



# The higher the better? The limits of analytical resolution in conflict event datasets

Cooperation and Conflict

48(4) 567–576

© The Author(s) 2013

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/0010836713507670

[cac.sagepub.com](http://cac.sagepub.com)



**Nils B Weidmann**

## Abstract

The majority of conflict event datasets rely on media reports as their sole source of information. Because of the various difficulties associated with media reports, it is useful to compare conflict events based on them with those obtained from other observers. A paper published in 2010 by O’Loughlin and colleagues makes a first attempt to do this by using (1) a media-based event dataset and (2) military records on Afghanistan. While the authors conclude that the level of agreement between the two datasets is high, my results show that this goes away once we aggregate to finer analytical resolutions – those that are typically used in micro-level conflict analyses. Thus, rather than giving us the ‘all-clear’ for the accuracy and quality of media-based conflict data, my results once again point to the importance of robustness tests in quantitative conflict research, but also to the need to study the discrepancies in different reporting mechanisms to find out what they can and what they cannot tell us.

## Keywords

Conflict event data, geographic information systems, spatial resolution

Electronic means of data collection and processing continue to make their way into the social sciences. There are plenty of reasons for researchers to be excited about this. Most importantly, new technologies allow for an exploration of unprecedented amounts of data that record the dynamics of social phenomena at high levels of precision. Event datasets on violent conflict represent an instance of this growing trend. While previous conflict databases such as ‘Correlates of War’ or the Uppsala-PRIO ‘Armed Conflict Dataset’ on political conflict code violence at aggregate levels (typically, country-years), more recent event data focus on individual violent incidents during episodes of ongoing war.

Conflict event data hold a lot of promise. Not only is it possible to trace how violence unfolds over space and time, but also how different types of violence are related to each

---

## Corresponding author:

Nils B Weidmann, Department of Politics and Public Administration, University of Konstanz, 78457 Konstanz, Germany.

Email: [nils.weidmann@uni-konstanz.de](mailto:nils.weidmann@uni-konstanz.de)

other, or how different actor constellations play out on the battlefield. In short, detailed conflict event datasets are at the core of a newly emerging research agenda – the micro-level analysis of violent conflict. By pinpointing the dynamics of conflict at a new level of detail, these datasets promise to push the study of political violence to a new level.

At the same time, however, the increased precision and high level of detail that event datasets aim to achieve warrants a closer look. For obvious reasons, the social phenomena we as civil war scholars are interested in do not really lend themselves to detailed empirical analysis: it is extremely difficult to accurately collect data in conflict regions. All we can do is rely on ‘observers’ such as journalists or military personnel who provide us with accounts of the events, which we can then assemble into more comprehensive datasets suitable for empirical analysis. This type of data collection has many potential problems that could occur on the way, ranging from omission to biased selection or presentation (Öberg and Sollenberg, 2011).

One way to find out about these potential problems with micro-level event datasets is to cross-compare (or ‘triangulate’) them with others coding the same type of social phenomenon, analyzing the degree to which they overlap. O’Loughlin et al. (2010) conduct such an exercise, comparing a media-based conflict event dataset to one assembled by military personnel. While not meant to be an exhaustive analysis, their finding is that the two datasets overlap to a great extent. This is interpreted as lending confidence to the micro-codings of violence provided by each observing mechanism.

Below, I scrutinize this exercise and find that the correlation between the two datasets decreases rapidly when we move away from aggregate comparisons. While this finding may initially give cause for concern, I argue that it need not. Rather, the differences I find once again emphasize the need to conduct robustness checks to see if analytical results hold when using alternative data sources. In addition, a systematic examination of the differences across datasets is interesting in itself, as this can tell us more about the problems and gaps inherent in the respective data collection mechanism and how to fix them.

## **Triangulating event data**

Violent conflict is probably one of the social phenomena that are most difficult to study empirically. Yet, with increased media coverage and electronic dissemination of news, our knowledge has increased dramatically, to the extent that conflict researchers are now able to assemble detailed lists of events of violence. The information in these datasets is usually based on one of three sources (see Otto, *in press*): (1) news reports, typically those provided by the major news agencies; (2) military records collected by one of the fighting organizations; and (3), post-conflict surveys and data collections such as those carried out by truth commissions.

The first source, news reports, is usually the preferred way to go, as it has global coverage and allows for an up-to-date coding of ongoing conflicts. However, media-based datasets may suffer from biases introduced by omission, selective reporting or distortion (Davenport and Ball, 2002; Galtung and Holmboe Ruge, 1965; Öberg and Sollenberg, 2011). In short, there may be a great deal of uncertainty associated with the use of media-based datasets alone. Consequently, researchers have sought to validate this source of

information, in order to find out whether it provides a sufficiently detailed empirical foundation on which to base analysis.

One type of validation can be carried out by means of data triangulation. Triangulation in the social sciences is the process of utilizing multiple procedures to derive a result. While mostly used to refer to different (statistical) methods, we can also triangulate the empirical *data* on which our results are based. To be sure, I see data triangulation as a more fundamental validity check than what is typically done as a robustness test in empirical work on conflict. In a robustness test, one conflict dataset is typically swapped for another, because there may be different ways in which conflict is coded (for example, different battle death thresholds, see Sambanis, 2004). However, these datasets are almost always created by the same mechanism, so a typical robustness test cannot take into account systematic difficulties associated with this mechanism. Triangulation, on the other hand, involves empirical data collected in fundamentally different ways. A true triangulation for conflict event data, then, would have to involve *different sources* or *probes* into the dynamics of violence on the ground.

There exist few opportunities for true triangulations to be done for micro-level conflict data. Since most of the current datasets are based on media reports, we are not able to assess whether reliance on this particular source systematically affects the accuracy and quality of the recorded data. Recently, however, a rare opportunity to do exactly that arose with the release of military event datasets for Iraq and Afghanistan by the WikiLeaks organization. The datasets constitute copies from an internal military database system, which is used to record a series of different event types for the two conflicts. The Afghanistan part of the release is called the 'Afghanistan War Diary' (WD henceforth), and includes more than two years of coverage of the still ongoing conflict, with more than 70,000 individual entries.

Using data from military sources also raises questions of accuracy and bias. Some scholars have pointed at the incentives that the military may have in underreporting collateral damage and civilian killings. In the case of the WD dataset, these concerns are partly alleviated by the fact that the data were never collected for public distribution and represent a documentation effort *internal* to the organization. Even if biases in the nature and number of casualties existed, they are not terribly relevant for the present analysis, since the main focus is on whether an event happened at a particular location and time, and how different observer mechanisms are able to capture that. For this purpose, relying on an organization involved in the fighting is likely to give us a more precise picture of events on the ground, even if this picture is itself not perfectly accurate in each and every detail.

Access to an event dataset created by a different reporting mechanism allows for a validation of media-based event datasets that rely on media records. O'Loughlin et al. (2010), one of the first academic articles using the WD data, conduct a test of this kind. The authors compare event counts from the Armed Conflict Location and Events Dataset (ACLED, Raleigh et al., 2010) in Afghanistan to those from the WD. Since the latter includes a huge number of events that refer to non-lethal violence, the authors restrict their analysis to only the lethal ones, which should correspond to those included in ACLED.

The authors conduct two kinds of comparison. First, they show that ACLED monthly event counts from the entire country track those in WD reasonably well (Figure 2 in O'Loughlin et al., 2010: 478). Of course, the number of WD events is much higher, but ACLED seems to be capturing major escalations during the conflict. Second, a similar analysis is conducted by aggregating all events over the entire period of analysis (2008–2009) to the provinces in Afghanistan (Figure 5 in O'Loughlin et al., 2010: 481). Both ACLED and WD identify the same provinces as conflict hotspots, such as Helmand and Kandahar. The authors take all these results as support for the accuracy of ACLED, and conclude that,

[w]hile this comparison for conflict in Afghanistan does not assure that other databases drawn from media sources are equally as reflective of the civil war occurrences, it nevertheless offers an important confirmation of the value of this type of event collection and analysis. (O'Loughlin et al., 2010: 482)

I believe that a more thorough comparison should be conducted. After all, O'Loughlin et al. (2010) employ high aggregations, either over the entire country, or over the entire period of two years. Can we reach a similar level of agreement if we use, say, monthly observations by province, thus disaggregating both by space and time? What if we refine the analytical resolution, which is what event datasets should allow us to do to a certain limit? In the next section, I replicate and expand the dataset comparison, with somewhat different results.

## A closer look

In order to illustrate the complexities that can arise when using event datasets, this section repeats and expands the comparison between two event datasets presented by O'Loughlin et al. (2010), ACLED and WD released by WikiLeaks. Both datasets focus on the ongoing conflict in Afghanistan, and record individual events of violence. In order to ensure comparability, all other events that do not correspond to (deadly) violence between government/coalition troops and insurgents were dropped.<sup>1</sup> Like O'Loughlin et al., I limit my analysis to the years 2008 and 2009, for which we have complete data in both datasets. This results in two subsets of the original datasets, with  $N = 2568$  (ACLED) and  $N = 3559$  (WD). Note that while the filtering of the data eliminates relatively few events from ACLED, it drops the vast majority of cases in WD. This is because the latter includes many non-deadly incidents that while observed by troops on the ground, never make it to the international news and will therefore not appear in any event dataset relying on media reports. Still, when we restrict the analysis to deadly events, the number of incidents we get from ACLED and WD is in roughly the same order of magnitude.<sup>2</sup>

With two comparable event lists in place, I follow the typical approach and aggregate the events to externally defined spatial–temporal units of analysis. However, since the main purpose of this exercise is to examine how well the datasets overlap as we refine the analytical focus, I use three different resolutions of the analytical unit, each of which is defined by a spatial unit and a temporal period to which the events aggregated. My first dataset follows O'Loughlin et al. (2010) and uses the 34 provinces of

Afghanistan as the spatial unit, and calendar months as the temporal one. The second dataset narrows the focus down to administrative districts and calendar weeks. Finally, I examine a high-resolution dataset that employs quadratic spatial cells of one-sixth of a degree in size, which corresponds roughly to 20 km. These cells are observed over three-day periods.

The first two choices of spatial units (provinces and districts) are oftentimes useful, since many other covariates that political scientists are typically interested in are measured at the level of administrative units (see e.g. Weidmann and Callen, 2012). Of course, the choice of the temporal period is somewhat arbitrary, but calendar units such as months or weeks are frequently selected (similar to cross-national research, where countries are oftentimes measured over annual periods). The last analytical resolution I choose (spatial cells observed over three-day periods) is particularly suitable to study the impact of variables that vary independently of political boundaries, such as terrain ruggedness or land cover. With new advanced measurement techniques being developed, more and more of these variables will become available to social scientists, which will make the cell-based approach increasingly important.

For each case (a particular spatial unit observed during a particular time period), I code two binary ‘conflict’ variables, one for ACLED and one for WD. Each variable is coded as 1 if at least one incident of the respective dataset takes place within the given spatial unit and time period.<sup>3</sup> Table 1 shows the three datasets and their properties.

The table clearly shows what micro-level researchers typically experience in this sort of analysis: Not surprisingly, there is a tremendous increase in the number of cases as we narrow the analytical solution. The cell-based dataset has more than 600 times as many cases as the province/month dataset, mostly driven by the vast increase in the number of spatial units (from 34 to 2249). As a result, however, conflict becomes increasingly rare as we narrow the analytical resolution: While in the first dataset, more than 50% of all province/months exhibit violence, this fraction goes down to less than 1% in the cell-based dataset. While not of key interest here, case selection techniques and/or statistical fixes exist to deal with this issue.

More important for us, however, is the extent to which the two event datasets – ACLED and WD – agree in pinning down the location and timing of violence in Afghanistan. As argued above, the rough comparison carried out by O’Loughlin et al. (2010) is only a first step. Rather than asking whether a media-based event dataset such as ACLED can accurately identify the most violent regions (aggregated over the entire study period) or the temporal escalation of violence (aggregated over the entire country), it is worthwhile to analyze if ACLED gets *both the time and place* of violence right. Using the three datasets described above, we can do this with different levels of precision.

With two binary conflict variables – one based on ACLED, the other based on WD – we do a first comparison by means of a truth table, listing the numbers of cases that ACLED, WD, neither or both categorize as ‘conflict’. Table 2 shows such a truth table for the province/month dataset.

The cases along the diagonal in the truth table are those where the datasets agree: 263 province/months are categorized as peaceful, and 429 province/months are conflict cases according to both datasets. We see that the numbers off the diagonal are relatively low in

**Table 1.** The three datasets used in the analysis and their properties.

	Number of spatial units	Number of time periods	Total number of observations	Proportion of cases with ACLED conflict	Proportion of cases with WD conflict
Province/months	34	24	816	0.587	0.616
District/weeks	398	104	41392	0.041	0.059
Cell/3-days	2249	243	546507	0.004	0.006

**Table 2.** Truth table for the ACLED-WD comparison at the province/month level.

	WD conflict = 0	WD conflict = 1
ACLED conflict = 0	263	74
ACLED conflict = 1	50	429

relation to the overall number of cases. All this shows that for a relatively coarse resolution, there is a good level of agreement between the two datasets.

However, do we really need to collect comprehensive and detailed event lists to find out that some provinces in Afghanistan were more violent than others? Provinces in Afghanistan encompass vast areas of land, and many of these provinces are larger than 20,000 square kilometres, with tremendous internal variation in demographic, geographic and economic variables. Thus, it seems that province as unit of analysis still over-aggregates tremendously. Luckily, with the precision of recent conflict event datasets, we should be able to do better.

Is the match between the two datasets equally good as we increase the analytical resolution? I subsequently repeat the analysis above for the other two datasets. However, in order to simplify the comparison, I focus only on conflict cases – that is, those cases where either ACLED or WD codes at least one event in the respective spatial unit and time period. These are the numbers shown in the grey-shaded boxes in Table 2. Comparing these numbers allows us to find out the extent to which the predicted conflict cases overlap, but for the sake of simplicity, we ignore the (huge) number of non-conflict cases. I present the number more intuitively in Figure 1. The size of the circles represents the number of conflict cases for each dataset, and the area of overlap is proportional to the number of cases where they agree – where both datasets code conflict. Thus, the larger the intersecting area, the better the fit of the datasets.

For comparison, the left panel of Figure 1 shows again the numbers presented above in Table 2. However, the high degree of fit disappears once we increase the resolution to the district/week level (centre panel). We can see, for example, that the majority of cases ACLED codes to be in conflict fail to be identified by WD. Similarly, more than half of the WD conflict cases do not show up in ACLED. Clearly, the mismatch is not due to a huge difference in the number of events contained in ACLED and WD, which is actually similar. This trend is even more pronounced when we narrow the resolution down to the cell/three-day level, where the degree of overlap becomes very small in relation to the overall number of conflict cases.

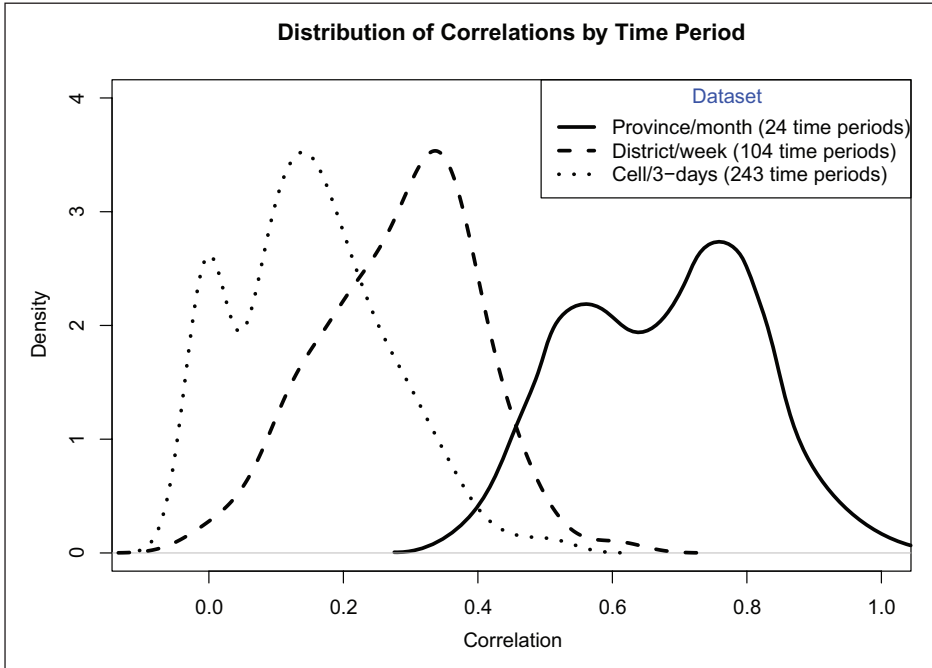


**Figure 1.** Number of conflict cases according to ACLED and WD. Top left: province/month aggregation. Bottom: district/week aggregation. Top right: cell/3-day aggregation. In each diagram, the overlapping area represents those cases where both ACLED and WD identify violence. The size of the circles is not to scale across the three panels.

We can do a similar comparison by computing the correlation coefficient between the ACLED and WD conflict codings.<sup>4</sup> Over the entire duration of the study period, this coefficient is 0.68 for the province/month dataset, 0.29 for the district/week dataset, and 0.15 for the cell/three-days dataset. In addition, I plot the distribution of the correlation coefficients computed by time period for each dataset. Essentially, Figure 2 shows for each of the datasets what range of correlations we get if we compute them separately for each time period (month, week or three-day).

The plot in Figure 2 again confirms the result presented above. While in the province/month dataset, the ACLED–WD correlations we get will be perfectly acceptable for social scientists (we rarely get below 0.5, with the majority ranging between 0.6 and 0.9), the district/week level dataset reveals a different picture. Although well within the precision range that media-based datasets should be able to achieve, the correlation with WD goes down to 0.2–0.4, which should raise concerns. For the cell/three-day dataset, correlations between 0.1 and 0.2 are too low to speak of a significant overlap.

One could argue that the reason for the mismatch of the two datasets is the difference in how geo-coordinates were obtained for events in the two datasets. While ACLED can only rely on the location mentioned in a news report, WD events are typically geo-referenced using GPS receivers operated by military personnel on the



**Figure 2.** Kernel density estimates for the distribution of correlations computed by time period.

ground. Thus, ACLED events will often be referenced in relation to cities and villages, while WD events can occur anywhere. The different geo-referencing procedure may partly explain the observed mismatch in the cell/three-day dataset, but it should not apply to the district/week one. The reason is that, even if news reports fail to include precise location information, they typically include the district in which an event occurred. This means that ACLED would still reference the event to the correct district, even if it cannot narrow the location down any further. Thus, despite the uncertainty inherent in media reports, ACLED and WD should still come up with district/week codings of conflicts that are relatively similar.

## Conclusions

As the above results have shown, our coding of violence depends to a large degree on the dataset in use. Counter to the conclusions drawn in O'Loughlin et al. (2010), I find that there is a substantial degree of mismatch between ACLED and the War Diary, at least when we move to higher resolutions. This may not be surprising to many, but it is a finding that must be taken into consideration by users of media-based event datasets. It is important to understand that we are not 'off the hook' when using media-based event datasets, which is what one could conclude from O'Loughlin et al.'s analysis. Rather, we



need to understand the issues inherent in each data collection mechanism, and reflect critically about its potential impacts on our research.

What does this mean for us as conflict researchers? I believe that there are at least two points to take home from this analysis. First, the mismatch between the two datasets I find above does not automatically invalidate findings based on one of them. After all, just like media-based event datasets may have their shortcomings, this could also apply to military records. In a sense, both datasets represent incomplete views on violence, each with its own problems and gaps. Is one more accurate than the other? I have no prior expectation. Thus, we must give each of them equal weight in our analysis. A simple solution is to replicate one's analysis using each of the datasets, which would then constitute a true data triangulation as defined above (e.g. Weidmann and Callen, 2012, online appendix). In a more complex analytical approach, we could even explicitly model the uncertainty associated with the individual codings. Events 'confirmed' by both datasets would have a lower uncertainty, and events coded by only one dataset would have higher values.

The second take-home point is rather an encouragement, namely to take the different perspectives on violence – and in particular their differences – seriously, and give them more attention in our research. Under what circumstances does reporting fail? How do different factions portray the same event? The 'observers' we employ to measure violence for us are social actors and organizations that respond to strategic incentives. In making these actors and the way they report about conflict an object of study, we can both learn a lot about their behaviour and at the same time improve the empirical basis on which we base our research.

## Acknowledgement

I am grateful to Andrew Linke and Sabine Otto for their comments.

## Funding

This research was supported by the Alexander von Humboldt Foundation (Sofja Kovalevskaja Award 2012) and the EU FP7 Marie Curie Zukunftscolleg Incoming Fellowship Programme, University of Konstanz (grant no. 291784).

## Notes

1. This includes, for example, reports of non-violent rebel activity in ACLED, and many incidents of non-deadly insurgent attacks in WD such as small arms fire without casualties (which are not reported in ACLED).
2. Both datasets include mostly battle-related incidents, but also events of one-sided violence. The criterion for inclusion in the comparison is that at least one casualty occurred, on whatever side.
3. The binary conflict coding is arguably a simplified way of measuring violence; however, it avoids a comparison of two highly skewed variables (very few cases have more than two ACLED or WD events). At the same time, it establishes a lower boundary on the conflict coding derived from both datasets, since not the number, but only the occurrence of violent events (yes/no) has to match for the datasets to agree.
4. With two binary variables, the ordinary correlation coefficient ('Phi-coefficient') needs to be interpreted with caution, since it is not bound by  $-1$  and  $1$  like the regular Pearson coefficient.

However, this does not apply if the proportion of 1s is similar in both variables, which holds in our case. Thus, all reported correlations can be interpreted as usual.

## References

- Davenport C and Ball P (2002) Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977–1995. *Journal of Conflict Resolution* 46(3): 427–450.
- Galtung J and Holmboe Ruge M (1965) The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. *Journal of Peace Research* 2(1): 64–91.
- Öberg M and Sollenberg M (2011) Gathering conflict information using news resources. In: Höglund K and Öberg M (eds) *Understanding Peace Research: Methods and Challenges*. New York: Routledge, 47–73.
- O’Loughlin J, Witmer F, Linke A, et al. (2010) Peering into the fog of war: The geography of the Wikileaks Afghanistan war logs, 2004–2009. *Eurasian Geography and Economics* 51(4): 472–495.
- Otto S (in press) Coding one-sided violence from media reports. *Cooperation and Conflict*.
- Raleigh C, Linke A, Hegre H, et al. (2010) Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research* 47(5): 651–660.
- Sambanis N (2004) What is civil war? Conceptual and empirical complexities of an operational definition. *Journal of Conflict Resolution* 48(6): 814–858.
- Weidmann NB and Callen M (2012) Violence and election fraud: Evidence from Afghanistan. *British Journal of Political Science* 43(1): 53–75.

## Author biography

Nils B. Weidmann is Professor of Political Science at the University of Konstanz, Germany. His research interests include violent and non-violent contestation, with a particular focus on the impact of communication and information technology.