The Uncounted Dead: Statist Bias and Civilian Targeting in Conflict Data

M. P. Broache ¹, Kate Cronin-Furman², Milli Lake³, and Agnes Yu³

¹University of North Carolina at Greensboro, USA, ²University College London, UK, and ³London School of Economics, UK

Abstract

Quantitative research on violence against civilians relies extensively on "off-the-shelf" data, such as the widely cited Uppsala Conflict Data Program's (UCDP) One-Sided Violence dataset. We show that, due to data collection and coding protocols that privilege government narratives of violence, such data often reproduces statist biases pervasive in the international system. These dynamics are particularly visible when civilian deaths result from airstrikes, shelling, and other forms of long-range bombardment. Such capabilities are disproportionately possessed by states, yet conservative coding practices, combined with government control over information and access restrictions, dictate that UCDP consistently codes civilian deaths at the hands of governments as "battle-related" or incidental rather than deliberate targeting. We analyze patterns in the UCDP data release, version 23.1, using evidence from Sri Lanka and Ethiopia to illustrate these patterns.

Resumen

La investigación cuantitativa en materia de violencia contra la población civil se basa, en gran medida, en datos "listos para usar", como el frecuentemente citado conjunto de datos sobre violencia unilateral de la universidad de Uppsala (UCDP, por sus siglas en inglés, Uppsala Conflict Data Program). Demostramos que, debido a los protocolos de recopilación de datos y de codificación, los cuales privilegian las narrativas gubernamentales en materia de violencia, dichos datos tienden a reproducir, con frecuencia, los sesgos estatistas dominantes en el sistema internacional. Estas dinámicas pueden observarse más claramente cuando las muertes de civiles son consecuencia de ataques aéreos, bombardeos y otras formas de bombardeo de largo alcance. Los Estados poseen estas capacidades de forma desproporcionada, sin embargo, las prácticas conservadoras en materia de codificación, combinadas con el control gubernamental sobre la información y las restricciones de acceso, dictan que el UCDP codifique, de manera sistemática, estas muertes de civiles a manos de los Gobiernos como "relativas a la batalla" o como incidentales en lugar de considerarlos como objetivos deliberados. Analizamos los patrones de los datos presentes en la versión del UCDP (versión 23.1), utilizando evidencia procedente de Sri Lanka y Etiopía con el fin de ilustrar estos patrones.

Résumé

La recherche quantitative sur la violence à l'encontre de civils repose en grande partie sur des données standard, comme l'ensemble de données One-Sided Violence de l'UCDP, une source très souvent citée. Comme la collecte de données et les protocoles d'encodage privilégient les récits gouvernementaux sur la violence, nous montrons que ces données reproduisent souvent les biais étatistes

Broache, M. P. et al. (2025) The Uncounted Dead: Statist Bias and Civilian Targeting in Conflict Data. *Journal of Global Security Studies*, https://doi.org/10.1093/jogss/ogaf013

[©] The Author(s) (2025). Published by Oxford University Press on behalf of the International Studies Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

omniprésents dans le système international. Ces dynamiques sont particulièrement visibles quand des civils décèdent à la suite de frappes aériennes, de tirs d'obus ou d'autres formes de bombardements à longue distance. Ces capacités sont bien souvent l'apanage des États. Pourtant, les pratiques d'encodage prudentes, combinées au contrôle des informations et aux restrictions d'accès imposés par le gouvernement, font en sorte que l'UCDP encode systématiquement les morts civiles du fait du gouvernement comme étant "liées à un affrontement" ou accidentelles, et non des ciblages délibérés. Nous analysons les schémas présents dans les données publiées par l'UCDP (version 23.1), à l'aide d'éléments probants provenant du Sri Lanka et d'Éthiopie pour illustrer ces schémas.

Keywords: conflict data, civilian targeting, state violence
Palabras clave: datos sobre conflictos, ataques contra civiles, violencia estatal
Mots clés: données relatives aux conflits, ciblage de civils, violence étatique

Introduction

The Uppsala Conflict Data Program's (UCDP) One-Sided Violence (OSV) dataset purports to identify all instances of "the use of armed force by the government of a state or by a formally organized group against civilians which results in at least 25 deaths" (Pettersson 2023c, 3). This dataset is used extensively in quantitative research on violence against civilians and informs forecasting efforts and policy.¹ However, users may be surprised that the Syrian government does not appear as an actor in the OSV in 2018, despite reports that over 1,600 civilians were killed in its February-March 2018 campaign to recapture Ghout alone (Human Rights Watch 2019, 555). Likewise,

1 As of January 9, 2024, Eck and Hultman (2007), which introduced the OSV, had been cited 930 times, per Google Scholar. Notable works utilizing the OSV span studies examining rebel sponsorship and civilian abuse (Salehyan, Siroky, and Wood 2014), repression of unarmed protestors (Sutton, Butcher and Svennson 2014), civilian-targeted violence during civil war (Wood, Kathman, and Gent 2012), and the effects of peacekeeping on mitigating one-sided violence (Haas and Ansorg 2018 utilize the GED more broadly but focus on one-sided violence), among many others. We raise these articles as examples of the various importance of OSV/GED data for furthering conflict research. The centrality of the UCDP data for conflict research has also meant that subsequent datasets have been designed to incorporate or be compatible with such data as a baseline comparison, such as the Sexual Violence in Armed Conflict (SVAC) dataset by Cohen and Nordás (2014), the Konstanz One-sided Violence Event Dataset (KOSVED) by Schneider and Bussman (2013), and the Ethnic Onesided Violence dataset (EOSV) by Fjelde et al (2021).

while the OSV attributes 1,112 civilian deaths in Ukraine in 2022 to the Russian government, this excludes civilians killed in Russian airstrikes and shelling targeting a maternity hospital and theater during the siege of Mariupol. Instead, these civilian deaths are coded as "battle-related" (Davies et al. 2023, 698).²

While reliable information about violence against civilians is notoriously difficult to collect, and rigid coding protocols may be necessary to preserve consistency and precision, the absence of these high-profile episodes of state-perpetrated civilian killings in the OSV feeds a historical narrative wherein states are perceived as guarantors of rather than threats to the security of civilians. A comparison of the OSV to two other datasets, the State-Sponsored Mass Killings (SSMK) (Ulfelder & Valentino 2008) and the Targeted Mass Killings (TMK) dataset, version 1.1 (Butcher et al. 2020), suggests the above examples are not isolated episodes. We chose to benchmark the OSV against the SSMK and TMK because, while all three datasets purport to measure civilian killings in various forms, the OSV maintains the broadest inclusion criteria. SSMK restricts its focus to state-sponsored mass killings, and TMK explicitly requires the targeting of specific communal or political groups. We would expect episodes included in the SSMK and TMK, therefore, to also appear in the OSV.

The SSMK identifies 62 episodes of state-sponsored mass killings ongoing in at least one year from 1989–2006, the period for which the SSMK and OSV overlap. Yet, for at least 12 (17.74 percent) of these episodes, there is no corresponding OSV entry for *any* of the years

2 Unless otherwise noted, all references to UCDP data are to version 23.1 of the relevant dataset (Pettersson 2023a, 2023b and 2023c) in which the SSMK records an active state-sponsored mass killing. For the TMK, there is no corresponding OSV entry for 14 of 235 actor-years from 1989 through 2020 (5.96 percent). Of these, nine involve government killings, while the remaining five involve pro-government militias or groups that received foreign government support.³

While the inclusion criteria in the OSV are the most expansive, the SSMK and TMK each consider expert testimony in addition to media reporting in compiling data on mass killings. We argue that UCDP's more conservative data collection and coding protocols amplify statist bias already found in the source reports used to construct UCDP datasets. Specifically, government efforts to limit information access, such as entry and movement restrictions affecting journalists and human rights actors, and to control information outflows by shaping narratives of violence, create pro-state biases in source reporting. Then, conservative coding protocols employed by UCDP to avoid erroneous accusations launder these biases in ways that further obscure the culpability of governments and reinforce states' legitimacy in their targeting of civilians (Weber 1965; Campbell et al. 2011; Gordon 2014; MacGinty & Firchow 2016). Classifying civilian deaths in "indiscriminate" attacks as "battle-related" and relying on parties' stated intent where there is uncertainty results in systematic undercounting of state-perpetrated one-sided violence. After developing this argument theoretically, we use illustrative examples from Sri Lanka and Ethiopia to document these patterns.

Theoretical Argument

Critical and feminist scholarship has long directed attention towards the privileging of state security over the well-being of civilians (Barkawi 2011; Sjoberg 2016). Existing literature has also documented challenges in collecting accurate information on violence against civilians and conflict more generally (Weidmann 2016; Dietrich and Eck 2020; Parkinson 2024). In addition to access restrictions caused by poor infrastructure, perpetrators may deliberately obscure information from journalists and civil society actors. We argue that states have particular advantages in shaping our knowledge of violence. States more frequently control borders, allowing them to restrict media and human rights access, affecting event reporting by preventing journalists from entering atrocity sites and by threatening future access denial. States are additionally more able to *control* fatality narratives,

3 The online appendix details our procedures for constructing these comparisons.

The UCDP Georeferenced Event Dataset (GED), on which the OSV and other aggregated UCDP datasets are based, relies on secondary sources—and predominantly media coverage—to identify individual events of organized violence (Sundberg & Melander 2013; Högbladh 2023).⁴ We show how data collection strategies like those used to compile the GED, which rely extensively on media reporting and other secondary sources and that omit expert testimony or human rights documentation, are prone to amplifying these biases.

Importantly, UCDP defines "battle-related" deaths as those "caused by the warring parties... directly related to combat" (Pettersson 2023a, 4). This includes civilian deaths resulting from "collateral damage in the form of civilians killed in the crossfire, indiscriminate bombings, etc.," provided that "the target for the attacks is either the military forces or representatives for the parties." This definition includes deaths resulting from "bombardments of military bases, cities, and villages" (Besaw et al. 2023; Pettersson 2023a, 4). Because UCDP treats battle deaths and one-sided violence as mutually exclusive (Eck and Hultman 2007, 235), coders must determine whether civilians were *intentionally* targeted or their deaths resulted from military clashes. Government restrictions on information access and control mean these determinations often rely on information produced or heavily shaped by state perpetrators, posing particular challenges for coding.

For instance, where there is uncertainty about intent, violence is coded "based on the stated intention of the parties" (Eck and Hultman 2007, 235). ⁵ However, threats of legal liability and political backlash create incentives for perpetrators to exploit ambiguity about whether targets are civilians or combatants (Kinsella 2011; Cronin 2013). Adhering conservatively to coding protocols in order to avoid incorrectly attributing intent means that government accounts are often privileged and civilian deaths are more likely to be coded as "battlerelated" when states are perpetrators.

Because states are more likely than non-state actors to possess the airpower and heavy artillery to conduct "indiscriminate" attacks (Qiu 2022), these dynamics are further amplified in cases of long-range attacks, airstrikes, or shelling. Governments are both more likely to possess

- 4 See the online Appendix for further detail.
- 5 Eck and Hultman (2007) note "exceptions" to this rule in cases with disproportionate ratios of civilian to military fatalities. This is discussed further on p. 9.

distinctive military infrastructure and dedicated bases (Jones & Johnston 2013) than insurgents. By virtue of the fact that government infrastructure and military targets are easier to identify than infrastructure of rebel groups, the inclusion criterion that any attack plausibly aimed at "military forces or representatives" of a combat party qualifies as "battle-related" necessarily renders state targeting as strategically and militarily more ambiguous than attacks perpetrated by non-state actors.

Meanwhile, strategic targeting by rebel groups perpetrated using small arms does appear in the OSV. A February 24, 2003, attack on Bogoro village in DRC was perpetrated in order to neutralize enemy forces and secure a strategically important crossroads.⁶ The event appears in the GED as part of the "FNI, FPRI vs Civilians" dyad. Given that OSV data is cited to underpin conclusions that rebel groups are more likely than state forces to commit attacks on civilians,⁷ we should be concerned about disparities in the treatment of strategic attacks on civilian populations perpetrated by state and non-state actors.

This coding approach seems to introduce artificial distinction between nearly identical events. For instance, human rights organizations, humanitarian workers on the ground, and public health scholars have extensively documented the Syrian government's use of aerial bombardment as part of a broader strategy of targeting the healthcare sector, which has also included detention, torture, and forced disappearance of healthcare workers. Per a 2016 Lancet article: "Analysis of attacks over several years in important opposition-held areas of Aleppo, Hama, Idlib, eastern Ghouta, and Homs reveals a pattern of repeated targeting with intention to shut access to health care, whether to impede opposition forces or to force civilian displacement." One study found that in 2016 alone, 297 civilians were killed in attacks on healthcare infrastructure, 91 percent of which were committed via aerial bombardment. Some of these deaths do appear in the OSV. For instance, the GED records the 58 deaths on April 27, 2016, in an airstrike on Al-Quds hospital in the "Government of Syria Civilians" dyad, which has a total of 135 deaths recorded for 2016 (matching the Government of Syria's 2016 entry in the OSV). But the 13 civilian fatalities (including 4 newborn babies in the ICU) recorded by the GED for remarkably similar attacks on hospitals in July and August are attached to the "Gov-

6 https://www.icc-cpi.int/sites/default/files/Transcripts/ CR2009 08585.PDF.

7 "rebels tend to be more violent on the whole, while governments commit relatively little violence except in those few years which see mass killings" (Eck & Hultman 2007) ernment of Syria-Syrian Insurgents" dyad, and classed as

"battle-related." The UCDP's tendency to categorize civilian deaths resulting from aerial attacks (capabilities disproportionately possessed by states) as "battle-related" reflects and reinforces broader trends in the discursive treatment of state violence. This presumption of unintentional collateral damage pursuant to the legitimate use of violence affords states the opportunity to cloud their culpability for intentionally indiscriminate attacks on civilian targets. We show that where civilian deaths occur in long-range attacks, governments frequently manipulate information access and control to ensure civilian killings are framed as targeting "military forces or representatives." When early media reports establish dominant, state-centric narratives, journalists and human rights investigators may be more likely to focus on subsequent events conforming to these narratives (Dawkins 2021; Gibilisco & Steinberg 2022).

Therefore, where state perpetrators restrict information flows, discredit victims' claims, and promulgate alternate narratives, conservative data collection and coding practices create conditions ripe for interpreting incidences of civilian killings by governments as "battlerelated." Indeed, because government perpetrators are disproportionately able to deploy these tactics as compared to insurgent groups, information sources available to desk researchers invariably privilege states (Davenport & Ball 2002; Clark & Sikkink 2013; Hendrix & Salehyan 2015; Fariss et al. 2020). In the next section, we turn to patterns in the GED that illustrate statist biases at play.

Coding Long-Range Attacks

To identify the extent of statist bias in UCDP data on civilian killings, we reviewed version 23.1 of the GED (Högbladh 2023). The dataset includes 316,818 events, with 1,370,447 total civilian deaths across 96,603 events. Table 1 presents their distribution across the three major categories of violence in the GED: state-based conflict, non-state conflict (Sundberg, Eck, and Kreutz 2012), and one-sided violence.

While the vast majority (1,124,466 of 1,370,447, or 82.05 percent) of civilian fatalities in the GED are classed as one-sided violence, 245,981 civilian deaths are classed as battle-related. Notably, in 41,858 of 54,239 (77.17 percent) events coded as battle-related but involving civilian deaths (183,910 in total), the GED reports *no* fatalities for belligerents in the relevant dyad. Such disparities might suggest that civilians were not killed in "crossfire" or in attacks on discernible military targets, but rather in

Table	1. C	ivilian	deaths	in	the	UCDP	GED,	version	23.1
-------	------	---------	--------	----	-----	------	------	---------	------

Category	Civilian death events	Civilian deaths tTotal
State-based conflict	50,205	223,558
Non-state conflict	4,034	22,423
One-sided violence	42,364 ⁸	1,124,466
TOTAL	96,603	1,370,447

areas populated by civilians. Although Eck and Hultman (2007, 235) specify that incidents with "a highly disproportionate ratio of military to civilian fatalities' are treated as 'exceptions' to the general rule of coding based on parties" stated intent, the substantial number of civilian fatalities coded as "battle-related" in events with no military fatalities leaves us unclear as to how consistently this "exception" is applied.

To explore our expectations about long-range capabilities, we further examined the subset of civilian deaths coded as battle-related where there were no belligerent fatalities to identify how many might have occurred in long-range attacks. We conducted a keyword search of the text of the "source_article" variable in the GED for terms relating to long-range bombardment, including iterations of the following: *air**, *aer**, *bombar**, and *shell**. We identified 6,617 events displaying these terms, plausibly involving long-range bombardment. These events involved 43,520 civilian deaths, all coded as battle-related with no belligerent fatalities. In comparison, when we conducted keyword searches of events coded as one-sided violence, we identified only 279 events and 2,135 civilian deaths.

To demonstrate how information *access* and *control* shape coding practices in cases of aerial bombardment and long-range shelling—modalities of war already disproportionately possessed by states—we turn next to illustrative examples from two GED conflict events in Sri Lanka and Ethiopia, respectively.

Information Outflows in Long-Range Attacks

This section examines two events in the GED involving (1) long-range bombardment in the latter stages of the Sri Lankan civil war in 2009 and (2) a government airstrike in Tigray, Ethiopia, in 2021. In each, the GED cited me-

8 There are 48,667 events coded as "one-sided violence" in the GED, version 23.1. However, for 6,303 one-sided violence events, the "best" estimate of deaths is reported as 0. dia reports wherein government authorities claimed the attacks targeted combatants. Both events were coded as "battle-related" despite the fact other sources indicated civilians were targeted deliberately, illustrating the legitimacy often afforded to government actors. The GED records *zero* belligerent fatalities in both events—an extreme case of a "highly disproportionate ratio of civilian to military fatalities" that would plausibly constitute an exception to Eck and Hultman (2007, 35)'s rule of coding based on parties' stated intent.

In both cases, government efforts to limit information *access* and *control* narratives of violence sufficiently distorted information outflows to create ambiguity around intent in source reporting. Consequently, civilian fatalities resulting from long-range attacks in both cases were conservatively coded as "battle-related" rather than one-sided violence. Indeed, the Sri Lankan government does not appear as an actor in the OSV in 2009 at all, illustrating how these dynamics can contribute to the omission of entire episodes of civilian killing. Ethiopia illustrates how coding practices that privilege government narratives can result in undercounting civilian deaths in the OSV, even where the relevant government actor appears in the dataset.

Sri Lanka

In the final stages of the conflict between the Sri Lankan government and the Liberation Tigers of Tamil Eelam (LTTE) in 2009, government forces repeatedly shelled civilian targets, including hospitals and their own declared "no-fire zones" (NFZs), where large numbers of civilians had fled, resulting in massive fatalities; a UN report (2011) cited sources estimating that as many as 40,000 civilians may have been killed by government forces in the latter stages of the war. The TMK similarly documents the targeted killing of 40,000 civilians in Sri Lanka in 2009. The GED, on the other hand, records just 6,830 civilian deaths in Sri Lanka in 2009.

Except for 239 deaths attributed to the LTTE, all remaining civilian deaths in Sri Lanka in 2009 in the GED are coded as "battle-related" and are therefore excluded from the OSV altogether. Accordingly, the Sri Lankan government does not appear in the OSV in 2009, leaving approximately 6,191 civilian battle deaths unaccounted for and almost 34,000 further civilian deaths missing altogether. What accounts for these discrepancies?

Credible evidence indicates that the government engaged in deliberate targeting as well as indiscriminate attacks on civilian populations. In May 2009, Human Rights Watch reported, "each time a hospital was established in a new location, the doctors transmitted GPS coordinates of the facility to the Sri Lankan government to ensure that the facility would be protected from military attack. Medical staff said that, on several occasions, attacks occurred on the day after the coordinates had been transmitted."⁹ Likewise, as the International Crisis Group concluded in 2010, the Sri Lankan military "repeatedly shelled the NFZs while making and then failing to respect several public and private commitments to stop using heavy weapons in civilian areas or to allow periods of "safe passage."¹⁰ In the aftermath of the war, deminers working in the NFZs discovered the remnants of cluster munitions,¹¹ a widely banned indiscriminate weapon.

Following the discussion laid out in the previous section, we posit that governments possess unique capabilities to control information outflows by restricting access and shaping media narratives. Indeed, throughout the war in Sri Lanka, the government took extensive steps to shape information about fatalities by restricting access to conflict zones for independent journalists, aid agencies, and human rights organizations. The government also undertook efforts to control information outflows by promulgating narratives denying or legitimating its use of violence against Tamil civilians, by questioning their civilian status, labeling them as "terrorists," and portraying casualties as collateral damage. To these ends, the government employed various coercive tactics, including threatening and arresting journalists who reported on government-perpetrated violence, while also actively cultivating relationships with journalists, including by hosting informal briefings with the president and granting access to reporters "embedded" with military units (Whitaker 2004; Rao and Pradeep 2011; Cronin-Furman 2020).

These efforts intensified in the last phases of the war. In 2009, the government effectively prohibited independent observers from entering the Vanni region, where the most intense violence occurred, and severely limited access to internally displaced persons camps in adjacent regions; as such, information from the Vanni was "extremely limited" at the height of the violence (Human Rights Watch 2009). While sources on the ground, including reporters for Tamilnet.com (a self-described "newswire service that provides up to date news with

- 9 https://www.hrw.org/news/2009/05/08/sri-lankarepeated-shelling-hospitals-evidence-war-crimes.
- 10 https://www.crisisgroup.org/sites/default/files/191war-crimes-in-sri-lanka.pdf pp. 21–22.
- 11 https://www.theguardian.com/world/2016/jun/20/ cluster-bombs-used-sri-lanka-civil-war-leakedphotos-suggest.

Tamil perspective"), provided reports on governmentperpetrated violence against civilians that were cited in international media, government sources regularly dismissed Tamilnet as pro-LTTE,¹² and international outlets historically qualified information sourced from Tamilnet with similar disclaimers (Whitaker 2004).

These dynamics were evident in an event recorded in the GED (ID 76,448), dated May 9-10, 2009, with an estimated 1,000 civilian-but no combatant-fatalities. The GED cites two media reports, published by Reuters (Sirilal 2009) and France24 (2009), as sources for this event. These reports detail civilian deaths resulting from shelling of NFZs in the last strip of LTTE-held territory. Both Sirilal (2009) and France24 (2009) cite Tamilnet sources claiming that civilians were deliberately targeted, but both qualify this reporting as from a "pro-rebel" website. Sirilal (2009) references a government minister's claim that the LTTE was responsible for the shelling, while France24 (2009) quotes a military spokesperson claiming that, "We have not used heavy weapons in the area where the Tamilnet claims civilians have been killed." Allegations of mass atrocities were dismissed by state officials as "baseless accusations whose sole purpose is to discredit Sri Lanka," fabricated by LTTE sympathizers in the international community.¹³

Notwithstanding these conflicting accounts, this event is classified as "state-based conflict" in the GED, and all 1,000 deaths are coded as "battle-related," effectively privileging the government's claims that civilians were not intentionally targeted. The incident illustrates how state perpetrators can leverage *access* restrictions and *control* fatality narratives with greater capacity and legitimacy than non-state actors.

Ethiopia

Credible estimates suggest that the 2020–2022 conflict in Tigray produced a civilian death toll in the hundreds of thousands.¹⁴ The number attributable to direct killings is difficult to estimate due to precisely the dynamics we identify in the Sri Lanka case; however, the most exhaustive research done to date puts this number between 50,000 and 100,000.¹⁵ From the start

- 12 Government-controlled media outlets, on the other hand, are rarely dismissed as "pro-state."
- 13 https://mfa.gov.lk/en/false-allegations-will-be-laidto-rest-with-the-factual-analysis-report-secretarydefence/.
- 14 https://martinplaut.com/2023/05/24/updatedassessment-of-civilian-starvation-deaths-duringthe-tigray-war/.
- 15 https://ethiopiatigraywar.com/methodology.php.

of the conflict in November 2020, the Ethiopian government took extensive steps to *control* information narratives and restrict *access*, shutting down Internet, phone service, and electricity in the region almost immediately (BBC News 2020). Several weeks into the fighting, as members of the victim community and international human rights groups sounded the alarm about massacres and intentional shelling of civilian populations by government forces, Ethiopia's prime minister insisted there had been no civilian casualties (DW 2020). A de facto humanitarian blockade and the expulsion of UN officials and journalists as the conflict unfolded limited access for outside observers (Endeshaw 2021), and government officials continued to deny reports of civilian killings (BBC News 2021).

It is therefore perhaps unsurprising that the GED records orders of magnitude fewer civilian deaths for the conflict than expert consensus currently suggests.¹⁶ Because the conflict is so recent, we expect future updates to the dataset to include significantly higher civilian fatality numbers. However, if we examine the data for one of the events currently included—a government airstrike on Togoga village—we can see evidence of patterns similar to the Sri Lankan case emerging.

In June 2021, a government airstrike on the market in Togoga village (GED ID 404,451) killed 64 civilians. The Ethiopian government's efforts to restrict access and control information flows are reflected in the source reporting for this event. The GED cites reports from Reuters (2021) and Reuters (Mersie and Endenshaw 2021), both of which quote government sources claiming that the airstrike targeted rebel fighters disguised as civilians. However, the GED reports no belligerent fatalities for this event, and other sources—including some quoted in the GED—suggest that civilians were directly targeted. Reuters (Mersie and Endeshaw 2021) quotes a Tigrayan health official stating, "Based on eye witnesses and the health teams we sent the dead, the dead are civilians... It

16 The GED records 5,128 civilian deaths on Ethiopian territory in 2021 in the context of multiple conflicts, including between the Ethiopian government (allied with the Eritrean government) and the Tigrayan People's Liberation Front (TPLF). Of these deaths, 4,407 are coded as one-sided violence, with 3,886 attributed to government forces; the remaining civilian deaths are classified as "battle-related" (Davies et al. 2023). In the latest version of the TMK (version 1.2), there are two events for Tigray in 2021 with 4246 civilian deaths in total. The first event includes the Ethiopian government as the key perpetrator of violence with 3662 deaths estimated, and the second event involves the TPLF with 584 deaths estimated.

Human rights organizations suggest that the type of violence deployed against Togoga was typical. Human Rights Watch research reveals numerous artillery attacks in the early weeks of the conflict that "did not appear aimed at specific military targets but struck generalized populated areas."¹⁷ Similarly, Amnesty International's report on the November 2020 assault on Axum chronicles how the Ethiopian and Eritrean forces used a strategy of indiscriminate shelling in tandem with extrajudicial executions of civilians to capture the city.¹⁸

As with the Sri Lanka case, while the sources disputing the Ethiopian military's account of the Togoga attack may have their own incentives to shape information flows, the classification of civilian deaths in this event as "battle-related" reflects the privileging of state narratives in UCDP source material, even when those narratives are disputed. Because the Ethiopian government disproportionately possesses the capacity to engage in indiscriminate shelling, and because this modality of violence definitionally fails to distinguish between civilian and military targets, ample opportunities are created for government actors to shape information outflows and distort the question of intent. While we expect that future UCDP releases will likely attribute more civilian deaths to one-sided violence by the government, the erasure of one-sided violence in Togoga and similar airstrikes suggests that biases produced by conservative data collection and coding practices will likely ensure that OSV fatality counts continue to underestimate the numbers of civilians killed by intentional government targeting during this conflict.

Conclusion

An aspiration of scientific objectivity underpins data collection efforts like the OSV. Yet, as documented in this note, the OSV—which purports to include "*all* [emphasis added] direct and deliberate killings of civilians" (Eck and Hultman 2007, 233), omits over 180,000 reported civilian deaths, classified instead as "battle-related," in incidents where *no* belligerent fatalities are recorded. As this figure includes only civilian deaths occurring in events with zero belligerent fatalities, it potentially represents a gross underestimate of the "uncounted dead"

- 17 https://www.hrw.org/news/2021/02/11/ethiopiaunlawful-shelling-tigray-urban-areas.
- 18 https://www.amnesty.org/en/documents/afr25/3730/ 2021/en/.

omitted from the OSV. Indeed, over 240,000 civilian deaths in the GED are coded as "battle-related."

While some civilian deaths omitted from the OSV may indeed have occurred "in the crossfire" of attacks against military targets, the exclusion of civilians killed in long-range attacks, including airstrikes and shelling, is not mere coincidence. Rather, coding rules *definitionally* categorize civilian deaths in such attacks as "battle-related"—and therefore as incidental or "collateral damage"—rather than intentional, one-sided violence. Because states disproportionately possess both the military *capabilities* to conduct such attacks and the ability to restrict information *access* and exert *control* over conflict narratives, OSV coding protocols necessitate the underreporting of state-perpetrated civilian killings.

We offer various future directions for research that could attend to some of these biases. The first recommends revisions to UCDP's coding protocols; the second involves revisiting extant scholarship utilizing UCDP through other UCDP datasets; the third involves triangulating the UCDP data family alongside other conflict datasets; and the fourth focuses on the potential for methodological innovation using techniques such as latent variable analysis.

First, the analysis presented in this article suggests there is scope for improving existing definitions, data sources, and coding protocols to mitigate state-centric bias in the coverage of civilian fatalities. In particular, we suggest that UCDP's definition of "one-sided violence" might be revised to include civilians killed in "indiscriminate" attacks against civilian population centers, or alternatively, UCDP could include a separate variable explicitly measuring "indiscriminate" civilian deaths in such attacks, distinct from "battle-related" fatalities. Insofar as states disproportionately possess the military capabilities to conduct such attacks, this would reduce statecentric bias and, more generally, provide a clearer picture of the dynamics of civilian victimization. We also suggest the need for clearer protocols for drawing inferences about intent when coding individual events involving civilian deaths; such protocols are particularly important for events where source reporting presents conflicting accounts, to avoid privileging narratives by actors (generally states) with disproportionate ability to shape information outflows.

Second, given that the OSV systematically undercounts violence against civilians by states, previous research based on this dataset—including findings that non-state actors have been responsible for most civilian killings in recent years (Davies et al. 2023, 697) might require revisiting. Extant theorizations of how governments and insurgent groups use civilian targeting to shape the strategic landscape (Wood, Kathman and Gent 2012) could be enriched with our theoretical intervention that certain tactics are more accessible to states (i.e., longrange bombardment). As a starting point, researchers might consider replicating such studies and other findings using data that incorporates "battle-related" civilian deaths in total civilian fatality counts and/or separately analyzing trends and determinants of "battle-related" civilian deaths. To this end, existing UCDP data might be leveraged to compile actor-year-level data on "battlerelated" civilian deaths by reviewing individual events in the GED to identify responsible actors. For this purpose, the newly released UCDP Country-Year Dataset on Organized Violence within Country Borders (Davies et al. 2024) provides a potentially useful model, as it reports annual counts of civilian death including both onesided and battle-related fatalities, although it does not attribute "battle-related" deaths to specific conflict actors and is therefore not directly conducive to replicating actor-level analyses. In principle, the GED might also be employed to generate aggregate data on the modalities of "battle-related" deaths, including airstrikes, artillery shelling, and other long-range attacks, which could provide insight into the frequency of "indiscriminate" violence against civilians. In this respect, the descriptive statistics on drone strikes reported by Davies et al. (2022), based on GED version 22.1, are a potentially useful starting point, although there is currently no publicly available, cross-national UCDP dataset measuring modalities of violence.19

Beyond considering what and how many "battlerelated" deaths are associated with state actors alone, future research could also consider utilizing the UCDP's External Support Dataset to examine whether cases where conflicts involving long-range bombardment and civilian-battle-related deaths are more broadly sponsored by state or non-state actors (and through what means). More broadly, efforts could analyze what patterns exist between rates of civilian battle deaths and certain (types of) external support actors.

19 The UCDP Country-Year Dataset on Organized Violence within Country Borders (Davies et al. 2024), which was first released in 2024, also provides a potentially useful resource, as this provides separate country-year estimates of civilians killed in both one-sided violence and battle events. However, civilian deaths in battle events are not attributed to specific actors (as they are coded in the relevant state-based or non-state conflict dyad); as such, this dataset still tends to obscure responsibility for civilian battle-related deaths.

We do not suggest that other datasets are immune to the dynamics we emphasize here. On the contrary, we anticipate all secondary source data relying on media reporting is vulnerable to similar statist bias. Our insights, therefore, also potentially extend to critically examining other datasets on civilian targeting, such as the TMK. The TMK reports that electoral autocracies and electoral democracies are "more prone to TMKs perpetrated by non-government actors," whilst "regime type is not significantly associated with TMKs committed by governments" (Butcher et al 2020, 1541). Cross-examining the missing cases in this article against the UCDP external support dataset, further work could explore whether Western/US intervention has a correlative effect on civilian deaths being defined as "battle-related," thereby complicating extant categorizing of violence committed by autocracies and democracies more broadly. This would allow us to further theorize the conditions under which statist bias may vary in magnitude and the pathways through which it manifests.

Third, moving beyond the UCDP data family, other conflict datasets may be useful for capturing dynamics obscured by UCDP's current coding rules. Of particular relevance, the Armed Conflict Location Events Dataset (ACLED, Raleigh et al. 2010) includes a category for "explosions/remote violence," defined as "incidents in which one side uses weapon types that, by their nature, are at range and widely destructive" (ACLED 2023, 15) This category includes artillery and air and drone strikes, inter alia, and ACLED additionally tags such events as "Civilians targeted" in cases where "civilians were the main or only target" (ACLED 2023, 15). However, ACLED's temporal coverage is currently more limited than UCDP's,²⁰ and as an events-based dataset, it is potentially subject to similar biases arising from using media and other thirdparty sources to identify and code individual events.²¹ As an alternative, the TMK and other datasets that identify broader episodes of violence-rather than aggregating from individual events-might be useful for capturing civilian killings omitted from the OSV due to restrictive coding rules.

Lastly, in addition to leveraging existing datasets, researchers might also consider replicating and extending previous studies using estimates of civilian deaths derived from latent variable models that triangulate information

- 20 For example, as of July 2024, ACLED does not cover Sri Lanka in 2009.
- 21 See Eck (2012) for a comparison of UCDP and ACLED, and ACLED (2023) for an overview of data collection procedures, including sourcing and protocols to minimize bias.

from multiple sources to produce more robust estimates of difficult-to-measure phenomena (Fariss and Lo 2020). Notably, Fariss et al.'s (2020) latent variable estimates of government one-sided violence, derived from various indicators of state repression, include several cases involving aerial bombing campaigns against civilians that are omitted entirely from the OSV (e.g., Government of Sudan, 2009).

We thus urge users of these data to attend to the above considerations when making claims based on them, alongside recommending that UCDP coding protocols be revised to include civilians killed in indiscriminate attacks, particularly when resulting from artillery fire that does not strike military targets or records no belligerent deaths. While seeking to methodologically innovate with regard to generating hard-to-observe data, we echo invitations from Hoover Green and Cohen (2021) that scholars attune themselves to a range of ethical issues affecting "desk research" on conflict, which are not exempt from the ethical complexities facing human subjects research. As this article has theorized with reference to civilian fatalities, such issues include how widely used data may reproduce and amplify statist bias in conflict reporting, unintentionally laundering incidents of stateperpetrated violence.

Supplemental Data

Supplementary information is available at the *Journal of Global Security Studies* data archive.

References

- Armed Conflict Location and Events Data Project 2023. Codebook. Madison: ACLED. Accessed July 12, 2024. www.acle ddata.com.
- Barkawi, T. 2011. "From War to Security: Security Studies, the Wider Agenda and the Fate of the Study of War." *Millennium: Journal of International Studies* 39 (3): 701–16.
- BBC News 2020. "Tigray Crisis: Ethiopia Orders Military Response after Army Base Seized." November 4. https://www.bb c.com/news/world-africa-54805088.
- . 2021. "Ethopia's Tigray Crisis: How a Massacre in the Sacred City of Aksum Unfolded." February 26. https://www.bbc.com/news/world-africa-56198469.
- Besaw, C., K. Ritter, and G.M. Tezcür 2023. "Beyond Collateral Damage: The Politics of Civilian Victimization in a Civil War." *Global Studies Quarterly* 3 (3): ksad050.
- Butcher, C., B. Goldsmith, S. Nanlohy, A. Sowmya, and D. Muchlinksi 2020. "Introducing the Targeted Mass Killings Dataset for the Study and Forecasting of Mass Atrocities." *Journal of Conflict Resolution* 64 (7-8): 1524–47.

- Campbell, S., D. Chandler, and M. Sabaratnam 2011. A Liberal Peace?: The Problems and Practices of Peacebuilding. London: Zed.
- Clark, A.M., and K. Sikkink 2013. "Information Effects and human Rights Data: Is the Good News about Increased human Rights Information Bad News for human Rights Measures?" *Human Rights Quarterly* 35 (3): 539–68.
- Cohen, D. K., and R Nordås. 2014. "Sexual Violence in Armed Conflict: Introducing the SVAC Dataset, 1989–2009." *Journal* of Peace Research 51 (3): 418–28.
- Cronin, B. 2013. "Reckless Endangerment Warfare: Civilian Casualties and the Collateral Damage Exception in International Humanitarian Law." *Journal of Peace Research* 50 (2): 175–87.
- Cronin-Furman, K. 2020. "Human Rights Half Measures: Avoiding Accountability in Postwar Sri Lanka." World Politics 72 (1): 121–63.
- Davenport, C., and P. Ball 2002. "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan state Terror, 1977–1994." *Journal of Conflict Resolution* 46 (3): 427–50.
- Davies, S., G. Engström, T. Pettersson, and M. Öberg 2024. "Organized Violence 1989–2023, and the Prevalence of Organized Crime Groups." *Journal of Peace Research* 61 (4): 673–93.
- Davies, S., T. Pettersson, and M. Öberg 2022. "Organized Violence 1989–2021 and Drone Warfare." *Journal of Peace Research* 59 (4): 593–610.
- . 2023. "Organized Violence, 1989–2022, and the Return of Conflict between States." *Journal of Peace Research* 60 (4): 691–708.
- Dawkins, S. 2021. "The Problem of the Missing Dead." *Journal* of Peace Research 58 (5): 1098–116.
- Dietrich, N., and K. Eck 2020. "Known Unknowns: Media Bias in the Reporting of Political Violence." *International Interactions* 46 (6): 1043–60.
- DW. 2020. Tigray leader urges PM Abiy to 'stop the madness.' November 30. https://www.dw.com/en/ethiopia-tigray-leader -urges-pm-abiy-to-stop-the-madness/a-55779609.
- Eck, K. 2012. "In Data We Trust? A Comparison of the UCDP GED and ACLED Conflict Events Datasets." *Cooperation and Conflict* 47 (1): 124–41.
- Eck, K., and L. Hultman 2007. "Violence against Civilians in War: Insights from New Fatality Data." *Journal of Peace Research* 44 (2): 233–46.
- Endeshaw, D.2021. "Ethiopia Expels Seven U.N. officials, Accusing Them of 'Meddling'." *Reuters*, October 6. Accessed July 12, 2024. https://www.reuters.com/world/africa/e thiopia-expels-seven-senior-un-officials-2021-09-30/.
- Fariss, C.J., and J Lo. 2020. "Innovations in concepts and measurement for the study of peace and conflict." *Journal of Peace Research* 57 (6): 669–678.
- Fariss, C.J., M. Kenwick, and K. Reuning 2020. "Estimating One-Sided Killings from a Robust Measurement Model of Human Rights." *Journal of Peace Research* 57 (6): 801–14.
- Fjelde, H., L. Hultman, L. Schubiger, L. E. Cederman, S. Hug, and M Sollenberg. 2021. "Introducing the Ethnic One-sided

Violence Dataset." Conflict Management and Peace Science 38 (1): 109–26.

- France24. 2009. Sri Lanka rebels say 250 civilians dead in govt attack. Accessed March 8, 2009. https://www.france24.com/en/20090308-rebel-counter-o ffensive-leaves-hundreds-tamils-dead-.
- Gibilisco, M., and J. Steinberg 2022. "Strategic Reporting: A Formal Model of Biases Conflict Data." American Political Science Review.
- Gordon, E. 2014. "Security Sector Reform, Statebuilding and Local Ownership: Securing the state or Its People?" *Journal of Intervention and Statebuilding* 8 (2–3): 126–48.
- Haas, F., and N. Ansorg. 2018. ""Better peacekeepers, better protection? Troop quality of United Nations peace operations and violence against civilians."." *Journal of Peace Research 55 (6)*: 742–758.
- Hendrix, C.S., and I. Salehyan 2015. "No News is Good News: Mark and Recapture for Event Data when Reporting Probabilities Are Less than One." *International Interactions* 41 (2): 392–406.
- Högbladh, S. 2023. UCDP GED Codebook, version 23.1. Uppsala: Department of Peace and Conflict Research, Uppsala University.
- Hoover Green, A., and D.K. Cohen 2021. "Centering human Subjects: The Ethics of "Desk Research on Political Violence." *Journal of Global Security Studies* 6 (2): ogaa029.
- Human Rights Watch. 2009. "War on the Displaced: Sri Lankan Army and LTTE Abuses against Civilians in the Vanni." 19 February. Accessed July 12, 2024. https://www.hrw.org/report/2009/02/19/war-displaced/sr i-lankan-army-and-ltte-abuses-against-civilians-vanni ______. 2019. World Report 2019: Events of 2018.
- Jones, S.G., and P. Johnston 2013. "The Future of Insurgency." Studies in Conflict and Terrorism 36 (1): 1–25.
- Kinsella, H.M. 2011. The Image before the Weapon: a Critical History of the Distinction between Combatant and Civilian. Ithaca: Cornell University Press.
- MacGinty, R., and P. Firchow 2016. "Top-down and Bottomup Narratives of Peace and Conflict." *Politics* 36 (3): 308–23.
- Mersie, A., and D. Endeshaw 2021. "Death Toll from Tigray Air Strike at over 50, Health Official Says." *Reuters*, June 24. Accessed July 12, 2024. https: //www.reuters.com/world/africa/death-toll-tigray-air-str ike-over-50-health-official-says-2021-06-24/
- Parkinson, S. 2024. "Unreported Realities: The Political Economy of Media-sourced Data." *American Political Science Review* 118 (3): 1527–32.
- Pettersson, T. 2023a UCDP Battle-related Deaths Dataset Codebook, Version 23.1. Uppsala: Department of Peace and Conflict Studies, Uppsala University.
- ——. 2023b UCDP Non-state Conflict Dataset Codebook, Version 23.1 Uppsala: Department of Peace and Conflict Research, Uppsala University.
- ——. 2023c UCDP One-sided Violence Dataset Codebook, Version 23.1. Uppsala, Sweden: Department of Peace and Conflict Research, Uppsala University.

- Qiu, X. 2022. "Rebel Strategies and the Prospects for Peace." *American Journal of Political Science* 66 (1): 140–55.
- Raleigh, C., r. Linke, H. Hegre, and J. Karlsen 2010. "Introducing ACLED: An Armed Conflict Location and Events Dataset." *Journal of Peace Research* 47 (5): 651–60.
- Rao, R., and N'Weerasinghe Pradeep 2011. "Covering Terrorism: Examining Social Responsibility in South Asian Journalism." *Journalism Practice* 5 (4): 414–28.
- Reuters. 2021. Ethiopian military says no civilians hit in latest Tigray air strike. June 24. https://reuters.screenocean.com/re cord/1623037.
- Salehyan, I., D. Siroky, and R. M Wood. 2014. "External Rebel Sponsorship and Civilian Abuse: A Principal-agent Analysis of Wartime Atrocities." *International Organization* 68 (3): 633–61.
- Schneider, G., and M Bussmann. 2013. "Accounting for the Dynamics of One-sided Violence: Introducing KOSVED." Journal of Peace Research 50 (5): 635–44
- Sirilal, R 2009. "U.N. blasts Sri Lanka Weekend Attack as 'Bloodbath.'" *Reuters*, May 11. https://jp.reuters.com/article /idUSCOL273563/.
- Sjoberg, L. 2016. "Centering Security Studies around Felt, Gendered Insecurities." *Journal of Global Security Studies* 1 (1): 51–63.
- Sundberg, R., K. Eck, and J. Kreutz 2012. "Introducing the UCDP Non-State Conflict Dataset." *Journal of Peace Research* 49 (2): 351–62.

- Sundberg, R., and E. Melander 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523–32.
- Sutton, J., C. R. Butcher, and I Svensson. 2014. "Explaining Political Jiu-jitsu: Institution-building and the Outcomes of Regime Violence against Unarmed Protests." *Journal of Peace Research* 51 (5): 559–73.
- Ulfelder, Jay, and B. Valentino 2008. "Assessing Risks of State-sponsored Mass Killing." Social Science Research Network SSRN-ID_23003048. https://papers.ssrn.com/sol3/pap ers.cfm?abstract_id=1703426
- United Nations 2011. "Report of the Secretary-General's Panel of Experts on Accountability in Sri Lanka.United Nations. Accessed April 11, 2025. www.securitycouncilreport.org/at f/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7 D/POC%20Rep%20on%20Account%20in%20Sri%20La nka.pdf.
- Weber, M. 1965. Politics as a Vocation. Philadelphia: Fortress Press.
- Weidmann, N. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." American Journal of Political Science 60 (1): 206–18.
- Whitaker, M.P. 2004. "Tamilnet.Com: Some Reflections on Popular Anthropology, Nationalism, and the Internet." *Anthropologcal Quarterly* 77 (3): 469–98.
- Wood, R. M., J. D. Kathman, and S. E Gent. 2012. "Armed Intervention and Civilian Victimization in Intrastate Conflicts." *Journal of Peace Research* 49 (5): 647–60.

[©] The Author(s) (2025). Published by Oxford University Press on behalf of the International Studies Association. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com