A latent variable approach to measuring wartime sexual violence

Jule Krüger and Ragnhild Nordås

University of Michigan

Word count: 6512

September 14, 2020

Abstract

Conflict-related sexual violence is an international security problem and is sometimes used as a weapon of war. It is also a complex and hard-to-observe phenomenon, constituting perhaps one of the most hidden forms of wartime violence. Latent variable models (LVM) offer a promising avenue to account for differences in observed measures. Three annual human rights sources report on the sexual violence practices of armed conflict actors around the world since 1989 and were coded into ordinal indicators of conflict-year prevalence. Because information diverges significantly across these measures, we currently have a poor scientific understanding with regard to trends and patterns of the problem. In this article, we use an LVM approach to leverage information across multiple indicators of wartime sexual violence to estimate its true extent, to express uncertainty in the form of a credible interval, and to account for temporal trends in the underlying data. We argue that a dynamic LVM parametrization constitutes the best fit in this context. It outperforms a static latent variable model, as well as, analysis of observed indicators. Based on our findings, we argue that an LVM approach currently constitutes the best practice for this line of inquiry and conclude with suggestions for future research.

Keywords: latent variable model, sexual violence, armed conflict, measurement, uncertainty, observational data

Corresponding author: julianek@umich.edu

Introduction

Wartime sexual violence can have a severe negative impact on survivors and affected communities. Policy and advocacy communities have called for improved documentation and analysis to curb abuses and mitigate their harm. We have also seen an upsurge in systematic empirical work on wartime sexual violence (Leiby, 2009; Roth, Guberek & Hoover Green, 2011; Cohen, 2013; Cohen & Nordås, 2014; Cohen & Nordås, 2015; Chu & Braithwaite, 2018).

Despite growing scholarly attention to wartime sexual violence, its empirical study is fraught with methodological challenges. For a range of reasons, sexual violence could be more difficult to observe and measure than other conflict-related violations. Historical data is largely missing and, if available, often contradictory, making it difficult to determine temporal trends. Even today, with international recognition in place, empirical data on victims, survivors, or perpetrators is extremely difficult to obtain (Roth, Guberek & Hoover Green, 2011).

Analyses of what drives wartime sexual violence can only be as good as their empirical basis. Without taking into account underlying data uncertainty, policy resources will probably be spent in inefficient ways, scientific conclusions might be inaccurate, and empirical findings from otherwise carefully conducted studies could be misinterpreted or misappropriated. More accurate data and better understanding of data uncertainty, will equip policymakers and practitioners with more appropriate tools to plan interventions, design preventative measures, and forecast the needs of survivor populations. A latent variable model (LVM) – designed to account for measurement uncertainty in light of observational challenges – will give us a much better understanding of uncertainties underlying observational data and enable a better identification of where more resources are needed to document abuses.

We develop and apply a novel LVM to estimate the prevalence of wartime sexual violence. This addresses an acute challenge for sexual violence research, but also a significant gap with respect to all conflict-related violence. LVMs have been used to study some unobservable concepts of interest to conflict scholars, such as, i.a., democracy (Pemstein, Meserve & Melton, 2010; Reuning, Kenwick & Fariss, 2019; Treier & Jackman, 2008), military alliances (Benson & Clinton, 2016), political-economic risk (Quinn, 2004), respect for human rights (Fariss, 2014; Schnakenberg & Fariss, 2014), treaty preferences (Fariss, 2018; Lupu, 2016), and voting behavior at the United Nations (Voeten, 2000). Many potential topics of interest suited for LVM remain to be explored. Now, LVMs are expanded to assess conflict behaviors, such as one-sided-killings (Fariss, Kenwick & Reunig, 2020), interstate hostility (Terechshenko, 2020), and international sovereignty of self-determination movements (Huddleston, 2020). Sexual violence may be among the most challenging of these, but it is also arguably a highly relevant conflict behavior to study using LVM methodology, given observational constraints.

Why sexual violence is difficult to measure

Observational challenges are common to all human rights violations. However, researching wartime sexual violence is particularly difficult given unique dynamics in the data generating process. Here, we briefly outline key issues that may cause systematic reporting problems.

For survivors and their families, reporting sexual violence can involve grave risks. For one, survivors often worry about social sanctions. In most cultures and countries, sexual victimization is associated with strong taboos. Reporting violations within a wartime context can result in retaliation and further abuses. Therefore, survivors navigate difficult social environments during and after war, choosing to (not) report their experience for strategic reasons (Utas, 2005).

Even if a violation is reported, survivors' and witnesses' descriptions of sexual violence may be less explicit than other types of conflict-related violence; especially in cultures, times, or settings of stronger taboos surrounding anything deemed sexual. For example, Leiby (2009: 89) found that the language and wording of sexual incidents was more convoluted compared to language surrounding killings or other violations. In particular, victims and witnesses engaged in self-censorship, used hard-to-decipher euphemisms, or relied on indirect language to describe experiences for a lack of relevant concepts in their native tongue (ibid.).

Trauma-induced memory loss can obscure specific details in victims' accounts. Inconsistencies in testimony and lack of a clear narrative are typical problems (e.g., International Protocol on the Documentation and Investigation of Sexual Violence in Conflict, 2017). Imprecise descriptions present considerable challenges for extracting systematic data from available source material. These limitations are compounded by the variety of categories included in most definitions of conflict-related sexual violence, as well as the different definitions and conceptualizations used by academics, NGOs, and local communities. This likely influences what is deemed worthy of reporting, i.e., local conceptualizations may drive what is considered as "sexual violence" and what behaviors are illegal.

Despite overwhelming impunity, sexual violations constitute war crimes. This disincentivizes perpetrators from admitting or reporting. Data based on soldiers' accounts likely contain general descriptions of whether or not their armed group committed sexual violence, rather than detailed information that could be used to establish prevalence rates (Baaz & Stern, 2009; Leiby, 2009).

Overreporting, and biased reporting generally, is another possible concern. For example, when attacks are public, victims might be less concerned with hiding their survivor status, whereas when violations happen in private, survivors may keep them hidden. In some cases, overreporting could occur if NGOs and others want to bring increased attention to the issue (Cohen & Green, 2012), or because they closely tie benefits and aid to sexual violence survivor status, which may incentivize desperate civilians to misreport.

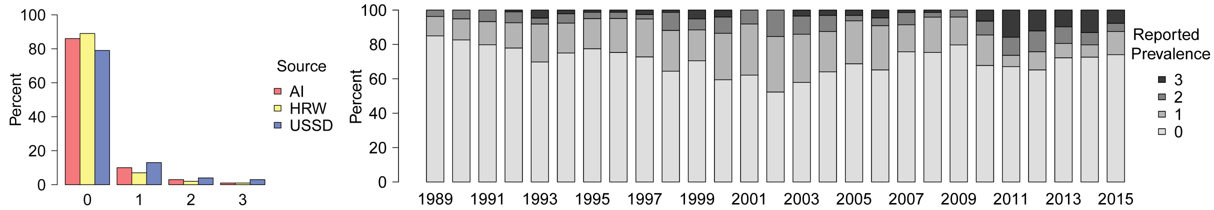
The reporting dynamics outlined thus far are likely exacerbated during war. Leiby (2009: 448) suggests that ongoing conflict enables "*conditions of anonymity and permissiveness that allow individuals to pursue their private interests without fear of detection or retribution.*"

State actors frequently perpetrate sexual violence but also constitute the institutions to which sexual violence survivors could report, i.e., the police, military, or security agents (e.g., Cohen & Nordås, 2014). In addition, many individuals may not survive a conflict to testify later on.

Lastly, both the perpetration and observation of conflict-related sexual violence may be time-dependent. Feedback loops are a possibility if field researchers and conflict monitors guide each others' attention to sexual violence "hotspots." In general, the level of political violence in a given year correlates with the level of violence in the previous year. It is therefore plausible that the observation of sexual violence is also correlated over time. This includes perpetrators' incentives to conceal information and evade accountability, the existence of slowly changing sociocultural norms and taboos, or the practices by which sexual violence events are worded and addressed. An approach to estimating sexual violence prevalence ideally accounts for time-dependent dynamics.

Reported prevalence of wartime sexual violence

Despite observational challenges, a variety of human rights reports are available that provide some insight on the use of wartime sexual violence over time. Three organizations–*Amnesty International* (AI), *Human Rights Watch* (HRW), and the *U.S. State Department* (USSD)–produce annual reports on the state of human rights, including mentions of sexual violations across countries. Scholars reviewed and systematically coded this information to create the *Sexual Violence in Armed Conflict* (SVAC) dataset (Cohen & Nordås, 2014). SVAC captures the reported wartime sexual violence practices of armed conflict actors around the world during the years 1989-2015. The dataset provides three categorical, ordinal-scale variables of the prevalence of sexual violence practices in each conflict-actor-year: "no reported incidents" (0), "some" (1), "several/many" (2), and "massive reports" of wartime sexual violence (3) (Cohen & Nordås, 2014: 420).



(a) Distribution of reported prevalence level, by source

(b) Highest reported prevalence level by year

Figure 1. Percentage distribution of reported level of wartime sexual violence as coded from three human rights sources for 2,092 conflict-actor-years between 1989 and 2015.

In this article, we only include SVAC observations for state forces and rebel groups, as data for militias is not available for the entire time period. For each categorical SVAC prevalence indicator (i.e., AI, HRW, USSD), we aggregate all conflict-actor-year observations across all rebel groups active in a given conflict into a single rebel-conflict-year observation by preserving the maximum reported value across the groups.[1] With this, we obtain a total of 2,092 conflict-actor-year observations for analytical purposes.

For a large majority of observations in our data (approx. 80%), the three sources do not report any (0) sexual violence (Figure 1a). Compared to the two non-governmental human rights sources AI and HRW, the USSD is more likely to report sexual violence. In Figure 1b, we show the distribution of reported prevalence over time by selecting the highest reported prevalence level across the three sources for each conflict-actor-year. For at least half of all conflict-actor-years no (0) codeable sexual violence is reported to have been practiced by armed conflict actors during any given year of the 1989-2015 observation period. While some (1) or several (2) reports of wartime sexual violence are common up until 2009, there seems to be a notable increase in widespread/massive (3) SVAC reports paired with a decrease in reports of level 1 prevalence since 2010. The proportion of conflict-actor-years with no reports of wartime sexual violence has however remained above 60% in this same time period.

Empirical information coded from available human rights reports contains unspecified

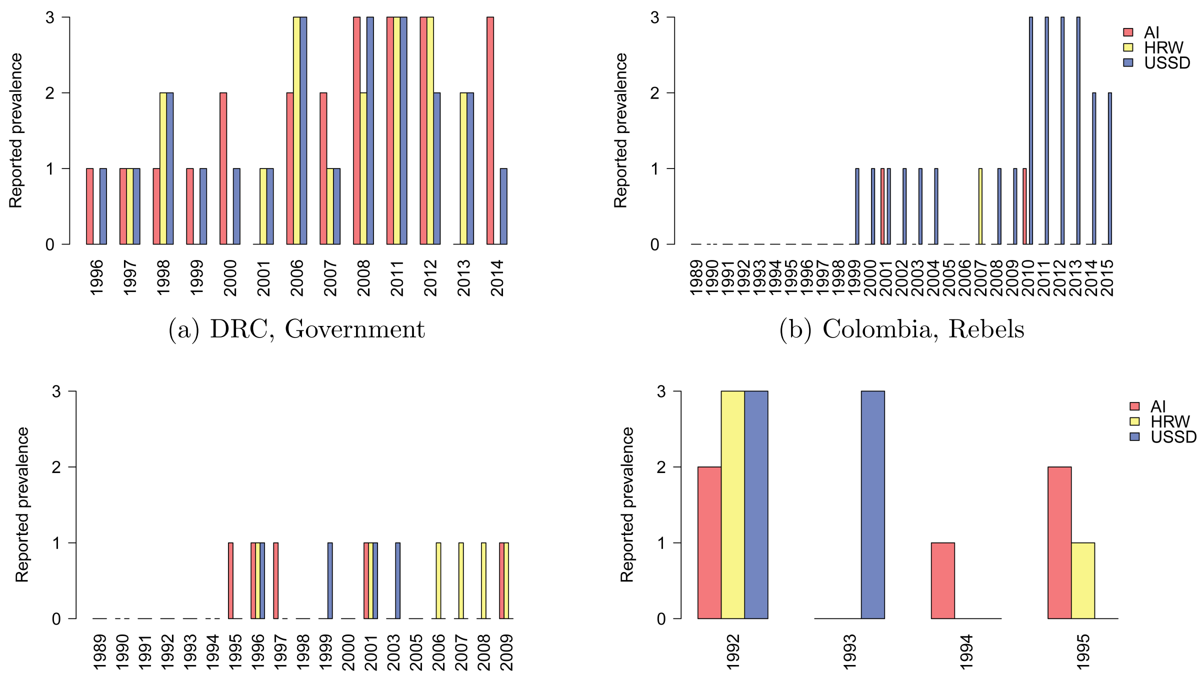
[1]We explain our reasoning for this aggregation strategy in the Online appendix (Section A).

measurement error, as well as, uncertainty with regard to the extent of underreporting (Brysk, 1994; Gohdes & Price, 2013; Goldstein, 1992; Krüger et al., 2013; Roth, Guberek & Hoover Green, 2011; Price & Ball, 2015; Weidmann, 2016). In Figure 1, we observe that prevalence level 3 (massive) is very rare in earlier years compared to later in the time series. This could be a result of changes in the reporting of sexual violence akin to what Fariss (2014) identified as the changing standard of accountability for human rights violations. Future research could help identify to what degree this observed temporal trend is due to changes in SVAC reporting. When taken at face value, categorical indicators that are coded from human rights reports do not allow for a quantification of the underlying uncertainty. Expressed in probabilistic terms, the human-coded ordinal measures provided in the SVAC data are to be understood as reporting the annual prevalence of wartime sexual violence practices with a probability of 1.

To illustrate the inherent uncertainty in observed accounts of wartime sexual violence, we focus on one case each across the four conflict regions Africa, Latin America, Asia, and Europe: the civil war between the government and various non-state rebel groups in the Democratic Republic of Congo (DRC); the war between the government and revolutionary guerrillas in Colombia; the armed confrontation between the government and insurgents over an independent Tamil State in Sri Lanka; as well as, the war between the state and non-state ethnic Serb forces in Bosnia-Herzegovina.[2]

The inherent uncertainty in observed accounts of wartime sexual violence is clearly visible when the three categorical SVAC measures are compared across conflict-actor-years for these specific cases (Figure 2). For example, the coded level of rebel groups' use of sexual violence in the Colombian conflict diverges considerably across sources. With no annual reports in 1990 and 2003, HRW reports no (0) use of sexual violence between 1989-2015, except for some (1) use in 2007, whereas the other two measures report no (0) SVAC use in that year

[2]Here, we only discuss one conflict actor type in each of the conflicts. We compare the reporting for both types in the Online appendix (Section B).



(a) DRC, Government

(b) Colombia, Rebels

(c) Sri Lanka, Government

(d) Bosnia-Herzegovina, Rebels

Figure 2. Reported prevalence of wartime sexual violence in four conflicts according to three human rights sources, by select type of armed group and conflict-year.

(Figure 2b). AI, in turn, reports some (1) use of sexual violence by Colombian rebel groups for the years 2001 and 2010, and no (0) use otherwise. USSD, agrees with AI in 2001 and those years in which neither of the sources report sexual violence (0), i.e., 1989-1998 and 2005-2006. However, USSD reports higher prevalence of wartime sexual violence than AI in all other years.

For Bosnia-Herzegovina (Figure 2d), HRW and USSD agree that there was massive (3) use of conflict-related sexual violence by rebel groups in 1992, while they disagree about the years 1993-1995. AI only notes several (2) reports for 1992, but reports more sexual violence than HRW and USSD in 1994 and 1995. While USSD reports widespread (3) prevalence of sexual violence in 1993, AI and HRW report no (0) SVAC use in that conflict-year. These patterns of low agreement across human rights sources are equally visible in the reporting of SVAC practices by state forces in the DRC (Figure 2a) and Sri Lanka (Figure 2c).

For a few conflict-actor-years, we observe agreement. This is the case for the DRC

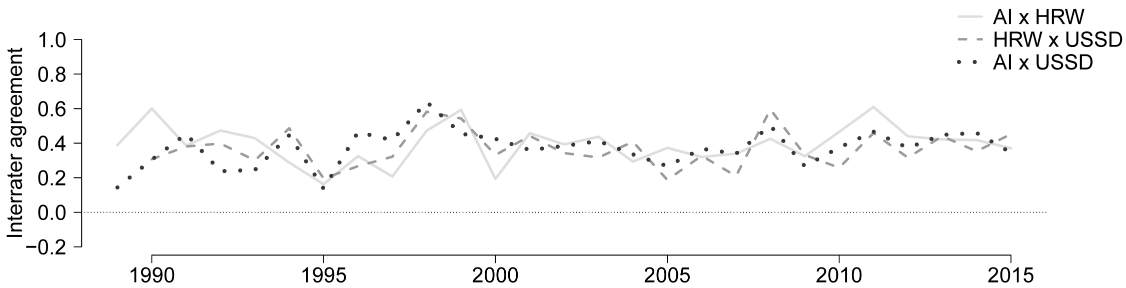


Figure 3. Temporal variation of interrater agreement (Cohen's weighted kappa) between human rights indicators, by year.

government in 1997 and 2011 (Figure 2a), and the government of Sri Lanka in 1989, 1991-3, 1996, 1998, 2000-1 and 2005 (Figure 2c). In some conflict-actor-years, two sources agree with regard to the reported level of SVAC prevalence, i.e., regarding state forces in the DRC (1996, 1998-9, 2001, 2006-8, 2012-3), and the Sri Lankan government (1990, 1994-5, 1999, 2003, 2006-9).

Table I. Cohen's weighted kappa coefficients (with p-values) as a measure of interrater agreement between pairs of human rights indicators for the entire SVAC dataset, and for government and rebel group observations, respectively.

Human rights indicators	All SVAC	p-value	Governments	p-value	Rebels	p-value
AI x HRW	0.408	0.000	0.353	0.000	0.466	0.000
AI x USSD	0.394	0.000	0.381	0.000	0.368	0.000
HRW x USSD	0.385	0.000	0.361	0.000	0.386	0.000

Generally, interrater agreement across the three indicators is weak. While the various human rights sources show positive agreement, Cohen's weighted kappa values are typically below 0.5 (Table I). AI and HRW agree most on wartime practices of rebel groups, and least on those of governments. This indicates relevant dynamics in the data generating process regarding the type of human rights source (non-governmental vs. governmental), and the conflict actor (governments vs. rebels). The level of agreement between human rights indicators varies considerably over time (Figure 3), with Cohen's weighted kappa values ranging between 0.1 and 0.6. In some years, there is more agreement between AI and HRW (1990, 1999, 2011), in others between HRW and USSD (1998, 2008), and yet in others between AI and USSD (1994, 1998, 2008).[3]

How are sexual violence scholars to reconcile diverging accounts of SVAC practices? Measurement options for observational data coded at the ordinal scale are statistically limited. The distance between ordered categories cannot be interpreted in a numerically meaningful way to allow for the calculation of a mean. Should scholars select one "best" source deemed most reliable according to some set of criteria? Or, should they choose the most commonly reported, lowest or highest level of observed abuse? Instead of either approach, we argue that an LVM offers a more fitting and sophisticated measurement strategy.

A latent variable model approach to measuring SVAC

Latent variable model approaches (LVM) offer an improved measurement strategy in research scenarios such as ours. At the conceptual level, LVM assume that there is some unidimensional latent construct (or "trait") of theoretical interest to researchers, denoted θ, that remains elusive in terms of accurate direct measurement. Rather, researchers are only able to obtain repeated measures of observed outcomes that approximate the latent construct to some degree. LVM enable scholars to estimate the associated uncertainty surrounding the observed level of the latent trait by averaging information across the available measures. The underlying rationale is intuitive: If x number of indicators repeatedly report the same or similar values for the construct of interest, we attribute more certainty to a cross-measure account. Alternatively, if x number of measures report x number of distinct values for the latent trait, we regard our final account of the latent construct more uncertain. LVM produce a probabilistic Bayesian estimate of the true value of the latent trait θ across the observed outcomes by calculating the mean value of all the draws taken from the assumed normal distribution of θ, and other parameters, given the available data. The

[3]In the Online appendix, we examine the disaggregated temporal variation of these relationships for government and rebel groups, respectively. This further illustrates differences between conflict actors.

standard deviation associated with those draws allows to express our uncertainty regarding the estimate in the form of a credible interval (Martin & Quinn, 2002; Quinn, 2004; Treier & Jackman, 2008; Pemstein, Meserve & Melton, 2010; Schnakenberg & Fariss, 2014). The resulting θ estimates are computed at the interval scale (vs. ordinal) which enables more advanced empirical analysis later on. Because the approach provides a latent variable estimate for each unit of analysis, scholars can engage in probabilistic comparisons across cases, i.e, in our research context with regard to SVAC prevalence across conflict-actor-years.

In this research, we parametrize a first LVM that was originally developed by Schnakenberg & Fariss (2014). We assume that the observed AI, HRW, and USSD indicators are functions of a unidimensional latent trait variable that represents the use of wartime sexual violence. To identify each conflict-actor-year as an individual unit of analysis, we index conflict-actors with i, and years with t. Our goal is to estimate the true prevalence of wartime sexual violence for each conflict-actor-year between 1989 and 2015, θ_{it}, using our three observed indicators (or "items") J, with $J = 3$. Each indicator is measured at the ordinal scale and can take on K_j values, here $K = 4$, i.e., it can take on levels $(0, 1, 2, 3)$. The observed values of each indicator are denoted as y_{itj} for a given conflict-actor-year and assumed to depend on θ_{it}.

For each indicator, we estimate an item discrimination parameter β_j and a set of $K_j - 1$ difficulty cut-points $(\alpha_{jk})_{k=1}^{k_j}$ (cf. Schnakenberg & Fariss, 2014: 7). An error term ϵ_{itj} captures measurement error stemming from observational challenges and human coding practices. We assume that the ϵ_{itj} error terms are independently drawn from a logistic distribution. This allows us to determine the likelihood of our LVM. Following this parametrization, we can derive a probability distribution for a given observed value to an indicator j. A likelihood function for β, α and θ given the data enables estimation of these unobserved parameters.

Bayesian item-response theory requires us to assume local independence. This means, we assume (1) local independence of indicators within the same conflict-actor-year, (2) local independence of indicators across conflict actors within years, and (3) local independence of indicators across years within conflict-actor units. In practice, each of these assumptions is likely violated. For example, (1) conflict observers may face the same challenges in a given war zone and start to collaborate, (2) in a given year, observers may focus on a particularly gruesome conflict actor at the expense of others, and (3) all conflict observers are likely to expand their efforts each time the international community increases its investment in SVAC documentation. Even though, in this article, we only address the violation of assumption (3), we believe that latent SVAC measures constitute an empirical improvement over purely observational data. In the conclusion, we offer suggestions on how violations to assumptions (1) and (2) could be addressed in future research.

Our first basic LVM is a *static* version because we place independent standard normal priors on each θ_{it}, i.e., $\theta_{it} \sim N(0, 1)$ for all i and t. This conforms to the (3) local independence assumption that sexual violence indicators are time-independent across years and conflict-actor units. As we argued above, both the perpetration and observation of conflict-related sexual violence may be time-dependent. To account for this temporal dynamic, we also parametrize a *dynamic* version of an LVM approach believing that it offers a better fit for measuring SVAC. In the dynamic model, we relax the (3) local independence assumption, specifying hierarchical priors for each θ_{it} that allow the estimated latent level of wartime sexual violence to vary with the previous year's value for a respective conflict actor, such that $\theta_{it} \sim N(\theta_{it-1}, \sigma)$ (cf. Martin & Quinn, 2002; Schnakenberg & Fariss, 2014). When $t = 1$, the standard normal prior is used. Like Schnakenberg & Fariss (2014), we estimate variance σ, modeling its prior as $U(0, 1)$.

Both the static and dynamic LVM were estimated using Markov Chain Monte Carlo simulation with 4 chains. Each chain was run with 2,000 iterations. The first half of iterations (1,000) was used for burning in the model, the second half for statistical inference. Plots of the Gelman-Rubin statistic confirmed model convergence (Gelman & Rubin, 1992).[4]

[4] We provide graphs of the Rubin-Gelman statistic in the Online appendix.

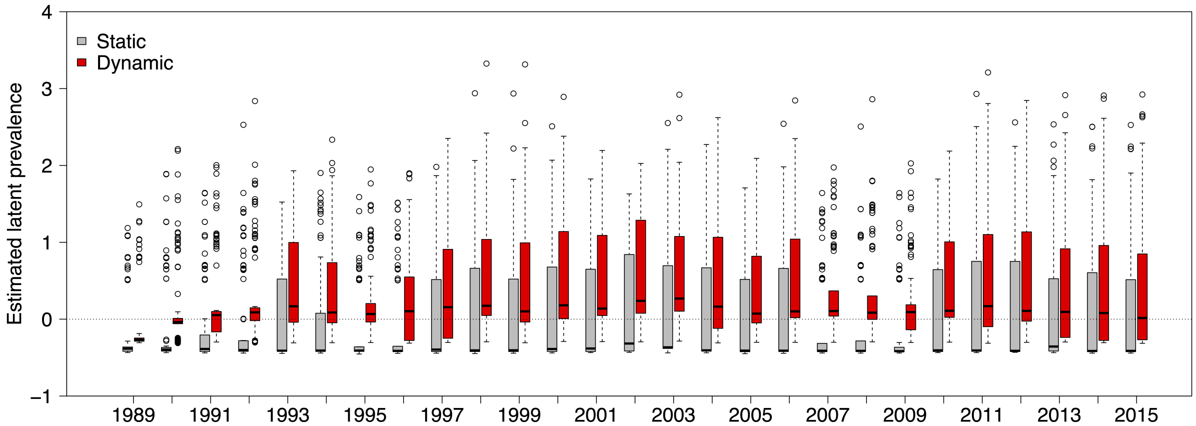


Figure 4. Distribution of static (gray/light) vs. dynamic (red/dark) estimates of latent SVAC prevalence for 2,092 conflict-actor-years between 1989 and 2015, by year.

Estimates of latent prevalence of sexual violence in war

We compute two sets of Bayesian estimates of the latent prevalence of conflict-related sexual violence across 2,092 conflict-actor-years between 1989 and 2015. The first set of estimates results from a static specification of our LVM, the second from a dynamic version of the same approach. Figure 4 plots the temporal trend of the static (gray/light) vs. dynamic (red/dark) estimates of latent SVAC prevalence for the period of observation. Most of the estimates naturally cluster around zero given that no sexual violence is reported for the majority of conflict-actor-years. Viewed in comparison, it is clear that a dynamic parametrization, which accounts for the latent prevalence of wartime sexual violence in the previous conflict-actor-year, corrects the θ_{it} estimates upward.

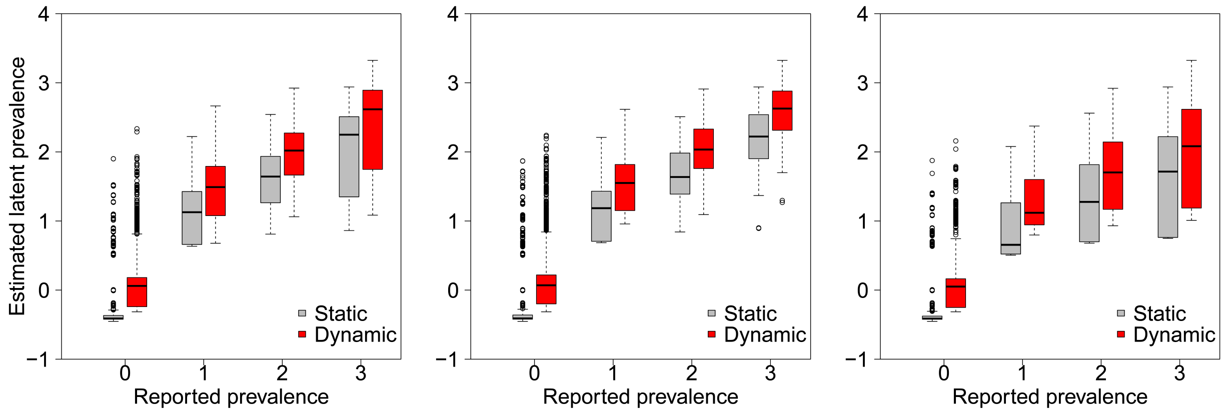
Empirically, the dynamic model fits the observed data better than the static model as its Widely Applicable Information Criterion (WAIC) score of 5,302 is smaller than the static model's WAIC score of 5,316. We believe the dynamic LVM specification is also more plausible given that conflict processes and our observation of them are highly time-dependent. For both sets of estimates, it is noticeable that the within-year range of latent SVAC prevalence changes over time. During the years 1989-1996, 2001-2, and 2007-2009, respectively, estimated latent prevalence ranges between -0.5 and 2, and from -0.5 to 3 for all other years. This reflects the observational pattern shown in Figure 1b. It suggests that limited historical information on sexual violence practices is available for the early years of the observation period. Furthermore, there are none or few conflict-actor-years of a reported widespread/massive (3) prevalence level. The estimated low prevalence levels during the years 2007-2009 suggests that the underlying data-generating processes here could be scrutinized in future research, which is beyond the scope of this article.

Figure 5 displays the distribution of static (gray/light) vs. dynamic (red/dark) estimates of latent SVAC prevalence for all conflict-actor-year observations relative to the observed values coded from each source. Conflict-actor-year estimates relative to observations in the reported level (3) category in AI and HRW are pulled upward, whereas they are pulled downward for USSD observations in the same level (3) category. This is due to USSD, on average, reporting more observations at the highest prevalence level compared to the other two sources (cf. Figure 1a). This pushes the latent variable estimates relative to AI and HRW values upward towards (3) on the interval scale (Figures 5a and 5b), while it lowers the estimates relative to the USSD level (3) category (Figure 5c).

Figure 6 displays the difficulty cut-points (α) for each ordinal human rights indicator for the static and dynamic model, respectively. Since the boxes do not overlap for either indicator in each model specification, we conclude that both models discriminate well between the human-coded categories of each item.[5]

Let us turn to the DRC, Colombia, Sri Lanka, and Bosnia-Herzegovina to illustrate our estimates of latent prevalence of SVAC (Figure 7). The point estimates, represented by dots, are the means of the marginal posterior densities of latent SVAC. The uncertainty with each point estimate is represented by 95 percent credible intervals, computed from the 0.025 and 0.975 percentiles of the marginal posterior densities. For both model specifications, the parameter estimates range from -2 to 3 (Colombia, Sri Lanka), or -2 to 4 (DRC, Bosnia). Static estimates are shown in gray/light, dynamic estimates in red/dark. Again, it is clearly visible

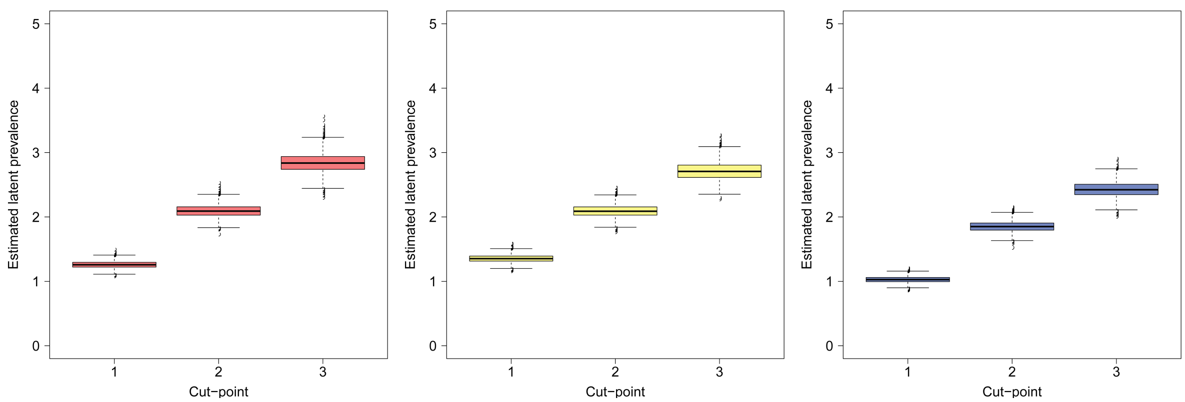
[5] We provide the estimates of the α and β parameters in the Online appendix.



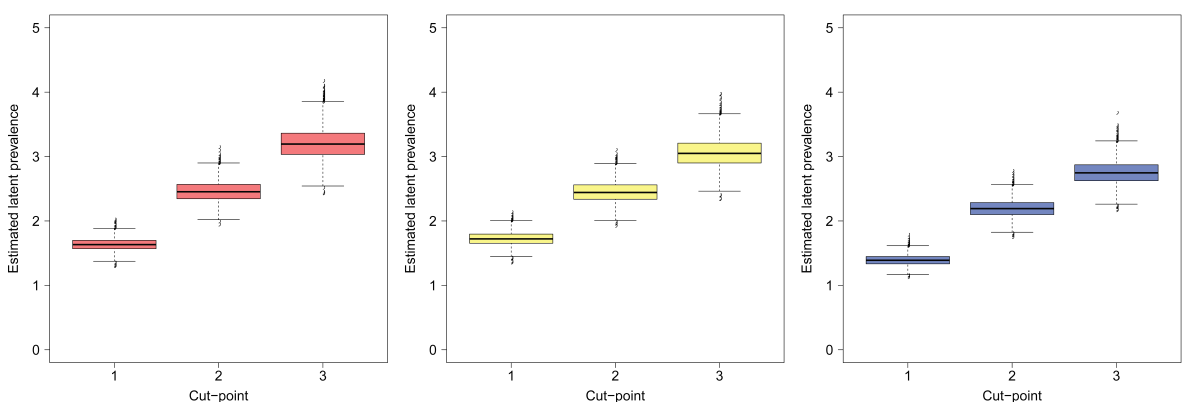
(a) AI (b) HRW (c) USSD

Figure 5. Distribution of static (gray/light) and dynamic (red/dark) latent variable estimates of conflict-actor-years relative to observed categorical values in three human rights indicators for the same conflict-actor-years. Because USSD, on average, reports higher levels of SVAC prevalence than AI and HRW, the latent estimates are distributed lower relative to the observed USSD categorical values, and higher relative to the observed AI and HRW categories.

that a dynamic specification of the latent variable model pulls the latent SVAC estimates upward. The size of the credible intervals varies quite a bit across conflict-actor-years. This is indicative of the uncertainty associated with relevant observations (cf. Figure 2). The intervals are smaller for conflict-actor-years that show more agreement across indicators, e.g., DRC 1996-8, Colombia 2001, Sri Lanka 1996 and 2001, and Bosnia-Herzegovina 1992. Larger credible intervals represent less agreement across the three sources. Many credible intervals include the value zero, which is due to the majority of zero-level observations.

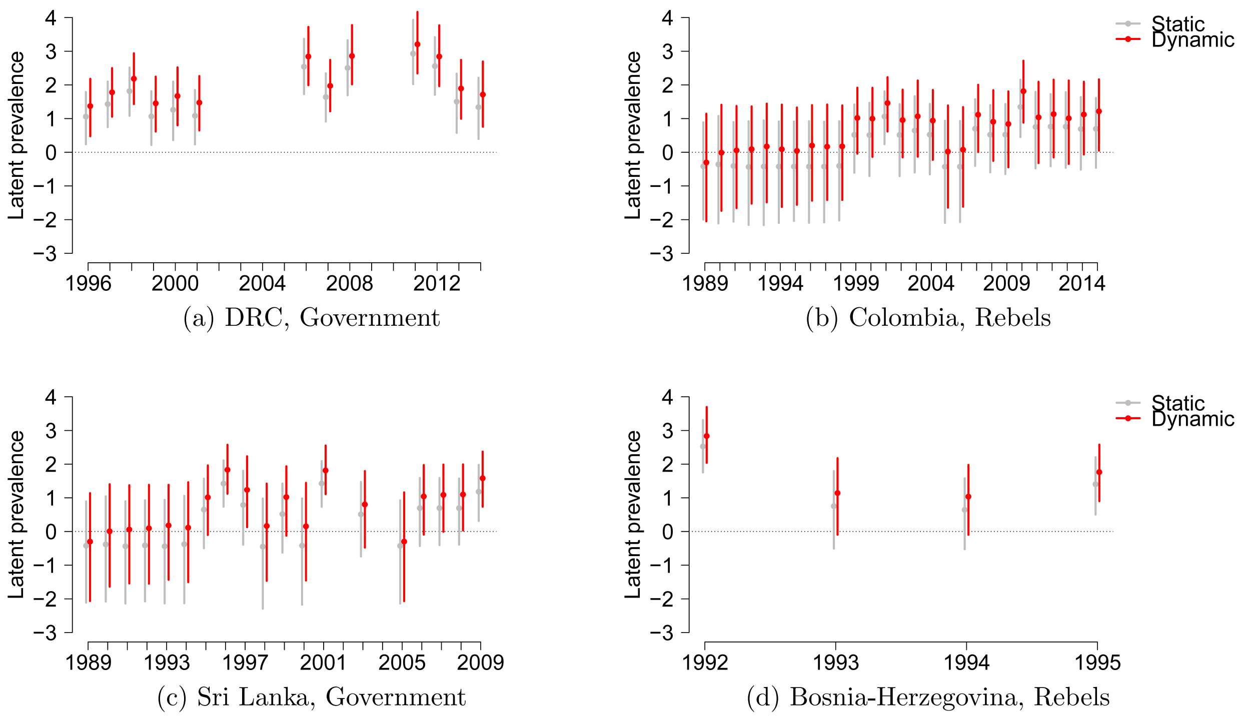


(a) AI: Static (b) HRW: Static (c) USSD: Static



(d) AI: Dynamic (e) HRW: Dynamic (f) USSD: Dynamic

Figure 6. Posterior draws for the scaled difficulty cut-points (α) for each human rights indicator for the static and dynamic latent variable models, respectively. The α parameters are transformed to the same scale of the latent variable by dividing each item difficulty parameter by its corresponding item discrimination (β) parameter.



(a) DRC, Government

(b) Colombia, Rebels

(c) Sri Lanka, Government

(d) Bosnia-Herzegovina, Rebels

Figure 7. Latent variable estimates of wartime sexual violence for four cases of armed conflict, by select armed group and conflict-actor-year. Static estimates are shown in gray/light, dynamic estimates in red/dark. Dots indicate point estimates (posterior means), lines are 95 percent credible intervals. Years with missing estimates are non-conflict years that were not estimated.

Latent estimates express observational uncertainty

While assessing the predictive validity and performance of our latent estimates, we encountered (somewhat) unexpected results in comparison to other applications of latent variable approaches to measurement.[6] Consistently, models computed with the observed indicators provide more statistically significant effects of a larger effect size, as well as better goodness-of-fit statistics, than models run with the latent estimates.[7] We attribute this to features of the research context of wartime sexual violence. In particular, we found a much weaker relationship, i.e., higher levels of disagreement across the indicators (cf. Table I), even though they are repeated measures of the same latent construct. Low levels of agreement may reflect the restricted capabilities of human rights monitors to register and document sexual violence events, if indeed sexual violence is particularly hard to measure.

In a research context of considerable observational challenges and inherent uncertainty, a statistical modeling approach to measurement is indispensable. LVM methodology enables us to statistically account for limitations in the underlying data-generating processes, a strategy proposed by Brysk (1994: 692) to improve measurement accuracy. LVMs directly express the inherent observational uncertainty with each latent estimate of SVAC prevalence in the size of the associated credible interval (Martin & Quinn, 2002; Quinn, 2004; Treier & Jackman, 2008; Pemstein, Meserve & Melton, 2010; Schnakenberg & Fariss, 2014). Substantive empirical analyses that use latent estimates instead of purely observational data prevent the analyst from overestimating individual effects and sizes, which could ultimately mislead scholarly understanding, or policy initiatives designed to address wartime sexual violence.

[6] We detail the analyses and results discussed here in the Online appendix (Section H).

[7] The LVM-based models are an average of 1,000 draws from the posterior distribution to rigorously incorporate the latent estimates' uncertainty.

Conclusion

In this article, we demonstrate how a latent variable approach to measuring wartime sexual violence outperforms empirical analysis with purely observational data. We show the advantage of measuring *latent* SVAC prevalence, as it allows the analyst to leverage information across multiple sources of information in a statistically advanced way without having to select or aggregate information across available measures in a more simplified form. Additionally, a latent variable approach expresses the uncertainty associated with the (dis)agreement across multiple observational measures by calculating a probabilistic credible interval. This uncertainty can then be incorporated in subsequent analyses, enabling a more accurate estimate of substantive effects given the underlying observational data. Future scholars will be able to use our LVM estimates to refit substantive models by drawing from the posterior distribution, thereby enhancing empirical research.

We hope that our findings encourage scholars to consider a latent variable approach to studying wartime sexual violence. There is considerable potential for further improving this measurement and modeling strategy. One concern are the many conflict-actor-years for which the available annual human rights sources report no (0) sexual violence practices, or the available information is too vague to code reliably. Zero-level observations are challenging in that they remain unclear with regard to the true level of abuse: was there really no practice of sexual violence, or did actual abuses go unreported? In the latter case, LVM estimates are biased towards zero for a lack of better information. While latent variable models mitigate uncertainty in the observed data, they are not designed to quantify the amount of violent practices that are entirely omitted from reports (i.e., *unobserved*) as is, for example, the case with multiple systems estimation.

We see the following opportunities for advancing this line of inquiry. One possibility is to acquire additional observational SVAC indicators in hopes that they will augment our understanding of currently zero-level conflict-actor-year observations. Incorporating more indicators could also help address the violation of local independence assumption (2). Another avenue is to identify empirical measures to account for "observability" in a given conflict-actor-year (cf. Fariss, 2014), which could address violations to model assumption (1). For example, the level of "openness" for information on sexual violence to become available could be approximated by including measures of press freedom, the state of women's rights, the total annual number of sexual violence related news reports in a given conflict-country, the extent of the local civil society, and/or the density of locally based NGOs focusing on SVAC. Select case studies of specific conflicts could further inform the explicit modeling of observational challenges. This could include, i.a., practitioners surveys as Clay et al. (2020) use for overcoming reporting biases in measures of civil and political rights. Another research opportunity is to model the local dependence of indicators within conflict-actor-years more explicitly to address assumption (1) violations. Sophisticated textual analysis of the annual human rights reports could assess the level of confirmation/complementarity regarding individually identified events, actors, and locations. For example, human coders may have derived a level (2) prevalence from AI and HRW reports based on the description of distinctly different events in both sources. Comparative evidence of "more violations" across available measures could be used to adjust estimates of the latent trait accordingly. With ample future research opportunities remaining, this article makes a first step towards more accurate measurement of wartime sexual violence using latent variable methodology.

Replication data

The replication data and Online appendix can be found at https://github.com/juleka/JPR-LVM-SVAC.

Acknowledgements

We are grateful to Christopher Fariss for his guidance at various research stages. We thank three anonymous reviewers for their comments on earlier versions of this manuscript.

Funding

This research was funded by the Norwegian Research Council, project number 262425.

References

Baaz, Maria Eriksson & Maria Stern (2009) Why do soldiers rape? Masculinity, violence, and sexuality in the armed forces in the Congo (DRC). *International Studies Quarterly* 53(2): 495–518.

Benson, Brett V & Joshua D Clinton (2016) Assessing the variation of formal military alliances. *Journal of Conflict Resolution* 60(5): 866–898.

Brysk, Alison (1994) The politics of measurement: The contested count of the disappeared in Argentina. *Human Rights Quarterly* 16(4): 676–692.

Chu, Tiffany S & Jessica Maves Braithwaite (2018) The effect of sexual violence on negotiated outcomes in civil conflicts. *Conflict Management and Peace Science* 35(3): 233–247.

Clay, Chad K; Ryan Bakker, Anne-Marie Brook, Daniel W Hill Jr & Amanda Murdie (2020) Using practitioner surveys to measure human rights: The human rights measurement initiative's civil rights metrics. *Journal of Peace Research* 57(6): xxx–xxx.

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 & Ragnhild Nordås (2015) Do states delegate shameful violence to militias? Patterns of sexual violence in recent armed conflicts. *Journal of Conflict Resolution* 59(5): 877–898.

Cohen, Dara Kay & Ragnhild Nordås (2014) Sexual violence in armed conflict: Introducing the SVAC dataset, 1989–2009. *Journal of Peace Research* 51(3): 418–428.

Fariss, Christopher J (2014) Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review* 108(2): 297–318.

Fariss, Christopher J (2018) The changing standard of accountability and the positive relationship between human rights treaty ratification and compliance. *British Journal of Political Science* 48(1): 239–271.

Fariss, Christopher J; Michael R Kenwick & Kevin Reunig (2020) Estimating one-sided-killings from robust measurement model of human rights. *Journal of Peace Research* 57(6): xxx–xxx.

Gelman, Andrew & Donald B Rubin (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4): 457–472.

Gohdes, Anita & Megan Price (2013) First things first: Assessing data quality before model quality. *Journal of Conflict Resolution* 57(6): 1090–1108.

Goldstein, Robert Justin (1992) The limitations of using quantitative data in studying human rights abuses. In: Thomas B. Jabine & Richard P. Claude (eds) *Human rights and statistics: Getting the record straight*. Philadelphia: University of Pennsylvania Press, 35–61.

Huddleston, Robert J (2020) Continuous recognition: A latent variable approach to measuring international sovereignty of self-determination movements. *Journal of Peace Research* 57(6): xxx–xxx.

International Protocol on the Documentation and Investigation of Sexual Violence in Conflict (2017) Best practice on the documentation of sexual violence as a crime or violation of international law, 2nd Ed. (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/31

Krüger, Jule; Patrick Ball, Megan Price & Amelia Hoover Green (2013) It doesn't add up: Methodological and policy implications of conflicting casualty data. In: Taylor Seybolt; Jay D. Aronson & Baruch Fischhoff (eds) *Counting civilian casualties: An introduction to recording and estimating nonmilitary deaths in conflict*. New York: Oxford University Press, 247–264.

Leiby, Michele L (2009) Wartime sexual violence in Guatemala and Peru. *International Studies Quarterly* 53(2): 445–468.

Lupu, Yonatan (2016) Why do states join some universal treaties but not others? An analysis of treaty commitment preferences. *Journal of Conflict Resolution* 60(7): 1219–1250.

Martin, Andrew D & Kevin M Quinn (2002) Dynamic ideal point estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* 10(2): 134–153.

Pemstein, Daniel; Stephen A Meserve & James Melton (2010) Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis* 18(4): 426–449.

Price, Megan & Patrick Ball (2015) Selection bias and the statistical patterns of mortality in conflict. *Statistical Journal of the IAOS 31*: 263–272.

Quinn, Kevin M (2004) Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* 12(4): 338–353.

Reuning, Kevin; Michael R Kenwick & Christopher J Fariss (2019) Exploring the dynamics of latent variable models. *Political Analysis* 27(4): 503–517.

Roth, Francoise; Tamy Guberek & Amelia Hoover Green (2011) Using quantitative data to assess conflict-related sexual violence in Colombia: Challenges and opportunities.

Schnakenberg, Keith E & Christopher J Fariss (2014) Dynamic patterns of human rights practices. *Political Science Research and Methods* 2(1): 1–31.

Terechshenko, Zhanna (2020) Hot under the collar: A latent measure of interstate hostility. *Journal of Peace Research* 57(6): xxx–xxx.

Treier, Shawn & Simon Jackman (2008) Democracy as a latent variable. *American Journal of Political Science* 52(1): 201–217.

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.

Voeten, Erik (2000) Clashes in the assembly. *International Organization* 54(2): 185–215.

Weidmann, Nils B (2016) A closer look at reporting bias in conflict event data. *American Journal of Political Science* 60(1): 206–218.

Biographical statement

JULE KRÜGER, b. 1982, PhD in Government (2014); Program Manager, Center for Political Studies, Institute for Social Research, University of Michigan (2018–).

RAGNHILD NORDÅS, b. 1977, PhD in Political Science (2010); Senior Researcher, Peace Research Institute Oslo (2010–), Assistant Professor, University of Michigan (2017–).