Electoral Contention and Violence (ECAV): A New Dataset[1]

Ursula Daxecker
Elio Amicarelli
University of Amsterdam

Recent elections in Afghanistan, Bangladesh, Cote D'Ivoire, Egypt, Iraq, Kenya, Nigeria, and Pakistan have displayed substantial contestation and violence. A rapidly growing literature explores the causes and consequences of electoral contention and violence, yet a major limitation of this literature is the lack of comprehensive and disaggregated data on its incidence. The ECAV data project conceptualizes electoral contention as nonviolent or violent events of contestation by state or nonstate actors related to the electoral process. The data contain more than 17,000 events of election-related contention covering 135 countries holding competitive elections between 1990 and 2012. The paper describes the scope of ECAV, presents the project's definition of electoral contention and the variables included, and outlines the coding procedure. We then compare ECAV to other datasets on electoral contention. A subnational analysis of electoral competition and violence in India shows that the data are useful for assessing the subnational implications of existing theories.

Word count: 6387
Key words: dataset, events data, disaggregation, elections, violence, protest

[1] Corresponding author: u.daxecker@uva.nl. Earlier versions were presented at the Electoral Integrity Project Workshop, July 23, 2016, Poznan, Poland, and the European Conflict Network meeting, October 16-18, 2014, Uppsala, Sweden. We thank Sarah Birch, Jorgen Elklit, Aurelien Evequoz, Jennifer Gandhi, Corinna Jentzsch, Lee Seymour, and conference participants for comments. Funding for this project comes from a Veni grant from the Dutch Science Foundation (NWO) and a Marie Curie Career Integration Grant. We thank Alexander Jung and Aurelién Equevoz for excellent assistance with data cleaning, codebook revisions, and the preparation of case discussions for countries experiencing armed conflict. We are extremely grateful for research assistance from many student coders: Hanan Bahar, Emily van Bronkhorst, Mikaela Burch, Jonathan Busnelli, Henry Chow, Gabrielle Espina, Monika Gyuris, Doran Hanlon, Johan Hansen, Roos Hopman, Marthe Huigsloot, Eva Huis in 't Veld, Alexander Jung, Marketa Kachynova, Ernst Kuneman, Frank van Moock, Toby Murray, Ghis Noorlander, Saskia Olivier, Ebe Ouattara, Wessel van Proosdij, Marta Ill Raga, Isabella Rebasso, Ina Rüber, Niels Snoek, Hannah Soine, Martin Wesseling, and Mark Wittfoth.

Introduction

Recent elections in Afghanistan, Bangladesh, Cote D'Ivoire, Egypt, Kenya, Nigeria, and Pakistan have displayed substantial contestation and violence. An emerging literature explores the causes and consequences of electoral contention, examining how the competitiveness of elections, the response of losing parties, the quality of the elections, armed conflict, and the role of international actors affects the incidence of electoral contention (Daxecker 2012, 2014; Straus and Taylor 2012; von Borzyskowski 2013; Rauschenbach and Paula 2017; Tucker 2007; Kuntz and Thompson 2009; Steele 2011; Hafner-Burton, Hyde, and Jablonski 2014, 2016; Hyde and Marinov 2014; Salehyan and Linebarger 2015; Norris 2012; Birch and Muchlinski 2017a; Flores and Nooruddin 2012; Matanock 2017; Staniland 2015; Kuhn, Patrick, and Inken von Borzyskowski). A major limitation, however, is the lack of comprehensive, disaggregated data on electoral contention—a lack that has been noted by scholars and practitioners alike (Fjelde and Höglund 2016b; European Commission and United Nations Development Programme 2011). Conceptualizing electoral contention as nonviolent or violent acts of contestation by state or nonstate actors that relate to the electoral process, the Electoral Contention and Violence (ECAV) dataset contains systematic, human-coded data on election-related contentious events for all unconsolidated regimes between 1990 and 2012. Because the unit of analysis is event–day–location, the data allow for spatially and temporally disaggregated analyses.

We begin by reviewing existing definitions and measurement of electoral contention, comparing them to ECAV. We then describe the dataset, define electoral contention, explain the variables included, and outline the coding procedure, before comparing ECAV to two other datasets. We conclude with a subnational analysis of electoral competition and election violence in India.

Why new data on electoral contention?

A rapidly growing literature explores the causes and consequences of electoral contention, but empirical assessments face two limitations: researchers lack data establishing a substantive link between elections and violence, despite its conceptual importance; and the absence of disaggregated information on the timing, location, and actors limits the development and testing of micro-level theories.

First, electoral contention is conceptualized as a subtype of political contestation that is substantively linked to electoral processes (Höglund 2009; Straus and Taylor 2009; Beaulieu 2014; Staniland 2014). For example, Höglund (2009: 415) highlights that actors engaged in electoral violence aim to "influence the electoral process" by objecting to elections or attempting to influence election outcomes, and others similarly stress a substantive connection between contention and the electoral process (Straus and Taylor 2012; Beaulieu 2014). Yet it is challenging to establish such a substantive connection empirically, which is why empirical assessments have often responded to this challenge pragmatically, considering all political violence for several months before and after elections as instances of electoral violence (Straus and Taylor 2012; Daxecker 2012). Omitting a substantive link to the electoral process, however, presumably means including events that would have occurred anyway, raising concerns about electoral contention as a distinct type of contestation (Fjelde and Höglund 2016b).

Second, theoretical accounts often highlight the micro-level processes motivating electoral contention, including the identity of actors or targets (Hafner-Burton, Hyde, and Jablonski 2014; Smidt 2016), the timing of elections (Daxecker 2014; Hafner-Burton, Hyde, and Jablonski 2014), or the geography of electoral contention (Lankina 2015). Without disaggregated data on elections, contention, and violence, however, empirical assessments are frequently limited to aggregate, election-level data. Global datasets on elections such as National Elections Across Democracy and Autocracy (NELDA), Quality of Elections (QED), Varieties of Democracy (V-DEM), Perceptions of Electoral Integrity (PEI), or the Countries at Risk of Election Violence (CREV) provide information on contention and violence at election- or election-month level (Kelley and Kolev 2010; Hyde and Marinov 2012; Coppedge et al. 2015; Norris et al. 2016; Birch and Muchlinski 2017b), but do not allow for temporally or spatially disaggregated analyses of contentious events relating to elections. CREV, a new dataset on election violence, uses automated event data to provide additional information on the actors, timing, and the intensity of violence, but nevertheless uses election months and elections as the units of analysis (Birch and Muchlinski 2017b). Global data on protests are also of limited use since they do not

identify whether contention was related to electoral processes and aggregate data to the country year, which omits many lower-level contentious events (Banks 1975).[2]

Recognizing these drawbacks, some researchers have used disaggregated event data on nonviolent and violent contentious events (Daxecker 2012, 2014; Fjelde and Höglund 2016b; Smidt 2016; Salehyan and Linebarger 2015), including the Armed Conflict and Events Data (ACLED), the Social Conflict Analysis Database (SCAD), and the UCDP Georeferenced Event Data (UCDP GED) (Raleigh et al. 2010; Salehyan et al. 2012; Sundberg and Melander 2013).[3] However, UCDP GED and ACLED focus mostly on violent events or even fatalities, missing lower-intensity events. Nor, with the exception of SCAD, do these datasets identify the issues around which actors' contestation revolve, thus including many events unrelated to the electoral process.[4] Existing event datasets also do not distinguish political parties as separate actors, despite their theoretical importance in election violence (Staniland 2014).

Unlike existing data, ECAV focuses exclusively on election-related contention, has clear substantive criteria to determine whether events are election-related, identifies the timing, geocoded location, and actors involved, and is available for a global sample.

Describing ECAV

Scope

ECAV includes data on the incidence of electoral contention and violence in all countries with unconsolidated regimes that held competitive elections between 1990 and 2012.[5] ECAV contains more than 17,000 events of electoral contention for 1,199 election rounds held in 135 countries.[6] Figure 1 shows the global incidence of events as 1x1 degree grid squares from 1990 to 2012. Only squares with at least one event are depicted on the map; darker squares indicate that more events occurred within that

[2] The Mass Mobilization in Autocracies Database (MMAD), which provides disaggregated information on global protests, is not yet publicly available (Rod and Weidmann 2013).

[3] Some research has used survey data to measure election violence (Kerr 2013; Bratton 2008), but such measures rely on self-reporting sensitive topics; they lack information on perpetrators and report fear of violence rather than actual experiences.

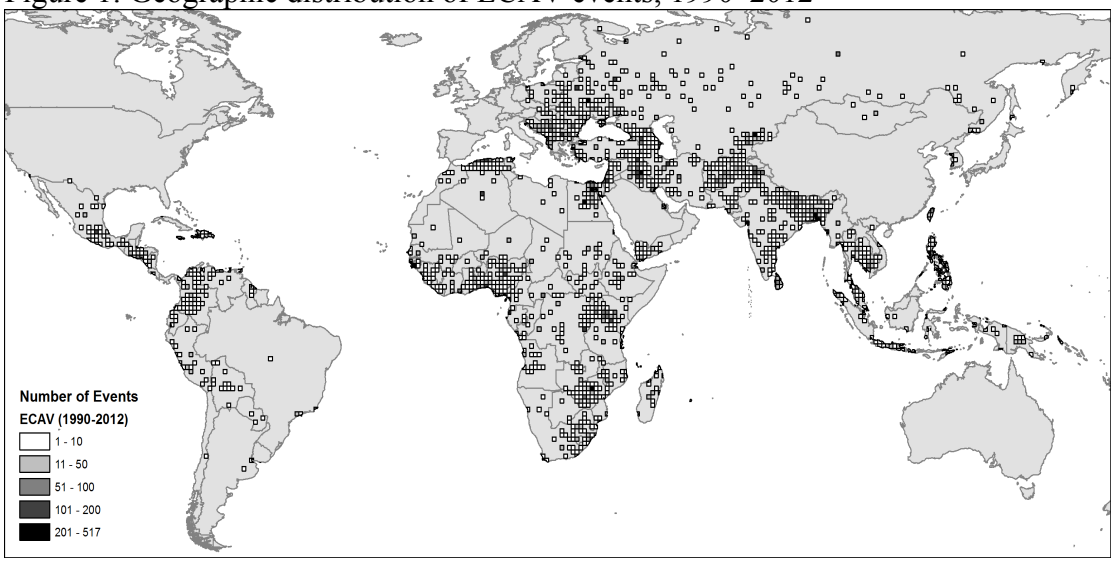
[4] While SCAD identifies the issues involved, including elections, the codebook does not offer clear criteria on how coders made their selections.

[5] Data collection focuses on post-1990 because almost all countries adopted competitive elections in that period.

[6] There was no election-related contention found in 6 of 135 countries: Estonia, Laos, North Korea, Oman, Panama, and Suriname. See appendix E for a list of countries and elections.

square. The largest number of contentious events took place in Asia (31%), followed by Africa (25%), the Middle East (20%), Eastern Europe (15%), and Latin America (8%). These regional patterns are notable because existing research has focused primarily on Africa, neglecting substantial electoral contention in Asia and the Middle East.

Figure 1: Geographic distribution of ECAV events, 1990–2012



To identify the set of competitive elections, the dataset uses the NELDA (version 4) definition, which requires elections to be direct and for a national executive or legislative office (Hyde and Marinov 2012).[7] ECAV data is thus limited to national elections.[8] Because we use NELDA, contentious events can be coded only if elections were actually held, omitting events in which contention demanding elections was unsuccessful.

The dataset excludes countries with consolidated democratic regimes, defined as states that were OECD members in 1990 (the first of year coding).[9] Consolidated democracies are excluded because the dataset is primarily concerned with elections in which the occurrence of electoral contention and violence is feasible. The risk of violence in consolidated democracies in the post-Cold War period is marginal.

[7] Data are available at http://hyde.research.yale.edu/nelda/.

[8] Dates for subnational elections are not available globally, making it difficult to identify the relevant subset of elections.

[9] ECAV therefore does not include events in Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States. Events in states joining the OECD after 1990 (Mexico, the Czech Republic, Hungary, Poland, South Korea, Slovakia, Chile, Slovenia, Israel, and Estonia) are included. Like NELDA, ECAV does not code events in states without competitive, direct elections (China, Eritrea, Somalia, Saudi Arabia, United Arab Emirates and Qatar) and excludes micro-states (see NELDA codebook for a list). There is one exception: we code events in Turkey, despite it being a member of the OECD since 1961.

Defining electoral contention

Electoral contention is defined as *public acts of mobilization, contestation, or coercion by state or nonstate actors used to affect the electoral process, or arising in the context of electoral competition*. A contentious event involves at least two actors who disagree on an issue and are thus seen as being on opposite sides.[10] The above definition implies that events of electoral contention are publicly observable, are linked to an electoral process in timing and substance, and can be violent or nonviolent in nature.

Publicness

- There is evidence of actual public contention in the form of arrests, arson, attacks, bombings, boycotts, clashes, killings, intimidation, protests, rioting, shootings, or strikes.[11]

Election-related

- We determine the relation to elections with *substantive* and *temporal* criteria.
- Regarding *substance*, contention must be linked to an ongoing electoral process: the election in question can be identified, and articles explicitly mention the electoral process as an issue around which contestation occurs. Events relating to other issues are thus included only if it can be inferred that actors choose to carry them out in relation to elections. For example, a strike over payment delays by health care workers a month before elections would not be included, unless the article mentioned that the proximity of elections was considered helpful in accomplishing the actors' goals. Consistent with this approach, events taking place near elections in states experiencing civil conflict are not automatically included. Rather, such events are included once the intent to affect (or respond to) electoral processes can be established.

[10] The logic behind "being on opposite sides" is as follows. A rally in support of a political candidate would not be coded as a contentious event, while an opposition party protest expressing disagreement with the government party would be coded.

[11] Events aimed at preventing or preempting contention, particularly curfews, are not coded as contentious events. However, events contesting the imposition of a curfew, or targeting curfew violators, would still be included. Verbal statements are also not coded.

Intent is established based on reporting in articles or statements by the actors involved.[12]

- Regarding *time*, events occurring between six months before and three months after the election are included. This timeframe may miss some relevant events, although we analyzed SCAD data to select this period before coding began (see appendix B). We use a common timeframe for all elections because the dataset includes all unconsolidated regimes and thus more than 1,000 election-rounds, making it practically impossible to research every single election to determine the most appropriate timeframe for analysis. For elections with multiple rounds, events were coded between six months before the first round and three months after the last election round.[13]

Nature

- Contention can be nonviolent or violent in nature. Events are coded as violent if they include the threat or actual use of force.[14]

Unit of analysis and coding procedure

The unit of analysis is the event–day–location. An event–day–location is an election-related contentious event reported in a media source on a single day in a particular location. This implies that events occurring in multiple locations and/or taking place over more than one day are coded separately.[15] We provide start and end dates for multi-day events to facilitate temporal aggregation. While spatial aggregation may also be possible, we caution that it can be difficult to establish whether similar events in different locations are part of the same, larger event.

[12] Because it can be challenging to substantively link elections to violence during ongoing armed conflict, appendix G includes additional information on coding decisions for each country experiencing UCDP armed conflict between 1990 and 2012 (Gleditsch et al. 2002).

[13] Elections with multiple rounds tend to occur in close proximity to each other. For elections with multiple rounds, we code events between six months before the date of the first round and three months after the last round. Since the unit of analysis in the data is the event-day-location, events can still be attributed to the correct election-round. When elections of a different type (i.e., presidential and parliamentary elections, thus not rounds of the same election) are held within nine months of each other, we split the time period in half. Regardless, coders are asked attribute event-day-locations to the correct election.

[14] We include threats in which actors publicly and physically threaten violence, such as armed groups making the rounds to threaten voters with violence if they participate in elections. However, verbal statements threatening violent action are not coded unless they actually take place.

[15] For multi-day events, coders identified the beginning, continuation, and end dates of the event in newspaper sources. For example, if a newspaper article mentioned a strike lasting two weeks, coders found articles reporting the onset of the strike two weeks before.

Information on electoral contention comes from news media reports in three newswire sources: Associated Press, Agence France Press, and BBC Monitoring available in LexisNexis.[16] The extraction procedure is mentioned in detail in the codebook included in appendix D. The project relied exclusively on human coders: graduate students in political science and other social sciences at University X. All coding took place in university computer labs under the supervision of the authors. Each coder was assigned a country and (as far as possible) coded all elections for this country. Before coding, students created a country fact file identifying election dates, main actors, notable political events, and administrative divisions.

The coding of events from news reports consists of two steps. First, coders identified the events in news reports; and second, they encoded events according to the variables listed above. Coding reliability thus required coders to identify the same events and interpret them similarly (Ruggeri, Gizelis, and Dorussen 2011). We conducted a detailed assessment of the reliability of event identification and event encoding of all coders in the ECAV data project. For event identification, Precision rates showed that coders on average identified 66% of events correctly. For event encoding, Krippendorff's alpha showed very good or good results for event violence, participant death, event direction, actor and target types, and location precision, and fair to intermediate results for remaining categorical variables (participant number, actor and target side). The procedure and results are described in detail in AUTHORS (2018).

Variables

ECAV includes several variables for each election–day–location. A sample entry can be found in appendix A. The first set of variables provides identifying information for the *event*, *election*, and *country* in question. The next variables code the *date*, *location*, *location coordinates*, and a descriptor capturing the *location precision* level.[17] Coders identified precise locations for 66% of events, with the remainder coded near an exact location (4%), second-order administrative units (7%), first-order

[16] These sources will undoubtedly miss some events, particularly those reported in local news or occurring in rural areas. We hope to include more local sources in future updates.

[17] Articles may contain more than one event–day–location, for example mentioning events in multiple locations and/or on multiple days. Coders were instructed to code all event–day–locations listed in an article separately.

administrative units (10%), or the country level (13%).[18] ID variables, date, location, and location precision must be recorded for every event–day–location in the dataset.

The next set of variables focuses on the participants in an event–day–location. *Actor type* distinguishes between state and nonstate actors, further distinguishing whether nonstate actors were civilians, party actors, members of an armed group. An additional category involves events in which the actors are unknown. Finally, actors whose identity is known but who do not fit any other category (such as international actors) are coded as other. Of all events in which the initiator could be established (see below for *Event direction* variable), 17% of events involve the state as the actor, 38% citizens, 11% parties, 15% armed groups, and 2% other. The relatively large percentage of events involving citizens as actors might seem concerning, but results from their disproportionate involvement in protests. Excluding protests, citizens are actors in 21% of events. *Actor side* records whether an actor supported the government, opposed the government, or unknown (if actor side cannot be established).[19] *Actor name* records the name of the actor in the event (e.g. students, police, opposition party members).[20] The next set of variables records information about the opposing actor, i.e. the target, of a contentious event. *Target type*, *Target side*, and *Target name* are coded identically to actor variables.

Participant number and *participant deaths* code the number of participants and the number of deaths as categorical measures, respectively. *Participant deaths* is recorded only for events involving violence. Of all violent events, 54% involved non-lethal violence, 35% led to 1-10 deaths, 5% involved 10-20 deaths, and the number of deaths could not be established in 6% of events.

ECAV codes additional information on the event in question. *Event direction* codes whether it can be determined which actor initiated or "started" an event, distinguishing between directed (actor/target can be established) and undirected (actor/target cannot be established) events. In 90% of events, the initiator and target could be established, leaving 10% undirected events. *Event violence* codes whether an

[18] The codebook contains a detailed description of the geocoding procedure and assignment of location precision.
[19] Actor side is coded for each event-day-location, hence can accommodate side switching of actors over time.
[20] For events involving armed groups, we provide the name of the armed group whenever possible and use armed groups consistent with UCDP actor names. ECAV could therefore be used in combination with UCDP datasets or other data focused on elections and armed conflict, such as the Militant Group Electoral Participation Dataset (Matanock 2016).

event day involved nonviolent or violent contention. Of the events, 61% involved violence or the threat of violence, leaving 40% nonviolent events. *Violence initiator* then identifies which participant of a contentious event initiated the violence (this is missing if the initiator cannot be determined).

Event name records the type of event. Coders were instructed to choose event names from a list of 18 events (see table D1 in appendix D). The most frequent events are, in descending order: protests, attacks, clashes, bombings, arrests, killings, strikes, intimidations, blockades, and shootings (the remaining event types are less common). Occasionally, coders encountered events that did not fit preexisting categories, but these events constituted less than 2% of the total. Additional variables are *event description* (a brief description of the event) and *source* (article title, date, and source: AP, AFP, BBC).

Limitations

Event data are attractive because they allow for temporally and/or spatially disaggregated research designs, but have important limitations. A major concern is reporting bias, since what is reported in the news targets a particular audience, is biased in favor of particular countries or events, and is not representative of what is happening locally (Weidmann 2014; Hoglund and Oberg 2011; Chojnacki et al. 2012; Baum and Zhukov 2015). Reporting bias is particularly worrisome for research comparing countries and/or over time. ECAV uses newswires from the Associated Press, Agence France Press, and the BBC, which likely pay more attention to the developed world and areas of strategic political or economic importance. Reporting also differs across regimes, with authoritarian regimes underreporting nonviolent collective action and overreporting violent events, whereas the opposite patterns holds in democratic states (Baum and Zhukov 2015). Furthermore, reporting has increased over time, making it difficult to establish whether increases result from greater reporting or represent a true increase in events (Urdal 2008). In appendix C, we examine this concern by comparing news reporting over time to the number of ECAV events recoded. We do not observe a positive correlation between reporting and event coding.

Without access to the universe of true events, it is impossible to assess the extent of reporting bias in ECAV. In general, elections draw a significant amount of media attention, suggesting that events during elections may be more likely to be

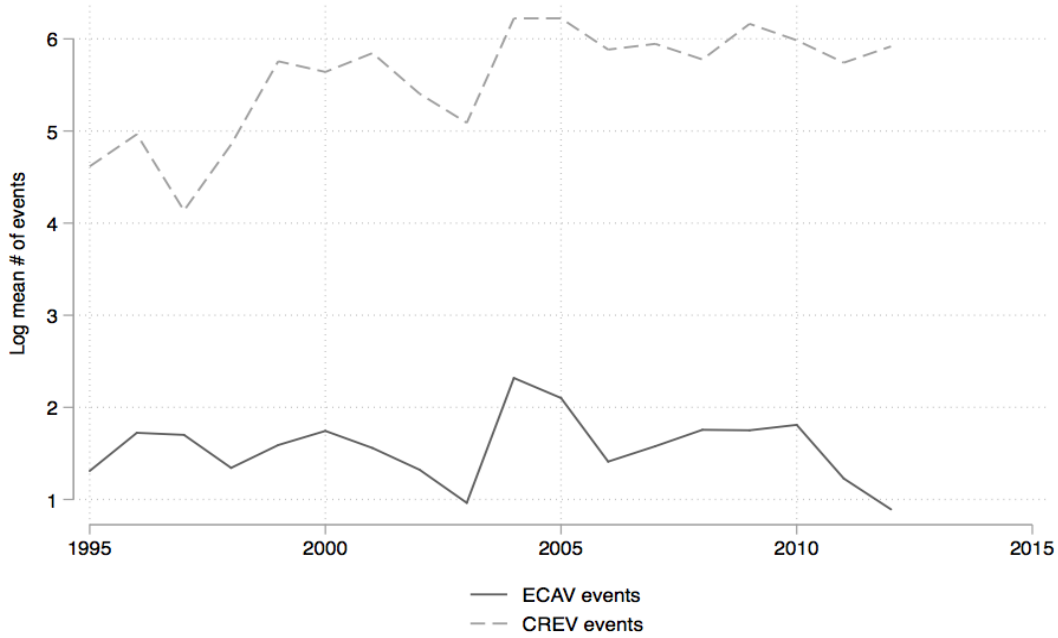
covered (McCarthy, McPhail, and Smith 1996; Fillieule 1997). Researchers can also attempt to correct for biases using new techniques to model over- or under-reporting (Cook et al. 2017), or control for factors associated with reporting bias.

Comparing ECAV to other data

We now compare ECAV to SCAD and CREV, the two datasets closest to ECAV. Because of their different geographic and temporal coverage, we present each comparison separately.

CREV is a new dataset on election violence coded from automated event data, available at the election and election-month level for 1995–2013. CREV includes 101 countries (compared to 135 in ECAV). Temporal coverage of both datasets overlaps for 1995–2012. We remove non-overlapping countries and years from the data. We also remove nonviolent events from ECAV, since CREV focuses on threats or acts of violence. Both datasets employ a similar timeframe before and after elections to identify events (10 months for CREV, 9 months for ECAV). CREV records more than 150,000 events, compared to almost 9,000 violent events in ECAV for the countries and period of overlap. Since the underlying events are not available for CREV, it is difficult to determine why such a large discrepancy exists. One possibility is that events related to issues other than elections are included in the data. CREV is based on ICEWS data events from several categories (*Threaten, Exhibit Military Posture, Coerce, Fight, Assault*) for 10 months around elections, but there is no discussion of additional pruning of events to determine the connection to election beyond the timeframe. Another possibility is the inclusion of duplicate events. If we compare the ten countries with the most election violence, CREV and ECAV agree on only three (Egypt, India, and Afghanistan). Calculating Spearman rank correlations at the level of the election, the correlation is fairly with corr=0.5. To assess whether time trends are somewhat comparable, however, Figure 2 shows trends of election violence in ECAV and CREV, using log mean number of violent events on the y-axis. Despite the large discrepancy in event incidence, there are some similarities, such as a drop in events 2003–2004 and a subsequent increase in 2005. However, the graph also shows divergence after 2010, with election violence increasing in CREV but decreasing in ECAV.

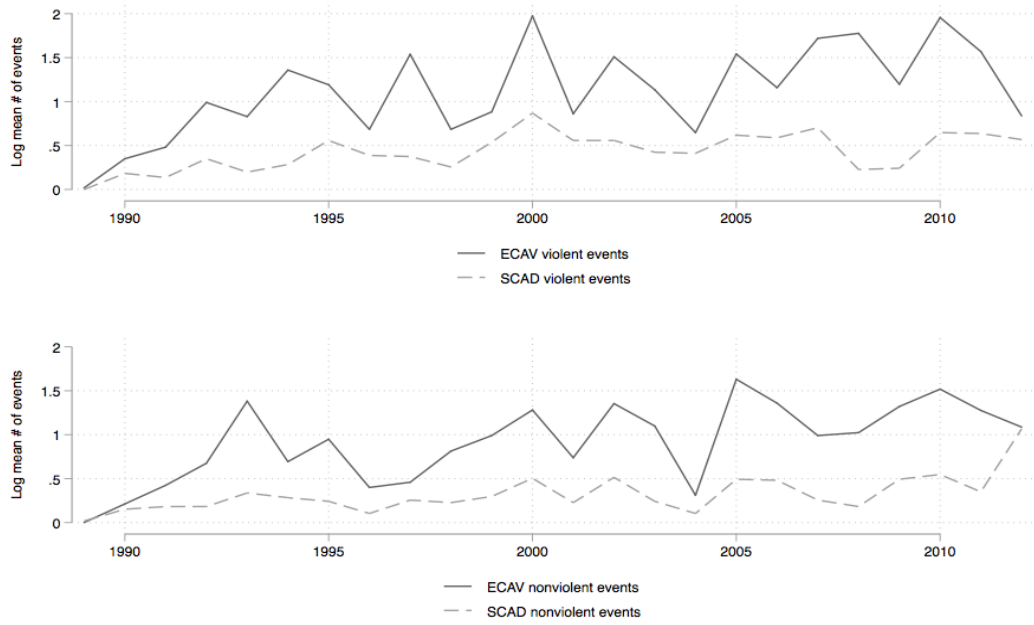
Figure 2: Comparing ECAV to CREV



We now compare ECAV with SCAD. Both datasets involve event data collected by human coders from similar sources (although ECAV also includes news reports from BBC newswires). The geographic scope for SCAD is more limited than ECAV; SCAD is available for Africa, Mexico, Central America, and the Caribbean. Our comparison is based on the countries and years (1990–2012) included in both datasets. Since SCAD focuses on general social conflict rather than electoral contention *per se*, we remove SCAD events in which elections were not one of the issues identified. Furthermore, because SCAD does not record events occurring as part of ongoing UCDP armed conflict, we remove ECAV events with armed groups as the actors or targets. We also collapse ECAV events lasting several days into single events. Because both SCAD and ECAV include nonviolent and violent events, Figure 3 compares both event types.[21] Based on the above criteria, ECAV contains 5,086 election-related events, compared to 1,300 in SCAD. Comparing the ten countries with most electoral contention in each dataset, ECAV and SCAD agree on eight (Egypt, Kenya, Zimbabwe, Mexico, Nigeria, South Africa, DRC, and Haiti). The Spearman rank correlations for all ECAV and SCAD events is corr=0.7, with corr=0.84 for nonviolent events. The time trends below also show that the data track reasonably well.

[21] For SCAD, we code demonstrations and strikes as nonviolent and remaining events as violent. For ECAV, we use the event violence variable.

Figure 3: Comparing ECAV to SCAD



To conclude this comparison, while both CREV and ECAV are global datasets on election violence, only ECAV is available at the level of the event, hence allowing for subnational analyses. Both SCAD and ECAV are events data, but for researchers interested in elections, ECAV's exclusive focus on electoral contention, clear procedures establishing the connection between elections and contention, and global geographic scope offer considerable advantages.

Electoral competition and violence in India

Existing theoretical accounts suggest that the risk of violence in unconsolidated regimes should be higher when elections are closely contested (Hafner-Burton, Hyde, and Jablonski 2014; Fjelde and Höglund 2016a; Salehyan and Linebarger 2015). Because arguments on competitiveness and violence have been tested primarily with cross-national data, however, we know comparatively little about the subnational implications of such arguments. Whether violence happens in the most competitive constituencies may depend on the electoral rule or the type of election. For example, it is not obvious that violence would be common in the most competitive constituencies when candidates compete in a single electoral constituency in presidential elections, since they may benefit more from deploying violence (and hence deterring turnout) in opposition strongholds. A global exploration of

subnational implications requires a large dataset consisting of electoral constituencies from many different countries, which is not feasible as part of this data feature.[22] Instead, we present models examining subnational patterns of electoral competition and violence in India. India should be a most likely case for finding such a subnational relationship because it has a parliamentary system with elections taking place in single-member districts. It thus seems plausible to expect that violence will be employed in the most hotly contested districts; deterring voters in such districts could swing the election in favor of the incumbent.

Results in table 1 are based on an analysis of election violence in all constituencies over six national elections held in India 1991–2010. Shapefiles for constituencies come from Sukhtankar (2011) for elections held before 2009, and Susewind (2014) for the 2009 elections. We use all 700 election-related violent events to create the dependent variable. We dichotomize election violence because few constituencies experience more than a few violent incidents per election. We create two measures of electoral competition (victory margins and party fractionalization), both commonly used measures (Wilkinson 2004; Salehyan and Linebarger 2015). We use data for election results from Bhavnani (2014) to create both. Party fractionalization is the effective number of parties calculated in line with the formula by Laakso and Taagepera (1979). For victory margins, we subtract the second-place party's votes from the winner and divide this number by total votes. We use margins from the preceding elections because we would otherwise assume that actors have perfect information about competitiveness months prior to the elections. We also control for gross cell product in 1990, electoral district size, distance from the capital, population, and previous conflict events in the district, and we include state fixed effects (controls omitted in table 1).[23]

[22] Projects such as the Constituency-Level Election Archive increasingly collect subnational data on election results across the world (Brancati 2014; Kollman et al. 2011), but are still incomplete and sometimes inconsistent across elections. Efforts to provide shapefiles delimiting constituencies, such as the Georeferenced Electoral Districts (GRED) project, are also underway (Kollman et al. 2017) but far from complete.

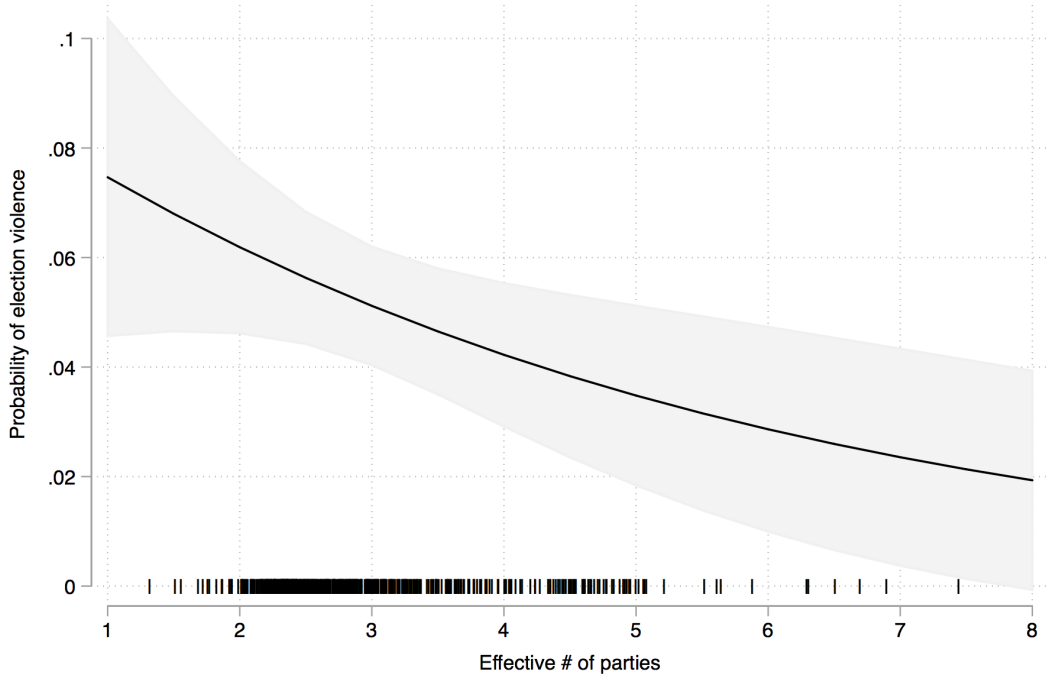
[23] Data for controls comes from the PRIO GRID, CIESIN, and UCDP GED (CIESIN 2005; Tollefsen, Strand, and Buhaug 2012; Sundberg and Melander 2013), and are merged with constituencies in ArcGIS.

Table 1: Logit regression of election violence in India, 1991–2010 (controls omitted)

	(1)	(2)
Effective no. of parties	-0.201*	
	(0.101)	
Victory margin in previous election		0.412
		(0.625)
State FE	yes	yes
No. constituencies	515	504
N	2,889	2,825
AIC	1,397.561	1,377.581
BIC	1,564.684	1,544.076

Models 1 and 2 confirm expectations for the party fractionalization measure (Wilkinson 2004), showing that violence is greater in the most competitive districts (i.e., those with two equally strong parties), but declines in less competitive (more fractionalized) districts. Figure 4 shows the substantive effect of party fractionalization on the probability of election violence. Districts with two effective parties experience more violence than those with three or more effective parties (and districts with fewer than two parties are empirically rare). In contrast, we find no effect for victory margins. This non-result may be because electoral competition in India is volatile, meaning that margins in previous elections are poor predictors years later. A detailed assessment, however, is beyond the scope of this paper. Taken together, diverging results show that scholars of election violence have yet to carefully consider the subnational implications of their arguments.

Figure 4: Predicted probability of election violence



Conclusion

ECAV data allow researchers to empirically examine a wide variety of arguments on the causes and consequences of electoral contention. The dataset covers a global sample from 1990–2012, including more than 17,000 nonviolent and violent contentious events related to elections in time and substance, and allows for disaggregation across actors, time, and space. A comparison between ECAV and other datasets showed considerable agreement between ECAV and SCAD regarding the countries experiencing most contention, and time trends also co-varied. An analysis of electoral violence in India's national elections showed that arguments on competitiveness and violence are partially supported using a subnational research design, which illustrates the importance of carefully considering and testing the subnational implications of theoretical arguments.

References

Steele, Abbey. 2011. "Electing Displacement: Political Cleansing in Apartadó, Colombia." *Journal of Conflict Resolution* 55 (3):423–45. https://doi.org/10.1177/0022002711400975.

AUTHORS. 2018. "Best Practices in Event Data Coding: Coding Quality in the Electoral Contention and Violence (ECAV) Data." Manuscript.

Banks, Arthur S. 1975. *Cross-National Time-Series Data Archive: User's Manual*. 1–2. Center for Comparative Political Research, State University of New York at Binghamton.

Baum, Matthew A., and Yuri M. Zhukov. 2015. "Filtering Revolution: Reporting Bias in International Newspaper Coverage of the Libyan Civil War." *Journal of Peace Research* 52 (3):384–400. https://doi.org/10.1177/0022343314554791.

Birch, Sarah, and David Muchlinski. 2017a. "Electoral Violence Prevention: What Works?" *Democratization* 0 (0):1–19. https://doi.org/10.1080/13510347.2017.1365841.

———. 2017b. "The Dataset of Countries at Risk of Electoral Violence." *Terrorism and Political Violence* 0 (0):1–20. https://doi.org/10.1080/09546553.2017.1364636.

Borzyskowski, Inken von. 2013. "Sore Losers? International Condemnation and Domestic Incentives for Post-Election Violence." *Working Paper*. http://www.polisci.wisc.edu/Uploads/Documents/IRC/Borzyskowski_Election%20Violence.pdf.

Brancati, Dawn. 2014. "Global Elections Database [Computer File]." *New York: Global Elections Database, Date Accessed* 1:20.

Bratton, Michael. 2008. "Vote Buying and Violence in Nigerian Election Campaigns." *Electoral Studies* 27 (4):621–32. https://doi.org/10.1016/j.electstud.2008.04.013.

Chojnacki, Sven, Christian Ickler, Michael Spies, and John Wiesel. 2012. "Event Data on Armed Conflict and Security: New Perspectives, Old Challenges, and Some Solutions." *International Interactions* 38 (4):382–401. https://doi.org/10.1080/03050629.2012.696981.

CIESIN. 2005. "Gridded Population of the World Version 3 (GPWv3): Population Grids." *CIESIN, Columbia University New York, Palisades, NY*.

Cook, Scott J., Betsabe Blas, Raymond J. Carroll, and Samiran Sinha. 2017. "Two Wrongs Make a Right: Addressing Underreporting in Binary Data from Multiple Sources." *Political Analysis* 25 (2):223–40. https://doi.org/10.1017/pan.2016.13.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, et al. 2015. *V-Dem [Country-Year/Country-Date] Dataset V5*. Varieties of Democracy (V-Dem) Project.

Daxecker, Ursula. 2012. "The Cost of Exposing Cheating: International Election Monitoring, Fraud, and Post-Election Violence in Africa." *Journal of Peace Research* 49 (4):503–16. https://doi.org/10.1177/0022343312445649.

———. 2014. "All Quiet on Election Day? International Election Observation and Incentives for Pre-Election Violence in African Elections." *Electoral Studies* 34:232–243.

European Commission, and United Nations Development Programme. 2011. "Elections, Violence, and Conflict Prevention." Summary Report.

Fillieule, Olivier. 1997. "Plus Ça Change, Moins Ça Change-Demonstrations in France During the Nineteen-Eighties."

Fjelde, Hanne, and Kristine Höglund. 2016a. “Electoral Institutions and Electoral Violence in Sub-Saharan Africa.” *British Journal of Political Science* 46 (2):297–320.
———. 2016b. “Electoral Violence: The Emergence of a Research Field.” *APSA Comparative Democratization Newsletter* 14 (2):8–11.
Flores, Thomas Edward, and Irfan Nooruddin. 2012. “The Effect of Elections on Postconflict Peace and Reconstruction.” *The Journal of Politics* 74 (2):558–70. https://doi.org/10.1017/S0022381611001733.
Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. “Armed Conflict 1946-2001: A New Dataset.” *Journal of Peace Research* 39 (5):615–637.
Hafner-Burton, Emilie M., Susan D. Hyde, and Ryan S. Jablonski. 2014. “When Do Governments Resort to Election Violence?” *British Journal of Political Science* 44 (1):149–179.
———. 2016. “Surviving Elections: Election Violence, Incumbent Victory and Post-Election Repercussions.” *British Journal of Political Science*, December, 1–30. https://doi.org/10.1017/S000712341600020X.
Hoglund, Kristine, and Magnus Oberg. 2011. *Understanding Peace Research: Methods and Challenges*. Taylor & Francis.
Hyde, Susan D., and Nikolay Marinov. 2012. “Which Elections Can Be Lost?” *Political Analysis* 20 (2):191–210.
———. 2014. “Information and Self-Enforcing Democracy: The Role of International Election Observation.” *International Organization* 68 (2):329–359. https://doi.org/10.1017/S0020818313000465.
Kelley, Judith, and Kiril Kolev. 2010. “Election Quality and International Observation 1975–2004: Two New,” no. 0550111.
Kerr, Nicholas. 2013. “Popular Evaluations of Election Quality in Africa: Evidence from Nigeria.” *Electoral Studies* 32 (4):819–837.
Kollman, Ken, Daniele Caramani, David Backer, Joel Selway, F Vasselai, and David Lublin. 2017. “GeoReferenced Electoral Districts Datasets (Beta) [Data Files and Codebook].” Ann Arbor, MI: Center for Political Studies, University of Michigan.
Kollman, Ken, Allen Hicken, Daniele Caramani, and David Backer. 2011. “Constituency-Level Elections Archive (CLEA).” *Ann Arbor, MI: University of Michigan Center for Political Studies*.
Kuhn, Patrick, and Inken von Borzyskowski. 2017. „Dangerously Informed: Christian Missionaries, Information, and Pre-Electoral Violence in Sub-Saharan Africa.” Manuscript.
Kuntz, Philipp, and Mark R. Thompson. 2009. “More than Just the Final Straw: Stolen Elections as Revolutionary Triggers.” *Comparative Politics*, 253–272.
Laakso, Markku, and Rein Taagepera. 1979. “‘Effective’ Number of Parties: A Measure with Application to West Europe.” *Comparative Political Studies* 12 (1):3–27.
Lankina, Tomila. 2015. “The Dynamics of Regional and National Contentious Politics in Russia: Evidence from a New Dataset.” *Problems of Post-Communism* 62 (1):26–44.
Matanock, Aila M. 2016. “Using Violence, Seeking Votes: Introducing the Militant Group Electoral Participation (MGEP) Dataset.” *Journal of Peace Research* 53 (6):845–53. https://doi.org/10.1177/0022343316668390.

———. 2017. “Bullets for Ballots: Electoral Participation Provisions and Enduring Peace after Civil Conflict.” *International Security*. http://www.mitpressjournals.org/doi/abs/10.1162/ISEC_a_00275.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. “Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991.” *American Sociological Review* 61 (3):478–99. https://doi.org/10.2307/2096360.

Norris, Pippa. 2012. “Why Electoral Malpractices Heighten Risks of Electoral Violence.” In *APSA 2012 Annual Meeting Paper*.

Norris, Pippa, Ferran Martinez I Coma, Alessandro Nai, and Max Grömping. 2016. “Perceptions of Electoral Integrity (PEI 4.0) Codebook.”

Raleigh, Clionadh, Andrew Linke, Haavard Hegre, and Joakim Karlsen. 2010. “Introducing Acled: An Armed Conflict Location and Event Dataset Special Data Feature.” *Journal of Peace Research* 47 (5):651–660.

Rauschenbach, Mascha, and Kathrin Paula. 2017. “Vote-Buying and Voter Intimidation: An Analysis of Strategic (De-)Mobilisation in African Elections.” Manuscript.

Ruggeri, Andrea, Theodora-Ismene Gizelis, and Han Dorussen. 2011. “Events Data as Bismarck’s Sausages? Intercoder Reliability, Coders’ Selection, and Data Quality.” *International Interactions* 37 (3):340–61. https://doi.org/10.1080/03050629.2011.596028.

Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. “Social Conflict in Africa: A New Database.” *International Interactions* 38 (4):503–11. https://doi.org/10.1080/03050629.2012.697426.

Salehyan, Idean, and Christopher Linebarger. 2015. “Elections and Social Conflict in Africa, 1990–2009.” *Studies in Comparative International Development* 50 (1):23–49.

Smidt, Hannah. 2016. “From a Perpetrator’s Perspective International Election Observers and Post-Electoral Violence.” *Journal of Peace Research* 53 (2):226–41. https://doi.org/10.1177/0022343315626240.

Staniland, Paul. 2014. “Violence and Democracy.” *Comparative Politics* 47 (1):99–118.

———. 2015. “Armed Groups and Militarized Elections.” *International Studies Quarterly* 59 (4):694–705. https://doi.org/10.1111/isqu.12195.

Straus, Scott, and Charlie Taylor. 2012. “Democratization and Electoral Violence in Sub-Saharan Africa, 1990–2007.” In *Voting in Fear: Electoral Violence in Sub-Saharan Africa.*, 15–38. Washington, DC: United States Institute of Peace Press.

Sukhtankar, Sandip. 2011. “India Assembly Constituency Maps.” http://www.dartmouth.edu/~sandip/data.html#maps.

Sundberg, Ralph, and Erik Melander. 2013. “Introducing the UCDP Georeferenced Event Dataset.” *Journal of Peace Research* 50 (4):523–532.

Susewind, Raphael. 2014. “GIS Shapefiles for India’s Parliamentary and Assembly Constituencies Including Polling Booth Localities.” https://pub.uni-bielefeld.de/data/2674065.

Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug. 2012. “PRIO-GRID: A Unified Spatial Data Structure.” *Journal of Peace Research* 49 (2):363–374.

Tucker, Joshua A. 2007. “Enough! Electoral Fraud, Collective Action Problems, and Post-Communist Colored Revolutions.” *Perspectives on Politics* 5 (3):535–551.
Urdal, Henrik. 2008. “Urban Social Disturbance in Africa and Asia.” Oslo: PRIO.
Weidmann, Nils B. 2014. “On the Accuracy of Media-Based Conflict Event Data.” *Journal of Conflict Resolution*, 0022002714530431.
Wilkinson, Steven I. 2004. *Votes and Violence: Electoral Competition and Ethnic Riots in India*. Cambridge University Press.