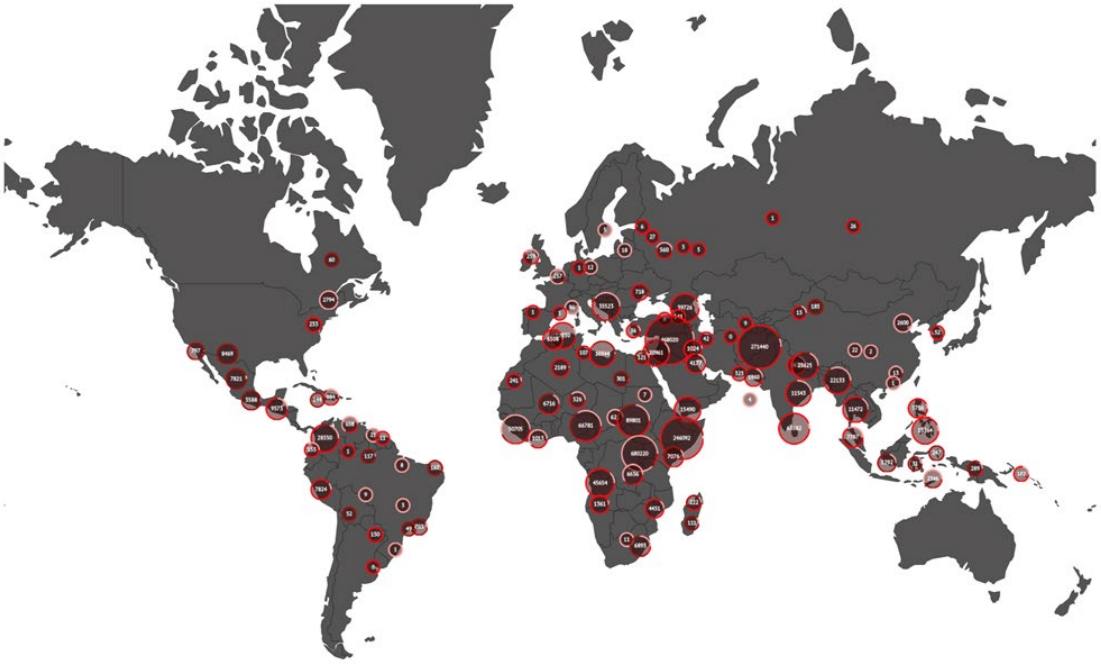
UCDP Paper Number 10



Measurement Sophistry

A response to Raleigh, Kishi and Linke "Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices" in *Humanities & Social Sciences Communications* (2023)10:74

Magnus Öberg
Director of the Uppsala Conflict Data Program
Department of Peace and Conflict Research
Uppsala University

Published by the Uppsala Conflict Data Program (UCDP)

Uppsala University, Sweden

ISSN: 1653-4573

ISBN: 978-91-506-3108-1

URN: https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-553982

UPPSALA UNIVERSITET

Abstract

In the target article Clionadh Raleigh, Roudabeh Kishi, and Andrew Linke set out to assess the validity and reliability of four conflict event datasets. To accomplish this would require a plausible reference data that at least approximates ground truth. Lacking this, the authors use their own Armed Conflict Location and Events Dataset (ACLED) as the baseline for comparison. A comparative evaluation of datasets could in principle bring to light differences, and perhaps identify strengths and weaknesses of the different datasets. It could help us improve and develop better data generation processes over time, and also provide a better understanding the limits of existing conflict event data so that it is used appropriately. However, the evaluation and analysis provided by the authors is based on false premises and mired by misrepresentations, factual errors and conceptual misunderstandings, to a degree that it becomes misleading rather than enlightening. In this paper I detail how the evaluation and analysis provided in the target article is based a factually false premise and mired by misrepresentations, factual errors and conceptual misunderstandings.

Introduction[1]

Comparing and analysing the data generation process is important for advancing knowledge on the informational landscape in wartime settings and how it influences the data generation process and the study of armed conflict (e.g. Baum & Zhukov, 2015; Croicu & Eck, 2022; Croicu & Kreutz, 2017; Davenport & Ball, 2002; Dawkins, 2021; Dietrich & Eck, 2020; Price & Ball, 2015; Weidmann, 2015, 2016). Comparisons to data that are approximations of ground truth are especially useful since they make it possible to assess validity and reliability of the datasets, and provides some insight into the data generation process in terms of what types of events and information makes it into the event datasets and what is lost along the way.

The article by Clionadh Raleigh, Roudabeh Kishi, and Andrew Linke (RKL for short) ostensibly engage in a systematic assessment of four different conflict events datasets. In contrast to the works cited above RKL does not use ground truth data as their reference point, instead they use their own dataset, the Armed Conflict Location and Events Data (ACLED), as a basis for comparison. This precludes a serious estimate of validity and reliability, but it is still possible to compare how different definitions, practices procedures, and sourcing strategies and to assess how they affect the resulting data.

Serious and sincere comparisons and critiques can be very useful for improving both data collection practices and to inform users about differences across datasets and what might explain these differences. The article by RKL, in contrast, starts from the false premise that the four datasets they compare are aiming to measure the same thing as ACLED – the occurrence of political violence very broadly defined – and proceeds with a series of misleading comparisons based on factually incorrect assertions about the contents, definitions, coding rules, sourcing and methods of the other datasets. I will deal here only with the comparison between ACLED and the Uppsala Conflict Data Program Geo-Referenced Events Data (UCDP GED) to show how the authors misrepresentations lead them to conclusions that are both misleading and unfounded.

False premises

The core stated purpose of the RKL article is to assess the reliability and validity of different datasets. In established parlance, reliability and validity is about how well a measure – in this case datasets – measure something. Validity refers to the accuracy of a measure and reliability refers to the consistency of a measure. Both are in reference to that 'something' which is to be measured. Thus, validity and reliability should be assessed in relation to what the dataset is aiming to measure.

Already in the abstract the authors claim that the examined datasets are "…all purporting to capture the same phenomena of political violence" (Raleigh et al., 2023, p. 1), and then go on to claim more specifically that: "We compare four datasets that collect data on the occurrence of political violence, defined broadly as the use of force by an organized armed group in the pursuance of a political goal, including replacing an agent or system of government; the protection, harassment, or repression of identity groups, and/or political groups/organizations; the destruction of safe, secure, public spaces; and contests to establish political authority over an area or community" (Raleigh et al., 2023, p. 3). The conceit here is RKL's claim that UCDP GED aims to measure political violence so defined.

The UCDP aims to collect data on three specific types of lethal organized violence. It is not an attempt to capture all 'political violence' as described above. It is an attempt to measure three specific categories of

organized violence: state-based armed conflict, non-state armed conflict, and one-sided violence (for definitions see Eck & Hultman, 2007; Gleditsch et al., 2002; Sundberg et al., 2012).[2] The UCDP data are designed to enable researchers to study the causes, dynamics, effects and resolution of particular forms of violence that are central to the literature on organized violence. We provide data on the events that make up these particular forms of organized violence, not on political violence writ large. For example, the UCDP definition of state-based conflict is an operationalization of war and civil war. War is conventionally understood as a large-scale organized armed contest over authority (See e.g. ICRC, 2008).[3] The UCDP definition thus focuses on the incompatibility, the organization and character of the parties, and the scale of the violence (at least 25 battle related deaths in a year) (Heldt et al., 1993; Sundberg & Harbom, 2011). Similarly, UCDP's data on one-sided violence (OSV) seeks to identify and measure violence by an organized actor that intentionally targets unarmed civilians, i.e. what is conventionally understood as massacres and genocides (Eck & Hultman, 2007). UCDP's data on non-state conflict (NSC) includes violence between two organized groups neither of which is the government of a state, e.g. violence between rebel groups, militias, ethnic actors, clans, drug cartels and criminal gangs (Sundberg et al., 2012). All three categories have a fatality threshold of 25 in a calendar to enter the data, and all three categories include civilian fatalities (OSV contains only civilian fatalities). [4] The three categories are mutually exclusive. Scholars will value different definitions differently depending on their objectives and research questions. That UCDP offers a set of concepts that are useful to scholars is borne out by the number of articles relying on UCDP data.

Contrary to RKL (Raleigh et al., 2023, pp. 3, 7) we do not think that including all forms of violence makes a dataset more valid. Researchers who are interested in analysing civil war will get a more valid measure using a dataset designed to measure civil war to the exclusion of other forms of political violence, and a less valid measure using data that does not exclude other forms of violence. Having a more inclusive definition may be useful for some purposes. However, for the purposes of studying the causes of civil war (as one example), including all forms of political violence only introduce noise in the measure.

At this point RKL may perhaps retreat from the bailey to the motte, and claim that all they are saying is that some datasets collect data on more forms of political violence broadly defined. This would be trivially true, but would also undermine their whole argument about validity – and it is far from all that RKL claim.

Definitions of reliability and validity

Comparing the reliability and validity of empirical measures is a reasonable endeavour if you use these concepts in their conventional sense and apply them accordingly. RKL does neither. They have their own persuasion definition of reliability and apply reliability and validity in inconsistent and misleading ways. If RKL had used standard definitions of reliability and validity, and applied them in the conventional way to assess how reliably and validly a dataset measures what it aims to measure their conclusions would not follow.

Specifically, RKL claim that they use two main concepts in their comparisons. The first, "…'Internal Reliability' is defined as the consistency of event information in the dataset over time and area." (Raleigh et al., 2023, p. 3). They go on to suggest that "… the information sourcing has constant reliability because the sources do not change" (Raleigh et al., 2023, p. 3). They thus define reliability as using the same sources over time and space. This definition of reliability entirely divorced from conventional definition of reliability in philosophy of science, and it is conveniently tailored to fit

factually incorrect claims they later make about UCDP's sourcing practices. If their claims about UCDP sourcing were true, this tailored definition would allow RKL to conclude – as they do – that UCDP prioritize high 'Internal Reliability.' However, their claims about UCDP sourcing are not true (see below), so their conclusions do not follow even if we were to accept their definition of reliability.

The second main concept RKL use for comparisons is 'external validity,' defined as "… an assessment of how accurate event information is, and how the collection of events represents political violence occurrence trends on the ground." (Raleigh et al., 2023, p. 3). In normal scientific lingo, external validity is understood as the extent to which the findings of a study are generalizable beyond the scope of the study. RKL's definition is more like what is normally just called validity, i.e. how accurately a measure measures what it is intended to measure – if what you intend to measure is 'political violence occurrence trends on the ground.'

Stable definitions and measurement

RKL claims that for datasets that collect data on a "wider, comprehensive remit, prioritizing stasis in definitions of a shifting phenomenon like political violence, or centering sources that are only intermittently stable, creates invalid and systemic biases…" (Raleigh et al., 2023, p. 3) pointing to the UCDP as an example. The 'centering sources that are only intermittently stable' is not a true claim about UCDP. Creating 'invalid and systemic biases' sounds really bad, even if invalid bias is not a known term in social science methodology. Systemic bias has a meaning (the inherent tendency of a process to support particular outcomes), but it is not clear how it applies here except as a pejorative – or perhaps as a substitute for validity.[5] Either way this is a peculiar critique of an effort to measure something. Stasis in definitions is normally considered a virtue in measurement.[6] Having yards and ounces of varying lengths and weights is not a hallmark good measurement, but RKL seem to imply that because political violence is 'a shifting phenomenon' we should have a flexible and adjustable yardstick. How would we be able to measure trends in 'political conflict' if what we measure at one time and place is not the same as what we measure at another time or place? Without a stable definition, how would you even assess if the measure were reliable or valid? Even so, the consistent application of definitions by UCDP is a major theme in RKL's critique of UCDP GED.

RKL argue that fitting "…events into existing definitions, … will affect which present conflicts are 'seen'" (Raleigh et al., 2023, p. 5). Yes, and applying definitions have precisely the same effect on past and future conflicts. That is the point of applying definitions consistently; it makes sure you measure the same thing over time and across cases. Researchers typically see this as an advantage, not a disadvantage, of a dataset.

Conceptual mischaracterizations of UCDP

A serious evaluation of the validity of a dataset would require that the evaluator have a reasonably correct understanding of the definitions and the conceptual apparatus of the dataset being evaluated. Yet, RKL asserts, "In UCDP GED conflict events are categorized as state-based, non-state-based, or one-sided. Firstly, it must be determined whether a country has a conflict, then that conflict can be populated by events. A country has a conflict if there is an 'active dyad' (which must involve a combination of states, non-state actors, or civilians) with a stated incompatibility (territorial or governmental), and that specific dyad must result in at least 25 battle-related fatalities per annum. … UCDP GED divides actors into the state [the governments of the state actually], non-state and civilian

groups. Requiring a 'stated incompatibility' privileges the violence of predetermined violent actors, and their agendas" (Raleigh et al., 2023, p. 7). This characterization of UCDP is factually incorrect and misleading in at least seven ways.

First, countries do not have conflicts in UCDP, actors do. Second, only state-based conflicts require a stated incompatibility, not non-state conflicts, nor one-sided violence. Third, one-sided violence does not require active dyads (clue is in the name). Fourth, fatalities in one-sided violence are not battle deaths (again, clue is in the name). Fifth, UCDP does not treat the civilians in one-sided violence as actors, and they need not be 'groups.' Sixth, new actors enter the UCDP data every year, we do not 'privilege the violence of predetermined actors.' All of the UCDP criteria are in keeping with most standard understandings of armed conflict – that an actor has to be organized and that it needs to have some political dimension. Finally, conflicts in UCDP data are not first identified and then populated with events. This is a misrepresentation. It is the other way around. Separate events are identified and then aggregate into conflicts. This has always been the case with UCDP data, even before UCDP started releasing event data. Moreover, this fact is known to at least one of the authors, Clionadh Raleigh, since she was provided access to the raw UCDP events data files when creating ACLED (Raleigh et al., 2010, p. 657 footnote 1; Raleigh & Hegre, 2005, pp. 7–9).

Distinctions are not distortions

Based on the incorrect premise that all datasets aim to measure the same very broad range of political violence, RKL also makes the claim that "the poorly recognized – yet intentional – biases and prioritization that underpins public, global data collections affect which conflict patterns and causes are seen and unseen. The reluctance to recognize the many forms of political violence creates and amplifies distortions of conflict levels" (Raleigh et al., 2023, p. 2). They go on to claim that "Conflict data collection projects that center on particular forms of political violence, however, do not capture all political violence with equal accuracy" (Raleigh et al., 2023, p. 7). These claims misleadingly suggests that data projects that collect data on particular forms of conflict somehow fall short in a way (accuracy) that would make them less valid. However, what counts as distortion is entirely dependent of what you aim to measure. All-inclusive definitions will distort the picture of war, and of genocidal violence, and so on. If you aim to measure the amount of violence without distinction, then a measure of war is not the right measure. If you want to measure war, a measure of violence without distinction is not the right measure. Validity depends on the correspondence between the concept and what is picked up by the measure.

Discriminating between different forms of violence as UCDP GED does is not a flaw; it provides additional benefits and, contrary to RKL, it makes the unseen seen and allows for identification of causes. Only if we assume that all 'political violence' has the same aetiology does it follow that measuring only some forms of violence rather than all forms might make causes unseen (See e.g. Vasquez, 2009, pp. 17–23). If that assumption is wrong, and not all 'political violence' has the same aetiology, then making distinctions and not including all forms of violence is key to discovering causes. To give one example, in ACLED conflict issues are unseen, whereas in UCDP GED they are required for one of UCDP's categories of violence (state-based violence). This criterion is criticized by RKL in several places, because it supposedly privileges "predetermined actors, and their agendas" (Raleigh et al., 2023, p. 7) and represents an "antiquated view of conflict agents" (Raleigh et al., 2023, p. 8). However,

distinguishing incompatibilities allow the user of the data to determine whether the observed violent events are part of one conflict or many different conflicts. We also know that conflicts over territory and conflicts over government have different aetiologies, and understanding the objectives of rebel groups can be useful as it affects things like external interventions, types of possible settlements, etc. (e.g. Buhaug, 2006; Maekawa, 2019). Thus, what may appear as 'antiquated' to RKL is foundational to being able to study many of the topics that researchers are interested in.

Sourcing

The sourcing requirements of a conflict data collection effort is a function of the exact type of data being collected. For example, the sourcing requirements for collecting data on interstate wars, as in the original Correlates of War project (Singer & Small, 1972), are far less demanding than collecting data on lethal events in armed conflicts like the UCDP GED (Sundberg & Melander, 2013), which in turn is far less demanding of sources than collecting data on e.g. small scale non-lethal protests like ACLED (Raleigh et al., 2010). The former is less demanding than the latter because of how information about events typically propagate through different channels including through the news ecosystem(Cf. Galtung & Ruge, 1965).

In general, the threshold for reporting on events goes down, and the granularity of the reporting goes up, the closer the reporting actor is to the event, and so does the likelihood of partisan bias (Demarest & Langer, 2018, 2022; Öberg & Sollenberg, 2011). The actors closest to an event are the actors participating in the events, and bystanders who themselves directly observe the events. They are the first-hand sources. News organizations (and other violence reporting organizations), sometimes have reporters or stringers on the ground that are first hand sources, but more often they get their information from talking to first hand sources or by relaying what other news organizations who have first-hand sources are reporting. So, news organizations who do not have to have their own reporters on the ground may still report on an event. Stringers and local news organizations typically sell their information to anyone who wants to buy it, and so news about events propagate up to national and international levels if news organizations at these levels think that their respective audiences and customers are interested (Öberg & Sollenberg, 2011). What is reported by news organizations at different levels depend in large part on what their intended audiences and customers are interested in. Thus, what is newsworthy varies by audience. Local audiences served by local media have lower thresholds for what is a newsworthy conflict event, but they also have an audience who is more likely to have a bias against or in favour of one side in the conflict. Catering to this audience, local news organizations are more likely to have partisan bias in their reporting, but they are also more likely to report on smaller events and provide more detail. Thus, for event data collectors the need for local sourcing is a function of how large the audience is for the type of events the project is collecting data on. Deadly events that are part of large scale organized violent conflicts have much larger audiences than peaceful protests about local issues and so the informational demands on datasets that aim to collect data on the latter (i.e. ACLED) are orders of magnitude higher than the demands on datasets that collect data on the former (i.e. UCDP GED).

It is difficult to measure how sourcing affects the missingness and bias in conflict datasets because researchers only rarely have access to anything approximating ground truth that can be used as a bench mark. Studies that do compare conflict event data to direct observation data collected using boots on

the ground (i.e. data approaching ground truth), show that non-lethal conflict events suffer from serious underreporting and also that lethal events are more likely to propagate up the news food chain (cf Demarest and Langer 2018; Croicu and Eck 2022). Croicu and Eck (2022) furthermore suggest that among non-lethal conflict events, the less coercive events are even less likely to be reported and hence picked up by conflict datasets. The difficulty of collecting data on low scale events like protests may also be reinforced by coder sampling error (Demarest and Langer 2022:648).

It stands to reason that overall, non-violent events should be more common than violent events, most likely orders of magnitude more common (Cf Croicu and Eck 2022:454). The fact that ACLED reported that only 6% of the events in ACLED data were non-violent events (Raleigh, 2012, p. 464)[7] is suggestive of how much more demanding the information requirements are for collecting data on non-violent conflict events. It also indicates that ACLED's sourcing strategy may be insufficient for the task that they have set themselves. Using data collected by the UNAMID Joint Mission Analysis Center in Darfur Croicu and Eck finds not a single overlap, a zero correlation, between troop movement events registered by UNAMID JMAC and troop movement events registered by ACLED (Croicu and Eck 2022:464-465).

The differences in the demands on information implied by the different types of events collected by ACLED compared to UCDP GED is not something RKL focus on in their article. They focus instead on the sourcing strategies and practices of ACLED compared to the alleged sourcing strategies and practices of other datasets. Here, RKL asserts that many datasets (unlike ACLED) "… do not take advantage of new and better data because of their reliance on subscription-based publication aggregators like Factiva (e.g. UCDP GED); and emphasis on international media sources (e.g. UCDP GED, ICEWS, GDELT), ..." (Raleigh et al., 2023, p. 10). This is a fundamental misrepresentation of UCDP sourcing.[8] UCDP identifies conflict events in more than one thousand different sources each year. Each year this includes identifying many new sources in addition to targeted searches using a curated selection of sources that have proven to be useful in the past.

In the set of top ten most used sources each year since 2018 we find eight conflict monitoring organizations,[9] six national news organizations,[10] four international news organizations,[11] two aggregators of local news,[12] one local newspaper,[13] and one ministry of defence.[14] If we expand the scope to the 50 most used sources, we see much more local and national media, IGO reports, blogs, twitter and Facebook accounts, and so on. While the thumbnail picture of UCDP sourcing painted here is just scratching the surface, it should suffice to put to rest the allegations that UCDP "… do not take advantage of new and better data because of their reliance on subscription-based publication aggregators like Factiva" (Raleigh et al., 2023, p. 10) and that UCDP has a 'static source list' (e.g. Raleigh et al., 2023, Table 1).

After incorrectly portraying UCDP GED as being based on what they term 'traditional media', RKL goes on to show how dramatically different the picture of conflict events in Syria 2017 become if you add what they call 'new media,' 'local partners,' and 'other' sources. Had they compared to UCDP sourcing, they would have found that UCDP GED used *all* the sources they mention below Figure 2 as sources ACLED use (Raleigh et al., 2023, p. 11).[15] They would also have found that UCDP GED data looks nothing like the left panel, but more like the right panel in Figure 2. Similarly, when discussing source numbers and scale, as well as what they call source dynamism, RKL makes no comparisons to UCDP GED (Raleigh et al., 2023, p. 11). Had they done so, it would have undermined the picture they convey that ACLED's

sourcing is categorically different from UCDP GED. It is not, and they omit this fact from their presentation.

Data Comparisons: Misleading and omitting inconvenient data

RKL asserts that their empirical comparisons and 'systematic investigation' demonstrates that there are 'significant discrepancies' in the picture of violence produced by the different data projects (Raleigh et al., 2023, p. 13). However, the comparison conducted by RKL is not systematic, nor is it comprehensive. Instead, it is an eclectic misrepresentation that seems selected to maximize the appearance of discrepancies between ACLED and UCDP GED, producing a misleading foil favourable to the author's dataset (ACLED). The alleged discrepancies RKL present in their analysis are not a result of "deficiencies across core criteria"(Raleigh et al., 2023, p. 13), but rather the result of deficiencies in the authors' representations of UCDP GED, omitting inconvenient data from the analysis.

One of the most egregious examples is the Mexico example that is presented in the introduction, setting the tone for the whole paper. Here is what RKL claim: "The Armed Conflict Location & Event Data Project (ACLED) identifies Mexico as the most dangerous place in the world for civilians in 2021, with more civilians killed than cartel militants as a result of gang battles —6739 civilian fatalities were documented that year, which represents over 81% of the country's annual violent fatalities—underlining the urgency of this case for domestic and foreign policy. Yet the Uppsala Conflict Data Program's Georeferenced Event Dataset (UCDP-GED) records only 28 civilian fatalities in Mexico in 2021 (or 0.15% of all violent fatalities reported by UCDP-GED in Mexico that year)" (Raleigh et al., 2023, p. 2).

This quote illustrates how the authors create an incorrect picture of UCDP data by misrepresenting what it measures. By claiming that One-Sided Violence is *the* measure of civilian casualties in UCDP GED they make it look like UCDP GED only register 28 civilian casualties in Mexico in 2021. This is a patently false claim. All categories of violence in UCDP GED contain civilian casualties as stated on the UCDP website, in codebooks and release articles (Eck & Hultman, 2007; Gleditsch et al., 2002; Högbladh, 2022; Pettersson, 2022d, 2022c, 2022b, 2022a; Sundberg et al., 2012; Sundberg & Melander, 2013)[16]. In UCDP, battle-related deaths "…refer to those deaths caused by the warring parties that can be directly related to combat. 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" (Pettersson, 2022a, p. 5).

So, what does UCDP data tell us about Mexico in 2021? UCDP records almost 19 000 fatalities in non-state and one-sided violence, of these 110 are identified as being civilian, 262 as gang members, and the remaining fatalities are recorded as 'unknown'. This is a statement of uncertainty about the nature of the fatalities, not a claim that they were not civilian. It reflects that sources did not allow UCDP to distinguish civilians among the more than 18000 casualties classified as 'unknown.' In Mexico, it is usually not possible to determine how many of the reported fatalities in non-state conflict events are civilian as opposed to combatants (e.g. members of rival gangs), certainly not with any degree of confidence. Thus, the discrepancy between UCDP GED and ACLED in Mexico is a) that ACLED claims to

be able to distinguish civilians from gang members to a much greater extent than UCDP, and b) that ACLED records far fewer fatalities over all.[17] This of course also disqualifies all the arguments RKL make about the supposed reasons for the alleged discrepancies in the Mexico example (Raleigh et al., 2023, p. 2). It also lends no support to RKL's general conclusion that the discrepancies are due to "deficiencies across core criteria – conflict definitions, boundaries, and source catchments…" (Raleigh et al., 2023, p. 13).

The rest of the empirical comparisons RKL include in the article are no more veracious. For example, RKL claim that the Pokot militia in Kenya is not included in UCDP GED, in spite of ACLED registering over 600 fatalities involving Pokot between 1997 and 2020 (including six years accumulating more than 25 fatalities) (Raleigh et al., 2023, p. 8). This claim is incorrect. Between 1997 and 2020, UCDP GED registers nine different non-state conflicts between Pokot and other ethnic groups inhabiting the administrative districts of West Pokot and Baringo of the Rift Valley province in Kenya, as well as the eastern Karamoja region in Uganda. This includes eighteen years accumulating more than 25 fatalities and the total number of fatalities registered in UCDP GED is more than twice the number of fatalities registered by ACLED. [18]

Another example of how RKL misrepresent UCDP GED data is provided in Figure 1 (Raleigh et al., 2023, p. 10). Here RKL compare a combination of *non-fatal and fatal* conflict events from ACLED to *fatal events only* from UCDP GED, but fail to acknowledge this difference between the two panels in the figure. Instead, they misleadingly use UCDP GED categorizations in the figure legends for *both* panels – as if they were measuring the same thing.[19]

Based on these examples RKL make large claims about how datasets prioritizing what they call 'external validity' (i.e. ACLED) are more accurate and comprehensive, while datasets using definitions rigorously (like UCDP does) make violence disappear, deaths vanish, and conflicts become unseen (Raleigh et al., 2023, p. 13). The evidence provided in the article provide no grounds for such claims and it rings rather hollow when the *ACLED Year in Review* fails to even mention the war in Ethiopia against the Tigray People's Liberation Front (TFLP) among the deadliest conflicts in 2022 (ACLED, 2023, pp. 11–14). This, in spite of the conflict in Ethiopia registering more fatalities than any other conflict in 2022, over 162000 fatalities in UCDP data (Davies et al., 2024) with other estimates going as high as 500 000 fatalities (Economist, 2023).

So, what overall conclusions might one draw from the empirics presented by RKL? First, we can conclude that RKL's assertion that they undertook a 'systematic investigation' demonstrating that there are 'significant discrepancies' is inaccurate. The evidence provided in the empirical investigation is anecdotal, and in all cases except Madagascar, also incorrect and misleading (i.e. anything but rigorous and systematic). Second, we may conclude that UCDP GED and ACLED produce somewhat different pictures of organized violence in the cases analysed by RKL, but not on the scale nor in the way they assert in the article. The Mexico and Pokot cases suggest that a dataset with narrower definitions more rigorously applied (UCDP GED) may even produce a higher number of observed fatalities than a dataset (ACLED) with a wider definition more loosely applied. The Philippines and Madagascar cases (Raleigh et al., 2023, pp. 8–9), shows that a wider definition more loosely applied (ACLED) allows for the inclusion of events and phenomena outside of the scope of a dataset (UCDP GED) with a narrower definition more rigorously applied. Taken together, the empirical cases analysed in the article thus suggests that validity

and reliability in the traditional sense of measurement accuracy and consistency are orthogonal to RKL's claims about 'external validity' and 'internal reliability.'

Conclusions

It is important to evaluate and scrutinize data, and to analyse the data generation process. This is how we can improve and develop data over time, and better understand the limits of existing data so that it is used wisely. However, this requires that evaluations are sincere. If evaluations and analyses are based on false premises and mired by misrepresentations and factual errors, they do not help advance or improve either data collection efforts or data usage.

If I were to offer advice to data users, it would be the following: Which data you should use depends on what kind of research and research question you are interested in. You will have to figure out which dataset best measures what you aim to measure. Make sure you know your data; quality varies across datasets and across measures within datasets. Understanding the data generation process is important. Do not base your assessment of a dataset's reliability and validity on claims by people who have replaced the conventional understanding of these terms with persuasion definitions to allow themselves to draw their preferred conclusions and promote their own dataset. I also do not think it is advisable to base your views on the qualities, strengths and weaknesses of different datasets on assessments by people who grossly misrepresent the datasets they claim to comprehensively compare.

Epilogue

This rejoinder, pretty much as it now stands,[20] was submitted to *Humanities and Social Sciences Communications* in early September 2023. It was sent two reviewers along with a reply to the rejoinder by the authors. The reviewers agreed with the points made in the rejoinder and both clearly recommended publication – one reviewer going as far as to say that the original article by RKL should never have been published. From this the editor of *Humanities and Social Sciences Communications* concluded that they could not publish the rejoinder because it did not challenge the core premises or the conclusion of the original article. I beg to differ.

References

ACLED. (n.d.). *ACLED Terms of Use, Version 2*. Retrieved April 19, 2023, from https://acleddata.com/download/35181/

ACLED. (2023). *ACLED Year in Review, Global Disorder in 2022. Escalating Violence and the Worsening Civilian Burden.* The Armed Conflict Location & Event Data Project.

Baum, M. A., & Zhukov, Y. M. (2015). Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war. *Journal of Peace Research*, *52*(3), 384–400. https://doi.org/10.1177/0022343314554791

Buhaug, H. (2006). Relative Capability and Rebel Objective in Civil War. *Journal of Peace Research*, *43*(6), 691–708.

Croicu, M., & Eck, K. (2022). Reporting of non-fatal conflict events. *International Interactions*, *48*(3), 450–470. https://doi.org/10.1080/03050629.2022.2044325

Croicu, M., & Kreutz, J. (2017). Communication Technology and Reports on Political Violence: Cross-National Evidence Using African Events Data. *Political Research Quarterly*, *70*(1), 19–31. https://doi.org/10.1177/1065912916670272

Davenport, C., & 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. https://doi.org/10.1177/0022002702046003005

Davies, S., Engström, G., Pettersson, T., & Öberg, M. (2024). Organized violence 1989–2023, and the prevalence of organized crime groups. *Journal of Peace Research*, *61*(4), 673–693. https://doi.org/10.1177/00223433241262912

Dawkins, S. (2021). The problem of the missing dead. *Journal of Peace Research*, *58*(5), 1098–1116. https://doi.org/10.1177/0022343320962159

Dawkins, S. (2022). Corrigendum for Dawkins S. The problem of the missing dead. *Journal of Peace Research*, *58*(5), 1098–1116. https://doi.org/10.1177/00223433221103919

Demarest, L., & Langer, A. (2018). The study of violence and social unrest in Africa: A comparative analysis of three conflict event datasets. *African Affairs*, *117*(467), 310–325. https://doi.org/10.1093/afraf/ady003

Demarest, L., & Langer, A. (2022). How Events Enter (or Not) Data Sets: The Pitfalls and Guidelines of Using Newspapers in the Study of Conflict. *Sociological Methods & Research*, *51*(2), 632–666. https://doi.org/10.1177/0049124119882453

Dietrich, N., & Eck, K. (2020). Known unknowns: Media bias in the reporting of political violence. *International Interactions*, *46*(6), 1043–1060. https://doi.org/10.1080/03050629.2020.1814758

Eck, K., & Hultman, L. (2007). One-Sided Violence Against Civilians in War: Insights from New Fatality Data. *Journal of Peace Research*, *44*(2), 233–246. https://doi.org/10.1177/0022343307075124

Economist. (2023, April 17). The world's deadliest war last year wasn't in Ukraine. *The Economist*. https://www.economist.com/international/2023/04/17/the-worlds-deadliest-war-last-year-wasnt-in-ukraine

Galtung, J., & Ruge, M. H. (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), 64–90. https://doi.org/10.1177/002234336500200104

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research*, *39*(5), 615–637. https://doi.org/10.1177/0022343302039005007

Heldt, B., Melander, E., Wallensteen, P., & Axell, K. (1993). *States in Armed Conflict 1992*. Uppsala: Department of Peace and Conflict Research, Uppsala University. https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-32935

Högbladh, S. (2022). *UCDP Georeferenced Event Dataset Codebook* (Version 22.1) [Dataset].

ICRC. (2008). *How is the Term "Armed Conflict" Defined in International Humanitarian Law?* (March 2008; International Committee of the Red Cross (ICRC) Opinion Paper, p. 5). International Committen of the Red Cross.

https://web.archive.org/web/20181101193920/https://www.icrc.org/eng/assets/files/other/opinion-paper-armed-conflict.pdf

Maekawa, W. (2019). External Supporters and Negotiated Settlement: Political Bargaining in Solving Governmental Incompatibility. *The Journal of Conflict Resolution*, *63*(3), 672–699.

Öberg, M., & Sollenberg, M. (2011). Gathering Conflict Information Using News Resources. In K. Höglund & M. Öberg (Eds.), *Understanding Peace Research: Methods and Challenges*. Routledge.

Pettersson, T. (2022a). *UCDP Battle-related Deaths Dataset Codebook* (Version 22.1) [Dataset].

Pettersson, T. (2022b). *UCDP Non-state Conflict Codebook* (Version 22.1) [Dataset].

Pettersson, T. (2022c). *UCDP One-sided Violence Codebook* (Version 22.1) [Dataset].

Pettersson, T. (2022d). *UCDP/PRIO Armed Conflict Dataset Codebook* (Version 22.1) [Dataset].

Price, M., & Ball, P. (2015). Selection bias and the statistical patterns of mortality in conflict. *Statistical Journal of the IAOS*, *31*(2), 263–272. https://doi.org/10.3233/sji-150899

Raleigh, C. (2012). Violence Against Civilians: A Disaggregated Analysis. *International Interactions*, *38*(4), 462–481. https://doi.org/10.1080/03050629.2012.697049

Raleigh, C., & Hegre, H. (2005). *Introducing ACLED: An Armed Conflict Location and Event Dataset*. Disaggregating the Study of Civil War and Transnational Violence, University of California Institute of Global Conflict and Cooperation, San Diego, CA,. https://www.files.ethz.ch/isn/38143/2005_Introducing%20ACLED.pdf

Raleigh, C., Kishi, R., & Linke, A. (2023). Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications*, *10*(1), Article 1. https://doi.org/10.1057/s41599-023-01559-4

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, *47*(5), 651–660. https://doi.org/10.1177/0022343310378914

Singer, J. D., & Small, M. (1972). *The Wages of War, 1816-1965: A Statistical Handbook*. Wiley.

Sundberg, R., Eck, K., & Kreutz, J. (2012). Introducing the UCDP Non-State Conflict Dataset. *Journal of Peace Research*, *49*(2), 351–362. https://doi.org/10.1177/0022343311431598

Sundberg, R., & Harbom, L. (2011). Systematic Data Collection: Experiences from the Uppsala Conflict Data Program. In K. Högrlund & M. Öberg (Eds.), *Understanding Peace Research: Methords and Challenges*. Routledge.

Sundberg, R., & Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*, *50*(4), 523–532. https://doi.org/10.1177/0022343313484347

Vasquez, J. A. (2009). *The War Puzzle Revisited*. Cambridge University Press. http://ebookcentral.proquest.com/lib/uu/detail.action?docID=452039

Weidmann, N. B. (2015). On the Accuracy of Media-based Conflict Event Data. *Journal of Conflict Resolution*, *59*(6), 1129–1149. https://doi.org/10.1177/0022002714530431

Weidmann, N. B. (2016). A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science*, *60*(1), 206–218. https://doi.org/10.1111/ajps.12196

---

[1] I am grateful for comments and suggestions on earlier versions of this text from Kristine Eck, Mert Can Yilmaz, Stina Högbladh, Lotta Themnér, Therése Pettersson, Margareta Sollenberg, Garoun Engström, and Erik Melander.
[2] UCDP definitions can also be found at https://www.pcr.uu.se/research/ucdp/definitions/
[3] This understanding of war, or armed conflict, is also evident in international law (see e.g. ICRC, 2008)
[4] For a more elaborate discussion battle-related deaths see https://www.pcr.uu.se/research/ucdp/definitions/#tocjump_8688934708187298_5
[5] I suppose what they meant to say was systematic bias, rather than systemic bias.
[6] There is a whole field of science devoted to consistency and precision in measurement – metrology.
[7] The ratio of non-violent to violent events may have improved since 2012, but I am unable to find out since I am unwilling to sign the user agreement required to download their data.
[8] https://www.pcr.uu.se/research/ucdp/methodology/
[9] Syrian Observatory for Human Rights, Violations Documentation Center, AirWars, Borderland Beat, Kivu Security Tracker, Crisis Watch, Reduction in Violence Monitoring, South Asian Terrorism Portal
[10] The Irrawaddy, Myanmar Now, Ukrinform, Khaama Press, Bakhtar News Agency, Pajhwok News
[11] AFP, Reuters, Xinhua, ITAR-TASS, NYTimes.com Feed
[12] All Africa, BBC Monitoring
[13] El Sol de Tijuana
[14] Ministry of Defense, Afghanistan

[15] The one exception is the Shaam News Network – which we do read, but did not happen to use for any event in 2017.
[16] https://www.pcr.uu.se/research/ucdp/definitions/
[17] This last point is based on RKL's statement that ACLED's 6739 civilian fatalities make up 81% of Mexico's annual violent fatalities, implying that all violent fatalities registered by ACLED in Mexico in 2021 comes to around 8320 (6739/.81) which is less than half of all fatalities registered in UCDP GED.
[18] To view the data, go to www.ucdp.uu.se and type Pokot in the search bar.
[19] The scheme used for translating ACLED to UCDP GED categories is also dubious in other ways, notably concerning civilian casualties, where RKL repeats the mistake they made in the Mexico case.
[20] The original submission included a critique of ACLED's user agreement which at the time prohibited users from "using ACLED's data or analysis in any manner that may harm, target, oppress, or defame ACLED" and "using ACLED's data or analysis in a manner that causes damage to the legitimate interests of ACLED" (ACLED, n.d.). I thought this put ACLED outside the scientific domain, especially since ACLED previously also enforced these conditions with authors having to write corrections satisfying ACLED (e.g. Dawkins, 2021, 2022), effectively preventing serious critique if it harmed ACLED's interests. I dropped this critique here because ACLED changed their user agreements in October 2023 softening the language somewhat.