Responses to Reviewers' Comments for Manuscript 2023-07372

Estimating Conflict Losses and Reporting Biases

Addressed Comments for Publication to

PNAS

by

Benjamin J. Radford, Yaoyao Dai, Niklas Stoehr, Aaron Schein, Mya Fernandez, and Hanif Sajid

To the editor,

Please find enclosed the revised version of our previous submission entitled "Estimating Conflict Losses and Reporting Biases" with manuscript number 2023-07372. We are grateful for the reviewers' very supportive and carefully considered feedback on our manuscript. It is clear that both reviewers engaged deeply with the paper and we value their thoughtful comments.

We have carefully considered all of the reviewers' comments and made changes to the manuscript where appropriate. In our responses to these comments, we include excerpts from the revised manuscript and highlight the relevant changes. We also appreciate that the reviewers have requested additional robustness tests and explanations in this response letter while recognizing that these may not fit in the manuscript itself. We have done our best to undertake all such robustness tests and report on them in this response letter. We address all comments in the order that they were given in the original review document.

Sincerely,

Benjamin J. Radford, Yaoyao Dai, Niklas Stoehr, Aaron Schein, Mya Fernandez, and Hanif Sajid

Authors' Response to the Editor

General Comments. See R2's major concern. Your revision should speak to that concern or to attach a memo explaining why your algorithm is unbiased. In any case, you will need to revise to speak to this concern. For all the other comments from both reviewers, please take them into account. I will send the revision and memo only to R2 before I make my final decision as editor.

Thank you for your handling of the review process. We have carefully considered all of the reviewers' comments and believe that they have resulted in a clearer and stronger manuscript.

While we provide a much more thorough response to R2's major concern later in this letter, we do want to take the opportunity to highlight how we have addressed this particular comment. Our prior on the bias terms should be understood as a conservative initial guess at the values of these biases and it does not enforce any strict cancellation of these parameters. We have modified the manuscript to clarify this.

We have made some changes to the language of the Materials and Methods section of our paper. This section now better explains our motivation for choosing a multilevel prior for the bias terms. We believe this choice to be both theoretically and statistically sound. In this letter, we have also clarified the role of the prior in question, explained the practical effect that this choice has on the bias estimates, and empirically demonstrated that it does not impose a perfect cancellation of the bias parameters (they do not average to zero). We take R2's concerns very seriously and understand why they may prefer other model assumptions. Therefore, we have estimated three additional models that incorporate these alternative assumptions. Two of these are essentially Bayesian implementations of fixed effects and one is similar to the original multilevel model but with a much more flexible prior placed on the prior bias means. We provide two figures to demonstrate that our results for quantities of interest are qualitatively insensitive to these changes.

In addition to thoroughly addressing R2's Comment 1, we engage with all other comments from both reviewers:

- R1 Comment 1: We discuss how our model handles class imbalance and modify the manuscript to report the number of observations per claim source.
- R1 Comment 2: We provide a figure that depicts all source-target-category bias terms with uncertainty.
- R1 Comment 3: We clarify that our model is multivariate and modify the manuscript to make this clear.
- R1 Comment 4: We add a sentence and a citation to the manuscript to support the claim that some loss categories are likely to be correlated.
- R1 Comment 5: We discuss the generalizability of our model and our estimated parameters.
- R1 Comment 6: We provide 8 robustness tests to demonstrate that our model parameters are robust to changes in class (im)balance.
- R1 Comment 7: We explain how our model accounts for autocorrelation and integration.
- R1 Comment 8: We discuss how we plan to allow exogenous factors to affect bias terms in future research.
- R2 Comment 1: We provide 3 robustness tests to demonstrate that our results are stable across alternative bias priors including a Bayesian fixed effects model. We also describe relevant changes to the manuscript.
- R2 Comment 2: We provide a robustness test in which we demonstrate qualitatively similar model results when civilian casualties are omitted.
- R2 Comment 3: We clarify in the manuscript that we are agnostic to the source of biases and that they may or may not be the result of intentional misrepresentation.

Authors' Response to Reviewer 1

General Comments. The authors design a model to use claims of losses in personnel and equipment in the Russian invasion of Ukraine to estimate the true daily and cumulative losses in the first year of the conflict, given the assumption of source based biases and correlations across loss categories. In other words, the central underlying argument is that the single source estimates are insufficient because they are either biased or missing and the Bayesian model employed aims to identify a) a more accurate estimation of the number of losses and b) the degree of bias of each of the warring parties. I expect this to be of high interest to readers, and the writing is clear and concise. However, there is some additional information that would be helpful for the authors to provide that would give the readers more confidence in the model's claims.

We thank the reviewer for their encouraging comments and are happy to hear that they think this paper will be of high interest to readers. We believe that we have addressed all of the reviewer's concerns either through changes to the manuscript or clarifications and robustness tests provided in this letter.

Comment 1

First, how many more reports about Russian losses are there than Ukrainian losses? Is there any reason why we should expect this to be the case? What are the implications for having significantly less data (i.e. fewer over all claims) about the Ukrainian losses? (See above comments about requiring additional modeling expertise). I would expect the results to be less reliable - the authors could provide more summary data (how many reports from which sources about which country), and an explanation as to why the significantly fewer reports about Ukrainian losses does not present a problem. Are Russians more likely to report on Ukrainian losses? Are Ukrainians more likely to report on Russian losses?

Overall, we have 4161 reports on Russian losses and 448 on Ukrainian losses. We think this is due, at least in part, to Ukraine's effective use of media to shape the conflict narrative. Our model is specifically intended to deal with class imbalances such as these; this was the primary motivation for selecting a multilevel ("partially pooled") model in which our theoretical understanding of the relationship between overall source-target ("dyad") biases and specific source-target-category ("dyad-category") biases could be leveraged to better utilize information from classes with very few observations. Furthermore, we allow the overdispersion factor of the Negative Binomial-distributed cumulative losses to vary across target-categories (e.g., Russian tanks and Ukrainian tanks both have their own overdispersion parameter).

In this letter we provide Table 1, a summary of the number of reports per dyad. We further explore the consequences of this class imbalance in response to Reviewer 1's Comment 6. This includes estimating eight additional models in which we experimentally manipulate the class imbalance and show that quantities of interest remain substantively unaffected.

Unfortunately, we have not included the above summary of observations in the manuscript due to the journal's limitation of one table and one figure per brief report. We have, however, modified the manuscript to include the number of observations from

Table 1: Observations per claim source per target

	GB	OSI	Other	RU	UA	UN	US
Russia	32	116	127	28	3804	0	54
Ukraine	0	53	27	219	54	78	17

each claim source as is shown below.

> We aggregate these sources into seven groups that we refer to as "claim sources": OSI ($n = 169$), Russian sources (247), UK sources (32), Ukrainian sources (3,858), the United Nations (78), US sources (71), and other sources (154).

Comment 2

Do we have a sense of whether non Ukrainian or Russian sources exhibit bias? The focus is on Ukrainian and Russian biases - these are important but not surprising. It would be helpful/interesting to know if other sources exhibited similar levels of bias (both from a policy and theoretical perspective).

We agree with the reviewer and we do have bias estimates for all dyads for all loss categories. However, we chose to include only a small portion of those in the manuscript due to space constraints given the *brief report* format. Please find Figure 1 in this letter, which visually depicts all dyad-category bias estimates. We omit dyad-categories for which there are zero observations (e.g., the U.N. does not make estimates about Russian tank losses) and for the idiosyncratic loss categories that we omit from the paper itself (some sources report combined categories, for example). The bias estimates in this figure are exponentiated and can therefore be interpreted as multiplicative scalars. Blue and red credible intervals (CI) indicate that the CI does not contain 1 ("no bias").

The only significant biases we observe other than those from Russian and Ukrainian sources are those from *open source investigators (OSI)* and *other*. While *other* is necessarily hard to interpret, we have reason to expect that OSIs would underestimate equipment losses. Some investigators or investigative groups (like the famous Oryx) only count equipment losses for which they have photographic or video confirmation. Therefore, they are bound to underestimate actual losses. We were thus encouraged to see several bias coefficients < 1 among the equipment loss estimates from open source investigators.

Comment 3

Additionally, given there are so few reports of daily losses in the data, it is not clear to me what the benefit of modeling these separately is (or at least not as an entirely separate exercise, since daily loss estimates must be dependent on the cumulative claims).

Our goal is (in part) to introduce a general purpose methodology for estimating losses given conflict scenarios using *both* cumulative and daily reports. While we have relatively fewer daily reports for this particular conflict, we (a) believe that these reports are likely to be more precise than cumulative reports and so want to use them to inform our estimates and (b) do not believe that the daily / cumulative imbalance will hold for all conflicts. We conceptualize both daily and cumulative reports as noisy realizations of an underlying process and therefore leverage them both in a single model, not two separate models.

This is a very helpful comment for us because it highlights that we must be more clear in our description of the model in the manuscript. As Reviewer 1 points out, we have very few daily observations (162). In order to effectively leverage all of our available information, we estimate only a single model with all $162 + 4447 = 4609$ observations combined (daily + cumulative). Our model is therefore multivariate in the true sense – we have two different outcomes that share many parameters between them. The only

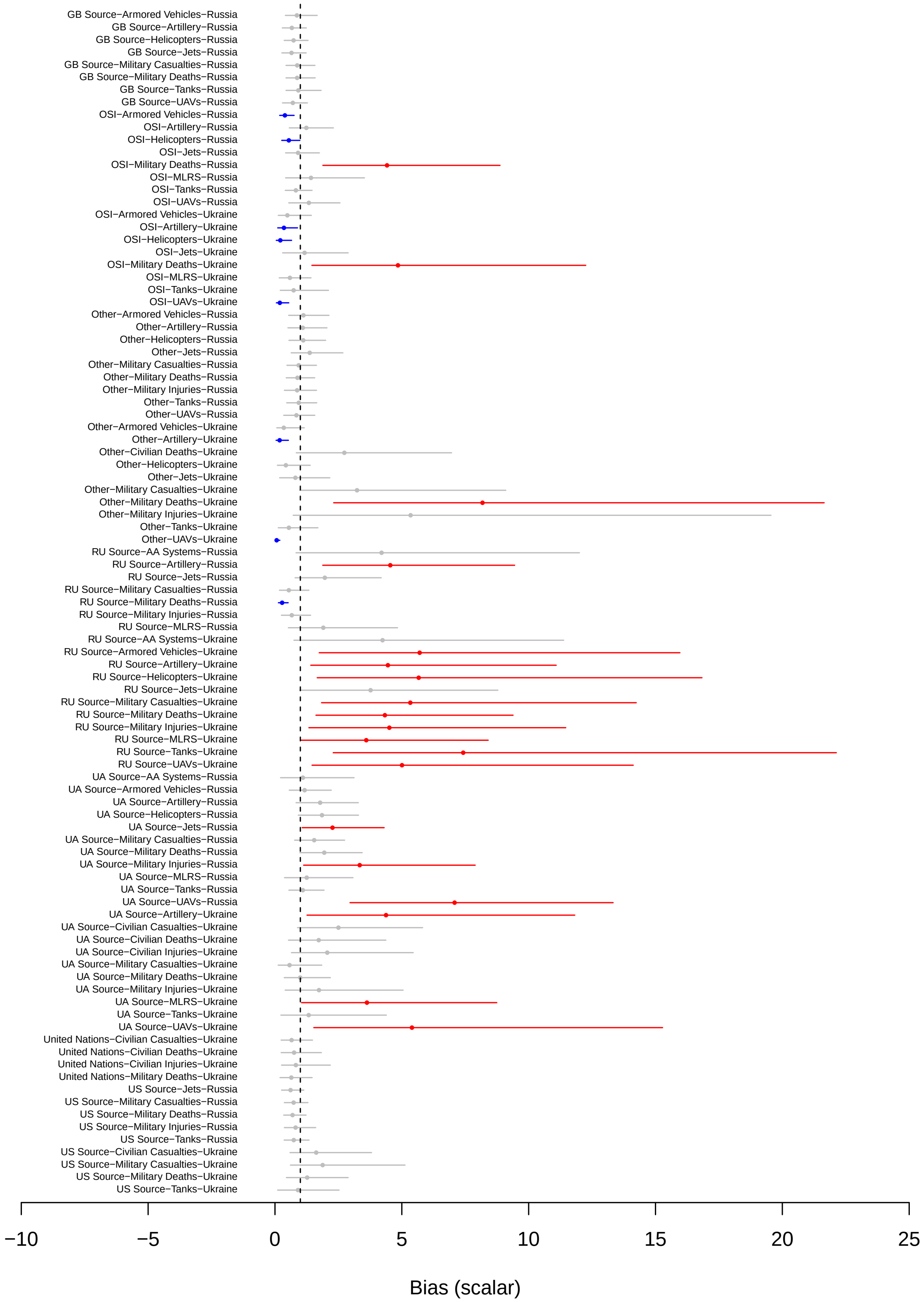

Figure 1: Biases for all reported loss types, by claim source and target country. Biases reported as multiplicative scalars.

difference between the means for daily and cumulative losses is that the cumulative means are the cumulative sums of the daily means. We can thus leverage all observations to inform the parameters of interest; e.g., the cumulative losses inform our estimates of daily losses. We go into a little bit more detail on this in our response to Reviewer 1's Comment 7.

To ensure that this is clear in the manuscript, we made the following highlighted change to the first sentence of the materials and methods section.

> We use a single multivariate Bayesian model with two outcomes: daily and cumulative loss counts.

Comment 4

> It would also be helpful to have a quick statement as to why we might expect the loss categories to be correlated (I can easily imagine the answer to this for some categories, but not necessarily for others.)

We have added the following simple example and a citation which we believe supports this reasoning. The citation is for a review of medical research on the types and amounts of human injuries incurred as a result of weapons designed for armored warfare.

> We also leverage reports of multiple loss categories to help fill gaps in reporting for any single loss category, under the assumption that losses in some categories will be correlated with others. For example, losses of armored vehicles are likely correlated with casualties (Khorram-Manesh et al., 2021).

Comment 5

Finally, the implicit claim here is that the authors can derive the true (or at least better estimate of the) number of losses on each side of the conflict. There is no theoretical claim here (to be clear this is not likely to be expected for a brief report), so the main contribution is a timely, relevant, new dataset of loss claims and the use of a Bayesian model to deduce some (close to) true value of the losses. To what extent is this approach specific to this conflict? Perhaps more directly, to what extent is it the case that the parameters assumed to be stable in this conflict will remain so in others? And to what extent is the estimation a function of the particular parsing of sources?

This is a good question that we will try to address in parts:

1. **...specific to this conflict?** We have no reason to suspect that this model is not generalizable to other conflicts. The model is extensible in the sense that it can accommodate an arbitrary number of dyads and categories, but we suspect the data collection itself may be more difficult in other settings for the reasons described here. First, there is a substantial fog of war with respect to loss counts; we see very different estimates from many different sources for quantities that should theoretically be countable. The high level of media coverage of this conflict also means that many estimates are reported, easing the data collection process. The conflict is between well-equipped militaries that use a variety of different technologies on the battlefield, allowing us to model losses across many different categories. Finally, the conflict has two (relatively) well-defined sides. This makes it easier to determine the "target" of losses. This may not be the case in, for example, civil wars in which parties may participate in complicated or ad hoc alliances. However, given adequate data, we believe that our model (with some modification) could be applied to any militarized conflict.
2. **...parameters assumed to be stable... in others?** On the other hand, all conflicts are unique and we do not anticipate that our estimated parameters will

describe other conflicts. For example, while we do not explore the motivations for reporting biases in this manuscript (which we discuss more in response to Reviewer 2's Comment 3), it could be that some reported figures are strategically chosen to shape narratives that result in greater material support from foreign allies. In wars where foreign powers are less likely to intervene directly (e.g., civil wars), it may be that belligerents make different strategic decisions in their communications due to the audiences they wish to target. We think this is a very promising avenue for future research and look forward to using our model in a comparative context across conflicts. As we discuss in response to Reviewer 1's Comment 8, we are especially interested in understanding the strategic misreporting of losses in response to external factors.

3. **...particular parsing of sources?** All parsing of sources was done by hand. We searched for and collected online reports that included claims of losses. Then our research assistants manually coded all claims using the context of the associated article or report. The model estimation itself is independent of parsing these texts. This means that the model should generalize across conflicts even if the coding rules need to be adapted for local contexts (if, for example, there are three major sides rather than two).

Comment 6

How are the authors accounting for the fact that there appear to be many fewer Ukrainian reports than Russian reports (drawn from Table 1)? Similarly, it appears that there are many more claims ABOUT Russian losses than Ukrainian losses (drawn from Figure 1). To what extent does the model take into consideration that there appear to be much more data to work with both from, and about, Russian losses?

Somewhat relatedly, how does the model handle the problem in which the Russians are more likely to report on Ukrainian losses (i.e. for every individual claim made by Russian authorities, there is a 75% probability that that claim is about Ukrainian troop losses, and only a 25% probability that that claim is about Russian losses)? The decision to report/claim and the content of that claim are linked - this could mean that the estimates of bias are an artifact of the number of claims made by each of the warring parties.

We believe our model to be robust to class imbalance; this was a primary motivator for our choice to use multilevel bias estimates. For dyad-categories that have few data points, the bias estimates are pulled towards the dyad bias mean. This allows each bias term to be informed by more information than would be possible in a non-multilevel approach. That said, because we are using a partially pooled estimator, we expect the bias estimates to change as more data are collected: as more data for a dyad-category are collected, the bias term will move away from the dyad mean bias and toward the dyad-category specific mean bias.

However, we chose to model bias terms at the dyad-category level specifically to avoid the situation that the reviewer has identified: over-reported categories may unduly influence our bias estimates. By grouping at the dyad-category level rather than simply the dyad level, dyad-categories with less data will receive their own unique bias terms even when other categories within that same dyad have many times more observations.

To evaluate our model's sensitivity to class imbalance, we estimated 8 additional models. In the first four of these, we duplicate the Russian source data 2×, 3×, 4×, and 5×. In the second four models, we duplicate the Ukrainian source data. We then computed the empirical distributions of quantities of interest and depict them in Figure 2. The model appears robust to even very large changes in class balance.

Comment 7

Being insufficiently familiar with the modelling approach: given that the data is cumulative deaths, these are likely to be correlated from one day to the next (within the same source - like a random walk). How does the model handle this?

This is an excellent question and is related to Revewer 1's Comment 3. We estimate only a single model with two outcomes (i.e., dependent variables): daily loss reports and cumulative loss reports. The latent values for losses are estimated at the daily level using smoothing splines. These smoothing splines help us to account for autocorrelation at the daily level. These estimates are then cumulatively summed over time to produce the cumulative loss estimates. In essence, we are modeling the first differences of the cumulative series as, by definition, the first differences of the cumulative time series *are* the daily time series.

Put another way, in order to model cumulative losses (an integrated time series), we compute the first differences of the integrated time series and model those differences instead. This is considered best practice for dealing with first order integrated time series. Where our model differs from (e.g.) a traditional ARIMA model is that we cannot *actually* compute the first differences of the cumulative time series because they are incomplete. There are many observations at time t that are missing observations at time $t + 1$ or $t - 1$. Therefore the series cannot be differenced. Our approach is to instead treat the first differenced series as a latent variable and to impute it (using smoothing splines). Where we have daily observations, we use those to inform these imputed daily values. Where we have a cumulative observation at time $T = t$, we use the cumulative

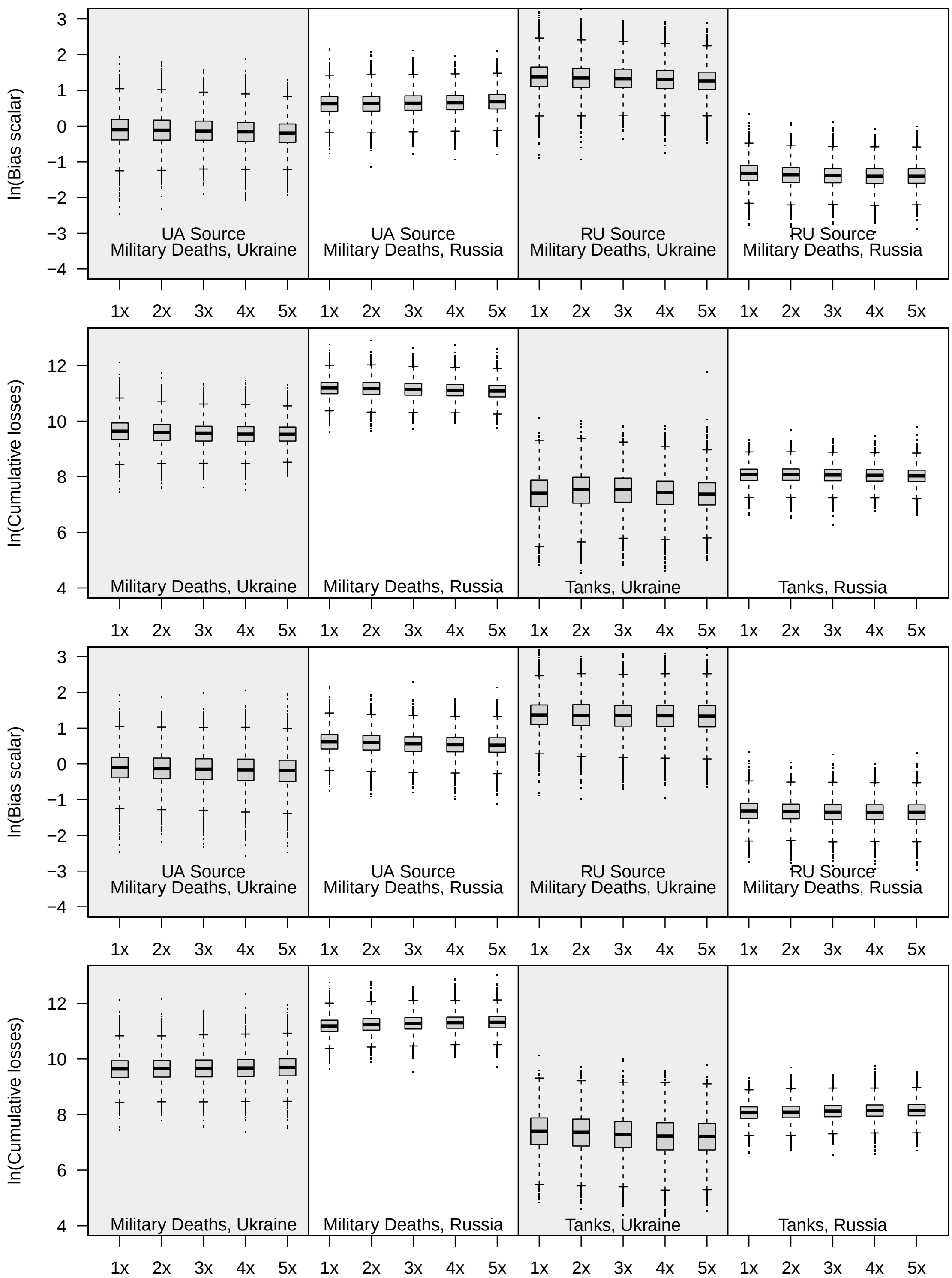

Figure 2: Selected empirical distributions of quantities of interest given data duplication. Top two panels: Russian source duplication. Bottom two panels: Ukrainian source duplication.

sum of the imputed daily losses from time $T = 0$ to time $T = t$ as our estimate.

Comment 8

To what extent does the model allow for the bias parameter to vary throughout the conflict? The authors are measuring bias as simply a source parameter that captures systematic over- or under-estimation by a particular source of a particular loss category. If bias is strategic in some way (actors are more wrong when they might believe it favors them, and less wrong when it does not - imagine Zalinsky overstating the number of Ukrainian deaths just ahead of a critical request for military aid, for example), we might expect its magnitude to change over the course of the conflict. This suggests that potentially exogenous events over the course of a conflict might shape the degree (and potentially the direction) of bias that particular sources exhibit, undermining the extent to which their bias can be helpfully captured by a single bias parameter. I do not have a good enough understanding of the modelling approach to evaluate whether this is likely to present a serious problem in using this data to estimate true values of losses, whether it presents a problem just for estimating a bias parameter (i.e. how much Russia is likely to understate its own losses), or neither.

This is a really fantastic suggestion and one that we hope to incorporate in follow-up work. Currently, we assume a single bias term per dyad-category for the duration of the first year of conflict. Therefore, the bias terms can best be thought of as averages over this period. It is possible to instead model the bias terms themselves by regressing them on exogenous factors such as nonlinear time trends or major events. This would be especially helpful when modeling very long conflicts or conflicts in which the leadership or strategy of one or more sides changes substantially.

That said, our team does not yet have strong beliefs about which events would matter and why. On the other hand, we do suspect that there is value in consistent messaging for the claim sources – sources may lose credibility if their estimates jump around in response to world events. If, for example, a single source reports cumulative losses that decrease over time rather than increase, it may diminish the credibility of their future reporting. For these reasons, we chose to err on the side of parsimony and assume time-invariant bias terms.

Authors' Response to Reviewer 2

We would like to thank Reviewer 2 for their very careful reading of our manuscript and especially for their careful consideration of our methodological choices.

Comment 1

My biggest concern is how the statistical model accommodates "source-specific biases" in the reporting of losses. Specifically, if I am reading the materials and methods correctly, then each bias term (for a category, source, and target) is drawn from distributions with means that vary according to the source-target dyad. While these means may differ across source-target dyads, averaging over dyads, the means are zero. I see no reason why this would be the case. For instance, fix a category, e.g., civilian deaths. I view the data generating process as a source (e.g., Russia, Ukraine, UN) as determining its own bias for each target. Why would these biases average out to 0? To address this concern, I think it is reasonable for the authors to show that their empirical findings are robust to an alternative model specification in which the mean bias terms (γ_{st}^{bias}) are fixed effects, not random effects. I do not know if such a model can be identified theoretically or fit to data practically, but I am skeptical of results that require assumptions about the distribution of biases across sources.

Thank you for your careful reading of our paper and consideration of our model. This is a very astute point that gets at one of our key modeling decisions. We would like to address this comment in several ways. First, we will clarify our modeling choices as they are presented in the paper. Then, we will discuss several robustness tests that we have undertaken in response to this particular comment. Finally, we will discuss how this comment has resulted in improvements to the manuscript.

As Reviewer 2 points out, we select a prior distribution for $\gamma_{st}^{bias} \sim Normal(0, \sigma_1^{bias})$, the source-target ("dyad") bias means. This is simply a prior and does not guarantee that the mean of these γ_{st}^{bias} terms is precisely zero. The prior is independent across γ_{st}^{bias} terms and they will therefore not necessarily perfectly cancel.[1] The mean value of these terms in our model is 0.05 (in the original logged scale of the parameters). We chose this prior deliberately – any other mean for this prior would indicate an a priori assumption about the direction and magnitude of bias for one or more source-target dyads. This prior represents our initial expectation for the bias of a given dyad and we intentionally chose it to be conservative, requiring that any bias estimated with the model is informed by the data and not by our own preferences. Another way of looking at this prior is as a regularization term: a zero-centered prior has the effect of "pulling" the estimates of bias terms for which we have little data toward the conservative estimate of 0, or "no bias." The multilevel bias estimates are then partially-pooled, a deliberate choice in that it prevents our model from estimating a very large dyad-category bias term when only a small number of anomalous reports exist to inform that particular bias term. If lots of data are present for every dyad, the model will not enforce that the bias terms are mean zero. Their values are free to reflect the bias present in the data.

In considering this comment, our team discussed how we might implement a fixed effects alternative to our current model. We agree with the reviewer that a traditional fixed effects parameterization of this particular model (such as an MLE FE estimator) would likely not be identified and would require substantial re-working. We have instead estimated three additional models that we think address the noted concerns, including one that is very similar to a FE model.

We first estimate a model with a set of dyad-category bias terms ($\beta_{c,st}^{bias}$ in the paper) but without the higher-level dyad γ_{st}^{bias} terms. In other words, we estimate the same model as that presented in the manuscript but omit the hierarchical/multilevel component. This is conceptually similar to fixed effects where the fixed intercepts are associated with source-target dyads and loss categories, all of which are independent from

[1] In fact, we did briefly experiment with a model that guaranteed a precise zero mean bias, but rejected it because we agree with the reviewer that such a strong assumption is unwarranted. It also required a substantially more complicated model.

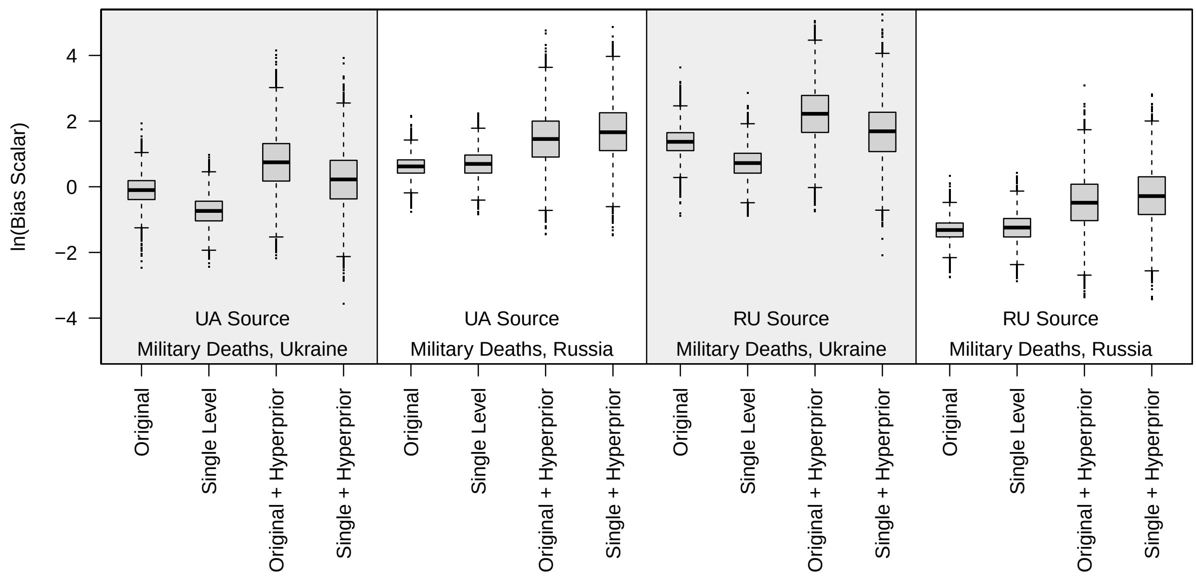

Figure 3: Selected source-target category bias terms (logged) under four different models.

one another. We call this the "single level" model as opposed to the "original" multilevel model.

Next, we re-estimate both the *single level* and *original* models allowing for random means in the prior for these biases. In other words, we have substituted $\sim Normal(0, \sigma_1^{bias})$ with $\sim Normal(\mu_1^{bias}, \sigma_1^{bias})$ where $\mu_1^{bias} \sim Normal(0, 1)$.[2] We refer to these models as "single level + hyperbias" and "original + hyperbias," where "hyperbias" is shorthand for the inclusion of a mean bias hyperparameter. These new priors replace the priors in Equation 9 of the manuscript.

Figures 3 and 4 depict the posterior empirical distributions of our primary quantities of interest under each of these four models. These distributions largely overlap and all models share similar central tendencies, though the uncertainty in estimates is greater when we include the random hyperprior on the prior mean. Most importantly, the estimates obtained in the multilevel and single level "FE" models are very similar, conditioning on the inclusion of the mean hyperparameter.

[2]Please note that the model is estimated in log parameter space and therefore this is not as strong a prior as it might otherwise appear to be.

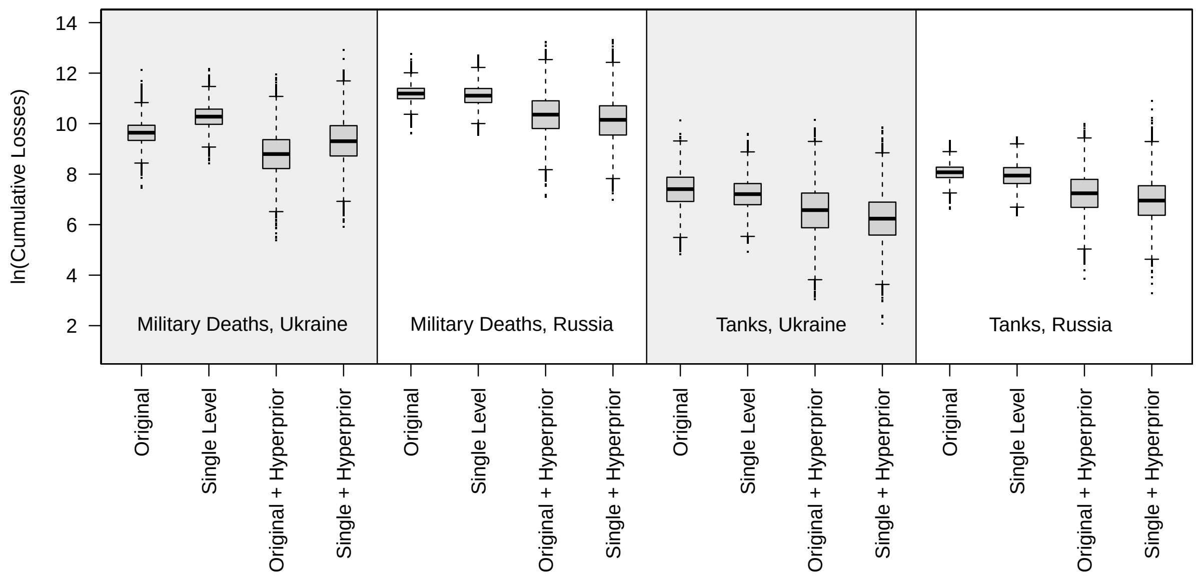

Figure 4: Selected cumulative loss estimates (logged) under four different models.

In practical terms, the value of μ_1^{bias} dictates the overall bias exhibited by all sources – systematic biases that impact all source-target dyads equally. Lacking any theoretical reason to expect such a bias, we prefer to assume that all biases are idiosyncratic to their dyads (i.e., there isn't a shared systematic bias across all dyads) and therefore prefer our original model for its parsimony.

In light of this comment, we have decided that the "random" versus "fixed effects" terminology is not appropriate here. "Random" and "fixed" effects lack consistent definitions across fields and modeling contexts. Furthermore, due to the cumulative latent variable, our model is not a typical panel regression model of the type that most scholars will associate with the terms RE or FE. We have therefore removed the reference to "random effects" from the paper and replaced it with the more accurate "coefficients" or "multilevel effects," where appropriate. We had originally used "random effects" to describe these multilevel effects. "Coefficients" better represents all scalar parameters in the model, both multilevel and single level. We believe this terminology is more familiar to readers, more precise, and more descriptive of the roles these parameters play in the model. The relevant changes to the manuscript are highlighted below.

Coefficients are denoted with β... Every observation is scaled by a multilevel bias coefficient that is specific to its source-target pair (indexed by st) and category, $\beta^{\text{bias}}_{c,st}$, to account for systematic over- or underestimation... These are normally distributed around source-target specific means which are themselves normally distributed with prior mean zero. The prior mean of zero for the bias terms encodes the conservative belief that a source is unbiased absent data to indicate otherwise. We elect to model bias terms as multilevel due to our theoretical expectation that source-target-category biases are likely correlated within source-target pairs and to mitigate source-target-category class imbalances through partial pooling.

Comment 2

My more minor concern is why include information from civilian casualties? Most of the categories in Table 1 are about military performance. For example, I view civilian casualties and combatant casualties as outcomes from different data generating process with different reporting incentives. This concern is more minor. I can see why the authors pursue a "more outcomes is better" approach. To address this concern, I think it is reasonable for the authors to show that their substantive conclusions are robust when they exclude civilian casualties.

Our team debated whether or not to include civilian losses in our model as well. On the one hand, we suspected that civilian losses would likely correlate with Ukrainian military losses to some degree. On the other, we also agree with the reviewer that the data generating processes for reporting on civilian losses may differ from those for other loss types. Ultimately, we chose to include civilian losses because we believe our model is flexible enough to accommodate them even if the biases associated with their reporting differ from biases associated with reporting of other loss types. Evidence for this can be seen in Figure 1 of this letter. In particular, our model allows for civilian losses to be either correlated or uncorrelated with all other loss types and allows for civilian loss bias terms to deviate from the other bias terms for a given dyad. Finally, we suspected that the omission of civilian casualties would lead to the opposite comment ("why not include them?") and that readers would be particularly interested in these estimates.

As suggested, we have also re-estimated our model omitting civilian loss categories and replicate Table 1 from the paper in Table 2 of this letter. The results are substantively similar to those presented in the paper.

Table 2: Model re-estimated without civilian loss categories.

ISO2	Loss type	n	Est.	CI
RU	AA Systems	233	331	[71–1052]
UA	AA Systems	13	1238	[101–6087]
RU	Armored Vehicles	400	6153	[2819–11223]
UA	Armored Vehicles	15	3220	[701–8418]
RU	Artillery	380	1438	[664–2655]
UA	Artillery	35	2269	[440–6914]
RU	Helicopters	389	167	[80–302]
UA	Helicopters	30	61	[12–174]
RU	Jets	409	141	[65–264]
UA	Jets	38	120	[31–361]
RU	Military Casualties	130	209510	[101692–383868]
UA	Military Casualties	16	94169	[23199–237425]
RU	Military Deaths	523	74309	[36916–133446]
UA	Military Deaths	67	21288	[7093–57316]
RU	Military Injuries	44	148532	[45864–379758]
UA	Military Injuries	8	37274	[5376–146964]
RU	MLRS	261	469	[140–1174]
UA	MLRS	27	526	[144–1516]
RU	Tanks	501	3287	[1618–6023]
UA	Tanks	33	2171	[416–6244]
RU	UAVs	292	329	[144–696]
UA	UAVs	40	1486	[262–3974]

Comment 3

The most minor concern, which is does not need to be addressed, is the following: why do countries like Ukraine and Russia lie? Suppose Ukraine and Russia know that one can tabulate this data and fit the authors' model to back out the true casualties.Then why lie in the first place? Or why not lie even more to influence the "true casualties" estimated with the pooled information?

We would like to clarify that we do not accuse any sources of lying, but rather of producing systematically biased estimates. These biases could be produced unintentionally through differences in procedures, differences in available information, misreporting at lower levels of command, or they could be produced intentionally. Nonetheless, there are many reasons countries might believe they would benefit from misreporting. For example, the benefits they receive from doing so may be accrued more quickly than researchers are able to produce better estimates (it has taken us over a year to produce the estimates in this paper). The general population receives most of its information on this topic from the news, not from academic articles. Furthermore, to the best of our knowledge, we are among the first to estimate conflict losses during an ongoing conflict from open source reporting and to adjust for these biases while doing so. In other words, states may have no reason to suspect that anybody can or would do this. Finally, we would also like to clarify that we do not estimate "true casualties," and have double checked the manuscript to ensure that we have not erroneously claimed to do so.

We clarify this in the manuscript as follows (changes in purple).

> We then develop a statistical model to better estimate losses for both sides given these reports.

> We also estimate the biases exhibited by claim sources with respect to loss categories. Bias does not necessarily imply intentional misrepresentation, but rather any systematic over- or underestimation relative to our estimated loss values.