British Journal of Political Science (2021), **51**, 230–252
doi:10.1017/S0007123418000509

British Journal of Political Science

ARTICLE

Systematic Measurement Error in Election Violence Data: Causes and Consequences

Inken von Borzyskowski[1] and Michael Wahman[2]*

[1]Florida State University and [2]Michigan State University
*Corresponding author. Email: wahmanmi@msu.edu

(Received 3 October 2017; revised 21 August 2018; accepted 5 October 2018; first published online 6 June 2019)

Abstract

What are the causes and consequences of systematic measurement error in violence measures drawn from media-based conflict event data? More specifically, how valid are such event data for geocoding and capturing election violence? This study examines sub-national variation in election violence and uses original data from domestic election monitor surveys as a comparison to widely used sources of event data. The authors show that conventional data under-report events throughout the election cycle, particularly in sparsely populated areas and outside anticipated violence hotspots. Moreover, systematic measurement error of media-based event data for measuring election violence can generate significant relationships where none exist, and can result in different effect magnitudes. The article suggests areas for future research and indicates ways in which existing work on election violence may have been affected by systematic measurement error.

Keywords: election violence; measurement; Africa; Zambia; Malawi

Since a wave of democratization swept the world in the 1990s, elections have become a global norm (Levitsky and Way 2010). However, many new democracies struggle with consolidation and holding elections that meet international standards (Hyde 2011; von Borzyskowski 2016). Elections have been marred by irregularities and repression. Most notably, violence has become a prominent feature of elections in many developing countries (Beaulieu 2014, 138; Collier 2009). When elections turn violent, it threatens their integrity, it tilts the electoral playing field towards actors with repressive capacity, and it limits participation, particularly among those unable to protect themselves from violence (Bratton 2008).

Election violence is a type of political violence that is aimed at influencing the election's process or outcome and occurs temporally close to elections.[1] A burgeoning literature in comparative politics and international relations examines the causes and consequences of election violence.[2] Some of the quantitative work in this area has been cross-national (for example, Birch and Muchlinski 2018; Daxecker 2014; Fjelde and Höglund 2016a; Hafner-Burton, Hyde and Jablonski 2014 and 2018; Kuhn 2015; Taylor, Pevehouse and Straus 2017; von Borzyskowski 2019a, 2019b), while other work has been sub-national (for example, Burchard 2015; Dercon and Guitérrez-Romero 2012; Goldring and Wahman 2018; Ishiyama, Pulido Gomez and

[1]Election violence can be directed against people (candidates, voters, election officials, external supporters) or objects (election facilities, party offices, material) and can happen before or after election day. Our definition of violence is fairly comprehensive and includes fatal and non-fatal events. A narrow focus on only the most extreme fatal incidents would underestimate the level of violence experienced in many African countries. Violence does not have to be fatal to reduce trust in political institutions or affect citizens' participation in the democratic process (Norris 2014).

[2]For recent reviews, see Dunning (2011) and Fjelde and Höglund (2016b).

© Cambridge University Press 2019.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

Stewart 2016; Linke 2013; Reeder and Bech Seeberg 2018; Weidmann and Callen 2013; Wilkinson 2004). Prior studies tend to assume that underlying data on election violence are free of systematic measurement error and reflect dynamics on the ground reasonably well; or more formally, that there is no systematic measurement error, and that the variables used are a valid measure of the underlying concept of election violence. We argue that this assumption is incorrect for some widely used datasets.

Most sub-national (and some cross-national) work on election violence has relied on media-based conflict event data.[3] A number of newer social conflict datasets, like the Armed Conflict Location and Event Data Project (Raleigh et al. 2010) and the Social Conflict Analysis Database (SCAD) (Salehyan et al. 2012), have enabled researchers to place violent events geographically and study variation both within and across countries. Such conflict event data have opened up a wealth of new opportunities for researchers interested in the causes and consequences of election violence. But how valid are such media-based conflict event data to measure election violence? Even more importantly, do these measures of violence have errors that may systematically skew empirical results? Indeed, some earlier work has argued that this may well be the case (Eck 2012; Weidmann 2016). Media logics inherent in domestic and international reporting are likely to create serious under-reporting and potential systematic measurement error. Several studies have acknowledged the limitations of media-based event data in their ability to capture violence or political mobilization (Davenport and Ball 2002; Earl et al. 2004; Mueller 1997; Weidmann 2016). Yet, we are the first to measure the extent of such systematic measurement error in relation to elections and at a low level of aggregation (constituencies). We document how serious this underestimate is, as well as its causes and consequences.

To deal with under-reporting and misclassification, researchers have proposed two main solutions. The first is to triangulate between multiple measures (Hendrix and Salehyan 2015; Weidmann 2016), for example with capture–recapture methods or merging. This solution is only possible when multiple sources exist and assumes that the measures do not suffer from similar systematic measurement error. A second solution is to fix the issue with statistical remedies, trying to assess the severity and direction of systematic measurement error through sensitivity analyses (Gallop and Weschle 2017) and trying to correct for it (Cook et al. 2017). However, as Weidmann (2016, 208) notes, 'there have been no tests on a real case of selectively reported data where the true values are known'. While we do not claim to have revealed the absolute 'truth', we provide two cases for which more accurate values are known, as data are systematically collected from domestic election monitors evenly spread across the country, regardless of remoteness and communication technology. The difference between the more accurate monitor measure of where violence happened and the media-based event measure is what we call under-reporting.

Measurement error concerns the relationship between an observed variable and a specific concept it is intended to capture, and can be random or systematic. Both our own data and event data may have some random measurement error. However, event data suffer from *systematic* measurement error due to the logic of reporting. Media-based incident data (such as ACLED and SCAD) are not clean measures of violence but measures of *reported* violence. This threatens measurement validity (Adcock and Collier 2001). Our method guarantees equal coverage across space and provides a baseline against which we can estimate the sources and effects of systematic measurement error in event data. Scholars wishing to use event data in their own research may use these results to account for systematic measurement error in modeling sub-national variation in election violence.

We make both methodological and substantive contributions to research on measurement error concerning media-based event data. Methodologically, we contribute to research on under-

[3]Some research also uses surveys (Bratton 2008; Dercon and Gutiérrez-Romero 2012; Gutiérrez-Romero 2014) such as the Afrobarometer. However, these survey data are usually only representative at high levels of sub-national aggregation (regions) but not politically relevant units (constituencies), and cover only a few constituencies in the country.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

reporting by comparing two frequently used media-based event datasets (SCAD and ACLED) to new datasets on constituency-level election violence. Our novel approach, captured in the Malawi Election Monitor Survey (MEMS) and Zambia Election Monitor Survey (ZEMS), uses systematic expert surveys with domestic election observers to map election violence. This new approach is designed to ensure consistent national coverage at the constituency level. Using two recent African elections, Malawi (2014) and Zambia (2016), we compare these constituency-level expert survey data to SCAD and ACLED. Based on these comparisons, we then sign the direction of the systematic measurement error that is likely to result from using media-based event data to capture election violence.

Substantively, we document under-reporting not only in rural areas but also outside expected violence hotspots and disproportionately in the pre-election period. We show that the consequences of systematic measurement error are severe, generating significant relationships where none exist and also different effect magnitudes. Our findings have important implications for scholars conducting empirical research on election violence, but also for those using other media-based event data on issues such as social mobilization or civil conflict.

To be clear, our argument is not that event data are entirely unsuitable for research on election violence. For large-N cross-national analyses, such data often remain the only easily accessible data source. However, scholars should be aware of systematic measurement error in these data, should be careful when interpreting results and use appropriate models to analyze the possible consequences of such error. Our substantive results cast doubt on some of the most important empirical work on the frequency, location and timing of election violence. Taken together, our findings raise a larger issue, as they suggest that the current state of knowledge on election violence may well be affected by systematic biases when using violence measures from media-based event data.

Potential Systematic Measurement Error when using Media-Based Event Data

Geo-coded data on social conflict creates new opportunities to conduct cross-national (for example, Burchard 2015; Daxecker 2014; Goldsmith 2015) as well as sub-national election violence research (e.g. Burchard 2015; Ishiyama, Pulido Gomez and Stewart 2016; Linke 2013; Reeder and Bech Seeberg 2018). We focus on two such datasets, ACLED (Raleigh et al. 2010) and SCAD (Salehyan et al. 2012), and concentrate particularly on their ability to capture sub-national variation in election violence.[4] In this article we compare these datasets to our own collected expert survey data (M/ZEMS).

Neither ACLED nor SCAD was designed with the primary purpose of capturing election violence, but both have been used for this purpose (for example, Burchard 2015; Daxecker 2014; Ishiyama, Pulido Gomez and Stewart 2016; Linke 2013; Reeder and Bech Seeberg 2018). We focus on ACLED and SCAD because of their popularity and their comparability to M/ZEMS data. ACLED, SCAD and M/ZEMS all aim to capture conflict events in African countries sub-nationally, include lethal and non-lethal incidents, and events inside as well as outside of civil wars.[5]

ACLED and SCAD are also similar to each other in that they both document conflict events in Africa based on media sources, and offer a geographic location indicator. They also differ in a number of ways. In terms of news sources, ACLED uses a variety of local, regional and national newspapers, and non-governmental organization reports to locate violent events. SCAD's data

[4]A third frequently used dataset, the UCDP-GED (Croicu and Sandberg 2015), could also possibly be used for this purpose but only includes fatal incidents, which makes it less comparable to the two datasets described above, and even less comprehensive for capturing violence.

[5]ACLED 2012 codebook pages 3, 16; SCAD 2014 codebook pages 1, 3. About 88 per cent of all ACLED and SCAD events recorded are non-lethal.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

collection is limited to two international news agencies: the Associated Press and Agence France-Presse.[6] Importantly, SCAD excludes all local newspapers. For the two countries included in this article, it means that SCAD excludes local newspapers such as Malawi's *Malawi Nation* and *Malawi Times* and Zambia's *The Post*, *Times of Zambia* and *Zambia Daily Mail*.

There are at least two preconditions for mapping constituency-level election violence: geo-precision and election specificity. The former requires being able to place each violent incident within the boundaries of an electoral constituency. One limitation of both ACLED and SCAD is that neither was designed to map incidents into electoral constituencies, and the precision of the geographic coordinates is often limited. For instance, a newspaper article may name the district of an election violence event, but a district (as in the case of Malawi and Zambia) may have several constituencies.[7] The second precondition for mapping election violence is the ability to distinguish election violence from other forms of social conflict that are not directly related to the electoral process. Although neither dataset is particularly focused on election violence, they provide ways to filter election-specific violence – with more or less objectivity and replicability. An election 'issue' filter exists in SCAD but not in ACLED; single-sentence incident narratives must be used for ACLED, which can be ambiguous in terms of electoral relevance, raising issues of replicability.

The fundamental problem associated with media-based event data is that they are likely to under-report violence, especially lower-scale violence that is overshadowed by other, more dramatic events. Some violent events are unlikely to be picked up by newspapers. For an event to appear in the media, a journalist needs to know and write about it, and an editor needs to decide to publish it. Yet both actors face constraints. Journalists face physical and economic constraints; they routinely make decisions about where to go and trade off one location over another. In deciding where to go, journalists consider where (more important) events are likely to happen, how to reach these places and how to publish stories from there. If notable events are likely to happen in both places A and B, but A is within easier reach or offers other conveniences, then journalists choose A over B. Such patterns are reinforced if journalists are mainly freelancers, paid per published story. Thus journalists tend to privilege towns over villages, central areas over more remote ones and places with access to better communication tools.[8] However, even if journalists see/hear and write about an event, it does not guarantee publicity. Editors also face constraints on what to publish; these can be space constraints due to paper/column space, or political constraints as to what is fit to print given the outlet's preferences and government policies. As a result of these constraints, many events do not make it into the media, and thus do not appear in event datasets based on media outlets.

This fundamental problem of media-based event data also applies to the case of election violence. Based on these reporting dynamics inherent in the media, we argue that under-reporting varies by election cycle, by connectivity and by expectations of violence. As for electoral cycles, general under-reporting should be especially noteworthy in the longer pre-election period, during which journalists are likely to be less alert to possible violence. After elections, however, journalists pay closer attention to reactions to the election results and process, and any contestation or unrest. We hence formulate the following hypotheses:

HYPOTHESIS 1 (Under-reporting Hypothesis): Media-based conflict event data tend to under-report incidents of election violence.

[6]SCAD codebook, page 1.

[7]ACLED codes the precision of the event. However, not even the highest geographic precision (spatial precision = 1) guarantees that an event is placed in the appropriate constituency. Precision = 1 means that a town is specified; however, we might not know the location within the town where the incident occurred. For instance, Zambia's capital Lusaka has fourteen constituencies.

[8]E.g., cell phone or internet connection; see Öberg and Sollenberg (2011, 55).

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

HYPOTHESIS 2 (Pre-Election Hypothesis): Media-based conflict event data tend to under-report incidents of election violence more in the pre-electoral than the post-electoral period.

These conjectures are important because under-reporting of election violence in media-based event data can paint a misleading picture of security concerns in relation to elections. A misrepresentation of the relative frequency of pre- and post-election violence can also distort research on the timing of election violence. For instance, using Reuters news database, Bhasin and Gandhi (2013) show that most government electoral repression happens after rather than before elections, but this result could have been affected by an increased propensity for the media to report on post-election violence. In addition to distorting studies on the timing of violence, under-reporting can also generate inference problems for violence studies more broadly if the under-reporting is systematically related to the independent variable (Weidmann 2016; Wooldridge 2006). Below we outline two different sources of possible under-reporting: (1) constituency connectivity and (2) expectations of violence. We argue that both of these sources may create media under-reporting, systematic measurement error, and thus biased estimates and inaccurate inference from (election) violence studies using media-based event data.

Constituency Connectivity

Sub-national units vary vastly in their level of connectivity. This is especially true in the developing world, where infrastructure is poor, mobility limited, and access to electricity, the internet and mobile networks is restricted. Journalists and editorial offices are based in cities and organizational life is much denser in major urban areas. We should hence see a significant urban bias in the reporting of election violence. Similar arguments have been made about the measurement of social conflict more generally (Weidmann 2016) and state terror (Davenport and Ball 2002). We formulate the following hypothesis:

HYPOTHESIS 3 (Connectivity Hypothesis): Media-based conflict event data tend to under-report incidents of election violence in less connected areas.

Connectivity driving under-reporting is important because much literature on election violence argues that it is a predominantly urban phenomenon (Burchard 2015; Wilkinson 2004). Possible systematic measurement error in relation to constituency connectivity could have significant ramifications, as the level of violence in less urban, more remote places is likely underestimated. Even more troubling, urban–rural contrasts in Africa are stark, and local political dynamics differ significantly between cosmopolitan African cities, regional commercial centers, densely populated commercial agricultural zones and semi-arid pastoral regions.

Many locations that have been predicted to be high-risk areas for election violence, including ethnically diverse cities (Straus 2011), agricultural settlement schemes (Boone 2011; Kanyinga 2009) and resource extraction sites (Bratton 2008), also have high population densities. Population density and electoral competition are highly correlated, with cities and highly populated rural areas showing significantly higher levels of political competition than low-density rural locations dominated by subsistence farming and pastoral areas (Wahman and Boone 2018). Population density also serves as a strong proxy for modernization and economic development. Figure 1 shows the relationship between constituency-level population density, political competition, ethnic fractionalization and economic development[9] in our sample. Population density is significantly correlated with political competition, ethnic fractionalization and economic

[9]Following a significant literature in development economics and geography, we approximate sub-national development with night light density (Weidmann and Schutte 2017). For each constituency we divide the mean score by the land area to get the mean night lights per square kilometer, add 1 to avoid losing observations that equal 0, and take the natural log.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

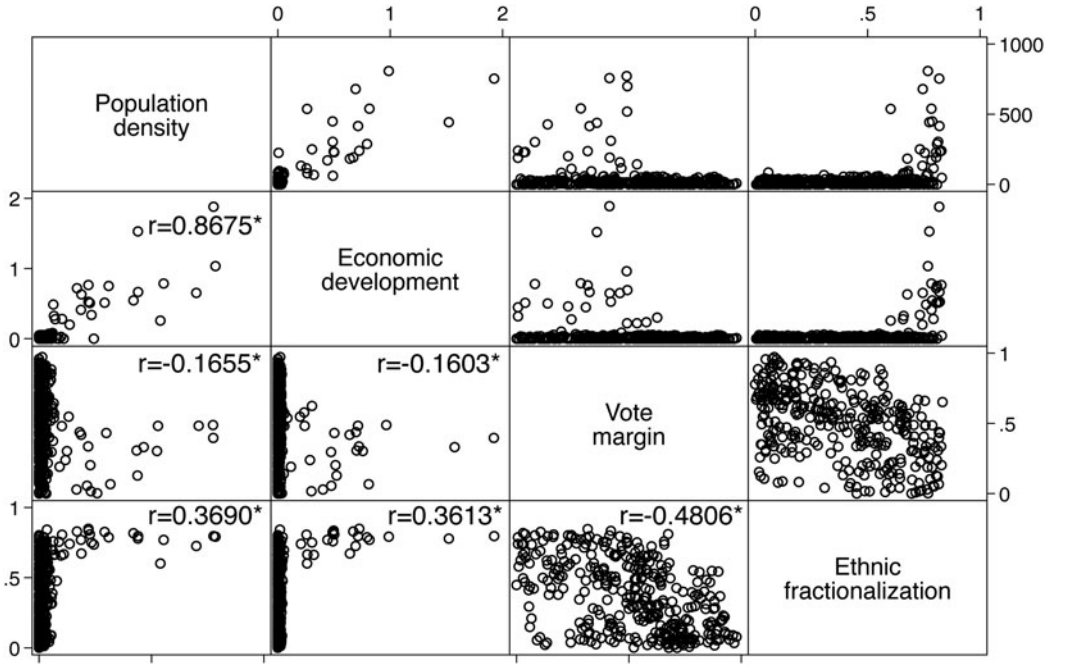


Figure 1. Correlation matrix (Malawi and Zambia)
Note: correlation statistics in individual plots. *p < 0.05

development (night light density). This is important for studies on sub-national election violence because relationships between socio-economic conditions, population density, political competition and election violence are difficult to disentangle if underlying data fail to record election violence equally in high and low population density areas. We return to this issue in the robustness section.

Expectations of Violence

In the run-up to elections, journalists and other country specialists often have expectations about where in the country violence is more likely to occur. These locations may be politically contested (high competition) or have a history of election violence. Election observers often invest more effort into monitoring such election hotspots. Civil society also tends to increase efforts to prevent violence, and representatives of international organizations tend to be more present in such locations. The increased attention devoted to hotspots should increase the probability of incidents being captured and reported. Hence we formulate the following hypothesis:

HYPOTHESIS 4 (Hotspot Hypothesis): Media-based conflict event data tend to under-report incidents of election violence outside expected violence hotspots.

Such potential systematic measurement error from using media-based conflict event data is problematic because it reinforces our preconceptions of the causes of election violence. Wilkinson (2004), using data from Indian newspapers, finds that election-related ethnic riots[10] tend to recur in the same locations. However, the logic presented above suggests that there may be a tendency for media to increase reporting from areas with a violent history.

Data from United States National Oceanic and Atmospheric Association's (NOAA) National Geographic Data Center have been aggregated to the constituency level by Boone (2016).

[10]Wilkinson does not talk about election violence specifically. We address this issue below.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

AN ALTERNATIVE MEASURE: EXPERT SURVEYS FOR MAPPING CONSTITUENCY-LEVEL ELECTION VIOLENCE

As an alternative to media-based event data, we propose a new technique of mapping election violence based on systematic surveys of domestic election monitors. The advantage of this method is that it ensures consistent national coverage by systematically collecting data on a constituency-to-constituency basis regardless of the constituency's salience, connectivity or expected levels of election violence. This study builds on two original constituency-level surveys, the MEMS carried out in relation to Malawi's 2014 general election and the ZEMS carried out in relation to Zambia's 2016 general election. Descriptive statistics in the appendix document that Malawi and Zambia are representative of African countries in terms of election violence levels and drivers, and thus allow generalization.

We use our Malawi and Zambia surveys as our benchmark for empirical comparison with data from ACLED and SCAD. Both surveys were conducted in collaboration with leading, non-partisan domestic election observer organizations, following the same general principles. We carried out our Malawi survey with the assistance of the country's leading domestic election observation organization, the National Initiative for Civic Education (NICE). NICE engaged its full-time district civic education officers, regional civic education officers, and community volunteers in long-term monitoring and party mediation, and engaged a total of 4,500 stationary monitors in short-term observation on election day in every polling center. In Zambia, we collaborated with two leading domestic election observation organizations, the Foundation for Democratic Process (FODEP) and the Southern African Centre for Constructive Resolution of Disputes (SACCORD). They jointly covered all constituencies in the country, deploying a combined 9,000 election monitors throughout the election cycle. NICE, FODEP and SACCORD are all non-partisan organizations with international funding. They all have a clear organizational structure and provide their monitors with extensive training. In both Malawi and Zambia, we carried out phone interviews with three selected observers in each constituency.

Importantly, respondents were *not* a random sample but instead deliberately a sample of experts, as in other prominent expert surveys on election integrity.[11] The survey respondents for MEMS and ZEMS were recruited from the NICE, FODEP and SACCORD networks. The executive directors of all three organizations commissioned their regional and district coordinators to collaborate in identifying three suitable respondents in every constituency (193 constituencies in Malawi and 156 in Zambia). In particular, we asked for individuals who had appropriate monitor training, a thorough understanding of the *entire* electoral cycle for a particular constituency and no known partisan affiliation. The surveyed monitors include the constituency coordinators for each constituency as well as two additional suitable monitors.

In total, we conducted 579 interviews in Malawi and 464 interviews in Zambia. Overall inter-coder reliability scores are reasonable: they are high for Malawi but a bit lower for Zambia pre-election violence.[12] We would not expect perfect scores since constituency coordinators (coder 1) are higher in the hierarchy and thus receive more information than coders 2 and 3. Indeed, constituency coordinators (coder 1) generally report more violence than other coders.[13] To err on the side of caution given a lower score in inter-coder reliability for pre-election violence in Zambia,

[11]Norris et al. (2018); Coppedge et al. 2016. MEMS and ZEMS experts have knowledge about the entire constituency rather than particular polling stations. During the election cycle our experts received information about events in the entire constituency and the survey asked about events in the entire constituency.

[12]We use the Kuder-Richardson (1937) statistic, which is similar to Cronbach's alpha but designed for binary variables instead of Likert scales. Kuder-Richardson scores range from 0 to 1, with higher scores indicating higher reliability. Scores above 0.5 are usually regarded as reasonable. Pooling the two countries, the inter-coder reliability (ICR) score for pre-election violence is 0.56 and for post-election violence is 0.70. Malawi's scores are 0.74 for pre-election violence and 0.86 for post-election violence.

[13]In Zambia pre-election violence, the case with the lowest ICR score, we observe the largest deviation between coordinators and the other coders: coordinators (coder 1) reported violence in 25 per cent of constituencies compared to only 15 per

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

we conduct analyses of under-reporting for the two countries separately; we show pooled results in the appendix. Generally, similar patterns of under-reporting emerge in Malawi where inter-coder reliability is higher and in Zambia where inter-coder reliability is lower.

The original MEMS survey in June 2014 asked respondents whether any pre-election or election-day violence had occurred in their constituency.[14] To validate our findings and ensure consistency across respondents and consistency with our definition of election violence, we also asked respondents who indicated violence for a narrative of the event.[15] The narratives provide more precise information on the location, severity and actors involved in the violence. In some cases, the respondents' violence narratives did not meet our definition of election violence (such as tense rhetoric); here, we changed the coding to no violence for that respondent. Overall, our datasets represent a wealth of material at the constituency level with several monitor reports per unit and qualitative descriptions.

Our analysis hinges on the assumption that our indicator of election violence based on experts' evaluations is measured without systematic error. One line of criticism of this assumption would be that the roles our experts hold may give them an interest in either downplaying or exaggerating the extent of violence. While we cannot rule this out completely, several aspects – including the non-partisanship of the monitor organizations and monitors, the timeline and detail of the surveys – make this unlikely. The first potential concern is individual partisan biases among monitors. Although we would not expect any institutional bias within our non-partisan monitoring organizations, we cannot exclude the possibility of individual biases. However, asking for detailed event narratives makes it less likely that respondents fabricate incidents and also allows the reported events to be assessed for validity. A second concern is that monitors may themselves be affected by the media sources included in ACLED and SCAD.[16] However, our data would mitigate these systematic measurement errors because – in contrast to journalists – domestic monitors provide coverage across all constituencies, regardless of connectivity, violence history or electoral period. Lastly, there could be recall error if monitors forget about incidents that were less serious or happened earlier in the election cycle. However, the empirical results – showing a significantly higher number of election violence observations in our monitor data than in the media-based event data – give us confidence that recall error is not a serious issue.

RESEARCH DESIGN

To test the hypotheses about systematic measurement error that results from using media-based conflict event data to capture election violence across space, we use expert surveys as a benchmark and then compare them to media-based event data in bivariate graphs and multivariate regressions. That is, we measure election violence before/after elections in each constituency in each of the four datasets (ACLED, SCAD, MEMS/ZEMS), and then generate three sets of outcome measures: (1) *consistency* between media and monitor datasets to test the under-reporting and pre-election hypotheses (Hypotheses 1 and 2) in bivariate graphs; (2) *under-reporting* in media as compared to monitor datasets to test the connectivity and hotspot hypotheses (Hypotheses 3 and 4), using measures for our hypothesized drivers of under-reporting; and (3) 'raw' violence

cent for coders 2 and 3. In Malawi pre-election violence, the same pattern persists but the difference is smaller: coordinators (coder 1) reported violence in 17 per cent of constituencies compared to 13 per cent for coders 2 and 3.

[14]The exact survey question was: 'Thinking only about the election in [Name of constituency]: To what extent have you personally experienced or received credible reports of pre-electoral violence (i.e., physical violence targeted at voters, party officials, candidates, monitors or election officials).'

[15]In Malawi, we conducted a second round of surveys in November 2014 during which we re-contacted all the respondents who had indicated violence in the first survey round. In Zambia, we did not perform a follow-up survey, but instead asked for detailed narratives in the first round of surveys.

[16]To be precise, probably more the sources in ACLED than the sources in SCAD: SCAD is based on two international media sources, and few monitors in Zambia and Malawi are likely to be consumers of these sources.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

measures from each dataset to investigate the *consequences of under-reporting* for inference, by emulating a canonical study of election violence in Africa.

Empirical Strategy To Assess the Consequences of Under-reporting

To assess the consequences of under-reporting for inference, we run identical model specifications on different dependent variables, that is, violence measures drawn from the monitor surveys and media event data. From each of the expert surveys (MEMS, ZEMS) we code two variables capturing election violence at two different stages of the electoral cycle. For example, the variable *MEMS Pre-election Violence* is a binary variable based on the Malawi expert survey that is coded 1 for any electoral violence before or on election day, and 0 otherwise. We code violence if one or more of the constituency's observer respondents reported election violence. Similarly, the variable *MEMS Post-election Violence* is coded 1 for electoral violence after election day and 0 otherwise in Malawi. Pre-election violence occurred in 22 per cent of constituencies in Malawi and 51 per cent of constituencies in Zambia. Post-election violence occurred in 4 per cent in Malawi and 9 per cent in Zambia. Figures 2 and 3 illustrate the spatial distribution and frequency of violence in the pre- and post-election periods in the two countries.

We source media-based event data from SCAD and ACLED, using their geocoded information to map these events to constituencies. To filter out *electoral* violence from other violence that happens to occur during the election period, we use dataset-specific strategies. For SCAD, we filter electoral violence by using their issue variable, which categorizes events as election related or not (issue = 1). For ACLED, we filter by narrative because no objective filter is provided. These strategies follow previous studies using these datasets for information on electoral violence (Daxecker 2014). We have also tried, as much as possible, to match the time frames of pre- and post-election violence in the three sources. The MEMS/ZEMS surveys asked for 'pre-election violence during the general election campaign'. To approximate this time period, we follow the convention in the literature (for example, Birch and Muchlinski 2018; Daxecker 2014; von Borzyskowski 2019a) and collect all events in ACLED and SCAD that occurred within six months before the election. While the official campaigning periods are only 2–3 months in Zambia and Malawi, the longer 6-month cut-off effectively gives ACLED/SCAD more chances to report an event, thus reducing the risk of documenting under-reporting in media sources.[17] For post-election violence we include the period from election day until 2 weeks after the final results are announced (which is about 3 weeks after the election). This is the time when we start conducting the monitor surveys. In both Malawi and Zambia post-election violence had settled at this point.[18]

We create the 'raw' violence variables *ACLED Pre-election Violence, ACLED Post-election Violence, SCAD Pre-election Violence,* and *SCAD Post-election Violence.* According to ACLED (SCAD) data, pre-election violence affected 4 (1) per cent of constituencies in Malawi and 17 (0.6) per cent of constituencies in Zambia. Also, according to ACLED (SCAD) data, post-election violence affected 2 (0.5) per cent of constituencies in Malawi and 9 (0.6) per cent of constituencies in Zambia. Note that both media-based event datasets record significantly lower rates of violence than our expert surveys. We also combine information from ACLED and SCAD to emulate what an analyst of electoral violence might do, that is, combining all available media-based event data. Thus, we generate variables called *ACLED/SCAD Combined Pre-election Violence* and *ACLED/SCAD Combined Post-election Violence*. They are coded 1 when either or both media datasets indicate violence and 0 when none does. These are our dependent variables when analyzing the consequences of systematic measurement error.

[17]Empirically, the vast majority of incidents in SCAD and ACLED cluster close to election day within the official campaign period. The longer 6-month window for ACLED/SCAD also accommodates the possibility that monitors may refer to events that happened before the official start of the campaign.

[18]The last recoded event for Zambia was 11 days after the election in ACLED and 5 days after the election in SCAD. The last event in Malawi was 10 days after the election in both ACLED and SCAD.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

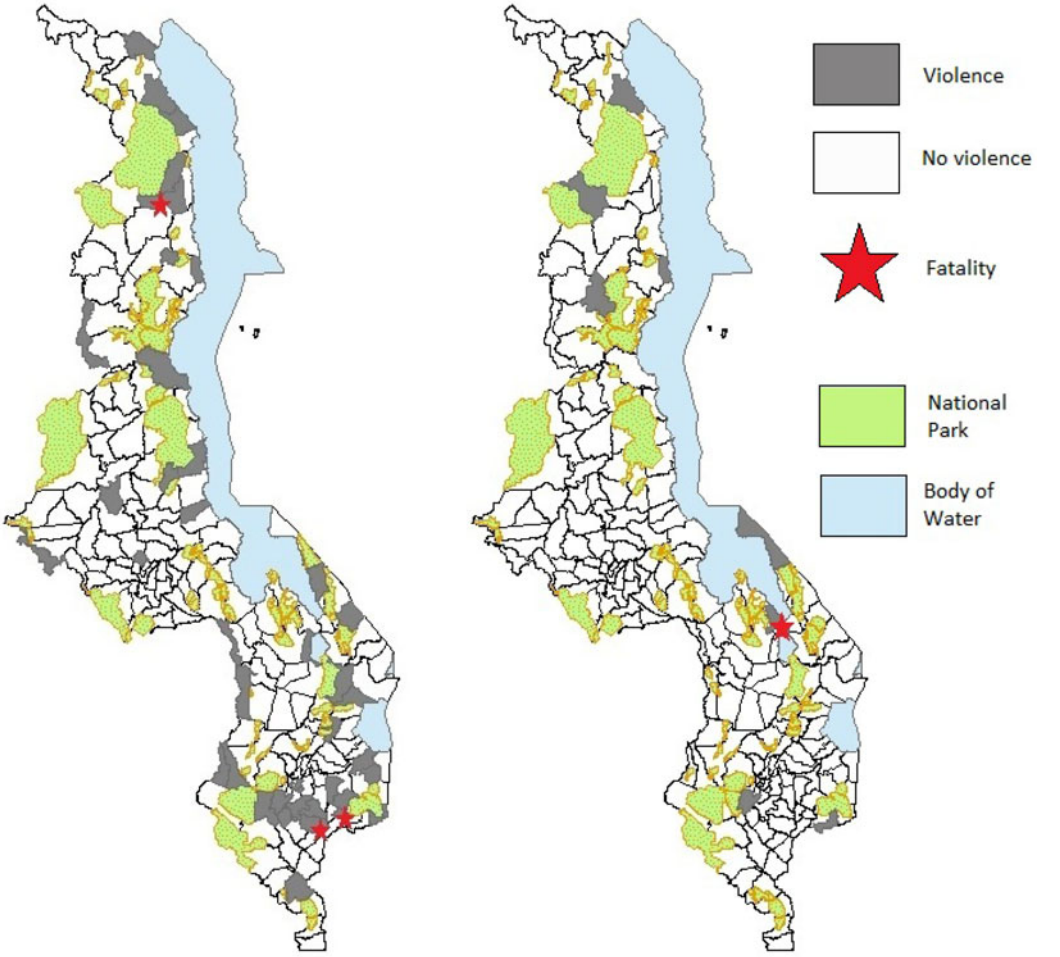


Figure 2. Constituencies with election violence in Malawi 2014
Source: MEMS data
Note: pre-election violence (left) and post-election violence (right)

For model specification, we emulate a recent study on constituency-level election violence in Africa by Reeder and Bech Seeberg (2018).[19] Like Reeder and Bech Seeberg, we control for urbanization,[20] night lights (economic development),[21] competitiveness[22] and democracy.[23] Further, we

[19]Reeder and Bech-Seeberg are particularly interested in differences between intraparty and interparty violence and hence use a dummy for that as well as measuring time until election. We do not include these variables in our models as we (like most of the election violence literature) do not distinguish between these types of violence.

[20]Reeder and Seeberg use grid data and measure the percentage of each grid cell covered by urban areas, averaged over each constituency. This measure yields little variation in Malawi and Zambia, as it suggests that 90 per cent of all constituencies are less than 1 per cent urban, and even urban constituencies are estimated to be less than 10 per cent urban. Instead we use logged population density, which varies widely and better captures differences between semi-urban and rural locations. Data comes from the latest available census (2008 Malawi, 2010 Zambia).

[21]Night lights data from Boone (2016) and NOAA, as described above.

[22]Competitiveness is measured as 1 – vote margin in the last presidential election (Malawi Electoral Commission; Electoral Commission of Zambia).

[23]Following Reeder and Seeberg, we use national-level Polity IV scores. This control is redundant in the models using country-fixed effects.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

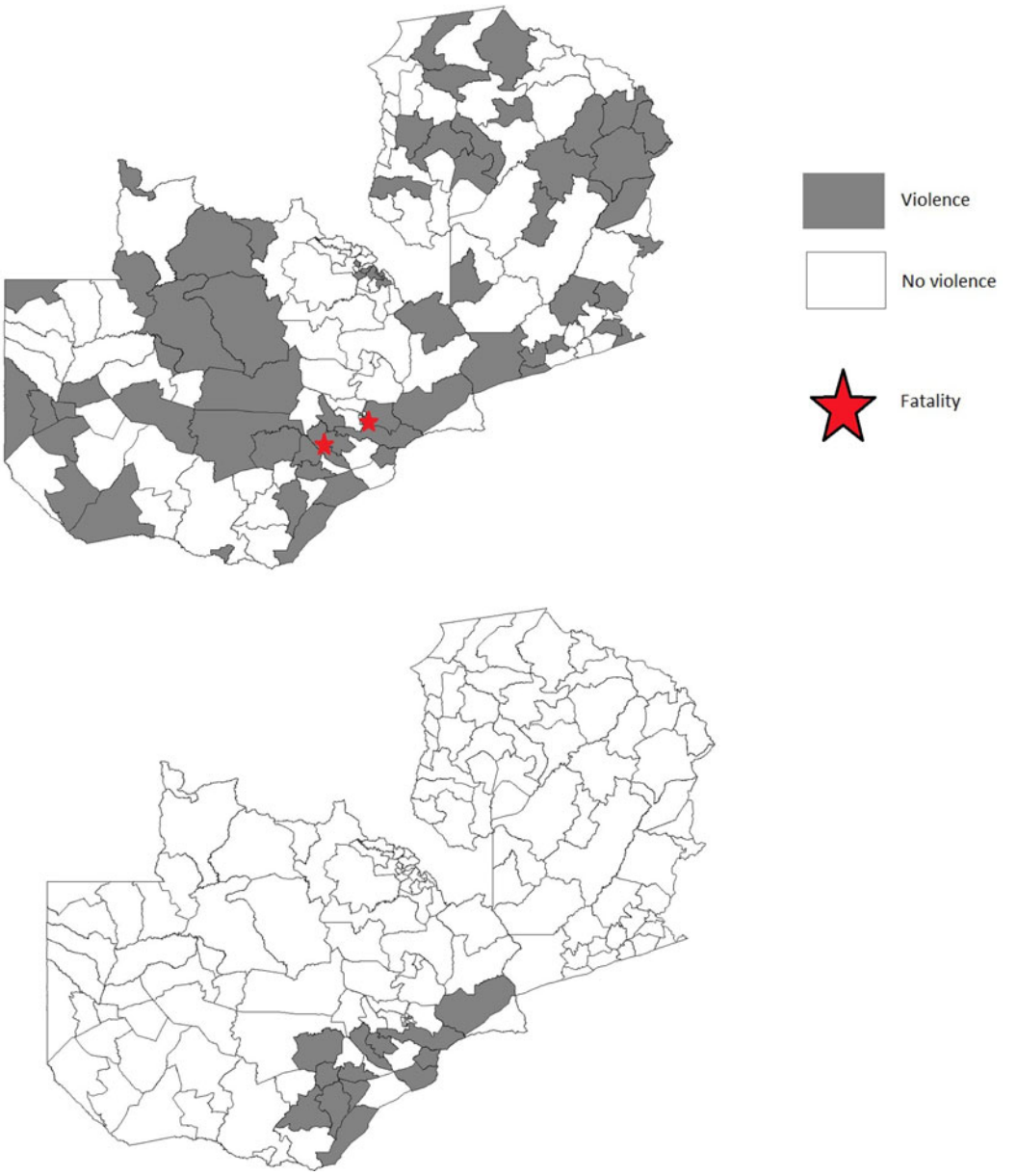


Figure 3. Constituencies with election violence in Zambia 2016
Source: ZEMS Data
Note: pre-election violence (top) and post-election violence (bottom)

use multiple model specifications (logit/OLS, country fixed effects, region clustered standard errors) to ensure that it is indeed the varying data source (monitor vs. media) rather than varying model specification that drives the differences in results.

Empirical Strategy To Test Hypotheses 1 and 2

Using the raw violence variables from both the expert surveys and the media event datasets (as detailed in the previous section), we then generate *consistency* variables to test hypotheses about

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

the existence of under-reporting generally and particularly during the pre-election period (Hypotheses 1 and 2). *Consistency* measures agreement between the two types of datasets about violence reporting for any given constituency and electoral period. For example, the variable *Consistency SCAD Pre-election Violence* for Malawi is coded 1 when **both** MEMS and SCAD reported pre-election violence, or when **neither** dataset reported pre-electoral violence for a given constituency. It is coded 0 when MEMS reported violence but SCAD did not; and when SCAD reported violence but MEMS did not. We code equivalently the variables *Consistency SCAD Post-election Violence, Consistency ACLED Pre-election Violence* and *Consistency ACLED Post-election Violence* for both countries. Similar to the violence measures, we also combine the media datasets and generate *Consistency ACLED/SCAD Combined Pre-election Violence* and *Consistency ACLED/SCAD Combined Post-election Violence*. These are coded 1 when monitor reports are consistent with either ACLED or SCAD, and 0 otherwise. We test hypotheses about the existence of under-reporting generally and particularly in the pre-election period (Hypotheses 1 and 2) using bivariate graphs of distributions and paired t-tests.

Empirical Strategy To Test Hypotheses 3 and 4

We next test Hypotheses 3 and 4 on connectivity and hotspots as driving under-reporting. We code under-reporting when monitor data indicate violence but media data do not. For example, *Under-reporting in Pre-election ACLED Data* for Malawi is coded 1 when MEMS indicates pre-election violence but ACLED does not, and 0 otherwise. We code equivalent variables for *Under-reporting in Post-election ACLED Data, Under-reporting in Pre-election SCAD Data, Under-reporting in Post-election SCAD Data, Under-reporting in Pre-election ACLED/SCAD Combined Data,* and *Under-reporting in Post-election ACLED/SCAD Combined Data.* We generate the same six under-reporting variables for Zambia.

To capture connectivity, we measure *population density* (number of people per km^2) using the most recent census data.[24] To capture anticipated hotspots, we measure *History of Election Violence* and *Vote Margin* from the presidential election. The variable *History of Election Violence* is sourced from the two media-based event datasets (ACLED and SCAD) for all prior elections. While media-based event data have significant blindspots (as documented above), we use these data to capture history because it is precisely the *widely* reported violence that might drive journalists' expectations of more violence and thus reporting dynamics. This variable is categorical and coded 0 when no violence was widely reported for a constituency in previous elections, 1 for widely reported violence in one previous election, and 2 for widely reported violence in multiple previous elections. In line with the Hotspot Hypothesis (Hypothesis 4), we expect a negative coefficient: hotspot constituencies – those with a history of election violence – should be less subject to under-reporting in the current election.

As another indicator of anticipated hotspots, we also use *vote margin* from the presidential election (from the Malawi Electoral Commission and Electoral Commission of Zambia, respectively). That is, we use presidential vote margin results from the current election in the models of post-election outcomes in both countries. We use vote margin results from the previous election (the 2015 presidential by-election) for pre-election outcomes in Zambia. However, in the case of Malawi we also use current results for pre-election outcomes. We do so because these dynamics (competitive vs. non-competitive constituencies) were widely anticipated and major realignment in the Malawian party system (the government party had split into two parties) made 2009 electoral results a poor approximation of 2014 competitiveness. Note that this could introduce post-treatment bias in the models of pre-election consistency for Malawi, since this variable is observed only after election day. However, the results are robust to dropping that variable in the pre-election period.

[24]We divide the variable *Population Density* by 10 to ease interpretation of the coefficient. The latest available census is 2008 (Malawi) and 2010 (Zambia).

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

We include several control variables to capture other potential causes of systematic measurement error. Since poorer locations may be less likely to attract media attention, we include economic development in terms of *electrification* (that is, the percentage of households with electricity in a constituency, from census 2008 and 2010)[25] and human development in terms of *literacy* (the percentage of literate citizens over the age of 15 in a constituency, from census 2008 and 2010). Descriptive statistics of all variables are available in Appendix Tables A1 and A2.

To test the Connectivity and Hotspot Hypotheses (Hypotheses 3 and 4), we estimate the following statistical model:

$$DV_{cr} = \beta_0 + \boldsymbol{C}'_{\boldsymbol{cr}}\beta_1 + \boldsymbol{H}'_{\boldsymbol{cr}}\beta_2 + \boldsymbol{X}'_{\boldsymbol{cr}}\beta + \boldsymbol{\epsilon}_{cr},$$

where β_1 and β_2 are the parameters of interest on the key explanatory vectors C_{cr} and H_{cr}, which capture connectivity and hotspot variables and thus test Hypotheses 3 and 4. DV are the respective dependent variables: *under-reporting* between survey and media data for analyses of the causes of systematic measurement error, and *violence* measures for analyses of the consequences of systematic measurement error. Further, the vector $\mathbf{X}'_{\mathbf{cr}}$ represents a set of control variables, ε_{cr} is the idiosyncratic error, and the subscript c stands for constituency, and r for region. The model is estimated using logit with region-clustered standard errors to account for the lack of independence between constituencies within the same regions.[26] Also, constituencies within one region might differ from those in another region because regions differ somewhat in remoteness/distance to the capital as well as language, which might influence reporting. Our estimates represent the average constituency-level variation within the same country and conditional on the included controls. We estimate models on each country – Malawi and Zambia – separately. We also provide pooled models in the appendix.

RESULTS

Reporting of Constituency-Level Violence

To assess the first two hypotheses, Figures 4 and 5 document *consistency* between the expert surveys and media-based event data in Malawi and Zambia, respectively. In each of these figures, we show four graphs (from left to right): monitor vs. ACLED pre-election, monitor vs. ACLED post-election, monitor vs. SCAD pre-election, and monitor vs. SCAD post-election. Each marker represents a constituency; we add random noise around the markers to visualize the relative frequency. The location of each data point on the graph denotes whether a constituency was coded consistently in the media-based event data and expert survey. Consistently coded constituencies are located in the upper right (1,1) if both types of datasets coded a constituency as violent, or the lower left (0,0) if both datasets coded it as non-violent. Inconsistently coded constituencies are located in the lower right (1,0) if media-based datasets coded violence but the expert survey did not, and in the upper left (0,1) if the expert survey reported violence but the media event data did not. Numbers indicate the number of constituencies in each group.

Figures 4–5 provide empirical support for the existence of general under-reporting and particularly in the pre-election periods (Hypotheses 1 and 2). Figures 4–5 show several constituency observations located off the diagonal, indicating disagreement between the datasets. In most cases, this disagreement involves constituencies in which the media-based event data reported no violence but the expert survey reported violence. Those cases are clustered in the upper left

[25]The average electricity prevalence is quite low in Malawi (under 5 per cent of households), so this coefficient has a rather large magnitude in the regression results.

[26]We cannot use region fixed effects in many of the models because several regions have no variation in violence. This is especially true for the post-election models and for SCAD.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

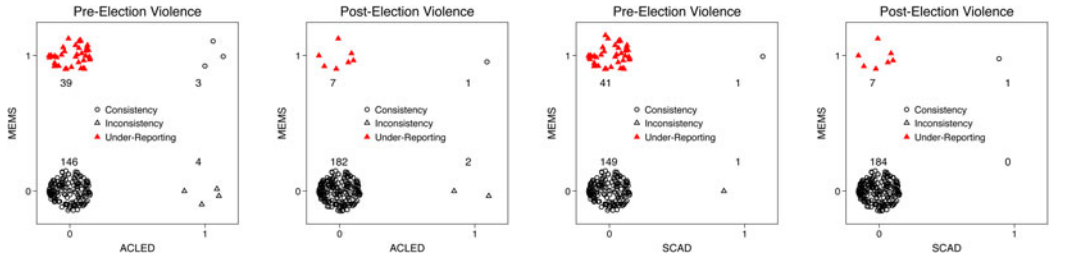


Figure 4. Consistency between ACLED/SCAD and Malawi survey (MEMS)

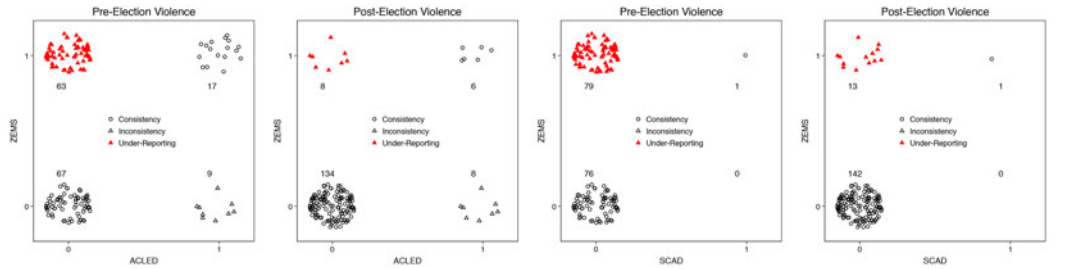


Figure 5. Consistency between ACLED/SCAD and Zambia survey (ZEMS)

corner of each graph where ACLED/SCAD = 0 and MEMS/ZEMS = 1 (marked by solid red triangles). In a minority of the inconsistencies, the relationship is reversed: there are a handful of cases where media data reported violence but the expert surveys did not (lower right corner of Figures 4–5). However, overall media data tend to under-report incidents of election violence compared to expert survey data.

The under-reporting of violence in media data compared to survey data is statistically significant. For example, for the pre-election period in Malawi, the survey data indicate that 22 per cent of constituencies experienced violence, while ACLED indicates 4 per cent and SCAD 1 per cent. Paired t-tests indicate that the difference between the survey measure and each of the media measures is highly statistically significant ($p < 0.001$).[27] For the post-election period, these differences are somewhat less severe (4 per cent vs. 2 and 1 per cent) but still statistically significant ($p < 0.09$, 0.01). From the eight comparisons, all but one[28] result in statistically significant differences between survey and media measures. This supports Hypothesis 1 about the existence of under-reporting in violence measures drawn from media-based event data.

Moreover, the under-reporting is significantly stronger before elections than afterwards, which supports Hypothesis 2. Figures 4–5 and paired t-tests indicate that media-based event data under-report violence more in the pre- than in the post-electoral period. Appendix Tables A1 and A2 show descriptive statistics: in the pre-election period, consistency between the media data and expert survey data ranges between 49 and 77 per cent, while it ranges between 90 and 96 per cent in the post-election period. Importantly, paired t-tests indicate that these differences between consistency before and after elections are highly statistically significant ($p < 0.001$). That is, consistency between survey and media data is significantly higher after elections than before elections.

The general trend shows that ZEMS/MEMS are more capable of picking up events than SCAD/ACLED. Nevertheless, in a handful of constituencies ACLED/SCAD found violence but

[27]We use paired t-tests because the same constituencies are measured in both survey and media data, so there *should* be a strong relationship between the scores. This test looks at the difference in scores for each constituency with the null hypothesis that the mean difference is zero. We reject the null hypothesis in almost all cases.

[28]The one exception is the paired t-test on post-election estimates between survey and ACLED in Zambia.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

MEMS/ZEMS did not. Looking at these cases we find a number of explanations as to why monitors may not have reported these incidents. In the case of Zambian pre-electoral violence, three of ten cases concern the burning of billboards. These may be examples of recall bias where monitors failed to remember this relatively mild form of violence. There are also cases for which we have reason to question the authenticity of the newspaper reports. The ACLED data include both heavily biased government newspapers and online publications (such as the *Zambian Watchdog*) without editorial control.

Causes of Systematic Measurement Error

What drives these substantial differences in violence reporting between media and expert survey data? We argue that under-reporting is contingent on constituencies' connectivity and expected violence hotspots, as stipulated in Hypotheses 3 and 4. We test these hypotheses through multivariate analyses. Tables 1 and 2 show estimated logit coefficients for the determinants of under-reporting (whether monitors reported violence but media sources did not) for Malawi and Zambia, respectively. Columns 1 and 2 show the determinants of under-reporting between MEMS and ACLED; Columns 3 and 4 show under-reporting for MEMS and SCAD; Columns 5 and 6 show under-reporting for MEMS with the combined ACLED/SCAD measure. Table 2 is structured the same way but for ZEMS data in Zambia. We remove the capital, Lusaka, from the sample;[29] we show results with Lusaka included in Appendix Table A3. We also provide results pooling both countries in the appendix (Table A4); the results are robust. An important caveat is needed here: there are fewer post-election violence events than pre-election violence events and also more consistency in the post-election period. The low number of inconsistencies in the post-election period makes the post-election results more unstable.

The results in Tables 1 and 2 provide support for the argument that connectivity drives under-reporting (Hypothesis 3) as well as some – concededly weaker – support for the argument that anticipated hotspots drive under-reporting (Hypothesis 4). The estimated coefficients on *Population Density* are negative and significant in almost all models of Tables 1 and 2. This suggests that higher population density (higher connectivity) is associated with less under-reporting. This supports the Connectivity Hypothesis (Hypothesis 3). While the coefficients on *Population Density* are highly statistically significant for Malawi ($p < 0.01$), for Zambia two of these coefficients are insignificant. All coefficients point in the hypothesized direction. As noted above, population density in the African context is highly correlated with political competition (a variable that has been associated with election violence in the literature). In Appendix Tables A5 and A6 we re-run the models without controlling for vote margin in the presidential election. This modified specification strongly reinforces the results related to population density. Without controlling for vote margin, the coefficient for population density is negative and significant for all post- and pre-models for both Malawi and Zambia.[30]

Surprisingly, SCAD and ACLED even missed some fatal cases of pre-election violence in rural areas. For example, lethal election violence in Malawi occurred in both Mulanje (where supporters of the incumbent party killed a Democratic Progressive Party (DPP) supporter for wearing DPP regalia) and Rumphi (where a person interrupting an AFORD meeting was killed). In Zambia, one person was stabbed to death in clashes between the Patriotic Front and the United Party for National Development in Magoye, Southern Province, again without the story reaching the news

[29]ACLED has a particular problem with geo-precision for the geocraphically small constituencies in Lusaka. Events are clustered in the most central constituency, Lusaka Central, although Lusaka is comprized of 13 constituencies. As a consequence, Lusaka constituencies that are not Lusaka Central have missing events as they have wrongly been given the geographic coordinates of Lusaka Central.

[30]We also replicate Tables 1 and 2 without controlling for population density; those results are reported in Appendix Tables A7 and A8.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

Table 1. Determinants of under-reporting between ACLED/SCAD and surveys – Malawi

	Under-reporting in ACLED data		Under-reporting in SCAD data		Under-reporting in ACLED/SCAD combined	
	pre-election 1	post-election 2	pre-election 3	post-election 4	pre-election 5	post-election 6
Population density	−0.011***	−0.033***	−0.005***	−0.033***	−0.011***	−0.033***
	(0.002)	(0.008)	(0.001)	(0.008)	(0.002)	(0.008)
History of election violence	−1.413***	2.316***	−1.824**	2.316***	−1.413***	2.316***
	(0.502)	(0.286)	(0.798)	(0.286)	(0.502)	(0.286)
Vote margin Presidential election	0.206	1.131	0.159	1.131	0.206	1.131
	(1.735)	(2.028)	(1.529)	(2.028)	(1.735)	(2.028)
Literacy	−0.650	2.525	−0.618	2.525	−0.650	2.525
	(1.237)	(2.903)	(1.490)	(2.903)	(1.237)	(2.903)
Electrification	8.161***	1.720	7.765***	1.720	8.161***	1.720
	(0.459)	(3.685)	(0.986)	(3.685)	(0.459)	(3.685)
Constant	−1.005***	−5.141***	−1.105**	−5.141***	−1.005***	−5.141***
	(0.358)	(1.565)	(0.556)	(1.565)	(0.358)	(1.565)
Observations	182	182	182	182	182	182
AIC	186.31	58.04	189.05	58.04	186.31	58.04
BIC	192.71	64.44	195.46	64.44	192.71	64.44
LL	−91.15	−27.02	−92.52	−27.02	−91.15	−27.02

Note: logit models with robust standard errors clustered on region in parentheses; two-tailed tests. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 2. Determinants of under-reporting between ACLED/SCAD and surveys – Zambia

	Under-reporting in ACLED data		Under-reporting in SCAD data		Under-reporting in ACLED/SCAD combined	
	pre-election 1	post-election 2	pre-election 3	post-election 4	pre-election 5	post-election 6
Population density	−0.015	−0.212***	−0.014**	−0.048***	−0.015	−0.092***
	(0.009)	(0.022)	(0.007)	(0.008)	(0.009)	(0.016)
History of election violence	−0.011	2.027***	0.240	0.535	−0.011	2.359***
	(0.773)	(0.535)	(0.669)	(0.369)	(0.773)	(0.594)
Literacy	−1.733	−1.238	1.108	−4.311*	−1.733	−4.853
	(2.586)	(2.946)	(1.975)	(2.380)	(2.586)	(3.304)
Electrification	1.665	25.241***	0.673	26.875***	1.665	23.554***
	(1.576)	(6.161)	(1.308)	(5.223)	(1.576)	(7.096)
Vote margin presidential election	1.389	37.553***	1.222	35.440***	1.389	37.752***
	(0.924)	(8.078)	(0.799)	(8.707)	(0.924)	(7.976)
Constant	−0.301	−35.778***	−1.562	−31.025***	−0.301	−34.167***
	(1.267)	(7.525)	(1.244)	(7.795)	(1.267)	(7.613)
Observations	142	142	142	142	142	142
AIC	195.52	31.22	202.16	42.30	195.52	28.07
BIC	213.26	48.95	219.90	60.03	213.26	45.81
LL	−91.76	−9.61	−95.08	−15.15	−91.76	−8.04

Note: logit models with robust standard errors clustered on region in parentheses; two-tailed tests. *p < 0.10, **p < 0.05, ***p < 0.01.

media. These events clearly show the problem of under-reporting violence outside major urban centers.

Among the control variables, the positive coefficient on *Electrification* was unexpected. However, *Electrification* is strongly correlated with *Population Density* (the correlation coefficient is r = 0.79 in Malawi and r = 0.52 in Zambia). When *Electrification* and *Population Density* are interacted (tables not shown), the coefficient on the interaction term is negative and significant in all models. In line with the argument, this indicates that more populated and developed places have less under-reporting in the media-based event data.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

In addition to finding evidence that connectivity generates under-reporting, we also find some evidence that expected hotpots receive differential attention. However, these results are weaker and not consistent. Recall that we use two variables to capture potential hotpots: competitiveness (*Vote Margin*) and *History of Election Violence*. *History of Election Violence* is negatively and significantly associated with under-reporting pre-election violence in Malawi (Table 1 Columns 1, 3, 5). This is evidence in favor of the Hotspot Hypothesis (Hypothesis 4). However, coefficients on *History of Violence* point in the opposite direction in Columns 2, 4, 6 of Tables 1 and 2, indicating that previous violence hotspots are more subject to under-reporting. This is surprising and counter to our expectation. However, as noted earlier, the small number of constituencies with post-election violence and the low under-reporting in the post-election period should caution strong inference from these post-election models.

For the second measure of anticipated hotspots, *Vote Margins*, the coefficients are consistently in the hypothesized direction (positive), indicating that larger vote margins (less competitive constituencies) tend to have more under-reporting. While the coefficients are consistently in the expected direction, the relationship only reaches statistical significance in the models for post-election violence in Zambia (Table 2, Columns 2, 4, 6). We thus conclude that while there are some signs that under-reporting is more pronounced in anticipated hotspots, that evidence is weaker and limited to particular cases.

These analyses provide support for the argument that election violence measures drawn from media-based event data have systematic measurement error. Both media-based event sources under-report violence; this under-reporting is more pronounced before than after elections; the systematic measurement error is driven by connectivity, with more densely populated areas receiving better coverage; measurement error also seems to be generated by journalists going to and reporting from highly competitive areas within the country, but this seems limited to post-election violence (recent competition) and the case of Zambia.

We now turn to the consequences of these measurement errors. Given that media-based event measures of violence are affected by systematic measurement error, how do they affect our inference? Do our substantive conclusions about the drivers of violence change when using different datasets?

Consequences of Systematic Measurement Error

To investigate these questions, we change the dependent variable from *Under-reporting* to the raw measure of *Violence* reported in each dataset and period. As outlined in the research design and shown in the descriptive statistics (Appendix Tables A1, A2), these variables include *MEMS Pre-election Violence* and *MEMS Post-election Violence* as well as the equivalents for ZEMS, ACLED, SCAD and ACLED/SCAD combined per country and electoral period. Recall that our model specification here emulates a recent constituency-level study on election violence in Africa by Reeder and Bech Seeberg (2018). Following the analysis of Reeder and Bech Seeberg, the models include measures of urbanization, night lights, competition and democracy. As in the prior analyses, we use logit and robust standard errors clustered by region.[31] The results for the pre- and post-election period are in Tables 3 and 4, respectively. In each of these tables, Column 1 shows the results for using the dependent variable from the monitor surveys, Column 2 is for ACLED, 3 for SCAD, and 4 for the combined ACLED/SCAD measure of election violence.

The results point to important differences for inference. Differences in measuring the dependent variable can generate significant relationships where none may exist. The coefficients on *Urbanization* and *Night Lights* suggest significant relationships with violence in media event

[31]We use several other model specifications for robustness (OLS instead of logit), adding country fixed effects, and not clustering standard errors. The Reeder/Seeberg study does not use clustering. Our results are robust to replicating Table 3 with different model specifications; Appendix Figures A1–A2 summarize these results.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

Table 3. Determinants of pre-election violence using different outcome measures

	M/ZEMS 1	ACLED 2	SCAD 3	ACLED/SCAD combined 4
Urbanization	0.190	0.478**	0.824**	0.401*
	(0.240)	(0.210)	(0.390)	(0.220)
Night lights	0.357	0.686	1.631**	1.820**
	(0.797)	(0.453)	(0.690)	(0.735)
Democracy	1.523***	2.290***	−0.057	2.214***
	(0.309)	(0.589)	(0.540)	(0.651)
Competition	−0.430	−0.322	−3.484	−0.369
	(0.678)	(0.659)	(6.132)	(0.744)
Constant	−10.743***	−18.445***	−5.991	−17.794***
	(2.214)	(4.272)	(4.195)	(4.751)
Observations	338	338	338	338
AIC	415.11	190.79	33.38	189.80
BIC	434.23	209.90	52.49	208.91
LL	−202.56	−90.39	−11.69	−89.90

Note: logit models with robust standard errors clustered on region in parentheses; two-tailed tests. *p < 0.10, **p < 0.05, ***p < 0.01.

Table 4. Determinants of post-election violence using different outcome measures

	M/ZEMS 1	ACLED 2	SCAD 3	ACLED/SCAD combined 4
Urbanization	0.409	0.478**	0.553*	0.476*
	(0.304)	(0.232)	(0.290)	(0.243)
Night lights	−1.822	−0.113	−9.849***	−0.079
	(1.608)	(2.241)	(1.993)	(2.305)
Democracy	1.129	2.204***	0.708	2.265***
	(1.014)	(0.649)	(0.636)	(0.685)
Competition	−1.633	−1.888	−2.528	−2.275
	(2.553)	(1.720)	(3.528)	(1.866)
Constant	−10.221	−17.898***	−9.834***	−18.092***
	(6.472)	(3.683)	(2.413)	(3.899)
Observations	338	338	338	338
AIC	165.05	127.31	33.39	131.25
BIC	184.16	146.43	52.50	150.36
LL	−77.52	−58.66	−11.69	−60.62

Note: logit models with robust standard errors clustered on region in parentheses; two-tailed tests. *p < 0.10, **p < 0.05, ***p < 0.01.

datasets, but not in the expert survey datasets. The estimated coefficient on *Urbanization* is not significant in the monitor data but is significant for ACLED, SCAD and the combined media data (Columns 2–4 in Tables 3 and 4). Using ACLED data, Reeder and Bech Seeberg (2018) find urbanization to be significantly correlated with election violence at the constituency level in their analysis of Kenya and Zimbabwe. Although this is not their main finding, our results here suggest that this particular positive association between urbanization and election violence in Reeder and Bech Seeberg may be, at least partially, an artifact of the event data.

Similar to urbanization, the coefficient on *Night Lights* (a proxy for development) is not significant for the monitor data but is significant for SCAD and the combined media data (Columns 3–4 in Table 3 and Column 3 in Table 4). The expert survey data suggest that there is no relationship between election violence and urbanization and development. In contrast, the media event data suggest significant relationships for these variables. Researchers might also come to different conclusions about the substantive magnitude of the effect. For example, the marginal effect of democracy on election violence is smaller in media event data than in monitor data, which is illustrated in Figure 6 (estimated from Table 3 columns 1 and 4).

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

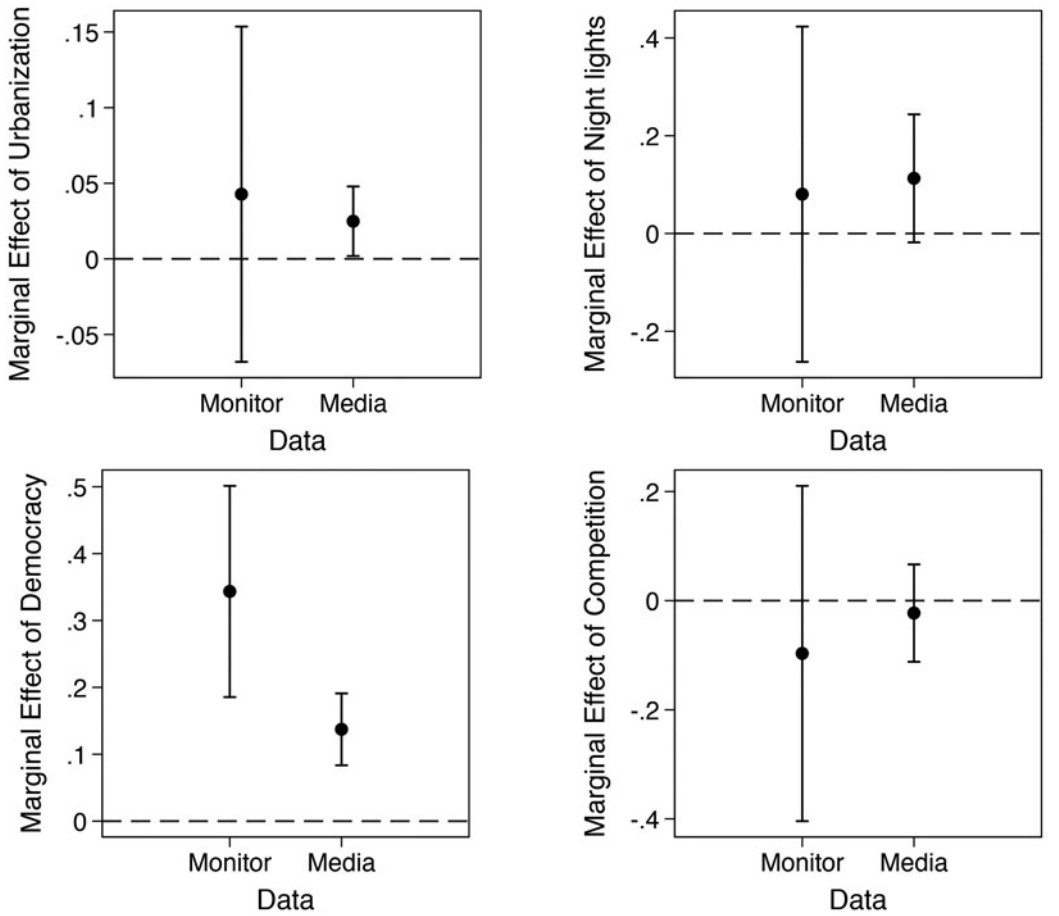


Figure 6. Consequences of reporting bias: different predictions depending on monitor or media event data

As an additional robustness test, we ran all our models again including all violent events in SCAD and ACLED (independent of whether they were directly related to the election or not) for the period under investigation. The results are shown in the appendix (Tables A9–A11). Including non-election-related violent events does not change our substantive results and adds surprisingly few additional events; in the case of Zambia and Malawi, most political violence in this time period was related to the election.[32]

Based on these results for Malawi and Zambia, we propose some ways in which election violence studies using media-based event data may over- and under-estimate relationships. Studies using geocoded media-based event data tend to skew the effects of population density and development, showing relationships where none may exist. These over-estimates seem largely driven by the selection of journalists into urban areas: they are more likely to reside and report from urban, more densely populated, economically vibrant centers as opposed to more remote areas with worse communication technology. Further, studies using media-based event data might find smaller effects for some variables (such as democracy) than they would find using higher-quality data.

[32]For violence after election day in Zambia, ACLED and SCAD do not record violent events of a non-electoral nature, so replicating Table 4 by including non-electoral violence yields identical results and is thus omitted.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

CONCLUSION

In this article, we have examined the prevalence, causes and consequences of systematic measurement error in violence measures drawn from media-based event data. From qualitative work on election violence we know that violence varies significantly across space (Boone 2011), but in this article we have discussed the potential limitations of existing conflict data for mapping such election violence systematically across space. As a way to assess and correct for under-reporting, we have introduced a new expert survey-based approach for measuring election violence at the constituency level. We have argued that the new approach is better designed to systematically measure election violence at the constituency level. Although restricted to two countries in this study, Malawi and Zambia, the method could be replicated in other places.[33] Finally, we used these new data to document the causes and consequences of under-reporting and systematic measurement error in the media-based civil conflict literature, pointing to problems in achieving consistency and geocoding accuracy.

So what should researchers interested in sub-national variation in election violence do with this information? Most research on sub-national variation in election violence is based on single-election studies (for example, Dercon and Guitérrez-Romero 2012; Ishiyama, Pulido Gomez and Stewart 2016; Linke 2013; Wilkinson 2004). When the country sample is small, our survey-based approach offers a realistic and preferable alternative to existing media-based event data. Other research on sub-national variation includes several countries. Here, the most radical solution would be to altogether abandon media-based event data as a source. This would, however, not be our recommendation. For multi-country studies, media-based event data are more cost effective and enable comparisons across time and space. Thus researchers may still wish to use event data but should be aware of their systematic measurement error, explicitly discuss them as weaknesses and try to remedy them.

As we demonstrate, violence measures from monitor and media data cannot be used interchangeably, so it is important for researchers to choose violence measures carefully and to demonstrate robustness across measures where possible. The results obtained in this study should help researchers account for measurement error in event data. This article has helped to identify some variables that may introduce systematic measurement error. It has also provided insights about the direction of such error. This information may be used to run sensitivity tests with simulated data (Gallop and Weschle 2017). Moreover, we have also shown that under-reporting is not equal across event data sources, but is higher in SCAD than ACLED. Researchers interested particularly in more low-scale incidents of election violence would do well to analyze data sources that do not solely rely on two or three international news outlets but a range of domestic and non-media sources as well. For sub-national studies, researchers should also investigate the accuracy of geocoded information in these datasets.

The results in this article also have important implications for prominent findings in existing work on election-related violence. Using data from Indian newspapers, Wilkinson (2004) shows, among other things, that ethnic riots are more common in large cities and in places with a history of violence. However, our results suggest that newspapers are better at picking up violent events in urban areas and in historic hotspots. Bhasin and Gandhi (2013), using reports extracted from the Reuters news database, show that government electoral repression is more common after than before elections. However, our results suggest that media-based data record events in the post-electoral period to a higher extent than events in the pre-electoral period.

This article focuses on systematic measurement error in sub-national data, but many of these measurement errors also have consequences for cross-national work. Media freedom and

[33]The approach has recently been replicated in Kenya (2017) and Liberia (2017); see Claes and von Borzyskowski 2018, pages 18 and 28.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

journalistic capacity vary significantly between countries, implying that countries with a better infrastructure and a more independent flow of information have more reports than otherwise similar countries where information is more restricted for political and/or capacity reasons. For instance, it seems likely that a higher share of incidents would be captured in a country like South Africa than in the Democratic Republic of Congo or Ethiopia, even if the same events happened.

Our discussions have focused specifically on violence related to elections. However, we contend that our findings have implications for violence event data more generally. There is no reason to believe that under-reporting of violence or systematic measurement error, particularly stemming from uneven reporting from urban and rural areas as well as geocoding problems, would be more pronounced in and around elections than during time periods disconnected from the electoral campaign. In fact, the opposite is likely true. Elections signify a focal event, as local and international media attention is high. The systematic measurement error and under-reporting we find on violence in media-based event data around elections is likely to be as high, or probably even higher, when studying violence unrelated to elections. Thus researchers have to take systematic measurement error seriously when studying violence at the sub-national level.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0007123418000509

Acknowledgements. We are indebted to the National Initiative for Civic Education, the Southern African Centre for Constructive Resolution of Disputes, the Institute for Policy Interaction, the Southern African Institute for Policy and Research, Boniface Chembe, Marja Hinfenlaar, Ollen Mwalubunju, Chimfwembe Mweenge and Nandini Patel for support in organizing our surveys. Invaluable research assistance was provided by Josephine Chanda, Felix Chauluka and Edward Goldring. We are grateful for comments and support provided by Catherine Boone, Johan Broché, Kristine Höglund, Holger Kern, Jonathan Krieckhaus, Patrick Kuhn, Bryce Reeder, Christopher Reenock, Nils Weidmann and Laron Williams. The study was supported by grants from the Swedish Research Council (VR DNR 2012-6653), the Magnus Bergvall Foundation (2015-00698) and the Economic and Social Research Council (ES/R005753/1). IRB#2004569, approved by University of Missouri Institutional Review Board Office. The two authors contributed equally to this article.

References

Adcock R and Collier D (2001) Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* **95** (3), 529–546.

Beaulieu E (2014) *Electoral Protest and Democracy in the Developing World*. Cambridge: Cambridge University Press.

Bhasin T and Gandhi J (2013) Timing and targeting of state repression in authoritarian elections. *Electoral Studies* **32** (4), 620–631.

Birch S and Muchlinski D (2018) The dataset of countries at risk of election violence. *Terrorism and Violence* **25** (3), 385–403.

Boone C (2011) Politically allocated land rights and the geography of election violence: the case of Kenya in the 1990s. *Comparative Political Studies* **48** (2), 173–202.

Boone C (2016) Spatial Inequalities in African Political Economy Dataset. LSE RIIF 2016 (#1-BRD-B076).

von Borzyskowski I (2016) Resisting democracy assistance: who seeks and receives technical election assistance? *Review of International Organizations* **11** (2), 247–282.

von Borzyskowski I (2019a) *The Credibility Challenge: How Democracy Aid Influences Election Violence*. Ithaca, NY: Cornell University Press.

von Borzyskowski I (2019b) The risks of election observation: international condemnation and post-election violence. *International Studies Quarterly* **63**.

Bratton M (2008) Vote buying and violence in Nigerian elections. *Electoral Studies* **27**, 621–631.

Burchard S (2015) *Electoral Violence in Sub-Saharan Africa: Causes and Consequences*. Boulder, CO: Lynne Rienner.

Claes J and von Borzyskowski I (2018) *What Works in Preventing Election Violence: Evidence from Liberia and Kenya*. Washington, DC: United States Institute of Peace.

Collier P (2009) *Wars, Guns, and Votes: Democracy in Dangerous Places*. New York: Harper Perennial.

Cook S et al. (2017) Two wrongs make a right: addressing under-reporting in binary data from multiple sources. *Political Analysis* **25** (2), 223–240.

Coppedge M, Gerring J, Lindberg S, Skaaning S and Teorell J (2016) Varieties of Democracy (V-Dem) Project.

Croicu M and Sundberg R (2015) *UCDP Georeferenced Event Dataset: Codebook Version 2.0*. Department of Peace and Conflict Studies. Uppsala: Uppsala University.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

Davenport C and Ball P (2002) Views to a kill: exploring the implications of source selection in the Case of Guatemala state terror, 1977–1995. *Journal of Conflict Resolution* **46** (3), 427–450.
Daxecker U (2014) All quiet on election day? International election observation and incentives for pre-election violence in African elections. *Electoral Studies* **34**, 232–243.
Dercon S and Gutiérrez-Romero R (2012) Triggers and characteristics of the 2007 Kenyan election violence. *World Development* **40** (4), 731–744.
Dunning T (2011) Fighting and voting: violent conflict and electoral politics. *Journal of Conflict Resolution* **55** (3), 327–339.
Earl J et al. (2004) The use of newspaper data in the study of collective action. *Annual Review of Sociology* **30**, 65–80.
Eck K (2012) In data we trust? A comparison of UCDP, GED and ACLED conflict events datasets. *Cooperation and Conflict* **47** (1), 124–141.
Fjelde H and Höglund K (2016a) Electoral institutions and electoral violence in sub-Saharan Africa. *British Journal of Political Science* **46** (2), 297–320.
Fjelde H and Höglund K (2016b) Electoral violence: the emergence of a research field. *APSA Comparative Democratization Newsletter*, July 2016, pp. 1–11.
Gallop M and Weschle S (Forthcoming) Assessing the impact of non-random measurement error on inference: a sensitivity analysis approach. *Political Science and Research Methods*.
Goldring E and Wahman M (2018) Fighting for a name on the ballot: a constituency-level analysis of nomination violence in Africa. *Democratization* **25** (6), 996–1015.
Goldsmith A (2015) Elections and civil violence in new multiparty regimes: evidence from Africa. *Journal of Peace Research* **52** (5), 607–621.
Gutiérrez-Romero R (2014) An inquiry into the use of electoral practices and effects of political violence and vote-buying. *Journal of Conflict Resolution* **58** (8), 1500–1527.
Hafner-Burton E, Hyde S and Jablonski R (2014) When do governments resort to violence. *British Journal of Political Science* **44** (1), 149–179.
Hafner-Burton E, Hyde S and Jablonski R (2018) Surviving elections: election violence, incumbent victory and post-election repercussions. *British Journal of Political Science* **48** (2), 459–488.
Hendrix C and Salehyan I (2015) No news is good news: mark and recapture for event data when reporting probabilities are less than one. *International Interactions* **41** (2), 392–406.
Hyde S (2011) *The Pseudo-Democrat's Dilemma: Why Election Monitoring Became an International Norm*. Ithaca, NY: Cornell University Press.
Ishiyama J, Pulido Gomez A and Stewart B (2016) Does conflict lead to ethnic particularism? Electoral violence and ethnicity in Kenya 2005–2008. *Nationalism and Ethnic Politics* **22**, 1353–7113.
Kanyinga K (2009) The legacies of the white highlands: land rights, ethnicity and the post-2007 election violence in Kenya. *Journal of Contemporary African Studies* **27** (3), 325–344.
Kuhn P (2015) Do Contentious Elections Trigger Violence? In Frank R, Martinez I, Coma F and Norris P (eds), *Contentious Elections*. London: Routledge, pp. 89–110.
Kuder GF and Richardson MW (1937) The theory of the estimation of test reliability. *Psychometrika* **2** (3), 151–160.
Levitsky S and Way L (2010) *Competitive Authoritarianism: Hybrid Regimes After the Cold War*. Cambridge: Cambridge University Press.
Linke A (2013) The aftermath of an election crisis: Kenyan attitudes and the influence of individual-level and locality violence. *Political Geography* **37**, 5–17.
Mueller C (1997) International press coverage of East German protest events, 1989. *American Sociological Review* **62** (5), 820–832.
Norris P, Wynter T and Cameron S (2018) Perceptions of Electoral Integrity (PEI) Project.
Norris P (2014) *Why Electoral Integrity Matters*. New York: Cambridge University Press.
Öberg M and Sollenberg M (2011) Gathering conflict information using news resources. In Höglund K and Öberg M (eds), *Understanding Peace Research: Methods and Challenges*. London: Routledge, pp. 47–73.
Raleigh C et al. (2010) Introducing ACLED-armed conflict location and event data. *Journal of Peace Research* **47** (5), 1–10.
Reeder B and Bech Seeberg M (2018) Fighting your friends? A study of intraparty violence in sub-Saharan Africa. *Democratization* **25** (6), 1033–1051.
Salehyan I et al. (2012) Social conflict in Africa: a new dataset. *International Interactions* **38** (4), 503–511.
Straus S (2011) 'It's sheer horror here': patterns of violence during the first four months of Côte D'Ivoire's post-electoral crisis. *African Affairs* **110** (440), 481–489.
Taylor C, Pevehouse J and Straus S (2017). Perils of pluralism: electoral violence and incumbency in sub-Saharan Africa. *Journal of Peace Research* **54** (3), 397–341.
Wahman M and Boone C (2018) Captured countryside? Stability and change in sub-national support for African incumbent parties. *Comparative Politics* **50** (2), 189–208.
Weidmann N (2016) A closer look at reporting bias in conflict event data. *American Journal of Political Science* **60** (1), 206–218.
Weidmann N and Callen M (2013) Violence and election fraud: evidence from Afghanistan. *British Journal of Political Science* **43** (1), 53–75.

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press

Weidmann N and Schutte S (2017) Using night light emissions for the prediction of local wealth. *Journal of Peace Research* **54** (2), 125–140.
Wilkinson S (2004) *Votes and Violence*. Cambridge: Cambridge University Press.
Wooldridge JM (2006) *Introductory Econometrics*, 3rd ed. Mason: Thomson Higher Education.

Cite this article: von Borzyskowski I, Wahman M (2021). Systematic Measurement Error in Election Violence Data: Causes and Consequences. *British Journal of Political Science* **51**, 230–252. https://doi.org/10.1017/S0007123418000509

https://doi.org/10.1017/S0007123418000509 Published online by Cambridge University Press