Sexual violence in armed conflict: Introducing the SVAC dataset, 1989–2009

Dara Kay Cohen

Harvard Kennedy School

Ragnhild Nordås

Peace Research Institute Oslo (PRIO)

Abstract

Which armed groups have perpetrated sexual violence in recent conflicts? This article presents patterns from the new Sexual Violence in Armed Conflict (SVAC) dataset. The dataset, coded from the three most widely used sources in the quantitative human rights literature, covers 129 active conflicts, and the 625 armed actors involved in these conflicts, during the period 1989–2009. The unit of observation is the conflict-actor-year, allowing for detailed analysis of the patterns of perpetration of sexual violence for each conflict actor. The dataset captures six dimensions of sexual violence: prevalence, perpetrators, victims, forms, location, and timing. In addition to active conflict-years, the dataset also includes reports of sexual violence committed by conflict actors in the five years post-conflict. We use the data to trace variation in reported conflict-related sexual violence over time, space, and actor type, and outline the dataset's potential utility for scholars. Among the insights offered are that the prevalence of sexual violence varies dramatically by perpetrator group, suggesting that sexual violations are common – but not ubiquitous. In addition, we find that state militaries are more likely to be reported as perpetrators of sexual violence than either rebel groups or militias. Finally, reports of sexual violence continue into the post-conflict period, sometimes at very high levels. The data may be helpful both to scholars and policymakers for better understanding the patterns of sexual violence, its causes, and its consequences.

Keywords

armed conflict, rape, sexual violence

Conflict-related sexual violence is now widely acknowledged as a problem of international security, and a potential weapon of war, ethnic cleansing, and genocide (e.g. Bloom, 1999; Carpenter, 2006; Cohen, 2013; Farr, 2009; Leiby, 2009; Skjelsbæk, 2011; Wood, 2006). Until recently, political scientists have tended to overlook or to minimize sexual violence, instead analyzing other types of violence, especially homicide (Cohen, 2013). Recent research shows that sexual violence likely occurs in all conflicts, but with immense variation in form and severity (Wood, 2010). However, most existing studies are case studies of specific conflicts where widespread violations are believed to have occurred. Many of these analyses select on the dependent variable and are not comparative in nature; there is little exploration of 'negative cases' where sexual violence has *not* occurred (Wood, 2006, 2009, 2010). A systematic comparison of conflicts with reports of massive sexual violence to those with little or no sexual violence could illuminate causal mechanisms and root causes (Wood, 2006, 2009). However, a lack of reliable cross-national data has hampered the quantitative study of wartime sexual violence (Cohen, 2013). Such data can also be a critical tool to improve policy initiatives geared towards decreasing sexual violence prevalence and mitigating its effects (Cohen & Hoover Green, 2012).

The Sexual Violence in Armed Conflict (SVAC) dataset is a systematic dataset on sexual violence during

Corresponding author:
ragnhild@prio.no

all 129 conflicts active in the period 1989–2009,[1] and the immediate post-conflict period. The dataset features annual data at the level of the armed actor, including all 625 active state militaries, pro-government militias (PGMs) and rebel groups.[2] The article proceeds as follows. The first section describes the SVAC dataset, including the definitions and the scope. In the second section, we discuss the sources used to code the data, as well as the coding rules and procedures. We also examine issues of reliability and validity, and how the SVAC data differ from other existing quantitative datasets on related themes. In the third section, we present a series of descriptive statistics. Finally, we outline research uses for the SVAC data.

The SVAC dataset: Definitions and scope

The definition of sexual violence used in the SVAC dataset builds on the International Criminal Court (ICC) definition, and includes (1) rape, (2) sexual slavery, (3) forced prostitution, (4) forced pregnancy, and (5) forced sterilization/abortion (ICC, 2000). Following Wood (2009), we also include (6) sexual mutilation, and (7) sexual torture.[3] Importantly, the definition is gender neutral and does not preclude the existence of female perpetrators or male victims. We focus on behaviors that involve direct force or physical violence. The definition reflects current legal understandings, but does not include acts such as sexual humiliation, sexualized insults or forced undressing, which some scholars have included in their definition (e.g. Leiby, 2009).

Scope
The SVAC dataset includes sexual violence by all conflict actors involved in intrastate, internationalized internal, and interstate conflicts in the period 1989–2009, as defined by Gleditsch et al. (2002) and Harbom, Melander & Wallensteen (2008).[4] 1989 is the conventional starting year for key datasets developed in recent years (e.g. Eck & Hultman, 2007), in part because data on earlier years are deemed more uncertain and sources are limited. We include reports of sexual violence by three actor types: (1) state forces, (2) rebel groups (both from Harbom, Melander & Wallensteen, 2008), and (3) PGMs (Carey, Mitchell & Lowe, 2013).[5] In addition to active conflict-years, we include up to five years in between active conflict-years – called 'interim years' – when lethal violence drops below 25 battle deaths but increases again before five years have passed. Finally, we include the first five post-conflict years after the last active year in the relevant conflict dyad.

Dimensions of sexual violence
Six dimensions of sexual violence are included: (1) prevalence, (2) perpetrators, (3) targeting, (4) form, (5) location, and (6) timing. The inclusion of these dimensions increases the specificity of the data beyond only measuring its occurrence, and allows for testing of features of sexual violence that are particularly relevant in the public debate and academic literature. Below, we discuss how each dimension is measured.[6]

Prevalence is measured as an ordinal scale estimate from 0 to 3, adapted from Cohen (2010, 2013), and presented in Table I. It captures the reported severity of sexual violence perpetration by an armed actor in a given year. The SVAC dataset does not include numerical estimates of sexual violence incidents or victims, due to data limitations and validity concerns (Peterman et al., 2011).[7]

Perpetrators include only organized armed actor groups, the main protagonists of armed conflicts – state forces, rebel groups or PGMs. We do not code perpetrator data at the individual level (e.g. a particular commander in an army), nor domestic violence, intimate partner violence, violence by peacekeepers or aid workers, or violence by civilians; collecting systematic and reliable data on these types of violence is not currently possible at the level and scale required for the SVAC dataset. There are, however, likely to

---

[1] See the Codebook and User Guide for detailed descriptions of scope, coding rules, and definitions, available at http://www.prio.no/jpr/datasets and http://www.sexualviolencedata.org. Active conflicts are defined according to the UCDP Dyadic Conflict Dataset v. 1-2010 (Harbom, Melander & Wallensteen, 2008). The dataset also contains post-conflict observations for conflicts that were active between 1984 and 1988; however, these observations are excluded from the discussion and analyses in the current article (n = 7,286).
[2] PGMs are defined as a group that is (1) pro-government or sponsored by the government (national or subnational), (2) not a part of the regular security forces, (3) armed, and (4) organized to some degree (Carey, Mitchell & Lowe, 2013: 250). Armed actors are defined by Harbom, Melander & Wallensteen (2008) and Carey, Mitchell & Lowe (2013).
[3] See the Online Appendix for details on the constitutive parts of the definition.

[4] The Carey, Mitchell & Lowe (2013) data end in 2007, and PGMs are therefore only included up to that point.
[5] PGMs are only included if they are involved in armed conflicts in the UCDP dyadic dataset.
[6] More detail on the coding is available in the Codebook and User Guide, available at http://www.prio.no/jpr/datasets and http://www.sexualviolencedata.org.
[7] However, in the absence of descriptive terms, coders used the number of reported incidents or victims to code prevalence (see Table I).

Table I. Summary of coding rules for reported prevalence of sexual violence

---

Prevalence level = 3 (Massive)
Sexual violence is likely related to the conflict, and:
- Sexual violence was described as 'systematic', 'massive', or 'innumerable'.
- Actor used sexual violence as a 'means of intimidation', 'instrument of control and punishment', 'weapon', 'tactic to terrorize the population', 'terror tactic', 'tool of war', on a 'massive scale'.
- A description of 1,000 or more victims of sexual violence in a given year.

Prevalence level = 2 (Numerous)
Sexual violence is likely related to the conflict, but did not meet the requirements for a 3 coding, and:
- Sexual violence was described as 'widespread', 'common', 'commonplace', 'extensive', 'frequent', 'often', 'persistent', 'recurring', a 'pattern', a 'common pattern', or a 'spree'.
- Sexual violence occurred 'commonly', 'frequently', 'in large numbers', 'periodically', 'regularly', 'routinely', 'widely', or on a 'number of occasions'; there were 'many' or 'numerous instances'.
- A description of 25–999 victims of sexual violence in a given year.

Prevalence level = 1 (Isolated)
Sexual violence is likely related to the conflict, but did not meet the requirements for a 2 or 3 coding, and:
- There were 'reports', 'isolated reports', or 'there continued to be reports' of occurrences of sexual violence.
- A description of 1–25 victims of sexual violence in a given year.

Prevalence level = 0 (None) Report issued, but no mention of sexual violence related to the conflict-actor-year.

'Likely related to the conflict' refers to events that seem to pertain to the conflict in question. The sources used to code the data also contain reports of sexual violations that are likely *not* related to the conflict, such as police abuse of civilians, child abuse or intimate partner violence.

be important linkages between conflict-related sexual violence as captured by the SVAC dataset and these other forms of sexual violence and sexual abuse (e.g. Nordås & Rustad, 2013; Peterman, Palermo & Bredencamp, 2011), and the SVAC data could be used in future research to analyze these connections.

Targeting of sexual violence can be directed towards particular groups (non-random/selective) or may be random/indiscriminate. We code a dummy variable to indicate whether the targeting is reported to be selective, as well as a series of subsequent variables indicating what types of groups are reportedly being targeted, including ethnic groups, religious groups, nationality groups, age groups, if victims are reported to be selected by assumed or real collaboration or affiliation with a fighting faction, or if there is another targeting logic reported for the given conflict-actor-year. Finally, we code a variable for whether there was no, some, or significant levels of sexual violence reported against males, children, detainees, and refugees, respectively – all victim categories of particular policy relevance.

Forms of sexual violence captured in the SVAC dataset are those included in the aforementioned definition of sexual violence. These forms are not mutually exclusive, as an armed actor may commit more than one form of sexual violence in a given year. In addition, we code a set of dummy variables indicating (a) whether there were reports of *gang rape* (rape by multiple perpetrators), (b) reports of *sexual violence by proxy*, where an armed actor has forced

someone to perpetrate sexual violence on her/himself or any third party, possibly to humiliate and terrorize both the perpetrator and victim(s),[8] and (c) whether *witnesses* were present at reported incidents of sexual violence, as well as which types of witnesses (family members, other victims, members of armed groups, or others).

Location is a text variable of keywords related to location found in the source material. Additionally, we include a series of dummy variables – which are not mutually exclusive – for frequently reported types of locations, due to their theoretical and policy relevance, including whether attacks occurred in or around a refugee/IDP camp, at a checkpoint, in a detention facility, in a private home/office, or in a school.[9]

Timing of sexual violence can be important for understanding the role sexual violence plays in the dynamics of war. For the SVAC dataset, data limitations preclude temporal specificity beyond the conflict-year. However, we code a text variable of keywords related to more fine-grained timing of sexual violence. In addition, five timing variables capture the month(s) of attacks (if reported), as

---

[8] Excluded from this variable are cases when commanders were reported to order soldiers to commit acts of sexual violence.
[9] The level of detail about location varies significantly in the sources used for data collection. The location variables are best utilized for contextual purposes. See the Codebook and User Guide for more information.

well as whether sexual violence is reportedly occurring immediately before, during, or after military operations, a political event, an errand or chore, or during the search of a private space, such as a home or office.[10]

Data sources, biases, and limitations

The SVAC dataset relies on the most commonly used sources in quantitative human rights scholarship: the US State Department, Amnesty International, and Human Rights Watch. The US State Department Country Reports on Human Rights Practices are published annually, and summarize the human rights record for every country, except the USA. Amnesty International (AI) and Human Rights Watch (HRW) both publish two types of reports that we use as sources: an annual report by country and periodic special reports by country and/or by human rights issue. We rely on these three sources because they provide annual global coverage, and are widely considered trustworthy and reliable sources for data on human rights violations.[11] These sources have been used by numerous other human rights data projects that provide the basis for a large number of quantitative human rights studies, including the CIRI Human Rights Data Project (Cingranelli & Richards, 1999) and the Political Terror Scale (Wood & Gibney, 2010). Like these other data collections, the SVAC dataset codes *reporting* by the three sources, which may be biased for a variety of reasons.

Possible biases
Underreporting by victims is the most frequently cited potential bias in large-scale data collection on sexual violence. Researchers often treat any estimation of the number of victims as conservative, assuming that many victims were either unable or unwilling to report. Reasons for victims not reporting sexual violence include the fear of stigmatization, shame, the fear of retributive violence, and the inability to reach authorities (Green, 2006; Wood, 2006; Leiby, 2009). Underreporting may be particularly prevalent where pregnancy outside wedlock is stigmatized and abortion is illegal (Wood, 2006), and in contexts where patriarchal norms are strong and virginity

is highly prized (Green, 2006). In addition, victims and witnesses of sexual violence may not survive the assault or the war in order to report the violation.

Agencies that collect reports of violence may focus on particular victim categories, which may result in systematic underreporting, for example, of male victims by organizations focused on violence against women (Carpenter, 2006). Male victims were reported in 33 separate conflicts, and in less than 1% (72 of 7,286) of the observations in the SVAC dataset. Sexual violence against men is likely to be especially under-reported, perhaps due to the minimal focus on male victims by NGOs and the policy community, as well to the severe stigma associated with reporting such violence (Sivakumaran, 2007). It is possible that the sources used for the SVAC dataset are biased against reporting of male victims, but the observations where male victims have been reported could serve as a starting point for future research.

Wartime, however, does not necessarily decrease the reporting of sexual violence. In some contexts, human rights groups and medical service organizations may actually be more accessible to war-affected populations than they were during peacetime (Wood, 2006). Scholars have also documented cases where armed groups, victims, advocates, and NGOs have sensed an advantage to emphasizing, or perhaps exaggerating, certain forms of violence in order to receive aid or donor funds (Cohen & Hoover Green, 2012; Peterman, Palermo & Bredencamp, 2011; Utas, 2005). In addition, as international attention to conflict-related sexual violence has increased in recent years, detection and reporting may also have increased as more resources were devoted to documentation and mitigation efforts (Cohen, Hoover Green & Wood, 2013). This could potentially introduce a temporal bias in the data, but one that is difficult to evaluate or correct.

Finally, biases may result from the process of data coding. Hathaway & Ho (2004) argue that errors can result from translating qualitative reports of difficult-to-measure data on human rights abuses into a quantitative dataset. In addition, Clark & Sikkink (2013) find that research assistants tend to code worse violations if reports are longer, regardless of the actual content of the report. Careful training of the coding team can address some of these potential biases.

Strategies to limit biases in the data/coding
We employed four measures to limit biases. First, we used multiple data sources as a form of triangulation to account for reporting biases from any one source. The

---

[10] As with location, the text variables are best utilized for contextual purposes (see Codebook and User Guide).

[11] Alternative sources of data, such as household surveys, are often limited in their coverage of countries and years, and do not inquire about the armed group affiliations of perpetrators, making them less useful for the SVAC dataset. See the Codebook and User Guide for additional details on other potential data sources.

correlations between the prevalence variables from the three main sources are between 0.47 and 0.50 (see Online Appendix). These correlations are reasonably high – but also suggest that the three sources are not simply an echo chamber, and are reporting different levels of violence by armed actors. Second, we tested an alternative data collection process on a subset of cases. For these cases, the coders performed a comprehensive search of all available online sources, including policy reports and academic articles. This considerably more time-consuming data collection process did not yield significant additional codeable information, nor did it reveal systematic biases in the information coded only from the three main sources. Third, to ensure intercoder reliability, we conducted detailed annual trainings of our coding team, and held weekly meetings between the coders and the project management. In addition, as part of our pilot study, we conducted intercoder reliability testing; the results increased confidence in the clarity and rigor of the codebook.[12] Finally, all coding decisions are documented in standardized 'Conflict Manuscripts' that make transparent the coding process and are available by request from the dataset authors.

Limitations

While the SVAC dataset follows the current recommended best practices for coding violence at the most disaggregated unit of analysis (Davenport & Moore, 2013), this strategy has a number of limitations. We have employed a conservative coding protocol, such that a source must identify the armed group by name and at least the year of the reported violation, in order to be coded. It was not possible to include in the conflict-actor-year structure those cases where the available information is not sufficiently specific about which armed group the reported perpetrators were, or in what year the reported violation occurred. In addition, the SVAC dataset is not geo-referenced and does not directly allow for spatial analysis of sexual violence. Although it may be possible to construct spatially disaggregated measures of some aspects of sexual violence for a limited set of cases, such a data collection was not deemed feasible or reliable for a global dataset.

Related quantitative datasets

The SVAC dataset builds on an existing data collection by Cohen (2010, 2013), the first systematic effort to compare the incidence and intensity of wartime rape across civil wars globally, by state and non-state actors. Cohen's (2013) data, coded from the US State Department reports, covers the 86 major civil wars between 1980 and 2009 (75 of which occur during the SVAC study period) and is based on cases from Fearon & Laitin (2011), an updated list of the Fearon & Laitin (2003) civil war cases. Cohen codes the relative prevalence of wartime rape (*not* sexual violence more broadly) on a four-point scale on the level of the group type-conflict-year. In contrast to the SVAC dataset, Cohen (2013) does not disaggregate rape perpetrated by PGMs, nor does the dataset include post-conflict years, interstate conflicts or low-intensity conflicts.

Basic patterns, both in terms of variation in perpetrating actors and temporal change, are similar across the datasets. A replication of the conflict-level analysis in Cohen (2013) largely confirms the findings in Cohen (2013) (see the Online Appendix for the results and discussion). However, the percentage of actors reported as perpetrators of wartime rape in Cohen (2013) is higher than in the SVAC dataset, for one main reason: Cohen (2013) codes reports of rape by armed group *type* (state actors or rebel actors), not by named armed actor; in contrast, the SVAC dataset requires the specific armed group to be named (e.g. 'rebels' reported as perpetrators would be coded in Cohen, but not in the SVAC dataset).

In addition to Cohen (2013), there are at least four existing cross-national studies of wartime rape or sexual violence.[13] Each of these has limitations in its utility for analyzing wartime sexual violence.[14] Green's (2006) dataset of 'collective rape' coded from newspaper articles covers 37 countries, but the data are not collected by conflict. Butler, Gluch & Mitchell's (2007) study of global variation in sexual violence by state security forces is limited to a single year (2003) and does not include violence by non-state actors. Farr (2009) focuses on 'extreme war rape' in 27 countries with recent conflicts, but does not include countries with more limited reported rape, or the post-conflict period. Finally, a report by Bastick, Grimm & Kunz (2007) includes 51 countries in an analysis of sexual violence in recent armed conflicts, but the data are not systematically collected for each variable of interest.

---

[12] See the Online Appendix for additional details on the testing and the results.

[13] See Cohen (2013) for a longer discussion of each study.
[14] See the Codebook and User Guide for a comparison to additional data projects, including the WomanStats project.
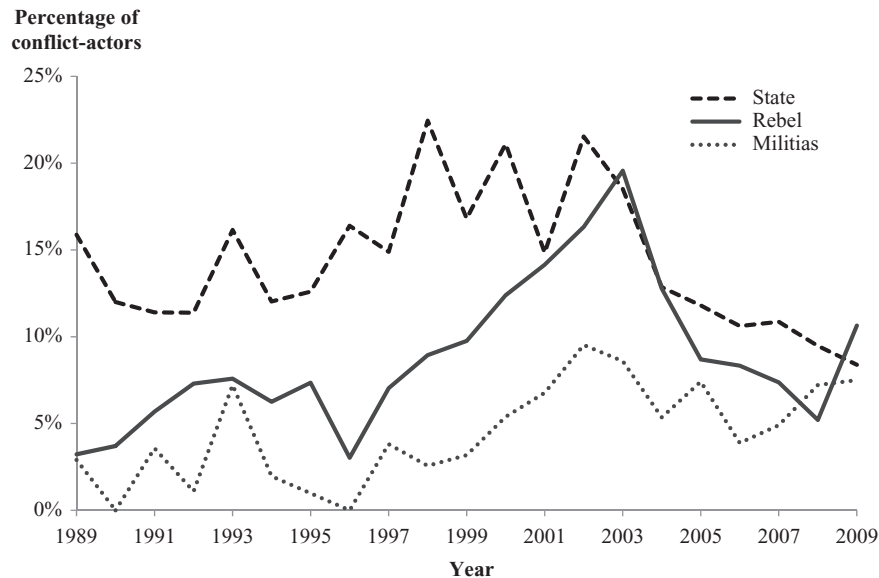
Figure 1. Percentage of conflict actor types reported as perpetrators of sexual violence

Descriptive statistics

In this section, we present key descriptive statistics of reports of sexual violence in the SVAC dataset: (a) variation by country and conflict for the highest reported prevalence; (b) variation over time for the highest reported prevalence; (c) variation in actor types reported as perpetrators of sexual violence; (d) distribution of selective targeting and indiscriminate violence, and which groups were most likely to be targeted; and (e) the incidence of sexual violence during post-conflict years.

Variation in highest reported prevalence by country and conflicts

Figure 2 shows the highest conflict-related sexual violence prevalence reported for each country in the sample. Of the 76 countries included in SVAC dataset, 17 countries reportedly experienced sexual violence at the highest prevalence level (coded as 3) by one or more conflict actors during at least one year of the study period. The map also shows countries with reports of sexual violence at the other levels of prevalence, and demonstrates that there is remarkable variation in prevalence.

In terms of individual conflicts, 14% experienced sexual violence at the highest prevalence level, whereas 43% had no reports of sexual violence. There is, however, considerable regional variation. For instance, 63% (26 of 41) of the active conflicts in Africa reported at least one year at either of the highest two prevalence levels of sexual

violence, while the comparable figures for Asia and Europe are 39% (15 of 38) and 26% (6 of 23), respectively.

Variation in highest reported prevalence over time

The SVAC data indicate considerable variation in sexual violence over time. Figure 3 shows a steady increase in reports of sexual violence from 1989 until the peak in 2002. In 2002, 57 armed actors reportedly perpetrated sexual violence, with two groups reported at the highest level of prevalence (coded as 3). In 2003, the number fell to 55, with 8 groups at the highest prevalence level – the highest frequency in the entire time series. Since then, the percentage of actors reported as perpetrators decreased globally, although by 2009 the number was 31, more than twice the number of groups in 1989 (15 groups). This may suggest that sexual violence has become more common in recent years than at the start of the study period.

Also shown in Figure 3, the percentage of armed actors reported to perpetrate sexual violence at a low level of prevalence (coded as 1) followed a similar pattern to the high prevalence cases. In these cases, there was a general increase over time with a peak in 2002, when 12% (42 of 353) of all actors were reported to have committed sexual violence.

Despite the recent downward trend, it remains unclear whether wartime sexual violence as a global phenomenon is getting better or worse (Cohen, Hoover Green & Wood, 2013). Importantly, the total number of victims per conflict-year is unknown, as these data
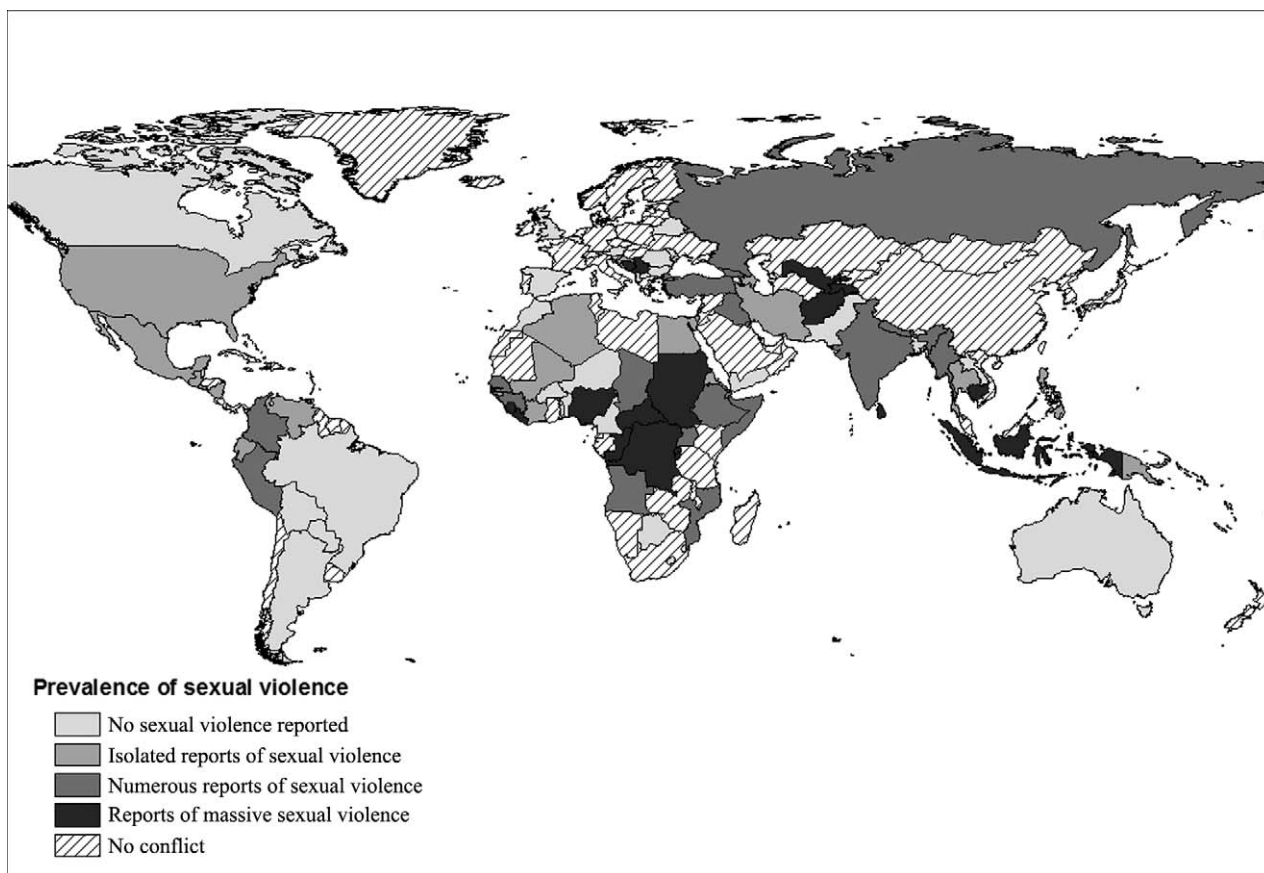
Figure 2. Highest reported prevalence of sexual violence by any conflict actor, 1989–2009

The map indicates the countries whose troops have been reported to commit sexual violence, not necessarily the location where sexual violence has occurred. In the case of the USA, for instance, the prevalence level refers to sexual violence by US forces in Iraq in 2003.
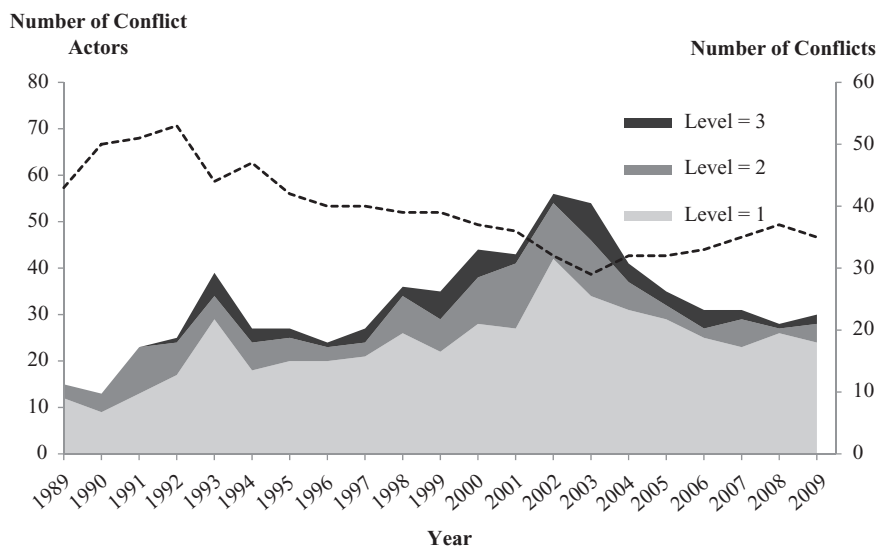

Figure 3. Frequency of armed actors reported as perpetrators, at different levels of prevalence

The dotted line indicates the number of conflict actors reported to perpetrate sexual violence in each year. Sexual violence prevalence levels: 1 = isolated reports of sexual violence; 2 = numerous reports of sexual violence; 3 = reports of massive sexual violence.

are not reported in the sources used in the study (nor are they available – at the level of detail required – in other sources).

Variation across actor types

The SVAC data challenges the conventional wisdom that it is unruly rebel groups and militias, not state militaries, that perpetrate the majority of conflict-related sexual violence. We find that the percentage of state actors reported as perpetrators of sexual violence is higher than that of both rebel groups and PGMs in all but three of the years covered by the SVAC dataset (see Figure 1). Indeed, 42% (56 of 132) of state actors were reported as perpetrators of sexual violence at some point during the study period. The equivalent numbers for rebel groups and militias were 24% (65 of 275) and 17% (38 of 218), respectively.[15] Furthermore, states may sometimes delegate sexual violence to militias, but the data show that state armies often commit sexual violence even when militias exist.

Variation in targeting

Reported incidents of targeting constitute a minority of the observations in the SVAC dataset. About 34% (130 of 381) of the state actor observations have reports of targeting, whereas the percentage for rebels and militias are 31% (64 of 205) and 33% (38 of 112), respectively; hence the difference between actor types is not substantial. The most commonly reported forms of targeting were ethnicity[16] (108 of 232), followed by association with a fighting faction (90 of 232) (denoted 'Actor' in Figure 4). Figure 4 summarizes the frequencies of reported types of targeting.

Post-conflict years

In recent years, there has been a growing concern that sexual violence continues in the post-conflict phase. Until now, however, there has not existed a systematic account of sexual violence by armed actors post-conflict. We find that 13% (58 of 446) of all actors were reported as perpetrators of sexual violence in the first five years post-conflict. For rebels, 13% (28 of 220) were reported as perpetrators in post-conflict years, with 6% (14 of 220) at the two highest levels of prevalence (coded as 2 or 3). For states, 21% (25 of 119) were reported as perpetrators in post-conflict years, with about 3% (4 of 119) of these reports at the two highest levels of prevalence (coded as 2 or 3). Finally, for PGMs 5% (5 of 107) reportedly perpetrated sexual violence in the post-conflict period, with about 3% (3 of 107) at the two highest levels of prevalence.[17] These patterns confirm that sexual violence by armed groups may continue post-conflict, sometimes at very high levels.

Research uses of the SVAC data

The SVAC dataset is designed to be compatible with other relevant and widely used datasets, such as the UCDP/PRIO armed conflict dataset (Gleditsch et al., 2002) and data on one-sided lethal violence (Eck & Hultman, 2007).[18] The ability to merge the SVAC data with other conflict-related data will contribute to more robust empirical analyses in several areas of research, including the repertoire of violent and nonviolent strategies against civilians during wartime, and analyses of conflict dynamics. For example, recent arguments about the causes of wartime rape and other forms of sexual violence focus on the characteristics of armed groups, including their internal norms, discipline, and recruitment practices (e.g. Cohen, 2013; Hoover Green, 2011; Wood, 2010). Coupling the SVAC data with data that capture characteristics of armed actors is necessary for testing such hypotheses.

The SVAC dataset could be used to investigate important policy-relevant consequences of sexual violence. The UN Secretary General, for instance, has stated that 'sexual violence in armed conflict hurts recovery and peace-building' (Ban, 2013), but this and similar claims remain to be systematically studied. The data might also be used to analyze variation in forms, location, and timing of sexual violence; for instance, whether and why certain forms of sexual violence are more prevalent in particular

---

[15] While the finding that states are more frequently reported as perpetrators is confirmed by previous studies (Cohen, 2013; Green, 2006), there may be bias in reporting based on perpetrator categories; for instance, state actors may be more recognizable than rebels because they wear uniforms.

[16] Targeting based on ethnicity is coded only when explicitly reported in the sources. Coders did not assume targeting by ethnicity based on location or other characteristics that might proxy for ethnic identification.

[17] About 5% (54 of 1,107) of the post-conflict observations (conflict-actor-years) for state actors had reports of sexual violence, while the equivalent numbers for rebel groups and PGMs are 4% (46 of 1,080) and 2% (9 of 486), respectively.

[18] Compatibility with existing datasets is assured through using standard country codes, conflict IDs, and actor IDs. Users of the SVAC data should consult the Codebook and User Guide to understand how post-conflict and interim years are coded, as these coding rules are unique to the SVAC dataset.
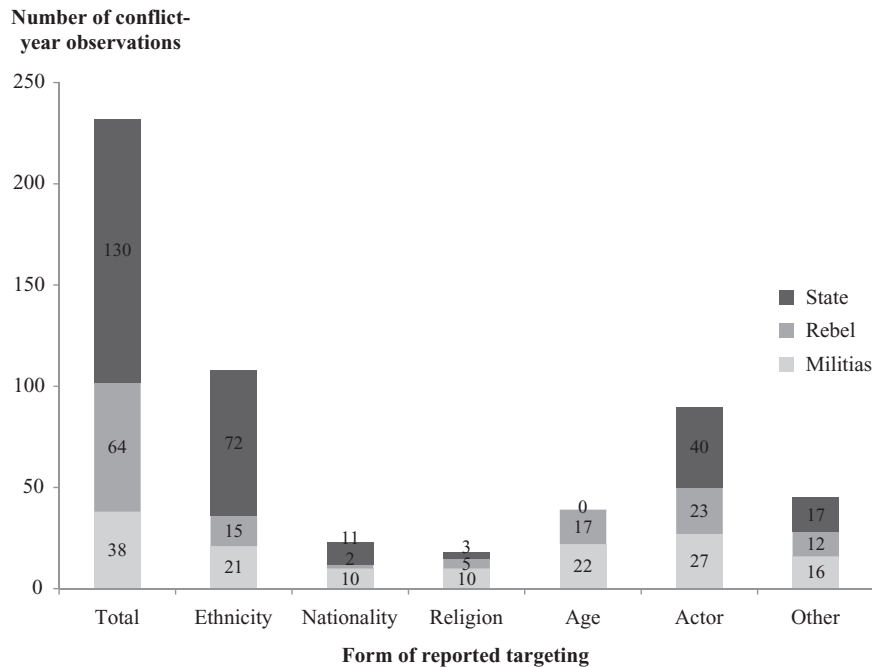
Figure 4. Frequency of forms of reported victim targeting, by perpetrator group type

subtypes of conflicts, or during certain phases of wars. Timing of violence may also shed light on causation; if sexual violence is being used strategically, it may serve as a brutal way for organizations that perceive themselves to be losing ground to gain concessions. If peace talks are imminent, a rebel organization might employ more violent tactics to guarantee its place at the negotiation table.

Conclusion

The lack of comprehensive data has hindered a systematic analysis of conflict-related sexual violence. In this article, we introduce a new dataset on sexual violence in all recent conflicts by all active armed actors. The data demonstrate significant variation in the prevalence of sexual violence, as well as variation over time, by actor group type, and in the forms and locations of the violence.

The SVAC dataset can help advance the understanding of patterns, causes, and consequences of conflict-related sexual violence. Both scholars and policymakers can make use of these data to select cases, to design robust studies, to test new and existing hypotheses, and to develop evidence-based policy solutions. The data show that sexual violence is not always part of conflict dynamics, a hopeful indication

that this terrible form of violence is not inevitable (Wood, 2006, 2009).

Replication data

The dataset and replication files for the empirical analyses in this article, along with the Online Appendix, Codebook and User Guide, can be found at http://www.prio.no/jpr/datasets and http://www.sexualviolencedata.org.

Acknowledgements

Funding

Academy, Norwegian Ministry of Foreign Affairs, and the Norwegian Research Council.

References

Ban, Ki Moon (2013) Remarks at Security Council Thematic Open Debate on 'Addressing impunity: Effective justice for crimes of sexual violence in conflict'. Security Council, New York. 24 June (http://www.un.org/apps/news/infocus/sgspeeches/statments_full.asp?statID=1902#.UfoSD209V3s).

Bastick, Megan; Karin Grimm & Rachel Kunz (2007) *Sexual Violence in Armed Conflict: Global Overview and Implications for the Security Sector*. Geneva: CDCAF.

Bloom, Mia (1999) War and the politics of rape: Ethnic versus non-ethnic conflicts? Unpublished manuscript.

Butler, Christopher; Tali Gluch & Neil Mitchell (2007) Security forces and sexual violence: A cross-national analysis of a principal–agent argument. *Journal of Peace Research* 44(6): 669–686.

Carey, Sabine; Neil Mitchell & Will Lowe (2013) States, the security sector, and the monopoly of violence: A new database on pro-government militias. *Journal of Peace Research* 50(2): 249–258.

Carpenter, R Charli (2006) Recognizing gender-based violence against civilian men and boys in conflict situations. *Security Dialogue* 37(1): 83–103.

Clark, Ann Marie & Kathryn 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): 539–568.

Cohen, Dara Kay (2010) *Explaining Sexual Violence during Civil War*. PhD dissertation, Stanford University.

Cohen, Dara Kay (2013) Explaining rape during civil war: Cross-national evidence (1980–2009). *American Political Science Review* 107(3): 461–477.

Cohen, Dara Kay & Amelia Hoover Green (2012) Dueling incentives: Sexual violence in Liberia and the politics of human rights advocacy. *Journal of Peace Research* 49(3): 445–458.

Cohen, Dara Kay; Amelia Hoover Green & Elisabeth Wood (2013) *Wartime Sexual Violence: Misconceptions, Implications, and Ways Forward*. USIP Special Report, February (http://www.usip.org/sites/default/files/resources/SR323.pdf).

Davenport, Christian & Will Moore (2013) Conflict consortium: Data standards and practices. Unpublished manuscript (http://conflictconsortium.weebly.com/up loads/1/8/3/5/18359923/cc-datastandardspractices3sept 2013.pdf).

Eck, Kristine & Lisa Hultman (2007) One-sided violence against civilians in war: Insights from new fatality data. *Journal of Peace Research* 44(2): 233–246.

Farr, Kathryn (2009) Extreme war rape in today's civil-war-torn states: A contextual and comparative analysis. *Gender Issues* 26(1): 1–41.

Fearon, James D & David D Laitin (2003) Ethnicity, insurgency, and civil war. *American Political Science Review* 97(1): 75–90.

Fearon, James D & David D Laitin (2011) A list of civil wars, 1945–2009. Revised and updated version of the list used for Fearon & Laitin (2003). Stanford University.

Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margit Sollenberg & Håvard Strand (2002) Armed conflict 1946–2001: A new dataset. *Journal of Peace Research* 39(5): 615–637.

Green, Jennifer L (2006) Collective rape: A cross-national study of the incidence and perpetrators of mass political sexual violence, 1980–2003. PhD dissertation, Ohio State University.

Harbom, Lotta; Erik Melander & Peter Wallensteen (2008) Dyadic dimensions of armed conflict, 1946–2007. *Journal of Peace Research* 45(5): 697–710.

Hathaway, Oona A & Daniel E Ho (2004) Characterizing measurement error in human rights: A Bayesian ordinal measurement model. Unpublished manuscript (http://www.law.yale.edu/documents/pdf/Faculty/Hathaway_measurement_error.pdf).

Hoover Green, Amelia (2011) Repertoires of violence against noncombatants: The role of armed group institutions and ideologies. PhD dissertation, Yale University.

ICC (International Criminal Court) (2000) International Criminal Court, Elements of Crimes, Article 8 (2)(e). UN document, PCNICC/2000/1/Add.2 (http://www1.umn.edu/humanrts/instree/iccelementsofcrimes.html).

Leiby, Michele (2009) Digging in the archives: The promise and perils of primary documents. *Politics and Society* 37(1): 75–99.

Nordås, Ragnhild & Siri Aas Rustad (2013) Sexual exploitation and abuse by peacekeepers: Understanding variation. *International Interactions* 39(4): 511–534.

Peterman, Amber; Tia Palermo & Caren Bredenkamp (2011) Sexual violence in the Democratic Republic of the Congo: Population-based estimates and determinants. *American Journal of Public Health* 101(6): 1060–1067.

Peterman, Amber; Dara Kay Cohen, Tia Palermo & Amelia Hoover Green (2011) Rape reporting during war: Why the numbers don't mean what you think they do. *Foreign Affairs* (online), August 1(http://www.foreignaffairs.com/articles/68008/amber-peterman-dara-kay-cohen-tia-palermo-and-amelia-hoover-gree/rape-reporting-during-war).

Sivakumaran, Sandesh (2007) Sexual violence against men in armed conflict. *European Journal of International Law* 18(2): 253–276.

Skjelsbæk, Inger (2011) *The Political Psychology of War Rape: Studies from Bosnia Herzegovina*. London: Routledge.

Utas, Mats (2005) Victimcy, girlfriending, soldiering: Tactic agency in a young woman's social navigation of the

Liberian war zone. *Anthropological Quarterly* 78(2): 403–430.

Wood, Elisabeth (2006) Variation in sexual violence during war. *Politics and Society* 34(3): 307–342.

Wood, Elisabeth (2009) Armed groups and sexual violence: When is wartime rape rare? *Politics and Society* 37(1): 131–161.

Wood, Elisabeth (2010) Sexual violence during war: Leveraging variation toward change. In: Alette Smeulers & Elies van Sliedregt (eds) *Collective Crimes and International Criminal Justice: An Interdisciplinary Approach*. Antwerp: Intersentia, 297–324.

Wood, Reed M & Mark Gibney (2010) The Political Terror Scale (PTS): A re-introduction and a comparison to CIRI. *Human Rights Quarterly* 32(2): 367–400.

DARA KAY COHEN, b. 1979, PhD in Political Science (Stanford University, 2010); Assistant Professor, Harvard Kennedy School (2012–).

RAGNHILD NORDÅS, b. 1977, PhD in Political Science (Norwegian University of Science and Technology, NTNU, 2010); Senior Researcher, PRIO (2010–).