





ARTICLE



<https://doi.org/10.1057/s41599-023-01559-4>

OPEN

Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices

Clionadh Raleigh ^{1,2✉}, Roudabeh Kishi ³ & Andrew Linke^{4,5}

Conflict event datasets are used widely in academic, policymaking, and public spheres. Accounting for political violence across the world requires detailing conflict types, agents, characteristics, and source information. The public and policymaking communities may underestimate the impact of data collection decisions across global, real-time conflict event datasets. Here, we consider four widely used public datasets with global coverage and demonstrate how they differ by definitions of conflict, and which aspects of the information-sourcing processes they prioritize. First, we identify considerable disparities between automated conflict coding projects and researcher-led projects, largely resulting from few inclusion barriers and no data oversight. Second, we compare researcher-led datasets in greater detail. At the crux of their differences is whether a dataset prioritizes and mandates internal reliability by imposing initial conflict definitions on present events, or whether a dataset's agenda is to capture an externally valid and comprehensive assessment of present violence patterns. Prioritizing reliability privileges specific forms of violence, despite the possibility that other forms actually occur; and leads to reliance on international and English-language information sources. Privileging validity requires a wide definition of political violence forms, and requires diverse, multi-lingual, and local sources. These conceptual, coding, and sourcing variations have significant implications for the use of these data in academic analysis and for practitioner responses to crisis and instability. These foundational differences mean that answers to “which country is most violent?”; “where are civilians most at risk?”; and “is the frequency of conflict increasing or decreasing?” vary according to datasets all purporting to capture the same phenomena of political violence.

¹University of Sussex, Global Studies, Brighton, United Kingdom. ²ACLED, Remote, USA. ³Princeton University, Bridging Divides Initiative, Princeton, NJ, USA. ⁴University of Utah, Salt Lake City, UT, USA. ⁵Present address: ACLED, Remote, USA. ✉email: clionadh.sussex@gmail.com

Introduction

“The science of statistics is the chief instrumentality through which the progress of civilization is now measured, and by which its development hereafter will be largely controlled.”

Which country is most violent? Where are civilians most at risk? Is the frequency of conflict increasing or decreasing? Answers to these questions depend on which conflict data are used to identify patterns. Public, open-source conflict datasets source, curate, and code violence differently, and the ramifications of these discrepancies are violence assessments that cannot be used interchangeably or compared directly. How we conceptualize conflict has drastic implications for how we measure information about political violence, and how these measures can be reliably used as valid assessments of conflict risk and severity.

Identifying levels of violence affecting communities has significant ramifications for humanitarian response, government policy, and academic research into patterns of conflict onset and duration (McGoldrick, 2015). Analysts in disciplines as broad as political science, geography, development economics, public health, education, environmental politics, ecological conservation, public policy, and medical sciences rely upon institutional conflict data collection programs for their research, and their conclusions have ramifications for a country's instability and development prospects (see for example, Thies and Baum, 2020; Ray and Esteban, 2017; Ekzayez et al., 2021). Assessments of how conflict causes civilian fatalities and injuries, undermines government institutions, and harms socioeconomic development, are the basis for international financial aid, disaster response, and political or diplomatic support.

If these conflict data and their differences have real-world implications, the reasons for those variations should be thoroughly interrogated. This article is about how we know, observe, and measure violent phenomena, and the practice of data collection for conflict specifically. The principles of epistemology—how we know what we know—are beyond the scope of this article, but we consider their application in recording and curating quantitative conflict event data, a relatively recent scholarly endeavor. We position our study in some foundational debates about the ascendance of statistics in areas of public policy and administration (Levy, 2001, on health) and whether the generalizations available through statistics outweigh the costs of simplifying complex social forces and political dynamics, like armed conflict.

Consider violence in Mexico: 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 battles¹—6739 civilian fatalities were documented that year,² 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).³ This stark difference is due to the integral role that unidentified or anonymous armed agents and gangs play in committing this violence, at the behest of cartels,⁴ with such groups remaining unnamed for several reasons, including the sheer number of sub-gangs that carry out contract civilian killings (Chaparro, 2021), and a reluctance amongst journalists to name groups due to the deadly costs of doing so (Dorff et al., 2022).⁵ If a dataset does not allow for violence to be captured if it is perpetrated by intentionally ‘unnamed, organized, armed groups,’ then such

violence is unaccounted for—resulting in starkly different representations of the same landscape. This can open the door to official deflections, such as the claim “*yo tengo otros datos*” (“I have other data”) by Mexico's President Andrés Manuel López Obrador when questioned about his administration's lack of response to the rising homicide rate during his term (Arista, 2021).

Conflict assessments from Mexico provide vastly different interpretations of the threat populations face, and these discrepancies result from how both datasets define political violence. Conflict event data⁶ are the main medium for understanding stability in the international system, but how datasets create and decide on collection parameters has not been previously interrogated,⁷ and the field lacks standards for reliability, validity, and comprehensiveness.⁸ In our empirical review, we consider how four public, regularly updated, global datasets define and track armed, organized political violence. We offer a critical reflection on conflict datasets' conceptual and collection practices and subsequent analyses. We present a comprehensive evaluation of conflict datasets that is valuable for people gathering and using these data, asking, “What does and doesn't count as conflict?” and “Why do data projects aiming to collect similar information fail to agree about conflict risks?” We locate the disparities across similar data projects in their conception of conflict, the measurement parameters for those concepts, and the practices of gathering information to reinforce both those concepts and measurements. The complexity of conflict zones and constantly evolving media environments makes achieving both perfect reliability and validity—concepts we define in detail below—impossible for conflict datasets, leading data providers to prioritize one over the other. We consider the variations in concept and measures in order to detail the tradeoffs in data-generating processes and to observe under which circumstances data projects obscure conflicts and their characteristics. In this way, our study contributes to the ongoing discussion around the validity and reliability of the use of such data in political science research (for example, Danziger, 1975; Franzosi, 1987), and in the study of conflict in particular, which has also drawn from similar debates in the social movement (for example, Earl et al., 2004; Hutter, 2014) and communications studies (Galtung and Ruge, 1965; Krippendorff, 2013; Harcup and O'Neill, 2016) fields.

There are several implications of this review and interrogation: in academic research and public practice investigating conflict patterns, early warning, and response, reliable and valid information about political violence is necessary. Indeed, ‘evidence based’ studies and programming are common features in international organizations and applied research. Further, this stage of public conflict data requires a reflection on standards in collecting and publicizing information on violence patterns and the ways in which these data are appropriately and correctly used, or misused. Finally, in addition to the development of standards of collection and evidence quality, 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. As we illustrate, varied representations of countries and conflicts in different datasets can have significant ramifications for how decisions are made by governments, humanitarian organizations, and multilateral institutions about who receives support, but also which policies come to define ‘conflict prevention’ measures within unstable states. Using data that are incompatible with events on the ground to inform policies can result in cycles of ongoing conflict.

Conflict datasets: conceptual and procedural overlaps and differences

We compare four datasets that collect information 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. All four datasets are public, open-source, and event-based, with a global remit and frequent updates,⁹ and all four use generally similar definitions of conflict as their organizing principles.¹⁰ We include the aforementioned ACLED and UCDP-GED, along with the Integrated Crisis Early Warning System (ICEWS), and the Global Database of Events, Language, and Tone (GDELT). By comparing datasets with public mandates, similar global scope and intentions, and a high volume of events, we expect some overlap. Indeed, overlap expectations are so common as to underscore several efforts to forcibly ‘join’ these data into larger, aggregated datasets (see, for example, Donnay et al., 2019); or common requests to perform replication studies on different conflict data for academic publications (see, for example, Berman et al., 2017).¹¹ However, content overlap expectations are exaggerated, and we document this incongruence and also embark on a comprehensive examination of *why* the datasets portray varying conflict patterns to users. Further, we only include from each of the four datasets the event types within each that would—by their own definitions—overlap with a common ‘political violence’ catchment.¹² These four datasets are identified in Table 1 alongside a description of their data-generating process, conflict catchment, and source catchment.

Additional conflict data are available but not considered here, as they may be infrequently updated, such as the Militarized Interstate Disputes dataset (MID; Maoz et al., 2018), Correlates of War (COW; Sarkees and Wayman, 2010), or the Conflict Barometer (Heidelberg Institute for International Conflict Research, 2021); or other datasets may focus only on specific types of conflict or violence, such as the Global Terrorism Database (GTD; LaFree and Dugan, 2007); or specific states or regions, such as the Yemen Data Project (YDP, 2022), Afghan Peace Watch (launched in 2020), or the Deep South Watch Database (DSW, 2004). Often more focused datasets will collect a range of event details that are deeply sourced, and can be integrated into the larger datasets.¹³

Following the presentation of the conceptual overlaps and differences across the data projects in the comparison above, in the following section we illustrate how variation across these attributes contributes to starkly varying conflict and violence patterns.

A comparison and empirical review of conflict event datasets

A critically understudied issue in conflict data and analysis is the degree to which similar data collection mandates and practices are affected by the choices of the dataset creators. These choices of conceptual prioritization and representation, frequency of release, and information source catchments have drastic implications for whether global conflict trends are accurately represented (for example, see Miller et al., 2022). Observing each dataset’s collection process—what we call ‘conflict catchment’ and ‘source catchment’ (see below)—we compare projects, allowing us to move forward with the establishment of standards of collection and transparency in conflict data use. Those standards can be summarized as a dataset’s reliability and validity, which in turn affect its comprehensiveness in terms of its corpus and sources of information.

We use two main concepts in our comparisons. ‘Internal Reliability’ is defined as the consistency of event information within the dataset over time and area. ‘External Validity’ is understood to be an assessment of how accurate event information is, and how well the collection of events represents political violence occurrence trends on the ground. Reliability concerns how well the data match internal project visions and definitions of conflict; and validity describes whether the dataset is effectively measuring what it purports to measure in a regional context. In turn, the range of collected conflict forms and sources, represented here by conflict and sourcing catchments, relays how comprehensive and thorough data projects are. Decisions made about the data generation process, which conflicts catchment to pursue and the breadth and depth of the sources result in a direct and indirect prioritization by datasets of reliability and validity. These choices create distinct dataset analysis outcomes and conclusions.

Reliability and validity are concepts regularly used in the evaluation of research quality (for example, see Drost, 2011). Ideally, a measure should strive to be both highly reliable and valid, though measures can be one, both, or neither. For example, a reliable but invalid measure collects select information that conforms to a set of project priorities: if a conflict dataset’s information sources are static and unchanging over time, or collected in one or limited languages, these practices prioritize ease of collection over accurately measuring violent trends on the ground. This creates a dataset of events according to specific media consistency rather than events that actually occurred. This is a pressing problem because neither information sources nor specific media are consistent. Media, especially foreign media, regularly run stories based on reader interest rather than the occurrence of conflict (Tai and Chang, 2002). Media patterns are also affected by internal workplace considerations, leading to a dearth of foreign conflict reporting in December because English-language journalists tend to be away on holiday leave (Raleigh and Kishi, 2019); and information environments change, influencing the ability of journalists to report in a consistent manner (see Dorff et al., 2022).

If the information sourcing has constant reliability because the sources do not change, but the conflict catchment is invalid, conflicts can ‘disappear’ by excluding entire forms of violence that change type or direction over time. Limited definitions that prioritize predefined contests (e.g. state vs. non-state), specific motivations (e.g. ‘terrorism’), group types, and events with high fatality levels may be reliable in capturing a single trend, but not valid because they do not accurately capture real-world conflict trends comprehensively. For some datasets whose mandate extends only to specifically defined catchments (e.g. the Global Terrorism Dataset), and do not purport to collect widely or for a full range of conflict, it is vital to prioritize a narrow, reliable strategy. For others with 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—as in the Mexico example above. Finally, an overly wide conflict catchment can ‘create’ conflicts through excessive and unregulated inclusion; these procedures also damage reliability and validity. Some examples of this are presented in the following section.

Conflict data creators choose one of three paths to reflect their priorities. The first path is to produce a dataset reflecting a wide range of event forms that are an accurate interpretation of political violence occurring across the world at any given time (privileging validity). The second is where a dataset produces a list of events that represent a continuum of how that data provider first defined conflict, which produces a static definition of violence and information from the past into the present (opting for

Table 1 Comparison of datasets based on information collection practices.

Dataset	Data-generating process	Conflict catchment	Source catchment
<i>ACLED</i> : Armed Conflict Location & Event Data Project (Raleigh et al., 2010)	<ul style="list-style-type: none"> • Researcher-led project • Remote team of over 200 based around the world • Inter-coder, intra-coder, inter-code, and inter-source reliability checked weekly • Process, methodology, and sources are public • Global coverage with data updated weekly • Coverage on select regions starts at 1997 • Over 1.5 million events 	<ul style="list-style-type: none"> • Wide inclusion catchment, collecting information on political violence, demonstrations, and non-violent strategic developments around the world via 6 event types, further disaggregated into sub-event types, where there is at least one armed, organized group engaging in extrajudicial and/or other forms of conflict. • Main events types include battles, violence against civilians, remote violence (e.g. bombings, etc.); rioting, including mob violence; protests; and strategic developments. Event subtypes specify forms within broader event types (e.g. battles with territorial gains; abductions; etc.) • Events can involve any combination of governments, rebels, militias, or other armed, organized groups with a political purpose/ effect, as well as civilians 	<ul style="list-style-type: none"> • Every event publicly associated with its source(s) • Multiple types and scales of sourcing, including both traditional and new media, and multiple local observatories • Over 13,600 sources in over 100 languages—including thousands of local and subnational sources, as well as partnerships with over 50 local conflict observatories—with local sources prioritized • New sources added after careful consideration and back-coding • Unique source lists based on each unique context • With expansive and comprehensive criteria and an information environment that is both changing and improving, conflicts constantly appear
<i>UCDP-GED</i> : Uppsala Conflict Data Project Georeferenced Event Data (Sundberg and Melander, 2013)	<ul style="list-style-type: none"> • Researcher-led project • Team of 10–15 based in Sweden • Process as desk research conducted in Sweden determines whether a conflict has met the threshold conditions for inclusion that maximize the reliability of the definitions • Global coverage with data updated annually • Coverage starts at 1989 • Over 293,000 events 	<ul style="list-style-type: none"> • Conflicts events are categorized as state-based, non-state-based, or one-sided • Conflicts are determined first based on whether a country has a conflict; then populated with events if the conflict demonstrates that an ‘active’ dyad (which must involve a combination of states, non-state actors, or civilians) have a ‘stated incompatibility’ (territorial or government), and that specific dyad reached a 25 battle-related (per annum) fatality threshold. • An event must result in at least one direct death at a specific location and date for inclusion 	<ul style="list-style-type: none"> • Emphasis on secondary/tertiary access to traditional international and national media found in media aggregators (e.g. Lexis Nexis) • Unknown number of sources, though fewer than 7000, including some non-English sources, based on media aggregators • Select reports from international organizations and national sources • Static source list with limited inclusion of new sources, unknown backcoding (some variance by context). Backdated information is included if the evidence is uncovered that conflicts reached fatality thresholds in the previous period. • Events reflect how a conflict has both adhered to the previously determined definitions and how it was reported in (often) externally based media
<i>ICEWS</i> : Integrated Crisis Early Warning System (Boschee et al., 2013, 2018)	<ul style="list-style-type: none"> • Automated collection process, using natural language processing (NLP) and a library of key terms to search internet media • Collection of ‘political events’ • Coverage starts at 1995 • Over 19 million events 	<ul style="list-style-type: none"> • Definitions of violence and conflict are determined by keywords and phrases found in internet searches; reports that contain search terms are categorized according to the CAMEO event system, and intensity is based on Goldstein classifications. The CAMEO codebook highlights over 15,000 phrases that form the backbone of the data (Schrodt, 2012)—e.g. ‘consult,’ ‘fight,’ ‘reduce relations,’ ‘express intent to cooperate’ • 300 different types of political events—though a fraction of which capture ‘conflict’ • Events are reports and reports are events 	<ul style="list-style-type: none"> • Sourcing from 6000 traditional media sources, in 4 languages (English, Spanish, Portuguese, and Arabic) from media aggregators (Factiva and the Open Source Enterprise) • Predetermined source list • Events are reports and reports are events. While “duplicate stories from the same publisher, same headline, and same date are generally omitted ... duplicate stories across multiple publishers (as is the case in syndicated new stories) are permitted” (Boschee et al., 2018). • No known review of sources or events

Table 1 (continued)

Dataset	Data-generating process	Conflict catchment	Source catchment
GDELT: Global Database of Events, Language, and Tone (Leetaru and Schrodt, 2013)	<ul style="list-style-type: none"> Automated collection using a library of key terms, the internet and predetermined scores Tracks human societal-scale behavior, capturing events, context, and tone Coverage starts at 1979 Global coverage with data updated every 15 min Over 250 million events 	<ul style="list-style-type: none"> Definitions of violence and conflict determined by keywords and phrases found in internet searches (e.g. 'clash,' 'dispute'), with categorizations from the Conflict and Mediation Event Observations (CAMEO) coding scheme and the Goldstein scale to gauge intensity (Schrodt, 2012) (Berman et al. (2017) conduct analysis in their study using ACLED data, yet must replicate their findings using data from UCDP-GED, found in an online appendix.) 300 different types of political events—though a fraction of which capture 'conflict' Reports with keyword/phrases become included events 	<ul style="list-style-type: none"> Reliant on keyword searches in traditional media reports No known review of sources or events (GDELT, 2021) Sourcing, including hundreds of thousands of traditional media stories, translated from over 100 languages Dynamic source, changes daily, without oversight

ACLED also includes gang violence, though only when it is used towards meeting overt political goals, as well as in cases where gang activity directly and fundamentally challenges public safety and security, in turn challenging the state's monopoly on power (ACLED, 2020a).

Demonstration and strategic development events are not considered in comparison cases in order to concentrate on the data variation collections between armed, organized groups engaging in political violence.

This was calculated based on the number of distinct sources named in data files.

For more on how and when new sources are added to ACLED coverage, see ACLED's 'Adding New Sources to ACLED Coverage' primer (ACLED, 2021).

This was calculated based on the number of distinct names of sources in the data file. However, it must be noted that *this number is inflation*: UCDP-GED does not standardize its source names so multiple spellings or abbreviations of the same source (e.g. 'Xinhua News Agency,' 'Xinhua,' 'X,' 'Xihua,' 'Xinhu,' 'Xinhu,' 'Xinhu,' to name a few) will be over-counted as more than one distinct source when calculating based on distinct names of sources in the data file (e.g. Xinhua News Agency double-counted as 6 distinct sources when the distinct names all refer to a single source). A more accurate count is not possible without a review of each unique event and meticulous re-coding—beyond the scope of this comparison table.

Launched in 2008 as a Defense Advanced Research Projects Agency (DARPA) program (of the United States Department of Defense), ICEWS is now maintained by Lockheed Martin, a large defense contractor. ICEWS data are publicly available on the Harvard Dataverse beginning for the period 1995 to the present, published on a weekly basis—though some of their event data are available 'For Official Use Only' (FOUO) and government-sponsored research activities (Bosch et al., 2018).

The codebook for CAMEO (Gerner et al., 2008) includes over 300 types of events of both cooperation and conflict. Each event is characterized by its type rather than its circumstances. These categorizations mean that a great deal of non-conflict events is captured as active conflict, and a great deal of actual conflict can also be missed. Despite the wide variation in event types captured by the CAMEO/Goldstein scales, many articles are miscategorized, and these mistakes are consistently in the direction of excessively violent designations: there is an over-inflation of '–10' classifications, which is the most serious form of conflict (for examples, see Raleigh and Kishi, 2019).

The Open Source Enterprise is a United States Government organization dedicated to open-source intelligence.

internal reliability). The final path is where a dataset trawls the internet for select terms or phrases and creates a dataset of conflictual events (here there is little preference for either reliability or validity). Each approach has advantages and disadvantages, but ultimately delivers portrayals of conflict that vary dramatically. The path of choice is partially determined by origin features: conflict event data collection and availability have grown substantially in just over a decade. Datasets that were designed with country-year or conflict-year as units of the collection have had to shift focus to become event-based. This has involved decisions of compromise between original and novel features (such as disaggregation scale and definitions) that other event-based datasets have not had to grapple with.

Data generation processes (researcher-led vs. automated projects), conflict catchment (i.e. inclusion criteria), and source catchment decisions (i.e. considerations of source diversity, language, number, scale, etc.) are what impact the external reliability, internal validity, and comprehensiveness of datasets. These factors are discussed in further detail below, with examples explored to demonstrate the impacts of different decisions.

Comparing data-generating processes. Characteristics of the data-generating process are fundamental and cannot be corrected in any way by data users *post hoc*.¹⁴ We identify benefits and tradeoffs associated with automated vs. researcher-led data projects and find that automated-coding designs continue to have vast disadvantages, despite hopes about their utility to come.

Both GDELT and ICEWS are automated datasets, and while they differ with respect to what they broadly seek to collect and how frequently, they both use natural language processing for their catchment collection and the CAMEO coding system for flagging

events of interest in the source information. CAMEO (Gerner et al., 2008) includes over 300 types of events of both cooperation and conflict, where each event is characterized by its type rather than its circumstances—though we only compare those events that fit within the 'common catchment' being explored in this study, as defined earlier (i.e. those events that would be considered to be 'political violence' or conflict, not cooperation). Both datasets also use the Goldstein Scale for the intensity of the reported event's characteristics, although previous reviews have indicated that events are frequently miscategorized with bias in the direction of excessively violent designations: there is an over-inflation of '–10' classifications, which is the most serious form of conflict (for examples, see Raleigh and Kishi, 2019).

Inclusion criteria. The remit of automated projects is to find keywords or phrases and categorize all as events, regardless of whether they are or are not conflict incidents. The inclusion criteria for the conflict catchment of machine automated datasets is therefore *the search itself*. Outside of the initial establishment of terms, there is no intervention or review by researchers, and word choice (pre-defined CAMEO terminology) within media reports is the *de facto* methodology¹⁵ and inclusion criteria. The event collection is not guided by concepts and strict standards, and the catchment is, in practice, a collection of news reports on conflict-related and adjacent words. Therefore, automated datasets are extremely unlikely to have validity in their conflict catchment.

Lack of oversight. The benefits of fully automated data projects are their speed and low labor cost, and hence perceived efficiency. In contrast, researcher-led projects have relatively higher costs, both financially (e.g. paying staff) and regarding time and nonmaterial

resources (e.g. training and managing researchers, etc.). The tradeoff for efficiency in an automated data collection effort is a lack of oversight. Automated data projects can return millions of events in a month. Such vast volumes of data resulting from duplicate, irrelevant, and inaccurately coded events that are undetected due to the lack of formal review procedures. False positives and event inflation are serious problems in automated datasets, leading to distortions of conflict patterns, conflicts appearing where none have occurred, or coding numerous conflict events when only one has occurred. Coding errors are rife, including crime and irrelevant events represented as if they were political conflict (e.g. sports ‘battles’ or slanderous ‘attacks’). It is largely impossible to identify what proportion of events are ‘true’ and what is ‘noise,’ with ‘noise’ comprised of both false positives (i.e. incidents that never happened) and duplicates (i.e. reporting a single event that occurred numerous times)—both are explored in further detail below.

This lack of oversight by automated datasets can also impact other facets of information collection, including the accuracy of geo-locations. In cases when a media report does not note the specific location of an event (e.g. notes an event happened “in Nigeria”), or if automated trawlers are unable to deduce more granular geographic information from the report (e.g. the report notes the name of a small, relatively unknown village in Nigeria), automated datasets like GDELT will default to the centroid (i.e. location center) of the country.

In one case, for example, the popular statistical analysis blog FiveThirtyEight assessed kidnappings in Nigeria, using data from GDELT (Chalabi, 2014a), though were forced to issue an apology for their conclusions shortly thereafter due to the fundamental issues with the data (Chalabi, 2014b). One of these issues was due to an over-inflation of events in the center-most governorate of Nigeria, Federal Capital Territory (Chalabi, 2014b), which resulted in the erroneous conclusion that this region was home to “unusually high numbers of kidnappings relative to their population size”—a finding that FiveThirtyEight was forced to admit was inaccurate (Chalabi, 2014b). Raleigh and Kishi (2019) note similar findings in their review of GDELT’s portrayal of the war in Yemen, finding that GDELT overrepresents the center of the country as a hotspot of the war when, in reality, violence is quite limited in the center of the country—a region of Yemen that is largely unpopulated.

False positives. Trend analysis using data populated by false positives depicts invalid patterns of violence where conflicts that never happened are created out of natural language processing mistakes. For example, in GDELT, thousands of incidents using ‘aerial weapons’ reported in the United States in 2019 were false positives, including reports of an actor falsely reporting on hate crimes (Moon, 2019); a weather report about a high wind warning (OC-Breeze, 2019); and sexual misconduct on crowded airlines (Martin, 2018). In another example, a report titled ‘Cows come home to haunt India’s Modi’ (France24, 2019) describes the state protection of cows in India, given their status under the Hindu religious tenets. And yet, this event is coded as ‘conventional military force’ involving the government, despite this report failing to mention actions by Indian military forces. To be clear, no conflict events occurred in these places or times, but they are reported and coded as conflict events.

Similarly, ICEWS data held that in June 2019, 25 events—all on par with a nuclear conflict (−10 on the CAMEO scale)—occurred in the United States between the US government and Iran, in Washington, DC, with the designation ‘fight with artillery and tanks,’ when in reality the events capture Iran engaging in a ‘war of words’ without any conflict or threats with the US (Havasi, 2019). In another example, ICEWS reports that 10 acts of

‘abduction, hijacking or hostage’ occurred in 2018 in Zimbabwe, such as one event recorded in Chikore, Zimbabwe, on 24 October 2018, involving Air Zimbabwe. But Chikore is not a place in Zimbabwe; nor was anyone associated with this business hijacked or abducted on this day. Mr. Simba Chikore is former President Mugabe’s son-in-law and used corrupt tactics during his tenure as chief operating officer of the carrier (Zindoga, 2019; Taruvinga, 2018). This story was recorded by ICEWS as a ‘−9’ level event on the Goldstein Scale, which is the designation for ‘severe repression’ and ‘assault.’ Most worrying is that false positives, like these examples, are common and easily identifiable, yet remain in these datasets, indicating that no (or quite limited) reviews and corrections are occurring at any stage of the collection (Weller and McCubbins, 2014; OpenNews, 2014).

These are easily identified mistakes, but determining which of these millions of events are true and what proportion are false is largely impossible. When using such datasets, many analysts simply use the data without making considerations for how much inaccurate information is included, including in published academic studies (for example, Mo and Yu, 2022, and in this journal, Greene et al., 2021), and within the world of policy-makers and practitioners, such as the EU (Halkia et al., 2019). This, in turn, creates a narrative of conflicts and crises where none appear and describes characteristics of ongoing conflicts that are false. The presence of false positives in ICEWS and other datasets is a glaring limitation because it creates a baseline that users can only correct with considerable time and effort, if at all. Absent an adjustment that somehow removes duplicates and false positives, analysts risk presenting an artificial trend manufactured by the data project.

Duplicates. Duplicate reports of the same event are rarely removed in automated systems, resulting in vast event inflation (Bond et al. 2003; Schrodt and Van Brackle 2013). Within ICEWS coding, for example, while “duplicate stories from the same publisher, same headline, and same date are generally omitted, ... duplicate stories across multiple publishers (as is the case in syndicated new stories) are permitted” (Boschec et al., 2018), as noted in Table 1. Given how news stories break and develop, get updated, and diffuse to other sources, this flagging system significantly inflates event counts.

GDELT’s duplication-checking indicates the number of repeated references to an event across all source documents and articles, but only during the 15 min window of the event first being reported (GDELT publishes data every 15 min, as outlined in Table 1). Again, given the nature of reporting in the media, stories are regularly updated, and rarely within 15 min. If all subsequent mentions of an event after the first 15 min in which a source reports it does not get flagged as duplicates, the result is a proliferation of duplicates.

Given automated data projects’ limited/lack of oversight, duplicate events are rampant. Compounding this issue is the fact that such duplicates (i.e. errors) are non-random. Certain types of events—for example, high fatality events—will be more susceptible to duplication since such stories are more often updated (as fatality counts are adjusted when more accurate information comes to light, and as more individuals succumb to injuries) and are also more likely to be picked up by other outlets. This propensity to further elevate large, high-fatality events further exacerbates the issue of such incidents being magnified while smaller-scale events get increasingly ignored.

Such was the case in the problematic FiveThirtyEight analysis introduced above, which purported to assess the rising rate of kidnappings in Nigeria, using data from GDELT in the aftermath of the highly publicized mass kidnapping of numerous girls from a school in Chibok, Nigeria (Chalabi, 2014a). In the apology they

were forced to issue, they noted that the conclusions they came to were compromised, due to their inability to distinguish ‘true’ events from ‘noise’ in the dataset (Chalabi, 2014b). In this specific example, given the many conflicting reports about the mass kidnapping of the ‘Chibok girls,’ coupled with the massive media coverage of the incident internationally, the ballooning of media reports had led to GDELT reporting *hundreds* of additional ‘kidnapping events,’ resulting in the author deducing that kidnappings in Nigeria had been escalating *exponentially*—much more significantly than in reality.

‘Fake news’. Further, distinct from event inflation and exaggeration concerns outlined above, there is also the issue of inclusion of false information (i.e. ‘fake news’). Information may be shared online that is purposefully fake (i.e. disinformation), created as propaganda or to mislead readers, or the media may report information after a reporter has been misled (e.g. misinformation).¹⁶ ‘False flag’ events too—i.e. events that have really taken place, yet have been disguised in a way so as to pin the blame on another party—can result in propaganda being taken as fact.¹⁷ Without oversight and contextual knowledge of the collected information, automated datasets may register all such ‘fake news’ as true events. Unverified information too is harmful: the inclusion of events from unverified or group-specific sources without due diligence about who, what, where, when, and how information was gathered is a similar act of inclusion without consideration, and can result in ‘fake news’ entering a dataset. Automated data projects that do not have oversight or systems in place to challenge such issues critically are at risk of integrating such false information into datasets as fact.

The combined outcome of these automated procedures integrates non-random errors, inflation, duplicates, mis-coding, and mis-information of report-based events; and ensures that: no baseline of conflict can be established; no clear proportion of correct and incorrect information can be identified; no comparisons can reliably take place; and no data are expected to be valid. Unlike automated datasets, researcher-led data collections often have a well-defined and articulated methodology and mandate, typically situating collection within a proscribed political context (instability, elections, social movements, etc.). Strict inclusion criteria make information more reliable and valid, and quality assurance through review and oversight ensures higher-quality data emerges from researcher-led projects. Conflict data collection projects that center on particular forms of political violence, however, do not capture all political violence with equal accuracy. The differences begin in how conflicts are defined. Because of the more stringent definitions for information inclusion (i.e. a strict conflict catchment) used by researcher-led datasets relative to automated datasets, as well as the significant data-generating process disadvantages of automated data projects (explored above), the case study comparisons in this section consider only researcher-led projects—i.e. ACLED and UCDP-GED.

Conflict catchment: Inclusion criteria and definitions. Each dataset’s conceptualizations and definitions of violence have direct implications for its data collection practices. The differences between ACLED and UCDP-GED are origin features: the move into global, event-based conflict data, rather than national, annual assessments, involved decisions about conflict definitions that could either prioritize and reflect ongoing conflict environments or generate new rules to fit into predetermined, prior collections. If the priority is to fit events into existing definitions of conflict, this will affect which present conflicts are ‘seen.’ Predetermined, strict inclusion criteria often require and enforce homogenizing standards. UCDP-GED prioritizes the reliability of

its own definitions of traditional intra-state civil conflict, resulting in a more narrowly defined conflict catchment than ACLED. ACLED presents conflict environments as heterogeneous, adaptable, and unstable: events differ by perpetrators, targets, frequency, intent, sequences, and intensity. ACLED prioritizes validity—data as an accurate reflection of reality—of conflict environments as its founding principle.

ACLED collects information on political violence,¹⁸ demonstrations, and non-violent strategic developments around the world via six event types, further disaggregated into sub-event types,¹⁹ where there is at least one armed, organized group engaging in extrajudicial and/or other forms of conflict: an act of political violence or a demonstration is the sole basis for inclusion within the dataset. ACLED classifies all violent actors into one of eight types, including: state forces; rebels; political, pro-government, and community militias; other security forces, including state forces engaging outside of their home country; coalitions and joint operations; peacekeeping missions; unnamed or anonymous armed groups; and spontaneous violent mobs, many with links to political parties or identity groups (e.g. ethnicity, religion, etc.).²⁰ This captures how several types of conflict can occur simultaneously in a state, without groups being linked through a political agenda or organization. Armed, organized agents are constantly appearing in conflict environments, and these groups are especially fractious and diverse.

In UCDP-GED, conflict events are categorized as state-based, non-state-based, or one-sided. First, it must be determined whether a country has a conflict; then that conflict can be populated with 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. An event must result in at least one direct death at a specific location and date for inclusion. UCDP-GED divides actors into the state, non-state and civilian groups. Requiring a ‘stated incompatibility’ privileges the violence of predetermined violent actors, and their agendas.

ACLED prioritizes validity in events as modern conflicts are less likely to look like civil wars (intra-state conflicts) with a typical government-insurgent competition, and more likely to include militia groups with diverse and flexible contestations and motivations. Further, conflicts and agents adapt to political circumstances, taking on new forms, agendas, and operations (see Daly, 2016). Political violence has changed considerably since the end of the Cold War, and now includes a far higher rate of activity by non-state actors against the state and civilians, and higher rates of state actions against civilians and non-state groups who are not seeking to replace the regime (see Stearns, 2022; Choi and Raleigh, 2015; Malu, 2019).

If the threat to states and civilians is rigidly defined, then new actors, factions, and local or small violent groups can easily be discounted, despite their risk to civilians and local government. This is problematic because previous research confirms that most violence against civilians—even in civil war environments—is perpetrated by small, often subsidiary groups (Boyle, 2014; Raleigh, 2012). Militias—non-state, armed, organized, politically violent agents—are widespread, existing in contexts as diverse as India, the United States, and Malawi (Raleigh and Kishi, 2020, 2023a). As a type of actor, these militias are absent in most conflict datasets, creating assessments of violence that are distorted and poor representations of local and national violence threats and potential. Because militias are missing, the resulting conflict tally can be distorted downwards, and the actor composition would reveal far less fragmentation than we know is a reality in most conflict settings. Fragmentation has multiple effects on increasing violence, explaining latent activity, localized instability, and kernels of new movements (Dowd, 2015, 2019).

Table 2 Researcher-led datasets’ conflict catchment strategies and their benefits and tradeoffs.		
Conflict catchment	Benefits	Trade offs
ACLED	Higher external validity: A wide range of conflict types ensures that new forms of violence are not missed	Lower internal reliability: More difficult to ensure consistency over time if catchment includes a wide range of types of conflict that can vary over time
UCDP-GED	Higher internal reliability: Easy to track the same narrow types of conflict consistently over time	Lower external validity: A narrow range of conflict types means that new forms of conflict that may appear will be missed

Who and what is included in a dataset goes beyond whether a conflict data project’s rules can accommodate changing forms of violence, and instead becomes a question of what is seen and unseen. It is true that stricter definitions of conflict can make developing instructions for human researchers less subjective, and can make comparative coding by researchers less error-prone. However, this comes at the cost of capturing vaguer and increasingly complex events (Demarest and Langer, 2022). Persistent low-level political violence is entirely missing in some datasets, for example, because of definition restriction rather than its occurrence being in doubt: Raleigh and Kishi (2019), for example, find that in Madagascar, in 2018, poll workers were attacked and party affiliates were assassinated throughout the election year. No such activity appears in UCDP-GED. Two years later, in 2020, at least four mayors were assassinated in the country, yet these violent events are missing in UCDP-GED. These fall outside their fatality criteria, despite clearly being common and important targeted forms of political violence.

Table 2 identifies broadly the benefits and tradeoffs associated with prioritizing different conflict catchments, specifically in regard to validity vs. reliability. If the political/conflict environment changes, conflict source information monitoring systems must adapt if they are to ensure the validity of new information; if the information/media environment changes, systems that do not evolve will miss important information. That missing source data reduces the validity of the collection. In short, while adapting a collection strategy to prioritize validity does alter the data-generating process, not adapting the strategy, and prioritizing only internal reliability to a predetermined conflict catchment, also results in other biases; *neither option is without bias* (Miller et al., 2022).

The Philippines, for example, is a case where conflict catchment priorities result in dramatically different assessments of violence patterns. The country’s various conflicts demonstrate the differences between datasets that prioritize external validity (e.g. ACLED) and internal reliability (e.g. UCDP-GED) when it comes to conflict catchments. The Philippines currently endures multiple concurrent conflicts: Islamist and communist separatist groups, including the Abu Sayyaf Group (ASG) and the New People’s Army (NPA) (the armed wing of the Communist Party of the Philippines, CPP), are active across the country, and especially so in the southern Mindanao region. Each of the global, public datasets under review captures this trend. ACLED, however, also captures the co-occurrence of wide-scale violence against civilians carried out by the state security apparatus under the guise of ‘targeting drug suspects’ across the country, though centered predominantly in the National Capital Region (i.e. metropolitan Manila) (see Fig. 1). In 2020, ACLED recorded over 1287 events and over 1496 fatalities in the Philippines, with at least 901 civilian fatalities (over 62% of all deaths recorded in the country that year).²¹ An analyst asking ‘Is inclusion of drug war related violence integral?’ should consider the fact that it is the source of most Filipino political violence, specifically that which targets civilians. The International Criminal Court (ICC) has

sought to investigate the state’s role in drug war killings in the country, underlining the significant risk such violence poses to the civilian population (Regencia, 2021) and to journalists.²² Yet, this violence is largely ignored by other datasets: in 2020, UCDP-GED records 157 events and 374 fatalities in the Philippines, including 37 civilian fatalities (<10% of all deaths recorded in the country that year, and approximately 10% of the similar events in the ‘conflict definition catchment’). Drug war targeting in the Philippines is mainly carried out by anti-drug ‘vigilantes,’ who have active links to the Philippine security infrastructure (Kishi and Buenaventura, 2021) and enjoy impunity and anonymity. But because these groups do not seek control of the central government, the perpetrators and violence are not captured by UCDP-GED.

These conceptual differences in conflict definitions create further incorporation barriers for event inclusion. Thresholds that limit inclusion criteria to predetermined incompatibilities, or arbitrary thresholds (e.g. 25 battle deaths per annum), obscure known violent contestation from view. This omission disregards actors because they neither have the structure nor objective that adheres to an antiquated view of conflict agents. Any resulting conflict dataset is incomplete at best, and, at worst, invalid vis-à-vis the widespread violence that is known to have occurred.

UCDP-GED includes events if one actor engages another known group, resulting in at least 25 battle-related deaths in a year. If dyads fail to reach this threshold, no events appear in the dataset. The implications are noteworthy: consider violence by a Kenyan Pokot militia (see Raleigh and Kishi, 2019), who engage widely but minimally, fighting with states and targeting civilians; less than 25% of their activity is with similar militia groups. ACLED codes over 110 fatalities from Pokot militias and Kenyan forces between 1997 and 2020; and 480 fatalities from the groups targeting civilians. In six of those years, ACLED codes over 25 battle-related deaths during the year, which would suggest that there were many chances for this conflict-dyad to be introduced to the UCDP-GED dataset, though it surprisingly was not. Such exclusive and relatively arbitrary coding decisions make the actions of some actors visible while obscuring others. UCDP-GED data users would conclude that Pokot militias have not been involved in violence for over three decades, despite the fact that they have created significant violence during the recent Kenyan elections.

Threshold decisions based on fatality counts are especially unreliable as death counts for individual events are often exaggerated or undercounted, and are consistently incorrect (for example, see Anyadike, 2015). Fatality information is the most biased and least accurate part of any conflict report, and extreme caution should be employed when using any fatality number from any source (Raleigh et al., 2017). All reported fatalities, from all forms of media and partners, are estimates. Some media are better at estimates, and some are worse—and this can vary too by country or conflict, with some regions easier or more difficult to access, or wars making it more difficult to collect accurate death counts altogether (Ball, 2016). Some (e.g. international media)

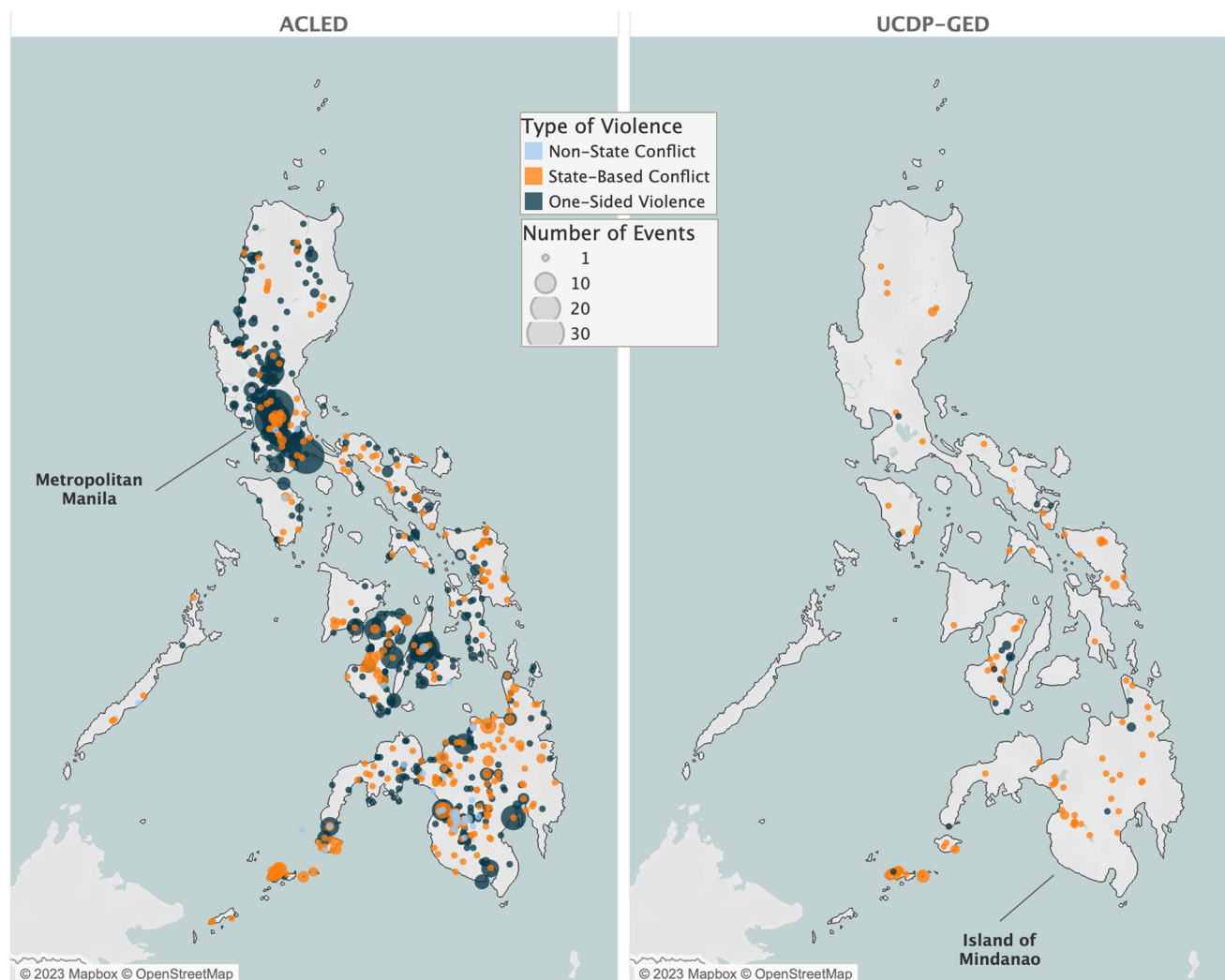


Fig. 1 Violence in the Philippines, 2020, according to ACLED vs. UCDP-GED. The categorizations used on these maps defers to UCDP-GED's categorization of conflict: state-based conflict, non-state conflict, and one-sided violence. ACLED data are mapped onto those categorizations as follows: firstly, determining whether an organized, political violence event involves targeting of civilians (i.e. an interaction code including a 7), and if so, categorizing it as *one-sided violence*; if not, determining whether an organized, political violence event involves the state (i.e. an interaction code including a 1), and if so, categorizing it as *state-based conflict*; and thirdly, determining whether an organized, political violence event involves non-state actors (i.e. an interaction code including a 2, 3, and/or 4), and if so, categorizing it as *non-state conflict*. Comparison of the global, public datasets under review for events captured in the Philippines in 2020. Figure created by authors.

will report on stories only if they reach a certain fatality threshold that elicits audience attention. Other information sources (e.g. in-depth human rights reporting by INGOs) will concentrate on attacks on civilians, regardless of fatalities (Davenport and Ball, 2002). Conflict groups exaggerate their own achievements, while playing down those of their opponents (Crawford, 2016), or may minimize their own violence to minimize international backlash (Shaver and Shapiro, 2021). Such biases must be considered, not only by conflict datasets but also by analysts. The outcome is that relying too heavily on fatalities as a precursor to the collection, as a comparator across conflicts and time, or as a measure of intensity, will be fraught with problems given the lack of consistency/reliability in its reporting across both time and space.

Threshold-based definitions also influence data inclusion in ways that have little to do with event occurrence: consider the implications for battle deaths as an inclusion criterion. If battle deaths decrease, one might assume that conflict, and associated events, have also decreased; or, more broadly, a lower number of battle deaths might suggest more overall peace. Such an argument

suggests that, over time, 'human nature' has changed (see Pinker, 2011). However, battle deaths have decreased as combat medical techniques have improved (Fazal, 2014), which negates any interpretation of battle deaths as indicative of cultural or political norms about the role of violence in society changing. Further, the most frequent attacks directed towards most people in most countries are of low lethality or are non-lethal. For example, in 2021, ACLED records over 93,700 political violence events, more than 89,500 (or over 95%) of which have fewer than 10 fatalities, and over half report no fatalities at all. Violent and destabilizing events are shifting and growing, but these incidents are largely absent from datasets with battle death thresholds.

In limiting what is 'seen,' conflict datasets that should capture similar events fail to do so. This erases the actions and violence of actors in a conflict, even in cases in which they may perform a large and destabilizing role. Despite these reflections on changing conflict environments and inclusion criteria, there should be considerable overlap in more traditional conflicts ('war') that are covered by all conflict datasets. Yet the differences across these

projects illustrate the consequences of adopting different conflict catchments. In short, conflict event datasets that adapt to allow for new forms of violence, and prioritize external validity, better reflect conflict patterns relative to those which defer to predetermined forms of violence and prioritize internal reliability, making conflict collections more linear.

Source catchment: Information on conflict events. Reliability and validity issues often emerge due to limitations in a project's source information. Here, we consider information from any origin that research teams or machines use to code conflict events in the dataset. The corpus of sourcing information for conflict event datasets varies by (1) source diversity (i.e. Are multiple forms of media and information used? Or is information limited to that which is produced by traditional media?), and, relatedly, (2) source languages (i.e. Sources in what languages are used? Are sources available only in local languages considered?); (3) source number and scale (i.e. Are many sources incorporated? Is local information integrated?); and (4) source dynamism (i.e. Are source lists dynamic? Are they updated and changed to reflect newly accessible/available information? Or do they rely on static source lists that do not change, or change very minimally, over time?)

Most datasets prioritize the 'most available' media, rather than the best information. While all conflict event datasets depend to some degree on traditional media forms,²³ some, like ACLED, increasingly engage in some 'new' or social media to supplement coverage. The media and information environments have never been richer, and improvements in access create multiple opportunities to harness better information. Yet, many datasets do not take advantage of new and better data because of their reliance on single or limited languages (e.g. ICEWS); subscription-based publication aggregators, like Factiva (e.g. UCDP-GED, ICEWS); and emphasis on international media sources (e.g. UCDP-GED, ICEWS, GDELT), to their detriment (Demarest and Langer, 2022).

Source diversity. Different types of sources often capture alternative events: local media can return more remote events than international media; international media prioritizes high fatality events relative to local media; local conflict observatories prioritize smaller-scale events (see, for example, Dowd et al., 2020; Day et al., 2015; Davenport and Ball, 2002; Barranco and Wisler, 1999; Baum and Zhukov, 2015; Clarke, 2021; Bueno de Mesquita et al., 2015; Demarest and Langer, 2018; Barron and Sharpe, 2008). Limiting one's sourcing strategy to only a small subset of types or scales of sources, therefore, can result in certain important conflict dynamics being missed. Several studies of traditional and new media suggest that, in combination, both source types cover much of the actively reported events in conflict environments (for example, Dowd et al., 2020). A diverse sourcing strategy—including partnerships, media, and private reports, coupled with a dedication to relaying the extent and form of all political violence—can ensure high validity. This is ACLED's strategy; others choose the integration of secondary sources with limited depth (e.g. UCDP-GED's use of BBC Monitoring or ICEWS' more limited subset of sources).

Information scale diversification strongly impacts the extent of a dataset's coverage. At the national and local level, ongoing conflict can limit correspondents' press freedom, access to zones of fighting, access to communications infrastructure, and the penetration of new media (for example, Kalyvas, 2006; Weidmann, 2016). National media may be integral in one context where there is a robust and free press, yet in a more closed press environment, local conflict observatories and new media sources play a more important, supplemental role. In short,

extensive and accurate information sourcing requires local partners and media in local languages.

Consider, for example, the widely covered Syrian conflict. Figure 2 depicts ACLED coverage of Syria in 2017²⁴ by source type, depicting the differences geographically (across governorates) and by violence type. Traditional media (*left*) is juxtaposed with information from local partners and sources (*right*). Traditional media focuses on the capital, Damascus, the southwest, and western Hama province. Conflict activity occurring in the desert (such as in Homs) where reporters were not present is limited, as is an activity in Al Hasakeh and Deir Ez Zor where Islamic State (IS) sleeper cells were active, targeting both civilians and Syrian Democratic Forces (QSD). Coverage along the M4 road (such as in Ar Raqqa), in the South (especially in Dara), and in the northwest (in Aleppo and Idlib) is also lacking in traditional media. National media best captures explosions and remote violence; local partners capture small-scale battles, violence against civilians, and large-scale arrests by state forces. In short, traditional sources offer a coverage baseline, which must be supplemented through targeted enrichment by sources specifically offering coverage of unique event types, actors, and geographies not captured elsewhere (Raleigh and de Bruijne, 2017).

Source language. Source diversity is not solely about the scale and new media integration, but language diversity. The four datasets explored here produce information in English, and most prioritize English language international media, despite its known limitations (Miller et al., 2022; Demarest and Langer, 2022; Schrodt et al., 2001; Boschee et al., 2013). International, English language media report on different types of violence compared to local media and reporting, and frame conflicts with certain prejudices (Raleigh and Kishi, 2019).²⁵ Second, reporting on peripheral events or events involving smaller groups is likely rejected in favor of high-fatality events, sensationalized events, and those involving large groups with an international profile (e.g., IS) (Galtung and Ruge, 1965; Harcup and O'Neill, 2001, 2016). Third, obvious idiosyncratic biases exist, including distortions in temporal patterns, such as less conflict being reported in December when English-language journalists tend to be away on holiday leave (Raleigh and Kishi, 2019), or foreign correspondents remaining in capital cities, which can contribute to increased urban bias (Weidmann, 2016). Overall, information on smaller, peripheral, nascent, complicated, or less deadly conflicts is poorly covered by national or international sources (Raleigh and de Bruijne, 2017). The narratives of violence from such sources create a stable, internally reliable record of events, but also distort conflict patterns, resulting in lower validity.

UCDP-GED and ICEWS both rely on media aggregators and prioritize English-language reporting at the national and international levels. For example, nearly 60% of UCDP-GED's sources are newswires that compile English-language reports (Högbladh, 2021). Yet, over half (58%) of single-source ACLED events are reported in languages other than English (ACLED does not rely on media aggregators). Further, aggregators that collate local language media, such as BBC Monitoring, make original reports only selectively available. These issues replicate known problems and biases, such as media blackouts (e.g. recent occurrences include Burundi in 2015; or Ethiopia between 2014 and 2018, and more recently in 2021) or poor coverage (e.g. Eritrea, Afghanistan, cartel violence in Central America). Aggregators can be helpful in bringing together numerous disparate sources, but they cannot be relied upon for a complete picture of particularly inaccessible or unstable contexts. Indeed, that is not the intended role of aggregators in the media environment. Those datasets reliant on aggregators and international media can therefore

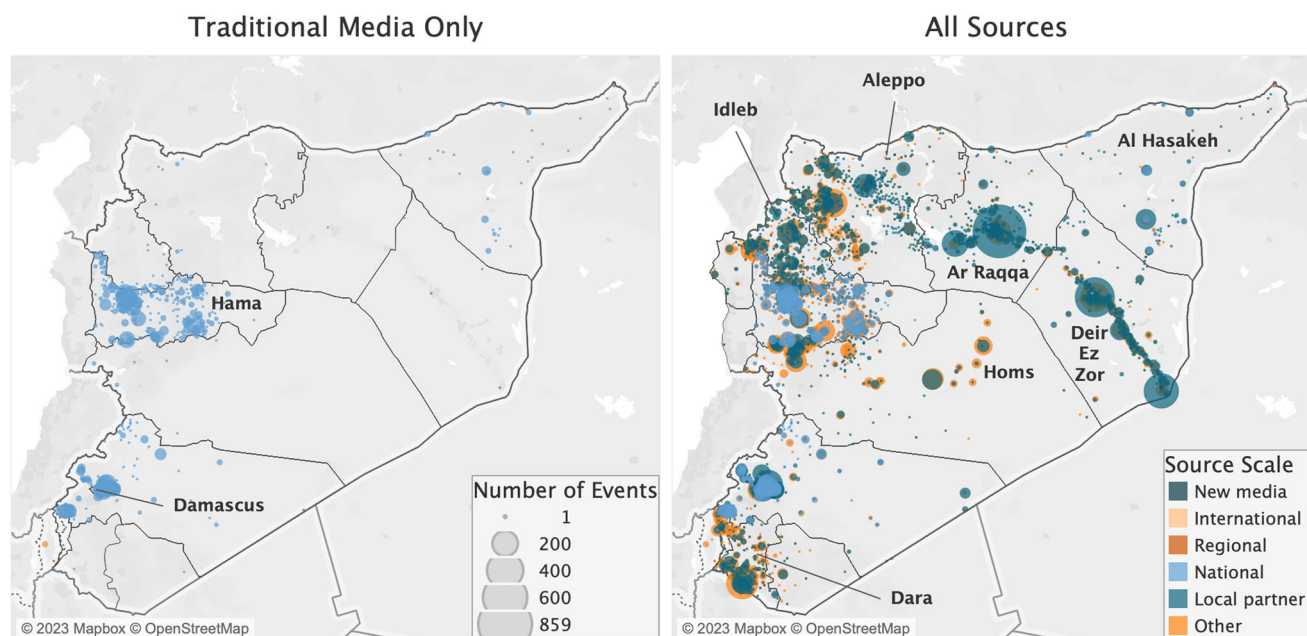


Fig. 2 2017 ACLED coverage of Syria distinguished by traditional media only vs. all source types. The categorization used on these maps relies on ACLED's *source-scale* variable. The raw ACLED data file has been restructured to allow for 'event-source scale' to be the unit of analysis; all 'local partner' source types have been grouped together. 'Traditional media' here refers specifically to international media (e.g. Agence France Presse), regional media (e.g. Israeli media outlets), and national media (e.g. Shaam News Network [Syrian opposition media]); all source types include the aforementioned traditional media types *as well as* new media (e.g. Twitter accounts of trusted sources), information from local partners (e.g. Liveuamap, the Syrian Network for Human Rights, Airwars; for more information, see ACLED's 'Syrian Partner Network'), and other non-traditional sources (e.g. the Syrian Observatory for Human Rights). ACLED coverage of Syria in 2017 by source type, depicting the differences geographically (across governorates) and by violence type. Figure created by authors.

produce a distorted assessment of conflicts (for example, see Bueno de Mesquita et al., 2015).

Source number and scale. There is tension between the number of sources required to record each event in the dataset, the scale of information available, and its veracity (Ortiz et al., 2005). Events reported at national and international levels (and often in English) are more likely to be accurate and frequently cited. Some consider the number of sources to be an indication of veracity (Jenkins and Maher, 2016), but high-intensity incidents do not represent the majority of conflict events. The opposite could be true, however: events harnessed from a single national or international media source are *more likely* to be unverified and questionable due to disinformation (national and international media will repeat known events many times over, rather than have one single report) (Raleigh and Kishi, 2019). Further, datasets that have a deep sourcing net find that in difficult reporting contexts, most unique conflict information comes from local sources, in local languages, and through networks designed to share conflict information that is un- or under-reported in media. These are most often single-source accounts, as the network exists to gather and share this information regionally, rather than to publish. The share of local reporting of single source events in select countries underscores the unique information that can be extracted from a deep sourcing strategy: in ACLED's coverage, 69% of Somali events come from a unique, local source network; in Guatemala, this represents 52% of ACLED's coverage; in Thailand, it is 44% of coverage; in Benin, 43% of coverage comes from such sources; and in Burundi, 36% of coverage comes from a single, local source, to list a few examples.

In short, as countries are unique in both their conflict and media landscapes, and all sources contain some biases (Miller

et al., 2022), the ideal constellation of sources captures how conflict varies across countries, and accounts for local information. Yet, the number of sources alone is not a metric of reliability, and no specific source type guarantees perfectly valid data for all events (Ortiz et al., 2005).

Source dynamism. The width of a sourcing catchment influences the scope of information a dataset might integrate. In countries where conflict reporting is neither consistent nor reliable—such as poorly developed peripheral areas; settings with limited media or government access (e.g. parts of Syria, Somalia, or Afghanistan); or places with high levels of violence against journalists (e.g. parts of the Philippines and Mexico)—adaptable strategies are necessary.

Dynamic or static sourcing refers to whether new information is integrated when available. While using similar sources over a long period of time can increase the internal reliability of data, it damages the external validity and comprehensiveness of data. An inflexibility or failure to manage a rapidly shifting information environment can result in sourcing practices that poorly reflect volatile conflict theaters. Changes to a conflict environment demand that sourcing strategies shift for information to remain valid and accurate, rather than solely reliable. While dynamic sourcing strategies may introduce certain biases because the data-generating process changes within time-series data, avoiding the integration of newly available information introduces other significant biases (Miller et al., 2022).

Source biases and tradeoffs. All sources have biases that have implications for reliability and validity and in turn comprehensiveness. Even the best sourcing strategy and information management cannot omit all biases (Miller et al., 2022). Foundational

concerns remain the tradeoffs necessary to prioritize reliability and validity. For example, if a data project chooses to hold firmly constant the definition of what counts as a conflict—as UCDP-GED does, for example—some degree of temporal reliability can be maximized. However, given that the information environment for singular countries and conflicts over time is *not* consistent (i.e. not reliable, though static boundaries of the conflict catchment assume they are), nor is information and media content under the control of a conflict dataset creator, this creates subsequent problems. The principal concern is that this misalignment of conflict and source catchments reinforces a rigid definition of violence, absent the ability to ensure consistent information about those conflicts. Hence, one may know less and less about the same phenomena over time, or more about one crisis but not about concurrent conflicts.

There are also shared problems of reliability when viewing this issue from a global perspective. The information environment is not consistent over time or space. For example, a conflict may garner significant attention making it possible to secure deep information about its agents and events—as a result increasing the validity of information on that particular crisis. But a neighboring conflict may not elicit the same attention, which could mean that a comparison of the two (or to any other ongoing conflict that also did elicit such reporting coverage) may be biased: one conflict may appear ‘worse’ than the other due to more conflict events being reported, yet the higher number of events may in fact be due to more attention having been paid to that conflict rather than actual activity. Expanding one’s sourcing strategy to capture the best information on a single conflict (e.g. the increased, in-depth reporting in the example above) may hence damage the consistency/reliability of coverage *across* conflicts while bolstering the validity of a *particular* conflict.

To mitigate this bias, one can maximize the information environment for *all* conflicts, despite the time, expense, and management that incurs. This would ensure maximum validity of conflict coverage (i.e. taking advantage of the best information available) while also maximizing reliability/consistency across conflicts (i.e. all conflicts are being compared to one another based on the best available information available on each). Ignoring the latter would mean that analysis and users of the data may confront high event counts in one case, and relatively lower counts in another, for reasons other than differences in conflict patterns. Local partnerships, observatories, and information stringers can mitigate these issues by helping to ensure that in-depth reporting is available for conflicts that do not necessarily capture media attention. Even in cases where media reporting is plentiful, such local sources can play an important role in the triangulation of the fuzzier elements of data collection, such as fatality counts.²⁶ This is the strategy that ACLED employs in curating unique sourcing profiles for each context (i.e. maximizing validity of conflict coverage as well as reliability across conflicts), supplemented by local partnerships.²⁷

For these reasons, datasets must be transparent about the strategies they have used and the effects they may have on both external validity and internal reliability—especially if these vary over time. For example, ACLED’s dynamic sourcing model may help to ensure that conflict coverage validity as well as reliability across conflicts is maximized, but given the changes over time in the information environment more largely, the comprehensiveness of its data collection can only be as good as the quality of information available at a given time. The availability of in-depth information, and the media environment more largely, in 1997 (the start of ACLED coverage for many African countries) is quite different from what that looks like in 2022. However, by that same token, using the same media in 2022 that was available in 1997 would significantly minimize the quality of information

being collected today, if consistency alone is prioritized. This is the reality of collecting event data: for an event to be collected, it must be reported in some way. Maximizing the diversity, scale, and number of sources, and adapting to newly available resources, will return very thorough information, but only if events were meticulously documented and reported in the first place.

Table 3 identifies broadly the benefits and tradeoffs associated with prioritizing different sourcing catchments. The most thorough sourcing strategy results in comprehensive data that prioritizes external validity. That is best achieved by integrating media and information forms, across several scales and prioritizing media and networks in local languages. Information can be triangulated and enriched by supplementing from various sources.

Why does data choice matter?

We began by asking several important questions about political instability, including: Which country is most violent? Where are civilians most at risk? Is the frequency of conflict increasing or decreasing? These questions are important for public awareness and for analysis, but answering them is not straightforward. We demonstrate that the measurement and aggregated trends of conflict by datasets purporting to capture these conflict trends are strongly influenced by the dataset’s (1) definition of conflict, (2) how it may capture the extent of conflict, and (3) from where it seeks information about conflict events. These choices are fundamental to whether the data collected by various global, public datasets are externally valid (emphasizing accurate patterns) or internally reliable (with static collection mandates). Indeed, these three dimensions determine the answer to the questions above, rather than the occurrence of conflict events and deaths.

These dimensions have such significant impacts on how we see the world and yet reflections around such decision-making rarely take place. This is the primary impetus for this article. For example, within academia, a review of articles over the past five years in three of the top political science journals focused on conflict—the *Journal of Conflict Resolution*, the *Journal of Peace Research*, and *Conflict Management and Peace Science*—found that the majority of articles using readily accessible data relied on UCDP datasets, including the UCDP-GED. This can be directly correlated with the lack of attention to the non-lethal consequences of conflict in the peace science literature, and that the majority of studies in these journals over the time period feature the state as a key actor, while only about 1% of articles focused on community groups (e.g. militias, clans, traditional governance structures, nomadic groups, etc.), despite conflict by these groups increasing precipitously over the past two decades.²⁸ These foci reflect UCDP-GED’s conflict catchment and results in the prioritization of certain conflicts or types of political violence to the detriment of others, as examples discussed above (e.g. the Philippines, Madagascar, Kenya) elucidated. This also aggregates conflict forms and practices that are not alike, nor display many similar causes or patterns, rendering the analysis performed across these forms questionable.

Another trend that the review of articles noted above highlights is that only 4% of the articles in the three top political science journals focused on the conflict over the past five years relied on data from automated datasets—suggesting that automated datasets like GDELT and ICEWS have not (yet) penetrated academic debates deeply. This, however, makes it so that those outside of academia who are considering the use of automated data sources struggle to learn about the potential benefits or drawbacks of such data, as they have a limited pool of studies to draw on. This makes conclusions such as, “peace keeping organizations such as

Table 3 Datasets' sourcing catchment strategies & their benefits and tradeoffs.

Source catchment		Benefits	Trade-offs
Source diversity & language	Very diverse	A wide range of source types, including in local languages, increases the likelihood of capturing all available information, both high-profile events as well as smaller-scale events	Integrating different source types to curate unique sourcing profiles for varying contexts can be costly and time-consuming, especially when in different languages, and can make comparisons across contexts more difficult
	Less diverse	Easy and low-cost to access and collect information in English (or a limited number of languages)	Certain types of conflict that (e.g. more remote; smaller scale) may be missed
Source number and scale	Expansive number and scale	Local reporting (which is often in local languages) provides information on incidents that might otherwise be missed; the use of more sources relevant to a specific context can increase the likelihood of capturing information	More difficult to access hyper-local information; review of more sources comes with higher cost (e.g. financial, time, etc.), especially when needing to employ researchers or training models to collect information in numerous non-English languages
	Limited number and scale	Easy to track the same more limited sources consistently over time	High-profile events (e.g. high fatality) or those in urban areas (i.e. easily accessible by foreign correspondents) prioritized over smaller-scale and more remote events
Source dynamism	Dynamic	Ensures that newly-accessible/available information is not missed	More difficult to make comparisons over time
	Static	Simplifies the use of control variables in analysis (i.e. controlling for sources over time)	Newly-accessible/available information is not included; this can result in its own form of bias

[the] United Nations and Red Cross ... can benefit from [GDELT] to make important decisions" (Keertipati et al., 2014) in published academic work worrying.

Calls for policymakers and practitioners to increase their reliance on such datasets seem to have been met—resulting in datasets with problematic trends being integrated into the workflows of such organizations. For example, UNICEF's *Horizon Scan* initiative, which seeks to identify which countries may be facing imminent increases in humanitarian need each month (AI for Good, 2022), relies on GDELT in its workflows, despite the conclusions presented earlier above. Problems stemming from such data sources—such as those explored in detail above—will impact which contexts are deemed most at risk, meaning that the stakes around these debates over conflict event data are higher than academic discourse alone.

Further, the ease and “intuitive” nature of public dashboards that data sources like GDELT may make available can even further attract conflict resolution experts, especially those who might consider themselves “non-technical experts” (Shields, 2014). Such individuals might not have relied on quantitative data in their workflows at all before—similar to that seen in academia, where less than a quarter of the studies in the review of over 500 research papers noted above relied on event data (Raleigh and Kishi, 2023b). In addition, engaging with conflict patterns solely via dashboards likely comes without considering what goes into the data generation process that underlies such public interfaces.

In short, use cases of datasets without regard for the impact of the decisions made by conflict data creators on conclusions are both numerous and high profile. The conclusions in turn stemming from such use contribute to why some policymakers or practitioners may view conflict contexts so differently.

Standards and transparency as guiding principles. We reviewed public, commonly used political violence datasets and exposed the effects of conflict scope conditions, sources, and coding choices on portrayals of conflict trends using each. Our empirical comparison—both automated datasets such as ICEWS and GDELT, as well as researcher-led datasets such as UCDP-GED and ACLED—demonstrates that there are significant discrepancies in narratives about violence by projects that are

ostensibly similar, and hence too often used interchangeably by the conflict analysis community. Ours is the first systematic investigation of the form and depth of quantitative conflict data measurement and has great significance for how we consider and respond to crises and instability worldwide. Rather than finding the substantial overlap between datasets, the review identifies several important gaps left by deficiencies across core criteria—conflict definitions, boundaries, and source catchments—for effective conflict data collection and analysis.

In short, the mechanics of the conflict data collection process matter: automated conflict data collections create artificial violent incidents due to duplicate reporting, false positives, an inability to deal with ‘fake news,’ a lack of valuable expert oversight, and no clear inclusion criteria. The result is that fictional conflicts may appear; cities may appear chronically unstable; mass populations may appear to move without evidence; countries may ‘attack’ each other, yet with words instead of weapons; and a wildly elevated, distorted assessment of risk may be accepted as accurate. Our subsequent detailed comparison of researcher-led conflict initiatives further shows that not all conflict datasets offer equal levels of coverage, depth, and content; and not all are equally externally valid, internally reliable, and, ultimately, comprehensively accurate. Depending on how conflict data projects define the boundaries of their remit, and where their information comes from, conflicts may be unseen; risks may be unaddressed; violence may disappear; deaths may vanish; and violent groups may be obscured. This occurs when boundaries and definitions are suited to theoretical criteria (prioritizing internal reliability) rather than the violence experienced by people across the world (prioritizing external validity). When local sources and non-English language reports are bypassed to privilege information from international media, data are distorted by media biases, significant holes in coverage emerge, and unstable places are unrepresented.

Our goal in this article is to follow other fields, ranging from global health (Davis, 2020), to urban planning and design (Calzada, 2017), to feminist critiques of international labor economics (D'Ignazio and Klein, 2020), where scholars interrogate problematic assumptions and data omissions that plague datasets (and methods) intended to solve important social problems. Based on the evidence noted here, if a researcher is

using data that intentionally or unintentionally miss conflicts, then subsequent analysis cannot accurately model, explain, or predict conflict. Further, the risks to citizens and states, the best practices for mitigating violence, and the weighing of causal inferences and explanations are all impacted by these choices. In addition to the fundamentals of dataset construction, there is a growing call for established datasets to undertake error identification and measurements as a component of the collection. Biermer and Amaya (2020), Amaya et al. (2020), and Biemer (2016) suggests there is a need to clearly indicate which errors are present and primary; what methods there are for assessing those errors within and across datasets; and what solutions can be brought to bear to mitigate those errors specifically and more broadly. In turn, we suggest that for automated datasets on conflict data, the issues are fundamental: we identify the primary errors in the collection as (1) coverage errors, (2) query errors, and (3) interpretation errors (as per Hsieh and Murphy, 2017). There is over-coverage, overly broad queries, and incorrect interpretations as the resulting corpus of ‘conflict events’ when in fact it is a corpus of online media reports with a word that is associated with conflict (or cooperation) in some way. In turn, we suggest that automated data are not a valid or reliable source of conflict data: they may be a reliable source for public ‘tone’ or media story focus, but not conflict data that are the subject of this piece.

Articles like Biermer and Amaya (2020), Amaya et al. (2020), and Demarest and Langer (2022) highlight a fundamental point that we pursue here: it is possible to enact solutions for datasets that collect material from sources with known baseline, characteristics, reach, and quality. These solutions can increase the reliability of datasets. However, they do little to improve the validity of these datasets, as the models they suggest for assessing error (whether the Total Error Framework or the Total Survey Error Framework, suggested by Demarest and Langer [2022]) operate from the assumptions that the collection model and static sources determine the outcome. For those datasets that are dependent on static sources—or, by extension, presume media at specific scales and languages constitute more static sources—it is possible to determine how the coding procedures and the definitions can lead to errors that can be systematically identified. For other datasets that are aware that the media landscape shifts—and that a large proportion of events are unreported or underreported in accessible media, and therefore can only be extracted from the conflict environment through local partnerships or in-country researchers—there is no static corpus to compare to. Instead, the onus is upon these datasets to be transparent about their collection methodology and how they have addressed errors and biases in their own collection, rather than relying on external frameworks that would limit their reach, remit, and validity, rather than enhance those qualities.

We suggest that the datasets with a public mandate and user base consider adhering to standards of collection, which begin with an accurate and honest appraisal of what they prioritize, what they do not, and how this affects the corpus of events that a user can extract. In addition, datasets should regularly and openly publish and discuss their information ecosystem, and what determines how deep, wide, and diverse it is. The reasons for variation in definition boundaries include limitations on costs or limitations on interest. Data user communities need to be aware of a set of standards to explain why datasets that claim to cover the same forms of conflict as each other do not do so in obvious ways, and why datasets that aim to capture the same events fail to do so. Acknowledging a prioritization of specific conflict forms, and privileging validity or reliability, are all acceptable choices. With greater transparency, the data-dependent communities can have more confidence in their data use and integration—

especially when it comes to important arenas such as conflict analysis.

Data availability

All data used for this study are fully and publicly available at their institutional websites for open access.

Received: 16 May 2022; Accepted: 14 February 2023;

Published online: 25 February 2023

Notes

- 1 As of mid-September 2022, Mexico is again on track for the most civilian fatalities (with over 4810 civilian fatalities reported), according to ACLED.
- 2 This is calculated based on fatalities in ‘violence against civilians’ events as well as fatalities in ‘explosions/remote violence’ events that target civilians (i.e. where ‘civilians’ are a primary actor).
- 3 This is calculated based on fatalities in ‘one-sided violence’ events; the ‘best estimate’ fatality count is equal to the ‘high’ and ‘low’ estimates for these events. In total, UCDP-GED records 18,812 fatalities in Mexico, per its ‘best estimate’ of fatalities. 28 civilian fatalities are 0.15% of 18,812 fatalities. UCDP-GEDs ‘high’ fatality estimate for Mexico 2021 is 19,573; its ‘low’ fatality estimate for that country year is 18,791.
- 4 ACLED, for example, finds that over 95% of civilian targeting in Mexico in 2021 was perpetrated by unnamed groups, including those working in coordination with cartels.
- 5 The World Press Freedom Index notes Mexico continues to be one of the deadliest countries for reporters in Latin America (Reporters Without Borders, 2021; for more, see Grillo, 2022).
- 6 Older efforts, such as Jackman and Boyd (1979), to stock-take what conflict datasets exist have been updated over time, such as by Anderton and Carter (2011), and more recently as part of the xSub project (Zhukov et al., 2019), which strives to provide a new portal for cross-national data on subnational violence.
- 7 A number of studies have compared empirical variations without conceptual interrogations, such as Donnay et al. (2019), Weidmann (2013), Ward et al. (2013), and Stundal et al. (2021).
- 8 Definitions of reliability and validity are in line with that used in the social sciences (for example, see Drost, 2011); these definitions are explored in further detail below.
- 9 These datasets are all widely used: per citation counts in Google Scholar, ACLED, and UCDP-GED have over a thousand citations, and ICEWS and GDELT have hundreds. These citations only reflect a fraction of the use cases of each dataset since these sources are used well beyond the academic community—with the audiences for some, such as ACLED and ICEWS, for example, being predominantly non-academic. Many users of these datasets, such as government entities, do so in private and confidential ways, making it impossible to have an accurate count of the number of use cases. For example, ACLED had nearly 9 million data downloads in 2021 alone (ACLED, 2022), suggesting a far higher use rate than the over 1600 academic citations in Google Scholar would indicate.
- 10 These data differ, however, with respect to where their boundaries for those definitions lie, and whether they collect information beyond conflict.
- 11 Berman et al. (2017) conduct analysis in their study using ACLED data, yet must replicate their findings using data from UCDP-GED, found in an online appendix.
- 12 For example, we do not include ‘strategic developments’ or demonstrations collected by ACLED as these are not within a common catchment of all reviewed datasets, and would bias the sample with inflated event frequency metrics.
- 13 We do not consider private or closed efforts, not only as neither we nor the public has access to such resources, but also because such data are often used specifically to monitor or evaluate practices or policies and have no role in global conflict pattern identification, research, or response.
- 14 On this point, we agree with Miller et al. (2022) that simple, empirical strategies to address bias post hoc “can provide a false sense of security and be quite detrimental ... [and can] introduce ethical consequences,” unlike Demarest and Langer (2022) who imply that certain empirical solutions are available to analysts. We argue that the empirical solutions that Demarest and Langer put forth can be largely impossible for data users to correctly calculate.
- 15 For example, a CAMEO code event type is ‘engage in mass killings,’ defined as the “Kill[ing of] a substantial number of people, typically with the intention of ridding a territory of a particular group of people” (Schrodt, 2012, p. 89). However, there is no clear definition of what “substantial” means in practice. All events labeled as “massacres” are coded under this category, “even if the language may be politicized;” ICEWS notes that “the use of this code in these cases reflects the severity of the language used and/or the extremity of the violence employed” (Boschec et al., 2018, p. 118).

- 16 Even if not purposefully shared as ‘fake news,’ sources themselves can also be impacted by their own ideological or political orientation, which in turn has effects on what and how events are reported (see Demarest and Langer, 2022).
- 17 The problem is often even worse for datasets tracking demonstration events, especially in cases where little oversight or inadequate inquiry occurs. People often will signal support for a cause rather than a witness or participate in a demonstration, or they may promote plans for a future demonstration that has not yet happened. Yet automated systems might assume all of the above are true ‘events’ that ought to be coded.
- 18 ACLED also includes gang violence, though only when it is used towards meeting overt political goals, as well as in cases where gang activity directly and fundamentally challenges public safety and security, in turn challenging the state’s monopoly on power (ACLED, 2020a, 2020b).
- 19 Demonstration and strategic development events are not considered in comparison cases in order to concentrate on the data variation collections between armed, organized groups engaging in political violence.
- 20 In addition to identifying actor names, ACLED introduced interaction codes in 2015, to note how different types of agents engage with one another in disorder. Classifying engagements amongst actors leads to accurate monitoring of changing dynamics.
- 21 ACLED data find that the civilian toll of the war on drugs in the Philippines, beginning in 2016 soon after Rodrigo Duterte assumed the presidency, perpetrated by the state and its supporters, is much higher than the official figures suggest at least 7742 Philippine civilians (Kishi and Buenaventura, 2021). The total of 1287 events does not include protests or strategic developments.
- 22 See UNESCO (2022).
- 23 Several reviews of ‘traditional’ media, including newspapers, have occurred in recent years, all concluding that scale (e.g. local vs. international media) can matter more than media form (see Demarest and Langer, 2022).
- 24 2017 was the most violent year in Syria. ACLED coverage begins in 2017.
- 25 See also Herkenrath and Knoll (2011) on differences in international vs. national media on protest reporting in Argentina, Mexico, and Paraguay; Bueno de Mesquita et al. (2015) on differences in national media reporting vs. an international wire aggregator on political violence in Pakistan; Demarest and Langer (2018) on differences between international vs. national news sources on the conflict in Nigeria; Barron and Sharpe (2008) on differences in district-level vs. provincial level reporting of violence in Indonesia).
- 26 Particular aspects of reporting are more susceptible to bias, such as whose ‘fault’ an event might be; who ‘started’ a violent encounter; and the number of fatalities. For example, armed groups may report more small-scale skirmishes or assaults in remote areas where more reputable sources lack access. Not including these may bias the dataset but inclusion without due consideration to the accuracy also creates biases.
- 27 For more on ACLED’s sourcing strategy, see ACLED (2020b).
- 28 See non-state armed group work by both academics, but more critically, by practitioners, about their engagements with groups in a conflict that are poorly covered by academic work on conflict (for example, International Review of the Red Cross, 2022).

References

- ACLED (2020a) Gang violence: concepts, benchmarks and coding rules. https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Gang-Violence-Concepts-Benchmarks-and-Coding-Rules_February-2020.pdf. Accessed 13 Apr 2022
- ACLED (2020b) FAQs: ACLED sourcing methodology. https://acleddata.com/acleddatanew/wp-content/uploads/dlm_uploads/2020/02/FAQs_ACLED-Sourcing-Methodology.pdf. Accessed 13 Apr 2022
- ACLED (2021) Adding new sources to ACLED coverage. https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Adding-New-Sources_May-2021.pdf. Accessed 13 Apr 2022
- ACLED (2022) ACLED 2021: annual report. <https://acleddata.com/about-acleddata/annual-report-2021/>. Accessed 5 Oct 2022
- AI for Good (2022) United National Children’s Fund (UNICEF). <https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/unicef/>. Accessed 7 Dec 2022
- Amaya A, Biemer P, Kinyon D (2020) Total error in a big data world: adapting the TSE framework to big data. *J Survey Stat Methodol* 8(1):89–119. <https://doi.org/10.1093/jssam/smz056>
- Anderton C, Carter J (2011) Conflict datasets: a primer for academics, policy-makers, and practitioners. *Def Peace Econ* 22(1):21–42
- Anyadike O (2015) Nigerian lives matter—the Baga controversy. *The New Humanitarian*. <https://www.thenewhumanitarian.org/feature/2015/01/15/nigerian-lives-matter-baga-controversy>. Accessed 5 Oct 2022
- Arista L (2021) Yo Tengo Otros Datos, Responde AMLO a Jorge Ramos Sobre Violencia en México. *Expansión Política*. <https://politica.expansion.mx/presidencia/2021/07/05/yo-tengo-otros-datos-responde-amlo-a-jorge-ramos-sobre-violencia-en-mexico>. Accessed 13 Apr 2022
- Ball P (2016) Why just counting the dead in Syria won’t bring them justice. *Foreign Policy*. <https://foreignpolicy.com/2016/10/19/why-only-counting-the-dead-in-syria-wont-bring-them-justice-hrdag-patternD@Sprng19-analysis/>. Accessed 5 Oct 2022
- Barranco J, Wisler D (1999) Validity and systematicity of newspaper data in event analysis. *Eur Sociol Rev* 15(3):301–322. <https://doi.org/10.1093/oxfordjournals.esr.a018265>
- Barron P, Sharpe J (2008) Local conflict in post-Suharto Indonesia: understanding variations in violence levels and forms through local newspapers. *J East Asian Stud* 8:395–423
- Baum M, Zhukov Y (2015) Filtering revolution: reporting bias in international newspaper coverage of the Libyan civil war. *J Peace Res* 52(3):384–400. <https://doi.org/10.1177/0022343314554791>
- Berman N, Couttenier M, Rohner D, Thoenig M (2017) This mine is mine! How minerals fuel conflicts in Africa. *Am Econ Rev* 107(6):1564–1610
- Biemer P (2016) Can TSE save survey science in a Big Data World? Keynote presentation at the International Total Survey Error Workshop
- Biemer PP, Amaya A (2020) Total error frameworks for found data. In Hill C, Biemer PP, Buskirk T, Japac L, Kirchner A, Kolenikov S, Lyberg L (Eds.). *Big Data meets survey science: a collection of innovative methods*, pp. 131–161. <https://doi.org/10.1002/9781118976357.ch4>
- Bond D, Bond J, Oh C, Jenkins JC, Taylor CL (2003) Integrated data for events analysis (IDEA): an event typology for automated events data development. *J Peace Res* 40:733–745
- Boschee E, Natarajan P, Weischedel R (2013) Automatic extraction of events from open source text for predictive forecasting. In: Subrahmanian VS ed *Handbook of computational approaches to counterterrorism*. Springer Science +Business Media, New York, pp. 51–67
- Boschee E, Lautenschlager J, O’Brien S, Shellman S, Starz J (2018) ICEWS weekly event data. *Harvard Dataverse*, V323. <https://doi.org/10.7910/DVN/Q12T9A>
- Boyle M (2014) Violence after war: explaining instability in post-conflict states. *JHU Press*
- Bueno de Mesquita E, Fair CC, Jordan J, Rais RB, Shapiro JN (2015) Measuring political violence in Pakistan: insights from the BFRS dataset. *Confl Manag Sci* 32:536–58
- Calzada I (2017) Metropolitan and city-regional politics in the urban age: why does “(smart) devolution” matter? *Palgrave Commun* 3(1):1–17
- Chaparro L (2021) Mexico’s war on cartels has created 400 new gangs that are taking on the police and cartels that are left. *Business Insider*. <https://www.businessinsider.com/fragmentation-in-mexico-war-on-drugs-created-400-new-gangs-2021-10>. Accessed 5 Oct 2022
- Chalabi M (2014a) Kidnapping of girls in Nigeria is part of a worsening problem (updated). *FiveThirtyEight*. <https://fivethirtyeight.com/features/nigeria-kidnapping/>. Accessed 7 Dec 2022
- Chalabi M (2014b) Mapping kidnappings in Nigeria (updated). *FiveThirtyEight*. <https://fivethirtyeight.com/features/mapping-kidnappings-in-nigeria/>. Accessed 7 Dec 2022
- Choi HJ, Raleigh C (2015) Dominant forms of conflict in changing political systems. *Int Stud Q* 59(1):158–171
- Clarke K (2021) Which protests count? Coverage bias in Middle East event datasets. *Mediterranean Politics* 1–27. <https://doi.org/10.1080/13629395.2021.1957577>
- Crawford N (2016) Update on the human costs of war for Afghanistan and Pakistan, 2001 to mid-2016. *Costs of War*
- D’Ignazio C, Klein L (2020) *Data feminism*. MIT press
- Daly S (2016) *Organized violence after civil war: the geography of recruitment in Latin America*. Cambridge University Press
- Danzger MH (1975) Validating conflict data. *Am Sociol Rev* 40:570–84
- Davenport C, Ball P (2002) Views to a kill: exploring the implications of source selection in the case of Guatemalan state terror, 1977–1995. *J Confl Resolution* 46(3):427–450
- Davis S (2020) *The uncensored: politics of data in global health*. Cambridge University Press
- Day J, Pinckney J, Chenoweth E (2015) Collecting data on nonviolent action: lessons learned and ways forward. *J Peace Res* 52(1):129–133
- Deep South Watch Database (2004) *Conflict Incident Database*. <https://deepsouthwatch.org>
- Demarest L, Langer A (2018) The study of violence and social unrest in Africa: a comparative analysis of three conflict event datasets. *Afr Aff* 117:310–25
- Demarest L, Langer A (2022) How events enter (or not) data sets: the pitfalls and guidelines of using newspapers in the study of conflict. *Sociol Methods Res* 51(2):632–666
- Donnay K, Dunford E, McGrath D, Backer D, Cunningham D (2019) Integrating conflict event data. *J Confl Resolution* 63:1337–1364
- Dorff C, Henry C, Ley S (2022) Does violence against journalists deter detailed reporting? Evidence from Mexico. *J Confl Resolution* 14–25. <https://doi.org/10.1177/00220027221128307>
- Dowd C, Justino P, Kishi R, Marchais G (2020) Comparing ‘new’ and ‘old’ media for violence monitoring and crisis response: evidence from Kenya. *Res Politics* 7:1–9

- Dowd C (2015) Actor proliferation and the fragmentation of violent groups in conflict. *Res Politics* 2(4):2053168015607891
- Dowd C (2019) Fragmentation, conflict, and competition: Islamist anti-civilian violence in sub-Saharan Africa. *Terror Political Violence* 31(3):433–453
- Drost E (2011) Validity and reliability in social science research. *Educ Res Perspect* 38(1):105–123
- Earl J, Martin A, McCarthy JD, Soule SA (2004) The use of newspaper data in the study of collective action. *Annu Rev Sociol* 30:65–80
- Ekzayez A, Alhaj Ahmad Y, Alhaleb H, Checchi F (2021) The impact of armed conflict on utilisation of health services in north-west Syria: an observational study. *Confl Health* 15(1):1–0
- Fazal T (2014) Dead wrong? Battle deaths, military medicine, and exaggerated reports of war's demise. *Int Secur* 39:95–125
- France 24 (2019) Cows come home to haunt India's Modi. <https://www.france24.com/en/20190405-cows-come-home-haunt-indias-modi>. Accessed 13 Apr 2022
- Franzosi R (1987) The press as a source of socio-historical data: issues in the methodology of data collection from newspapers. *Hist Methods* 20:5–16
- Galtung J, Ruge MH (1965) The structure of foreign news: the presentation of the Congo, Cuba and Cyprus Crises in four Norwegian newspapers. *J Peace Res* 2:64–90
- GDELT (2021) The GDELT story: about the GDELT project. <https://www.gdeltproject.org/about.html>. Accessed 13 Apr 2022
- Gerner D, Schrodt P, Yilmaz Ö (2008) Conflict and Mediation Event Observations (CAMEO): an event data framework for a post-cold war world. In: Bercovitch J, Gartner S (eds) *International conflict mediation: new approaches and findings*. Routledge, pp. 287–304
- Greene K, Tornquist C, Fokink R, Lindelauf R, Subrahmanian V (2021) Understanding the timing of Chinese border incursions into India. *Humanit Soc Sci Commun* 8(1):1–8
- Grillo I (2022) Who is really killing Mexican journalists? <https://ioangrillo.substack.com/p/who-is-really-killing-mexican-journalists>? Accessed 13 Apr 2022
- Halkia M, Ferri S, Papazoglou M, Van Damme MS, Jenkinson G, Baumann K, Thomakos D (2019) Dynamic Global Conflict Risk Index. JRC Technical Reports
- Harcup T, O'Neill D (2001) What is news? Galtung and Ruge revisited. *Journalism Stud* 2:261–80
- Harcup T, O'Neill D (2016) What is news? News values revisited (again). *Journalism Stud* 18:1–19
- Havasi A (2019) Iran vows to ditch more nuclear curbs in war of words with US. Agence France Presse <https://www.yahoo.com/video/iran-warns-us-attack-imperil-interests-across-mideast-000622035.html>. Accessed 13 Apr 2022
- Heidelberg Institute for International Conflict Research (2021) Conflict barometer. <https://hiik.de/data-and-maps/datasets>
- Herkenrath M, Knoll A (2011) Protest events in international press coverage: an empirical critique of cross-national conflict databases. *Int J Comp Sociol* 52:163–80
- Höglblad S (2021) UCDP GED Codebook Version 21.1. Department of Peace and Conflict Research, Uppsala University. <https://ucdp.uu.se/downloads/ged/ged211.pdf>. Accessed 13 Apr 2022
- Hsieh YP, Murphy J (2017) Total Twitter error: Decomposing public opinion measurement on Twitter from a total survey error perspective. In: Biemer PP, de Leeuw E, Eckman S, Edwards B, Kreuter F, Lyberg LE, Tucker NC, West BT (Eds.) *Total survey error in practice: Improving quality in the era of big data* vol 74. pp. 23–46. <https://doi.org/10.1002/9781119041702.ch2>
- Hutter S (2014) Protest event analysis and its offspring. In: Porta Donatelladella ed *Methodological practices in social movement research*. Oxford University Press, Oxford, UK, pp. 335–367
- International Review of the Red Cross (2022) Non-State Armed Groups. IRRC, no. 9115. <https://international-review.icrc.org/reviews/irrc-no-915-non-state-armed-groups>. Accessed 14 Dec 2022
- Jackman R, Boyd W (1979) Multiple sources in the collection of data on political conflict. *Am J Pol Sci* 23(2):434–458. <https://doi.org/10.2307/2111011>
- Jenkins JC, Maher TV (2016) What should we do about source selection in event data? Challenges, progress, and possible solutions. *Int J Sociol* 46:42–57
- Kalyvas SN (2006) *The logic of violence in civil war*. Cambridge University Press, Cambridge, MA
- Keertipati S, Savarimuthu BTR, Purvis M, Purvis M (2014) Multi-level analysis of peace and conflict data in GDELT. In: *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis (MLSDA'14)*. Association for Computing Machinery, New York, NY, USA, pp. 33–40. <https://doi.org/10.1145/2689746.2689750>
- Kishi R, Buenaventura T (2021) The drug war rages on in the Philippines: new data on the civilian toll, state responsibility, and shifting geographies of violence. In: *Armed Conflict Location and Events Data Project*. <https://acleddata.com/2021/11/18/the-drug-war-rages-on-in-the-philippines-new-acled-data-on-the-civilian-toll-state-responsibility-and-shifting-geographies-of-violence>. Accessed 13 Apr 2022
- Krippendorff K (2013) *Content analysis: an introduction to its methodology*. Sage, Thousand Oaks, CA
- LaFree G, Dugan L (2007) Introducing the Global Terrorism Database. *Terror Political Violence* 19:181–204
- Leetaru K, Schrodt P (2013) GDELT: Global Data on Events, Location, and Tone, 1979–2012. Paper presented at the International Studies Association Annual Convention, San Francisco, CA, USA. Retrieved from <http://data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf>
- Levy E (2001) Quantification, mandated science and judgment. *Stud Hist Philos Sci Part A* 32(4):723–737
- Maoz Z, Johnson P, Kaplan J, Ogunkoya F, Shreve A (2018) The Dyadic Militarized Interstate Disputes (MIDs) dataset version 3.0: logic, characteristics, and comparisons to alternative datasets. *J Confl Resolution* 63:811–835
- Malu LN (2019) *The International Criminal Court and peace processes*. Springer International Publishing
- Martin H (2018) Sexual misconduct on crowded airlines is happening more often. What airlines are doing about it. Orlando Sentinel. <https://www.orlandosentinel.com/travel/la-fi-harassment-flights-20181231-story.html>. Accessed 13 Apr 2022
- McGoldrick C (2015) The state of conflicts today: can humanitarian action adapt? *Int Rev Red Cross* 97(900):1179–1208. <https://doi.org/10.1017/S181638311600028X>
- Miller E, Kishi R, Raleigh C, Dowd C (2022) An agenda for addressing bias in conflict data. *Sci Data* 30 9(1):593–601
- Mo X, Yu D (2022). Analysis of interaction patterns of intercountry cooperation and conflict events based on complex networks. In: *Advances in intelligent systems and computing*. Springer, Singapore, pp. 229–237
- Moon E (2019) Hate crimes are on the rise. why do many still go unreported? Pacific Standard. <https://psmag.com/news/hate-crimes-are-on-the-rise-why-do-many-still-go-unreported>. Accessed 13 Apr 2022
- OpenNews (2014) GDELT and the problem of decontextualized data. <https://source.opennews.org/articles/gdelt-decontextualized-data/>. Accessed 13 Apr 2022
- Orange County Breeze (2019) Weather in Northwest Orange County. https://www.oc-breeze.com/2018/12/31/132798_weather-in-northwest-orange-county-for-tues-january-1-2019/. Accessed 13 Apr 2022
- Ortiz DG, Myers DJ, Walls NE, Diaz M-ED (2005) Where do we stand with newspaper data? *Mobilization* 10:397–419
- Pinker S (2011) *The better angels of our nature: why violence has declined*. Penguin Books
- Raleigh C (2012) Violence against civilians: a disaggregated analysis. *Int Interact* 38:462–481
- Raleigh C, de Bruijne K (2017) Where rebels dare to tread: a study of conflict geography and co-option of local power in Sierra Leone. *J Confl Resolution* 61:1230–1260
- Raleigh C, Kishi R, Russell O, Siegle J, Williams W (2017). Boko Haram vs. al-Shabab: what do we know about their patterns of violence. *Washington Post* 2
- Raleigh C, Kishi R (2019) Comparing conflict data: similarities and differences across conflict datasets. https://www.acleddata.com/wp-content/uploads/2019/09/ACLED-Comparison_8.2019.pdf. Accessed 13 Apr 2022
- Raleigh C, Kishi R (2020) Hired guns: using pro-government militias for political competition. *Terror Political Violence* 32(3):582–603
- Raleigh C, Kishi R (2023a) Pro-regime Militias. In: Johnson P, Wittels W (eds) *Militias, states, and violence against civilians: civic vice, civic virtue*. Routledge
- Raleigh C, Kishi R (2023b) Conflict event data. In: Mello P, Ostermann F (eds) *Routledge handbook of foreign policy analysis methods*. Routledge
- Raleigh C, Linke A, Hegre H, Karlsen J (2010) Introducing ACLED: an armed conflict location and event dataset: special data feature. *J Peace Res* 47:651–660
- Ray D, Esteban J (2017) Conflict and development. *Annu Rev Econ* 9:263–293
- Regencia T (2021) ICC to open full investigation into Duterte's 'war on drugs.' Al Jazeera <https://www.aljazeera.com/news/2021/9/15/philippines-icc-agrees-to-probe-dutertes-war-on-drugs>. Accessed 13 Apr 2022
- Reporters Without Borders (2021) 2021 World Press Freedom Index. <https://rsf.org/en/index?year=2021>. Accessed 3 Oct 2022
- Sarkees M, Wayman F (2010) *Resort to war: 1816–2007*. CQ Press
- Schrodt P (2012) CAMEO Conflict and Mediation Event Observations Event and actor codebook. <http://data.gdeltproject.org/documentation/CAMEO.Manual.1.1b3.pdf>. Accessed 13 Apr 2022
- Schrodt PA, Van Brackle D (2013) Automated coding of political event data. In: Subrahmanian VS ed. *Handbook of computational approaches to counter-terrorism*. Springer Science+Business Media, New York, pp. 23–49
- Schrodt PA, Simpson EM, Gerner DJ (2001) Monitoring conflict using automated coding of news wire reports: a comparison of five geographical regions. Conference paper, June 8–9, Uppsala, Sweden
- Shaver A, Shapiro J (2021) The effect of civilian casualties on wartime informing: evidence from the Iraq War. *J Confl Resolution* 65(7-8):1337–1377
- Shields CM (2014) GDELT Global Dashboard: Big data for conflict resolution and peacebuilding. *Peace Insight* <https://www.peaceinsight.org/en/articles/gdelt-global-dashboard/?location=&theme=research>. Accessed 7 Dec 2022

- Stundal L, Bagozzi B, Freeman E, Holmes J (2021) Human rights violations in space: assessing the external validity of machine-geocoded versus human-geocoded data. *Polit Anal* 31(1):81–97. <https://doi.org/10.1017/pan.2021.40>
- Stearns JK (2022) The war that doesn't say its name: the unending conflict in the Congo. Princeton University Press
- Sundberg R, Melander E (2013) Introducing the UCDP Georeferenced Event Dataset. *J Peace Res* 50:523–532
- Tai Z, Chang T (2002) The global news and the pictures in their heads: a comparative analysis of audience interest, editor perceptions and newspaper coverage. *Gazette* 64(3):251–265
- Taruvunga M (2018) Zimbabwe: Mugabe son-in-law granted \$30 bail on kidnapping charges. New Zimbabwe. <https://allafrica.com/stories/201810240687.html>. Accessed 13 Apr 2022
- Thies C, Baum C (2020) The effect of war on economic growth. *Cato Journal*, Vol. 40, No. 1, pp 199–212 (Winter, 2020)
- UNESCO (2022) UNESCO observatory of killed journalists - Philippines. <https://en.unesco.org/themes/safety-journalists/observatory/country/223790>
- Ward M, Beger A, Cutler J, Dickenson M, Dorff C, Radford B (2013) Comparing GDELT and ICEWS event data. *Analysis* 21(1):267–297
- Weidmann N (2013) The higher the better? The limits of analytical resolution in conflict event datasets. *Cooper Confl* 48(4):567–576
- Weidmann NB (2016) A closer look at reporting bias in conflict event data. *Am J Political Sci* 60:206–18
- Weller N, McCubbins K (2014) Raining on the parade: some cautions regarding the global database of events, language and Tone Dataset. *Political Violence at a Glance*. <https://politicalviolenceataglance.org/2014/02/20/raining-on-the-parade-some-cautions-regarding-the-global-database-of-events-language-and-tone-dataset/>. Accessed 13 Apr 2022
- Yemen Data Project (2022) Collating and disseminating data on the conduct of the war in Yemen with the purpose of increasing transparency and promoting accountability. <https://yemendataport.org>
- Zhukov Y, Davenport C, Kostyuk N (2019) Introducing xSub: a new portal for cross-national data on subnational violence. *J Peace Res* 56(4):604–614
- Zindoga T (2019) Zim Airways: the untold story. *The Herald*. <https://www.herald.co.zw/zim-airways-the-untold-story/>. Accessed 13 Apr 2022

Acknowledgements

The authors wish to thank Olivia Russell and Dara George-John for their assistance with the final article. The authors also wish to thank the many data creators and analysts who have shared their thoughts and research over the past years.

Competing interests

The authors declare that they created one of the datasets under consideration. They note that ACLED is a 501(c)(3) designated non-profit NGO. None of the authors profit from this or other articles concerning data comparisons.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Clionadh Raleigh.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023