead use the `fatalmin` and `fatalmax` categories to communicate this information and, with it, uncertainty around estimated fatalities in confrontations. Researchers can calculate their own version of this fatalities (with precision) variable. Cases where `fatalmin` equals `fatalmax` are cases of known fatalities, with precision.
- `revstate`, `revtype1`, and `revtype2`: We omit the revisionist state and revision type variables from this version of the data because we have not systematically reviewed these codings. Our initial reviews of these variables suggest they are not reliably coded and often provide little information (see, for example, ?). Instead, we are currently completing a dataset of all issues in militarized confrontations that includes issue positions for confrontation participants, changes in those issues positions, and changes in the status quo for each issue over time.

#### *A Discussion of Coding Differences with the CoWMID data*

CoWMID's last response to our suggested changes highlighted well several differences in how CoWMID coding rules are interpreted. We briefly reiterate those differences here:

1. Perhaps the biggest misunderstanding concerns the difference between clashes and attacks. CoWMID coding manuals are ambiguous when defining clashes, so we referenced ?, 589, the original article that introduced the dataset, and used their unambiguous definition of a clash as military hostilities in which the initiator could not be determined. These differ from attacks, which are coded as hostilities with a clear initiator. The CoWMID datahosts wrote that they were ignorant of this original definition; this suggests that CoWMID has substantially variation in how the majority of actions in its dataset are coded.
2. We did not code propaganda. If reports about militarized events were ambiguous, we consulted multiple sources and assessments of the information, including secondary sources and reviews.
3. We did not code shows of force against private citizens. These make no sense in light of the clear coding rules established in ? and their subsequent emendations.

4. We did not code cases for which we have no information. CoWMID continues to include 19 cases for which no source information exists. We did not code phantom cases of conflict.
5. We did not code border closures as blockades. A border closure denies entry into the initiator's own state; it does not necessarily deny the entry of goods and citizens into other states.

### *Treatment of the Conflicts Leading to the World Wars*

There has always been an underlying tension between how the CoWMID and the CoWWar datasets are coded. The datasets often do not agree on case values, including fatalities. As ? often imply in the "coding notes" sections of their case studies, it is difficult to merge the two datasets. Multiple disputes are sometimes associated with wars, or disputes precede but do not temporally overlap the war cases. This is strange if we assume, as most do, that wars are a subset of disputes with military fatalities greater than 1,000.

?, 176 include the following language when discussing wars:

When a dispute ends up in war, we treat the participants somewhat differently with regard to the aggregation of incidents. When two states go to war, all other ongoing disputes between these two states cease. Any dispute that erupts between a war belligerent and a non-belligerent state is treated as a separate dispute and only merged with the "core" war if the non-belligerent actively joins the war. If the entry into an ongoing war occurs within six months of its onset, then all sub-war militarized actions between a war belligerent and the third party entrant will be considered as part of incidents leading up to the intervention into the war. In cases when war intervention occurs six months or more after the start of the war, a separate militarized interstate dispute exists between the war belligerent and the other state up to its official entry in the ongoing war; thereafter, all actions are coded as part of the ongoing war. A state can be a participant in a war at a lower level of hostility only if its actions are fully coordinated with the war participants and its military combat falls short of the war threshold.

These coding rules are rarely followed by CoWMID. For example, MID#3702 is a bilateral dispute between Germany and Belgium that begins on November 5, 1939 and ends on May 10, 1940, with the German invasion. Belgium's highest action is an attack while Germany is coded as joining an ongoing war. Meanwhile, World War II (MID#0258) includes Belgian participation from August 25, 1939 until it leaves the system after German occupation on May 28, 1940. CoWMID codes Belgium's and Germany's highest actions as beginning an interstate war. These disputes effectively double count all the militarized incidents between Belgium and Germany during that period.

In another case, Luxembourg is coded as having a two-day, bilateral dispute with Germany, beginning on May 9, 1940 and ending with German occupation. Britain and France are excluded from the dispute even though France had more than 15,000 troops in the conflict, and the British RAF involved one squadron. Five French soldiers and one British soldier were killed in the invasion, but these are also ignored by CoWMID's coding of the case.

Perhaps more important than the inconsistency in coding rule application is the fact that the rules for wars differ conceptually between the CoWMID and CoWWar data. Italian participation in World War II is a case that highlights some of this conceptual confusion as a war is treated differently in the CoWMID data. CoWMID codes a separate dispute for Italian actions against

France, beginning on April 15, 1940. The dispute ends with an Italian declaration of war on June 10, 1940, and its highest action is coded as joins ongoing war. We could find no actions on April 15th, but we did find an Italian show of force against French and British forces on April 17th. Since Italy had started coordinating with Germany this show of force was Italy's entry into the larger conflict. CoWMID instead treats this action as distinct from the larger conflict, using the declaration of war to artificially create more conflicts in the data. Wars should simply be a subset of disputes with the highest fatalities—indeed, this was the *raison d'être* for collecting the original CoWMID data—, but instead the presence of war, especially the world wars, alters the data generation process and substantially inflates the number of conflict cases.

We change this type of treatment of the war cases and provide a unified coding of conflict cases in our data. We return to the base ? coding rules and define a confrontation as any threat, display, or use of force between two or more states. We define originator(s) on side A of the confrontation as the state that initiated the first militarized event, and, for multilateral conflicts, we identify parties to the confrontation as those states actively fighting together and/or coordinating their participation. We do not try to define when a confrontation becomes a war. However, with our events data, researchers can easily examine how conflicts unfold and, if necessary, define for themselves when escalation to war occurs—by fatality thresholds, actions, or another indicator.

This unification of conflict cases leads us to merge several conflicts with the world wars. We merge Belgian (#3701), Dutch (#3702), Luxembourgish (#3703), American (#0339 and #0414), Russian (#3822), and Thai (#1694) participation in World War II with the core militarized confrontation (#0258). We do not merge cases in which the parties were not coordinating actions or fighting together, which describes such cases as the Soviet takeovers of Lithuania (#0504), Latvia (#0022), and Estonia (#0505), or the British actions against Bulgaria (#0518) prior to Bulgaria's decision to join the Axis powers in 1941. Bulgaria was not yet coordinating with the Germans.

### *Reliability, Validity, and Robustness Checks*

We agree wholeheartedly with the CoWMID project that, with the exception of protest-dependent cases, MIDs are an excellent concept for identifying interstate conflicts. Unfortunately, we found consistent problems with the reliability of the application of those coding rules to widely available source information. That is why our dataset now exists.

Finally, a seemingly natural implication of the back-and-forth with CoWMID and our introduction of the MIC data is that both datasets can be used as robustness checks for empirical research. However, given what we know of the data and what the CoWMID datahosts have demonstrated, we think that use of the MID data for robustness checks is possible only for confrontations between 2002 and 2014. The error rate in prior versions of the CoWMID data is just too high (a lower bound of 2,588 errors at the dispute level and 5,913 at the participant level derived from CoWMID estimates). So, for example, consider a hypothetical in which estimates of the independent variable of interest are statistically significant using only our data. What does this mean? Does the result imply a lack of robustness to other datasets or simply that thousands of errors in the MID data are skewing estimates of the variable? It would be up to the researcher to determine which is true. CoW's own estimates suggests multiple thousands of errors in their data, so it would be impossible to determine without a comprehensive review whether *any* differences between estimates are due to coding rule or errors in their data.

## References

- Ghosn, Faten, Glenn Palmer and Stuart A Bremer. 2004. "The MID3 data set, 1993–2001: Procedures, coding rules, and description." *Conflict management and peace science* 21(2):133–154.
- Gibler, Douglas M. 2017. "What they fight for: Specific territorial issues in militarized interstate disputes, 1816–2001." *Conflict management and peace science* 34(2):194–211.
- Gibler, Douglas M and Steven V Miller. 2023. "The Militarized Interstate Confrontation Dataset, 1816-2014." *Journal of Conflict Resolution* p. 00220027221104704.
- Gibler, Douglas M. and Steven V. Miller. N.d. "Our Perspective: Thousands of Errors Remain in the MID Data." Posted response to CoW authors' rejoinder in ISQ, available at .
- Gibler, Douglas M, Steven V Miller and Erin K Little. 2016. "An analysis of the militarized interstate dispute (MID) dataset, 1816–2001." *International Studies Quarterly* 60(4):719–730.
- Gibler, Douglas M, Steven V Miller and Erin K Little. 2020. "The Importance of Correct Measurement: A Response to Palmer, et al." *International Studies Quarterly* .
- Gochman, Charles S and Zeev Maoz. 1984. "Militarized interstate disputes, 1816-1976: Procedures, patterns, and insights." *Journal of Conflict Resolution* 28(4):585–616.
- Jones, Daniel M, Stuart A Bremer and J David Singer. 1996. "Militarized interstate disputes, 1816–1992: Rationale, coding rules, and empirical patterns." *Conflict Management and Peace Science* 15(2):163–213.
- Palmer, Glenn, Vito d'Orazio, Michael Kenwick and Matthew Lane. 2015. "The MID4 dataset, 2002–2010: Procedures, coding rules and description." *Conflict Management and Peace Science* 32(2):222–242.
- Sarkees, Meredith Reid and Frank Wayman. 2010. *Resort to war: 1816-2007. Correlates of War*. Washington DC: CQ Press.