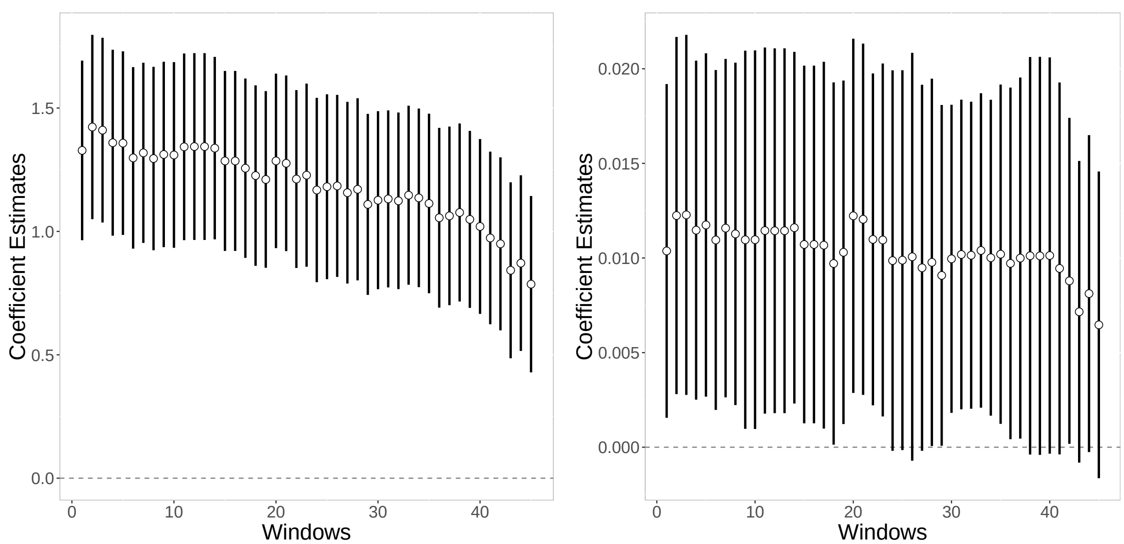
Supplementary Appendix for "A Re-Assessment of Reporting Bias in Event-Based Violence Data with Respect to Cell Phone Coverage"

July 6, 2017

Contents

1 Additional Figures – Cross-sectional Models

Figure 1 shows additional model specifications for the windowed analysis using the cross-sectional data. Panel (a) displays coefficient estimates for a logit model that includes the country mean of cell phone coverage (top control for unobserved, time-invariant factors). Panel (b) shows results for a linear probability model of the cross-sectional data that includes fixed effects for countries and a spatial lag of the dependent variable. In both cases the windows here are based on the high-fatality estimate to order events. For both model specifications, a slight decline of the marginal effect with event severity is visible. Yet generally, the vast majority of effects are still substantial and the 95% confidence intervals do not cover zero.



(a) logit with country mean, "high" estimate (b) spatial lag + fixed effects, "high" estimate

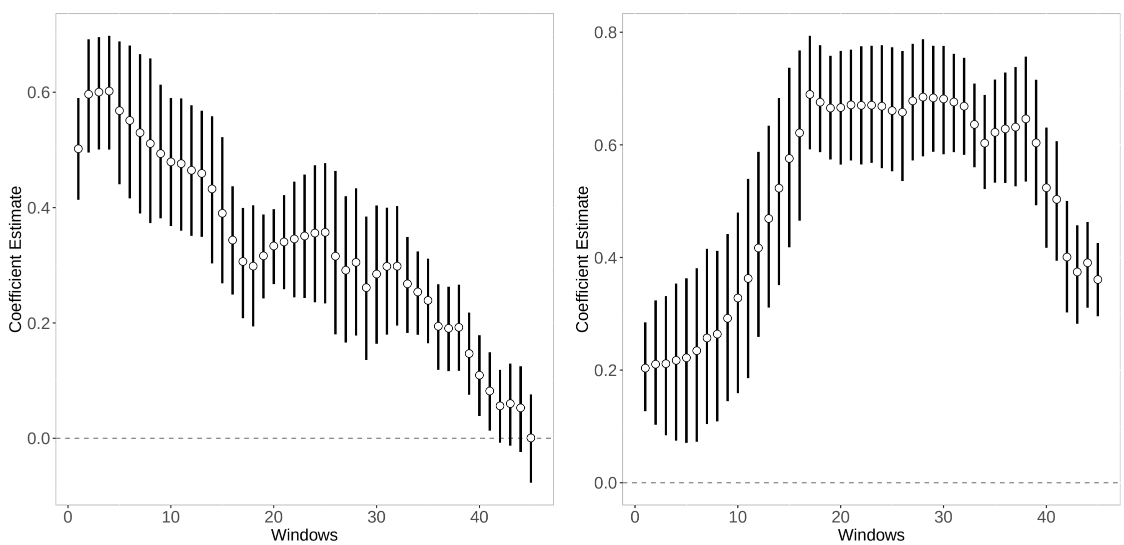
Figure 1: This figure shows coefficient estimates of cell phone coverage for the logit models including the country mean of cell phone coverage (left plot), as well as for a linear probability model with country fixed effects and a spatial lag. The windows here are based on the high-fatality estimate to order events by severity. The red horizontal line depicts the overall coefficient estimate using all violent events.

1.1 Ties in Ordered Event Windows

A potential reason for using the *high* estimate is that the *best* estimate has many events with zero fatalities, which cannot be ordered. Indeed, based on the *best* estimate, the UCDP-GED data set reports 161 events with zero fatalities and 164 events with one fatality for 2008.

However, using the *high* estimate does not change this problem, as it just shifts the number of events to higher-fatality categories. Yet within categories, events still cannot be ordered. Based on the *high* estimate, we now have 189 events with one fatality and 136 events with two fatalities. In total, the *high* estimate provides ordering based on 77 categories, while the *best* estimate allows ordering based on 66 categories.

It is unclear why the *high* estimate should be preferred over the *best* estimate when ordering events, but this again is a choice scholars would have to make when it comes to the sensitivity analysis. One additional way to deal with this problem, is to first order events based on the given fatality estimate and subsequently break ties based on random chance. We do so for both the *high* and *best* estimates.[1] Specifically, we do the random ordering 250 times and then estimate the windowed models. We then draw 300 coefficient estimates based on the coefficient and standard errors. Again, this allows us to account for estimation uncertainty in the different windows.



(a) ordering *high* estimate and ties ordered randomly

(b) ordering *best* estimate and ties ordered randomly

Figure 2: This figure shows coefficient estimates of cell phone coverage for the logit models when events are first ordered by the given fatality estimate and then ties are broken based on random draws.

AS one can see in Figure 2, this leads to similar results are before. Using the *high* estimate (left plot), the results are in line with reporting bias as suggested by Weidmann

[1] We thank an anonymous reviewer for suggesting this approach.

(2016). When using the *best* estimate to first order events (right plot), the results do not suggest reporting bias.

Lastly, we also present the windowed analysis when events are first ordered by the *best* estimate and then ties are broken by ordering according to the *high* estimate. Again, results do not change significantly, and are not suggestive of reporting bias.

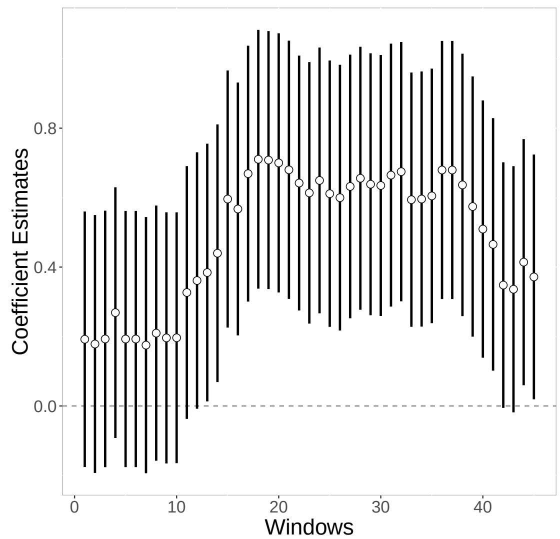
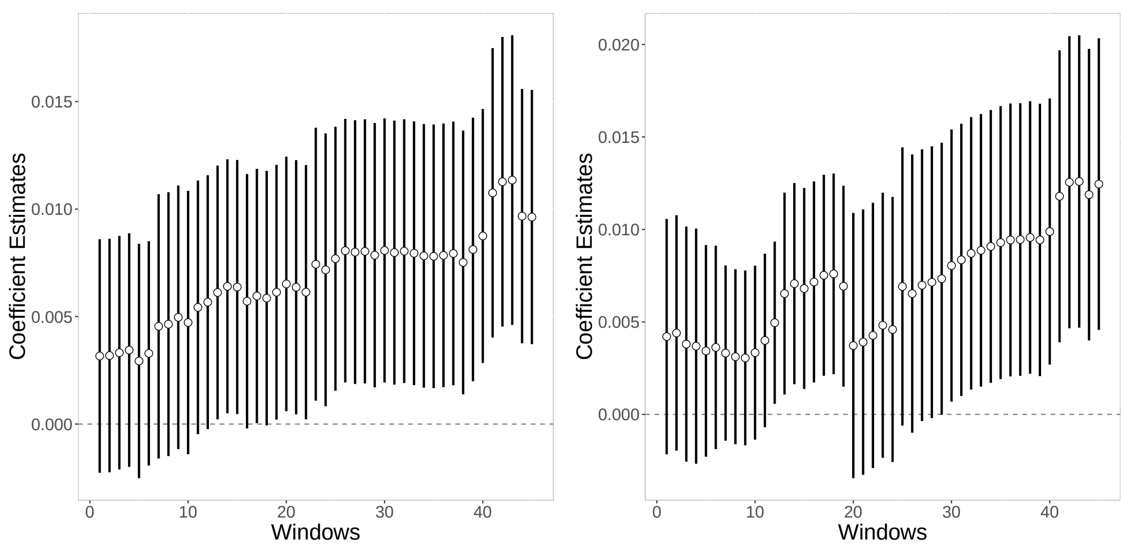


Figure 3: This figure shows coefficient estimates of cell phone coverage for the logit models when events are first ordered by the *best* fatality estimate and then ties are broken based on the *high* estimate.

2 Additional Figures – Panel Models

To implement the sensitivity analysis for the panel data, we follow Weidmann (2016) and create sliding windows for each year in our data set: 2007, 2008, and 2009. The step size by events for sliding the window is 10 for 2008 (as in Weidmann 2016). To create the same number of windows for 2009, however, the step size is increased to 15 (because of the larger number of events). This is done so that we can create 45 windows for each year.

Figure 4 shows coefficient plots for the panel data using the "best" fatality estimate. Panel (a) restricts the data to events that can be geo-located precisely. Panel (b) shows a coefficient plot for the panel data with the "best" fatality estimate, based on the updated version of the UCDP-GED data. The original study by Pierskalla & Hollenbach (2013) relied on an older release of the UCDP-GED data. The new version of the data includes a larger set of events. In both cases the coefficient estimates increase with the fatality estimate, the exact opposite patterns implied by reporting bias.

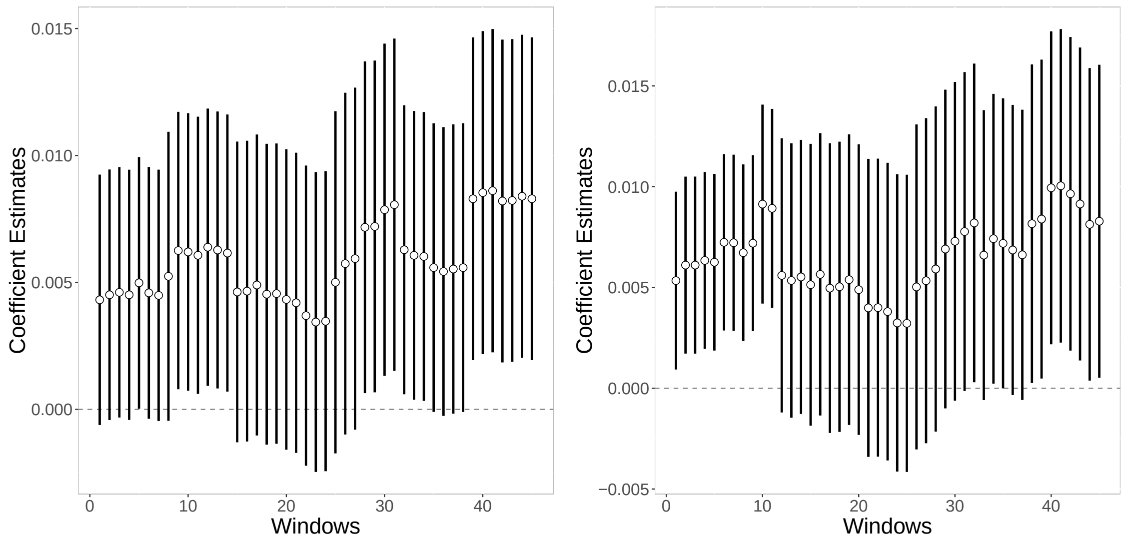


(a) Panel Data, Best Estimate, Precise Data (b) Panel Data, Best Estimate, UCDP 1.5

Figure 4

Figure 5 shows the same model specifications as Figure 4 but relies on the "high" fatality estimate. Again, the coefficients increase with the fatality estimate. This example also illustrates that it is not only the change from the "high" to the "best" estimate that produces patterns inconsistent with reporting bias.

Figure 6 shows coefficient plots for the panel data relying on the count of events per



(a) Panel Data, High Estimate, Precise Data

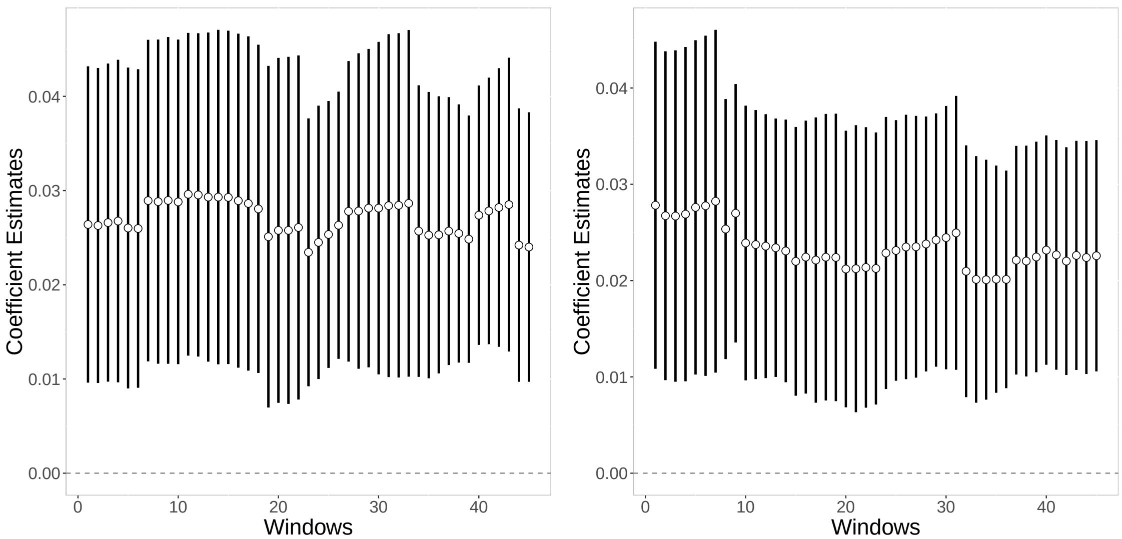
(b) Panel Data, High Estimate, UCDP 1.5

Figure 5

grid-cell-year (estimated via OLS). Panel (a) shows results for the “best” fatality estimate and panel (b) for the “high” fatality estimate. In both cases coefficient estimates are very stable across the event windows, providing no evidence of reporting bias.

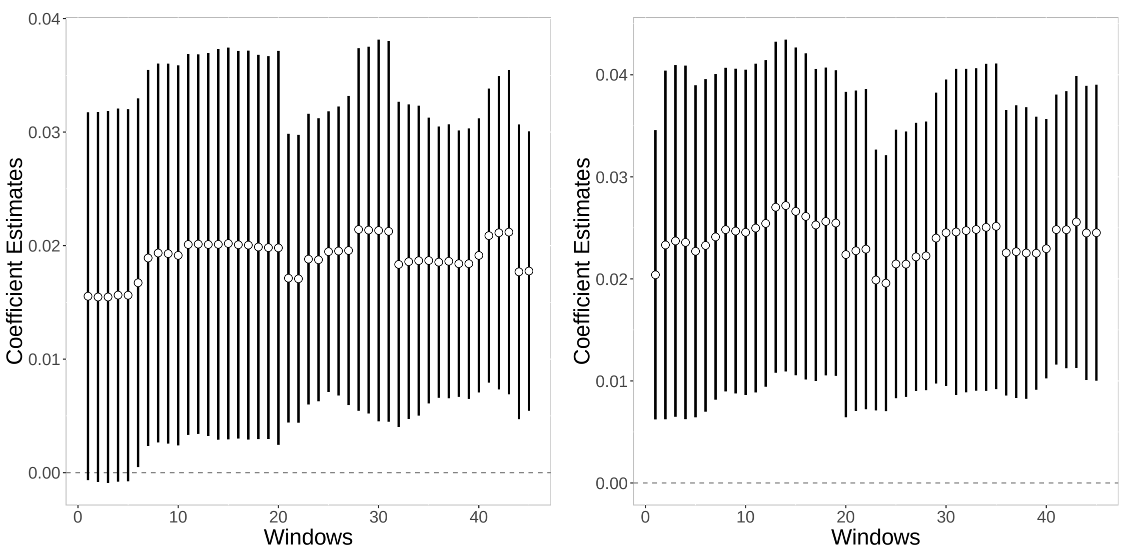
Figure 7 shows coefficient plots for the panel data and the count dependent variable, windows ordered by the “best” fatality estimate for precisely geo-located events (panel (a)) and data based on the new version of UCDP-GED (panel (b)). Coefficient estimates are stable across event windows, indicating no reporting bias.

Figure 8 shows coefficient estimates for the same models as in Figure 7 but ordering windows by the “high” fatality estimate. As before, there is no indication of an attenuation of coefficient estimates for higher fatality events.



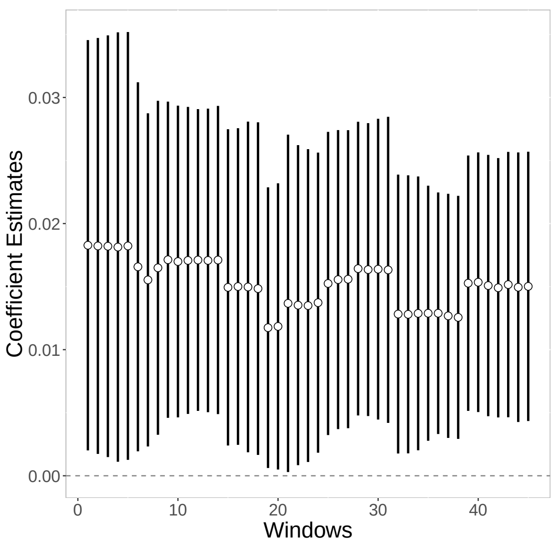
(a) Panel Data, Count Model – Best Estimate (b) Panel Data, Count Model – High Estimate

Figure 6

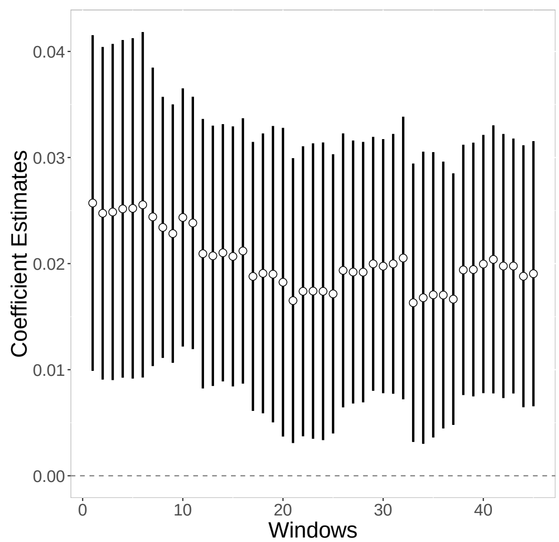


(a) Panel Data, Best Estimate, Precise Data (b) Panel Data, Best Estimate, UCDP 1.5

Figure 7



(a) Panel Data, High Estimate, Precise Data



(b) Panel Data, High Estimate, UCDP 1.5

Figure 8

3 Alternative Sensitivity Analyses

Creating event windows based on fatalities is one reasonable approach to test for reporting bias. We believe we can use two simple alternative sensitivity checks to further test for the presence of reporting bias. Two other factors that should mitigate reporting bias are distance to the capital and grid cell population counts. Locations close to the national capital will be more heavily scrutinized, and foreign press offices are often located in capitals. Violent events thus ought to be more likely to be reported close to the capital, independent of the presence of cell phone technology. Similarly, locations with larger populations are also more heavily covered by the international and national media. Hence, introducing cell phone coverage in urban areas is less likely to lead to the reporting of unreported violent events, since a large share of events has already been identified. One the other hand, introducing cell phone technology in more rural, remote areas is likely to have a larger marginal effect on the reporting of events. Thus violent events should be reported at a higher rate near the capital, and bias introduced by cell phones should be smaller. Hence, if reporting bias is driving the findings, we would expect that the effect of cell phone coverage is strongest in locations far away from the capital and with smaller population counts.[2] To test this idea, we first estimate a simple interaction model for the three-year panel data, interacting the time-varying cell coverage indicator with each grid cell's distance to the capital and population size (the constituent terms are being absorbed by the grid cell fixed effects). Table 1 reports the estimated coefficients and associated standard errors. As clearly shown, the interaction of capital distance is negative for both the binary and count models. This indicates that the effect of cell phones on violent events is stronger closer to the capital, the opposite of what one would expect if reporting bias were driving the result. For population size, the interaction term is statistically and substantively indistinguishable from zero, again not indicating a pattern of reporting bias.

[2]Note that this an assumption on our end. If cell phone technology has a larger effect on reporting in urban environments – either because networks are more reliable or because penetration of the population is deeper – then our expectations about reporting bias are reversed.

Table 1: Interactions, FE-OLS

	(1) Binary	(2) Count	(3) Binary	(4) Count
Cell Coverage	0.0177*	0.0631***	0.00877^{+}	0.0526**
	(0.00839)	(0.0184)	(0.00523)	(0.0172)
Cell Coverage × Capital Distance	-0.0000136	-0.0000284^{+}		
	(0.0000114)	(0.0000161)		
Cell Coverage × Population			4.27e-08	-3.49e-08
			(5.54e-08)	(9.97e-08)
Year FE	✓	✓	✓	✓
Grid Cell FE	✓	✓	✓	✓
Observations	32,022	32,022	32,022	32,022

Standard errors in parentheses, clustered at the grid cell level.

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

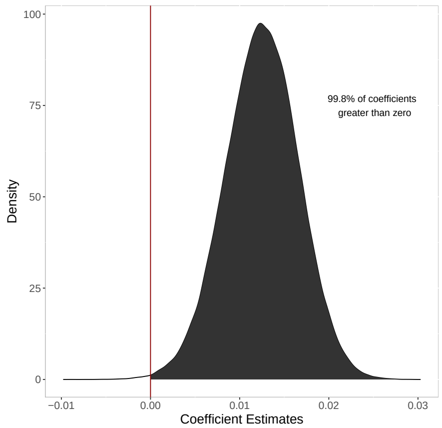
4 Additional Simulation Results – Panel Model

Figure 9 shows the distribution of coefficients for the panel data estimations with grid cell fixed effects. To add fake events for the panel model which does not include any covariates aside from the cell phone coverage variable, we do the following. First, we estimate cross-sectional models for each year with the same covariates as in the cross-sectional model in Pierskalla & Hollenbach (2013) but not including the cell phone coverage variable. Then we draw new "fake" event locations for each year based on the predicted probabilities from each of the cross-sectional models. We then estimate the panel models with grid cell fixed effects on the new simulated data.

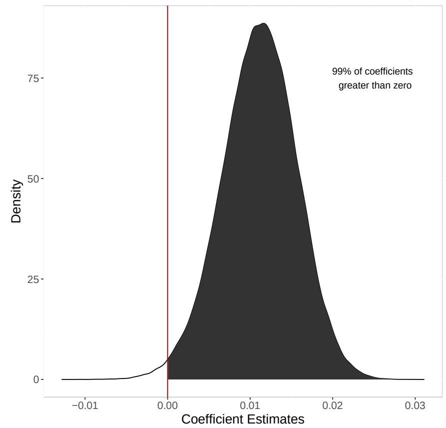
Panel (a) in Figure 9 shows, for the case when the observed data are only 95% of the "true" events in the non-cell coverage areas, the coefficient estimate is positive in 99.8% of cases across all 1000 simulations. As we move through the remaining panels (b)-(e), increasing the amount of added events, we start to observe more and more cases in which the coefficient for cell coverage comes close to or is smaller than zero. If the observed data are only 90% of "real" data we find that 99% of coefficients are still larger than zero. Even in the most extreme example though, our observed data represent only 75% of actual events, still 86.3% of coefficients are larger than zero. While in the majority of simulated datasets we would still find evidence in line with the original finding, at this simulated level of reporting bias, the level of uncertainty around the estimated coefficient becomes larger.

In the case of the panel data, there is a chance reporting bias could explain the original finding, depending on the magnitude of reporting bias in the African UCDP-GED data. If we assume the same first difference effect as in Weidmann's (2016) Afghanistan study[3], but take a baseline probability of reporting in non-cell areas of 30%, then panel (d) in Figure 2 in the main text and Figure 9 below are the appropriate references—showing that with this level of reporting bias the estimated effect of cell phone coverage is positive in 99.9% of cases in the cross-sectional models and 93% of the time in the panel data. If we assume that an event has a 50% chance of being included in the African UCDP-GED data in areas without cell phones and a 57% chance in areas with cell coverage, then our observed data represent 88% of observable events if cell phone coverage were present. Panel (c) in Figures 2 in the main text and Figure 9 indicate that this implies 97-100% of estimated coefficients across our simulations are larger than zero. In summary, only in the worst case scenario, in which

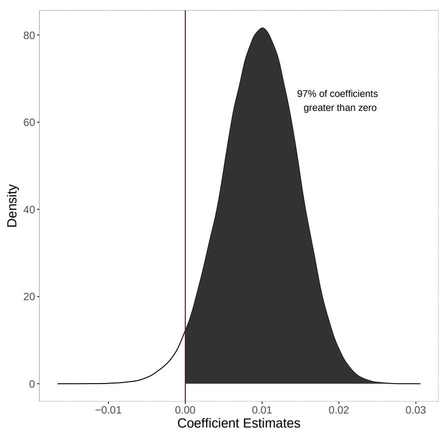
[3]It is plausible to believe that this effect might actually be smaller for higher baseline probabilities of reporting, since the marginal benefit of cell phone technology will be limited if nearly all events are already being accurately reported.



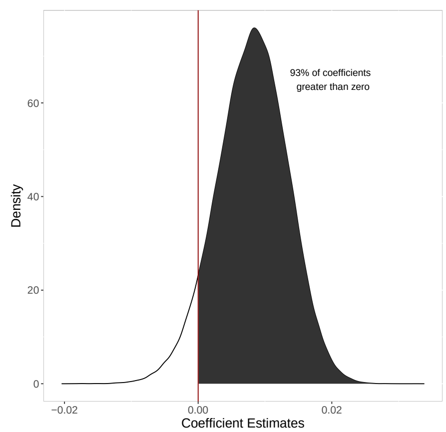
(a) Observed data 95% of "real" events



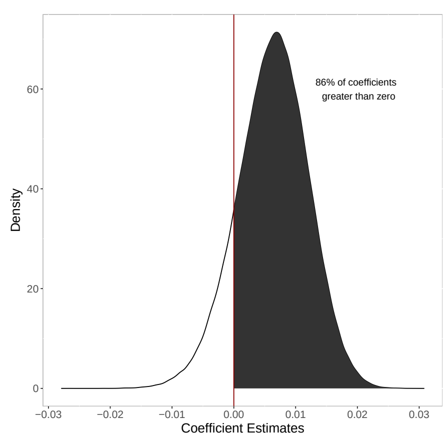
(b) Observed data 90% of "real" events



(c) Observed data 85% of "real" events



(d) Observed data 80% of "real" events



(e) Observed data 75% of "real" events

Figure 9: This figure shows the empirical distribution of coefficients for the cell coverage variable in our panel models for each level of added events for over 1000 randomizations. The shaded area indicates the share of coefficients that are larger than zero.

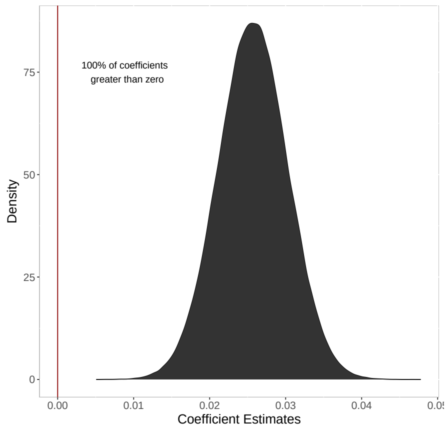
reporting in the UCDP-GED data as a whole is exactly like in the case of Afghanistan, is there a chance that reporting bias amounts to a meaningful threat to the initial findings – and even then only for some of the estimated models. If reporting in non-cell coverage areas for the UCDP-GED data in Africa is only marginally better (e.g., a baseline probability of reporting in non-areas of 30%), then reporting bias is highly unlikely to explain all of original finding in Pierskalla & Hollenbach (2013).

5 Additional Simulation Results – Completely random "fake" event assignment

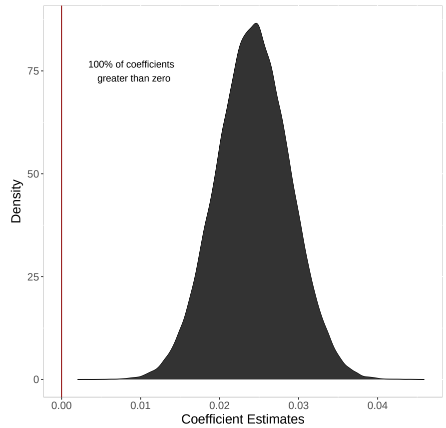
In the manuscript we use predicted probabilities from a linear model without cell phone coverage as the probabilities to randomly draw grid cells without cell phone coverage and reported violence to add "fake" events. This takes into account that based on the included covariates some cells are more likely to have exhibited violence than others. Instead, here we present the same model results when we draw the grid cells to which add "fake" events completely at random. Figure 10 shows these results for the cross-sectional model. The results are virtually the same as when events are added based on predicted probabilities.

As one can see in Figure 10, the even when events are added at random to non-coverage cells without violence, the results are very robust and nearly indistinguishable from the ones presented in the manuscript.

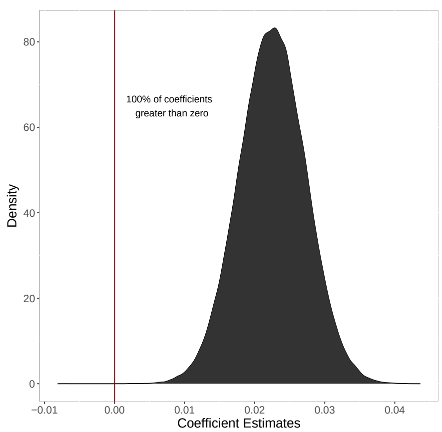
Figure 11 shows the results from estimating the panel model on the simulated data when events are added completely at random. As one can see, here the results are a bit more in line with what be evidence of reporting bias. Yet, even in the worst case scenario when the reported events are only 75% of the true events, we recover a positive coefficient with more than 65% probability.



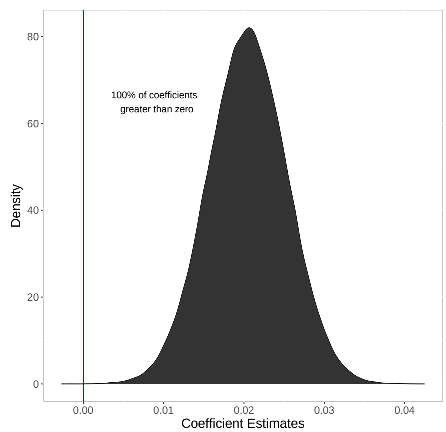
(a) Observed data 95% of "real" events



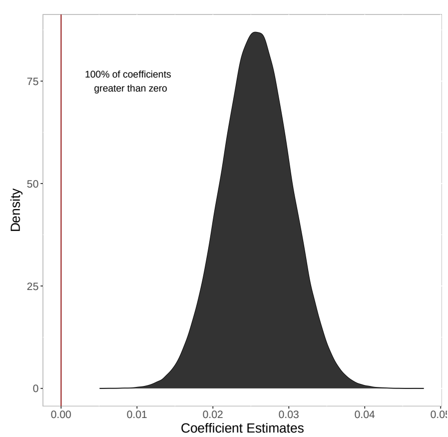
(b) Observed data 90% of "real" events



(c) Observed data 85% of "real" events

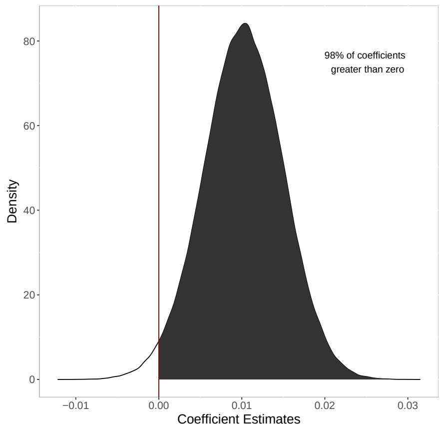


(d) Observed data 80% of "real" events

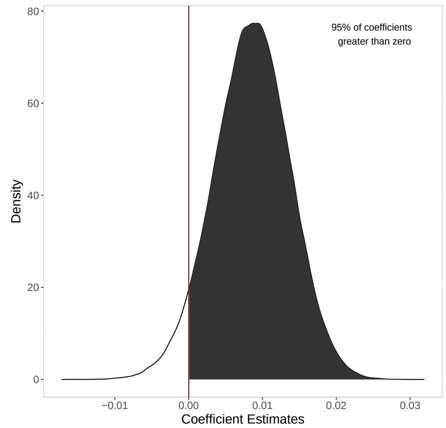


(e) Observed data 75% of "real" events

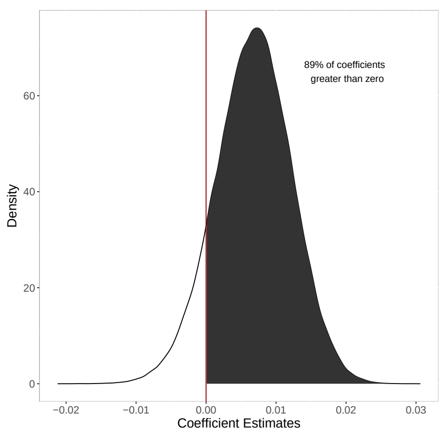
Figure 10: This figure shows the empirical distribution of coefficients for the cell coverage variable in the cross-sectional, country fixed effects models when events are added at random. The shaded area indicates the share of coefficients that are larger than zero.



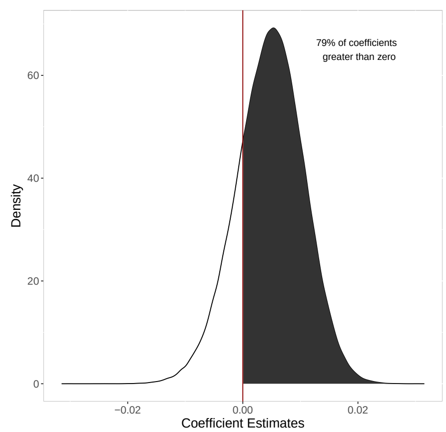
(a) Observed data 95% of "real" events



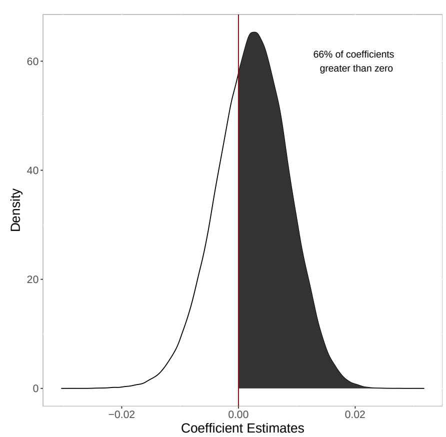
(b) Observed data 90% of "real" events



(c) Observed data 85% of "real" events



(d) Observed data 80% of "real" events



(e) Observed data 75% of "real" events

Figure 11: This figure shows the empirical distribution of coefficients for the cell coverage variable in our panel models for each level of added events for over 1000 randomizations. Here "fake" events again are added completely at random to cells without previously recorded violence or cell phone coverage. The shaded area indicates the share of coefficients that are larger than zero.

References

Pierskalla, Jan H., & Hollenbach, Florian M. 2013. Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa. *American Political Science Review*, **107**(2), 207–224.

Weidmann, Nils. 2016. A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science*, **60**(1), 206–218.