The subnational electoral coercion in India (SECI) data set, 1985–2015

Cécile Richetta [a,*], Imke Harbers [b], Enrike van Wingerden [c]

[a] *Political Science and International Relations, University of Geneva, Geneva, Switzerland*
[b] *Political Science, University of Amsterdam, Amsterdam, Netherlands*
[c] *London School of Economics and Political Science, London, UK*

ABSTRACT

This research note introduces the Subnational Electoral Coercion in India (SECI) Data Set, which provides comprehensive data on electoral coercion in 186 Indian Vidhan Sabha elections between 1985 and 2015. SECI draws on news reports to capture instances of electoral coercion, including coercive fraud, election boycotts by non-state armed groups, and deaths resulting from electoral violence at the assembly constituency level. SECI differs from existing data in its focus on subnational elections, its temporal coverage and its broad definition of electoral coercion, thus opening up new directions for research on electoral politics across Indian states.

1. Introduction

Recent research shows that elections – while central to democracy – can simultaneously be the locus of violence and coercion (Birch et al., 2020). In this research note, we present new data on electoral violence in Indian Vidhan Sabha (state legislative assembly) elections covering three decades (1985–2015). Building on previous literature, we conceptualize electoral violence as "coercive acts against humans, property and infrastructure" with the goal of influencing "the process or outcome of elections" (Birch et al., 2020: 4; also Birch and Muchlinski, 2020: 2; Höglund, 2009: 415).[1] Our data capture violence in the month leading up to and following polling days.

India is the largest democracy in the world, and even state-level elections far outstrip those of many countries in magnitude. In the 2012 assembly election in Uttar Pradesh, for instance, more than 125 million voters were eligible to cast their ballot, and more than 75 million people voted. Vidhan Sabha elections are administered by an independent constitutional body, the Electoral Commission of India (ECI). State legislative assemblies have a five-year term, though political factors or other circumstances may disrupt the electoral calendar.[2] Vidhan Sabha elections are fundamental to Indian political life, but they can also be sites of intense electoral coercion, including lethal violence (e.g., Iyer and Shrivastava, 2018; Harbers et al., 2022). In cross-national comparisons, India emerges as an outlier due to a comparatively high incidence of electoral violence in national-level elections (e.g., Daxecker and Jung, 2018). Yet, comparative data for analyzing electoral violence in state-level elections have been scarce or limited in temporal coverage.

In this research note we introduce the Subnational Electoral Coercion in India (SECI) Data Set, which provides comprehensive data on electoral coercion in 186 Indian Vidhan Sabha elections between 1985 and 2015. The unit of analysis is the assembly constituency-election. Assembly constituencies are the electoral districts within which members of Vidhan Sabhas are elected, under a first-past-the-post system. They are therefore the focus of electoral strategies and contention. Drawing on a national newspaper - *The Times of India (TOI)* – SECI captures which assembly constituencies witnessed electoral coercion. The data also identify whether coercive fraud was reported at polling stations, whether non-state armed groups called for an election boycott, and the overall number of deaths resulting from electoral violence in a constituency. The data set further contains event descriptions and the names of relevant *TOI* articles.

In the next section, we review existing data to outline the gap addressed by SECI. We then present our data, which is accompanied by an extensive codebook. We conclude by laying out potential applications and promising directions for future research.

2. Existing data on electoral coercion in India

Researchers interested in analyzing electoral coercion in India at the event-level have four major cross-national data sets at their disposal.

---

* Corresponding author.
  E-mail address: cecile.richetta@unige.ch (C. Richetta).
[1] We use the terms violence and coercion interchangeably.
[2] Article 374(1) of the Indian Constitution defines these rules.

The Electoral Contention and Violence (ECAV; Daxecker et al., 2019) data set records electoral contention (violent and non-violent) for national elections between 1990 and 2012. The Deadly Electoral Conflict Dataset (DECO; Fjelde and Höglund, 2022) is a global georeferenced data set of events with lethal outcomes, covering the period 1987 to 2017. DECO is a sub-set of the more comprehensive Uppsala Conflict Data Program (UCDP; Sundberg and Melander, 2013). The Armed Conflict Location Event Dataset Project (ACLED; Raleigh et al., 2010) offers data on political violence and demonstrations. For India, ACLED data are available starting in 2016. Data sets on riots provide a fourth, India-specific data source (Varshney and Wilkinson 2006; Mitra and Ray 2014).

Scholars of electoral politics can also draw on data sets at the election-level. Birch and Muchlinski (2020) Dataset of Countries at Risk of Electoral Violence, for instance, covers 642 elections across the world between 1995 and 2013. The Election Integrity Project's (PEI) data set captures the perceived quality of national elections based on expert surveys (Norris et al., 2014). These data sets are important resources for researchers interested in cross-national comparisons, but they do not speak to within-country variation. A partial exception is a study by Mahmood (2020) that extends the PEI (Norris et al., 2014) index to nine Indian states for elections conducted between 2015 and 2017.[3] The study documents substantial variation across states, highlighting the need for a closer examination of within-country variation.

Our data set differs from available sources in three critical ways. First, it focuses on subnational Vidhan Sabha elections in India, and not national ones, while still offering broad temporal coverage. Second, the data are at the assembly constituency-election level (and not the event or election level), allowing for insightful studies of spatial and temporal dynamics of electoral coercion. Finally, it covers multiple types of electoral coercion, including non-lethal ones, and is more comprehensive in its scope than data focusing solely on deadly violence or electoral integrity. Overall, our data fill a gap in the study of electoral coercion by providing novel data for a major federal democracy.

3. Electoral coercion identification procedure

Our primary data source is *The Times of India* (*TOI*), an English-language national newspaper with daily circulation across the country. This allows us to systematically track information over several decades. Since the *TOI* is a national newspaper, it provides better coverage of Vidhan Sabha elections than international newspapers, which tend to prioritize national-level elections (Von Borzyskowski and Wahman, 2019). Given its non-partisan and broad geographic coverage, the *TOI* has been used by researchers as a reliable source of violence data for India (e.g. Varshney and Wilkinson 2006; Mitra and Ray 2014). Yet, as with other media-based event data, there is a risk of under-reporting in rural areas (Von Borzyskowski and Wahman, 2019). As its readership is concentrated in urban areas, the *TOI* offers more comprehensive coverage of urban than rural issues. However, coverage of rural issues tends to focus on crime and political violence, which makes up the largest portion of rural news (Mudgal, 2011). This is consistent with Franzosi's (2014: 98) finding that violence and threats of violence attract media attention. Even though *TOI* reporting provides more fine-grained coverage of events in urban areas, it remains an invaluable source for reports of violence across the country.

We accessed *TOI* archives through news aggregation databases: ProQuest (1985–1997), Factiva (1998–2009), and Lexis Nexis (2010–2015). We extracted articles that were published in the period starting one month prior to the first polling date of an election and ending one month after the results were announced. If there was a repoll due to irregularities, we took the date when the results of the repolls

were made public as the result date.[4] We required articles to contain the name of the state and the word 'election', as well as at least one of the following key terms: 'repoll', 're-poll', 'irregularit*', 'violen*', 'unrest', 'attack', 'intimidat*', 'booth captur*' or 'boycott'.[5] For the 186 elections covered, we extracted 3587 articles (including editorials and news items), the majority clustered around election days. We then proceeded to manually code assembly constituencies affected by electoral coercion if the incidents reported had the following characteristics: (1) election-related; (2) in the state holding assembly elections; and (3) substantiated, i.e., either presented as a fact by the *TOI* or substantiated by the ECI.

4. Main features of SECI

4.1. Unit of observation

Each observation in the data set is a unique assembly constituency in a specific election. Whenever an assembly constituency was mentioned as a site of electoral coercion, we reported it, alongside incident description(s) and article name(s). If only polling station numbers were mentioned, we used the ECI Electoral Reports to identify the corresponding assembly constituency. Where possible, we complemented incomplete information about extracted incidents with alternative national news sources such as *The Hindu* and *India Today*. Overall, our main data set includes 1901 unique assembly constituency-elections affected by coercion. Incidents were sometimes mentioned at less precise administrative levels, such as district or state: we include these incidents in separate files, which can be used by researchers interested in higher levels of aggregation. The district level data set includes 84 unique district-elections, and the state level one 36 unique state-elections.

In the main data set, the variable *Constituency_Name* includes the name of the constituency as reported in the article. These names are not reliable identifiers as spelling may be inconsistent over time or across languages. Therefore, the data set also reports the *Constituency_ECI_ID*: the official assembly constituency identification number, as reported in the ECI's Electoral Report. These numeric identifiers can be used for merging the data set with other sources, such as electoral data (Jensenius and Verniers, 2017; Agarwal et al., 2021).[6]

4.2. Electoral coercion and its sub-types

We code incidents of violence if they are directly related to the election. This includes violent attacks against voters, poll workers, election officials, candidates or police forces deployed to safeguard the election, and clashes between party supporters. Violent incidents that occur prior to the election and that are debated during the campaign, but that do not appear to be directly related to polling or the election, are not included. The following example from the 2002 Manipur election illustrates the kind of events our data set reports:

> "Three persons - an India Reserve Battalion constable and two polling officials - were killed on Wednesday afternoon when militants

---

[3] The states covered are Assam, Bihar, Goa, Kerala, Manipur, Punjab, Tamil Nadu, Uttar Pradesh and West Bengal.

[4] We provide researchers with a state-election level data file that includes: the polling date (start and end); the result date; the number of articles extracted; the number of constituencies affected; and the overall number of constituencies in the state.

[5] The word 'riot*' was omitted due to being incorrectly recognized by ProQuest, yielding articles with the search term 'not' instead and therefore many false positives. We included the term "riot" in a separate search to extract articles through Factiva and Lexis Nexis: no incident was identified solely based on this key term, so we do not expect the exclusion of the term to influence results.

[6] Note that units can change over time, especially in the course of redistricting (Electoral Commission of India, 1976; 2008) or when new states are created.

attacked a polling party in the Saikul assembly constituency […]. In another incident, two presiding officials in Singhat constituency were kidnapped on Tuesday night along with the election material but released later." (*3 killed, BSF post attacked in Manipur*, 21 February 2002, *The Times of India*).

In addition, we further distinguish whether the constituency was affected by two specific sub-types of coercion: attempts at fraud and boycott calls by non-state armed groups. These constitute subtypes that are commonly reported in the *TOI*, and that may be of particular interest to scholars of electoral violence and subnational elections. *Fraud* at the polling booth implies the coercion of poll workers or security guards, and the tampering with either ballot boxes, electronic-voting machines, or ballots themselves. These events generally result in a re-poll at the polling station.[7] The following example from the Haryana 1991 election illustrates the reporting of booth-capturing, a type of fraud where party members or armed groups take over a polling station and prevent registered voters from voting and/or vote in their stead:

> "Belying earlier apprehensions, Haryana had an 'unexpectedly peaceful' poll with only three reported cases of booth-capturing in Adampur, Rohtak and Rewari constituencies." (*Large-scale violence in U.P., Bihar,* 21 May 1991, The Times of India).

The second sub-type of coercion we distinguish are *boycott calls by non-state armed groups*. There has been sustained activity by such groups in subnational elections—e.g., separatists in Jammu and Kashmir, Punjab and Assam, and Maoists insurgents in Andhra Pradesh, Bihar, Madhya Pradesh and West Bengal (Cline 2006; Staniland, 2016). These groups have regularly called on voters in their areas of influence to boycott elections and accompanied these calls by threats of violence towards voters, election workers and candidates. For some boycott calls, entire districts (and sometimes several districts) were affected. In such cases, we report each constituency in the district as affected. For instance, based on the following report on the 1994 election in Andhra Pradesh, we coded 44 constituencies in four districts as affected by a boycott call of People's War Group:

> "Andhra Pradesh is going to its second and final phase of polling tomorrow in 12 districts amidst unprecedented security arrangements under the shadow of PWG poll boycott. An estimated 2.2 crore voters are expected to decide the fortunes of 1358 contestants in 153 constituencies, of which 44 are naxalite-infested in north Telangana districts of Warangal, Karimnagar, Adilabad and Nizamabad." (*1 killed, 8 hurts in blasts on eve of A.P polls*, Dec 5, 1994, The Times of India).

Fig. 1 visualizes the number of assembly constituencies affected by electoral coercion, and its two sub-types, over our entire 30-year period. For the five largest spikes, we report the relevant elections.

In Fig. 2, we visualize how the percentage of assembly constituencies affected in each state changed over the period. The figure underlines the different trends present across Indian states: some states experience slow increases in electoral coercion (e.g., Arunachal Pradesh); others experience sudden spikes (e.g., Punjab, Uttar Pradesh); some experience consistently moderate levels of electoral coercion (e.g., Bihar); the majority experience low levels of electoral coercion (e.g., Gujarat, Haryana, Karnataka, West Bengal). The data demonstrate substantial variation between and within states.

4.3. Death toll

We coded the overall number of fatalities linked to electoral coercion, as reported by the *TOI*. Deaths are particularly unlikely to go unreported by newspapers (Franzosi, 2004: 168). They may result from partisan clashes, assaults on poll workers and candidates, and attacks from non-state armed groups.

Fig. 3 reveals that overall, there is a reduction of the number of fatalities related to electoral violence over the 30-year period, and that reporting becomes more geographically precise over time. If we relate Fig. 3 to Fig. 2, we further uncover that some states, such as Maharashtra, Orissa, and Assam, have a low proportion of constituencies affected by electoral coercion, but that reported incidents result in dozens of fatalities.[8] The severity of electoral coercion is therefore not always correlated with the number of constituencies affected by electoral coercion.

4.4. Descriptions and articles names

In the data set, we include descriptions of the incidents coded. For 91.3% of our observations, reports were made in a single article. For the rest, reports were found in up to five distinct articles. Descriptions are concise and include only the words relevant for the identification of electoral coercion. If several descriptions were in a single article, they are separated by the operator "[… and …]" in the event descriptions. Using regular expression in a statistical software, researchers can easily separate the different event descriptions if that is of interest.

5. Comparison of SECI and DECO fatalities

Among existing data sets on electoral coercion, partial overlap exists only with the Deadly Electoral Conflict Dataset (DECO; Fjelde and Höglund, 2022). DECO covers subnational elections with a temporal coverage (1987–2017) close to SECI's, with precise event dates, the location of events, and with event classification by human coders. To cross-validate our data, we compare *fatalities* for state-elections, as only lethal events are reported in DECO. We use the GADM India shapefile to match DECO event coordinates with Indian states. We select DECO classified as "high certainty".[9] Fig. 4 below visualizes the overall fatalities in DECO and SECI.

The comparison indicates similarity between the counts of fatalities, but also that there are some notable outliers. In four elections, DECO reports at least ten more fatalities than SECI: Jammu and Kashmir 2002 (+104 fatalities), Uttar Pradesh 1991 (+34), Chhattisgarh 2008 (+21), and Jammu and Kashmir 2008 (+12). These occur in 148 distinct events; out of these, 93 are "spells" (i.e., events that are part of an extended period of electoral violence). This suggests that DECO reports more fatalities in zones affected by armed conflict than SECI, particularly in Jammu and Kashmir. It is possible that the *TOI* does not explicitly link violence between the government and Kashmiri insurgents to elections. Alternatively, DECO might classify such violence as election-related, even when this link is difficult to establish for specific fatalities in the context of longer spells. Looking at elections where SECI reports at least 10 more fatalities than DECO, we identify 13 elections covering 11 states and various years (from 1989 to 2014). When examining event descriptions for these fatalities, we find clashes between partisans, murders of candidates and poll-workers, and non-state armed groups activities. Specifically, DECO appears to underreport violence that took place in Bihar (1990) (+82 fatalities in SECI), Bihar 1995 (+52), Assam 1996 (+46), and Punjab 1992 (+41). SECI coverage thus appears more comprehensive in earlier years, where information in the news aggregators used by UCDP might be less specific. Overall, the comparison suggests that SECI offers more complete coverage of fatalities outside the main zones of armed or insurgent conflict.

---

[7] We code only *reported* events of fraud and acknowledge that well-executed attempts at fraud may remain unreported. Unreported fraud might be detected by analyzing reported results (e.g., Rozenas, 2017).

[8] The official name of the state changed from Orissa to Odisha in 2011.

[9] We do so as our own data set only includes reports of events which are substantiated (see Part. 3).
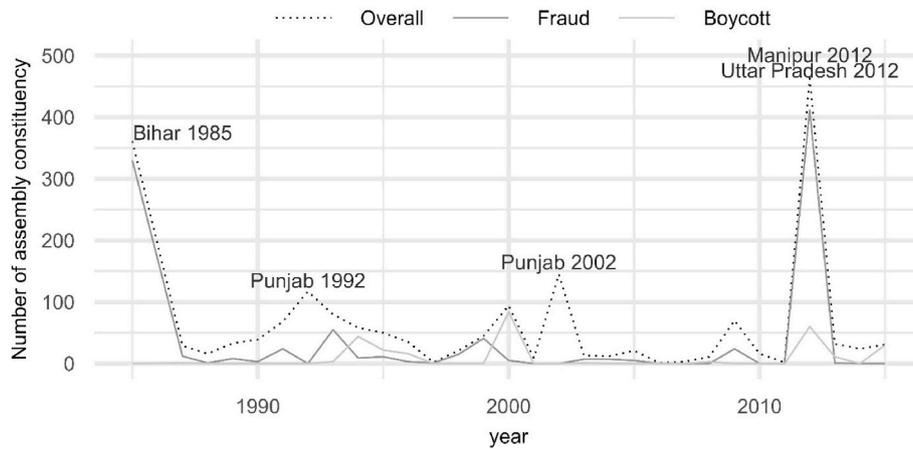
Fig. 1. Number of assembly constituencies affected by electoral coercion and its sub-types.
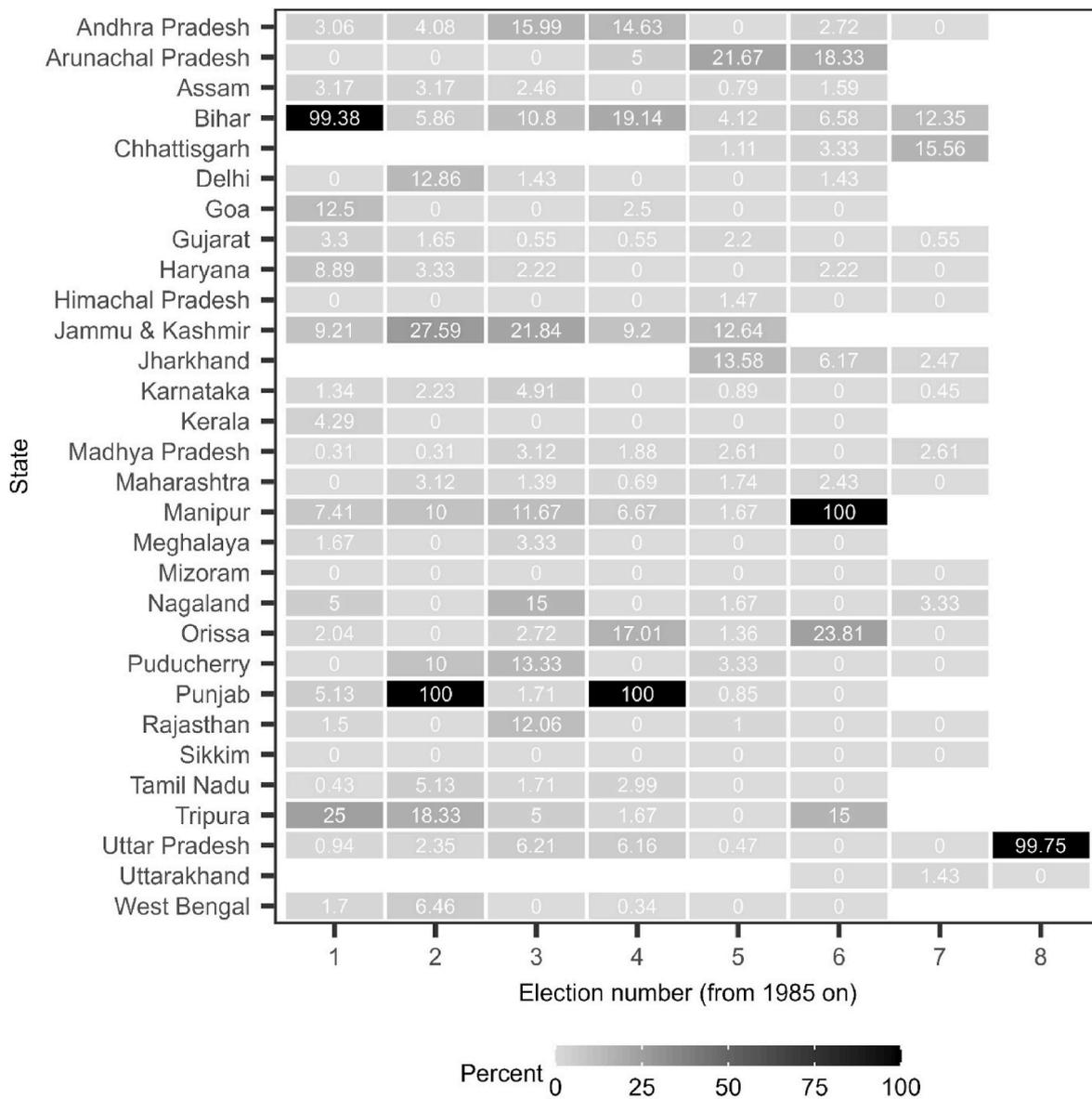

Fig. 2. Percentage of assembly constituencies affected, by state and election number starting in 1985.
Note: As Vidhan Sabha elections are staggered, we report election number rather than election years for this figure. Between 1985 and 2015, each state election is numbered. Three states were created in 2000: Jharkhand from Bihar, Chhattisgarh from Madhya Pradesh, and Uttarakhand from Uttar Pradesh. Their election number is matched with the numbering of the states from which they were created.
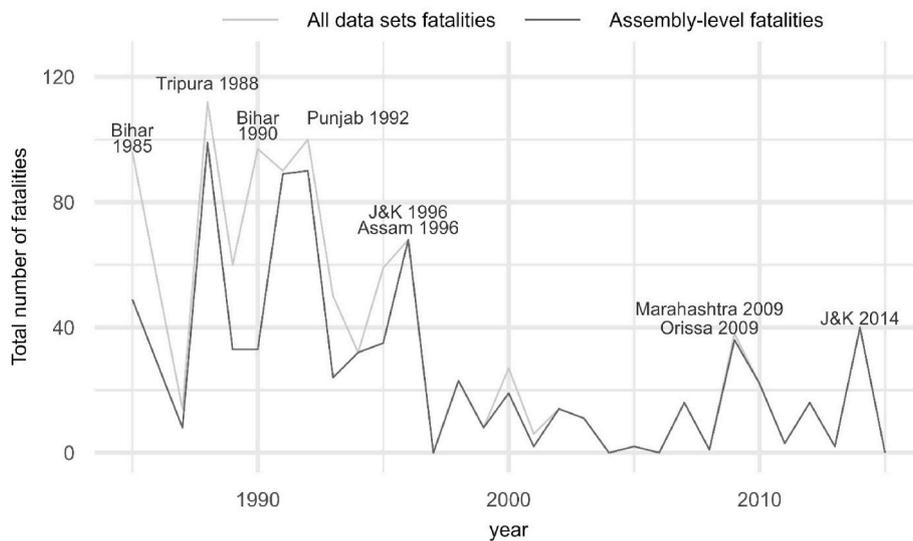
Fig. 3. Overall number of fatalities related to electoral violence, per year.
Note: "J&K" = Jammu and Kashmir. "All data sets fatalities" is the sum of fatalities by year in our constituency-level data set, and our two additional data sets (district level and state level). "Assembly-level fatalities" is the sum of fatalities by year reported in the constituency-level data set.
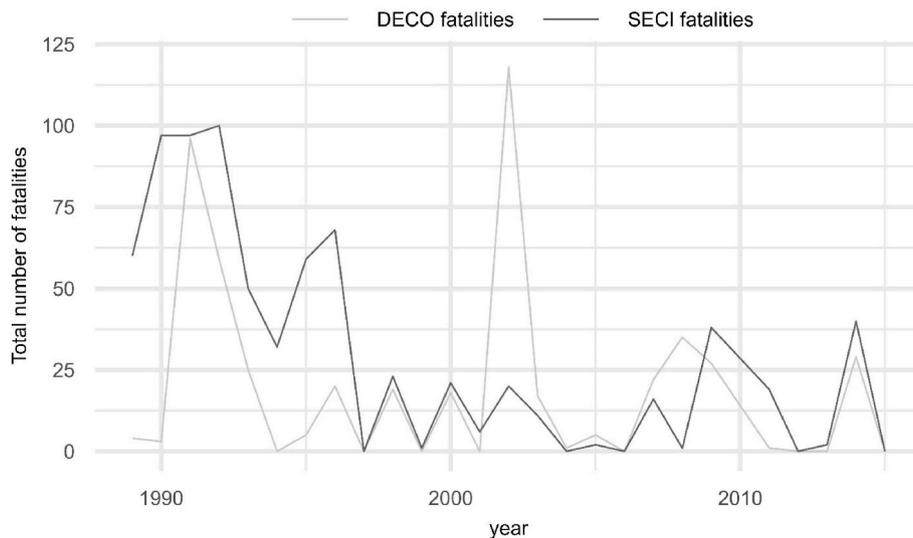

Fig. 4. Overall number of fatalities related to electoral violence in DECO and SECI, per year.
Note: "SECI fatalities" is the sum of fatalities by year reported at the constituency, district, or state level.

6. Potential applications and conclusion

Compared to existing data covering subnational elections in India, SECI has four key advantages. First, it covers an early (1985) and long period of time (30 years), so it can be leveraged to study change. The general trends in SECI (Figs. 1–3) suggest several broad lines of scientific enquiry: Why does the incidence of electoral violence vary over time? Why is there an overall reduction of fatalities related to electoral coercion? Researchers can also answer more specific questions, focusing on the sub-categories that we report. For instance, research on election management and integrity raises important questions about best practices. Our data coincide with the gradual introduction of Electronic Voting Machines (EVMs) across India. Did they succeed in reducing fraud and coercion at the polls? In addition, our data make it possible to examine when armed groups boycott, rather than participate in elections (Matanock and Staniland 2018).

Second, our data allow flexibility regarding the level of analysis. They can remain disaggregated at the assembly constituency level or aggregated to districts or states based on the needs of researchers. For instance, data can be leveraged to indicate how many constituencies were affected by electoral violence in a given election, and thus to speak to the overall integrity of subnational elections (e.g., Harbers et al., 2019; Mahmood, 2020).

Third, the data set can be merged with existing shapefiles of assembly constituencies (e.g., Jensenius, 2017). Geo-referencing the data makes it possible to study spatial processes, such as proximity or spill-over effects. Moreover, merging data with existing shapefiles facilitates integration of SECI with other available data sources. The National Family and Health Survey, for instance, provides spatial data for sampling clusters in recent waves, and includes data on subnational public goods provision.[10] Lokniti-CSDS conducts state-level election studies with pre- and/or post-poll surveys in selected assembly constituencies.[11]

---

[10] For more information see: http://rchiips.org/nfhs/index.shtml.
[11] https://www.lokniti.org/state-election-studies.

Finally, even though our unit of observation is the assembly constituency-year, SECI includes event descriptions so researchers can further investigate actors involved, the targets of electoral coercion, or identify additional sub-types of coercion based on their specific question or line of inquiry (e.g., Harbers et al., 2022). Studies of patterns and dynamics of subnational electoral coercion in India can be useful to researchers in the fields of electoral studies, political violence, policy-making and democratic regimes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data sets, codebook and R script are available on Harvard Dataverse: https://doi.org/10.7910/DVN/SCGY8A

References

Agarwal, A., et al., 2021. TCPD Indian Elections Data v2.0. Trivedi Centre for Political Data, Ashoka University. https://lokdhaba.ashoka.edu.in/.

Birch, S., Muchlinski, D., 2020. The dataset of countries at risk of electoral violence. Terrorism Polit. Violence 32 (2), 217–236.

Birch, S., Daxecker, U., Höglund, K., 2020. Electoral violence: an introduction. J. Peace Res. 57 (1), 3–14.

Cline, L.E., 2006. The insurgency environment in Northeast India. Small Wars Insur. 17 (2), 126–147.

Daxecker, U., Jung, A., 2018. Mixing votes with violence: election violence around the world. SAIS Review of International Affairs 38 (1), 53–64.

Daxecker, U., Amicarelli, E., Jung, A., 2019. Electoral contention and violence (ECAV): a new dataset. J. Peace Res. 56 (5), 714–723.

Electoral Commission of India, 1976. The Delimitation of Parliamentary and Assembly Constituencies Order.

Electoral Commission of India, 2008. Delimitation of Parliamentary and Assembly Constituencies Order.

Fjelde, H., Höglund, K., 2022. Introducing the deadly electoral conflict dataset (DECO). J. Conflict Resolut. 66 (1), 162–185.

Franzosi, R., 2004. From Words to Numbers: Narrative, Data, and Social Science, Vol. 22. Cambridge Times: Cambridge University Press.

Harbers, I., Bartman, J., van Wingerden, E., 2019. Conceptualizing and measuring subnational democracy across Indian states. Democratization 26 (7), 1154–1175.

Harbers, I., Richetta, C., van Wingerden, E., 2022. Shaping electoral outcomes: intra-and anti-systemic violence in Indian assembly elections. Br. J. Polit. Sci. 1–17.

Höglund, K., 2009. Electoral violence in conflict-ridden societies: concepts, causes, and consequences. Terrorism Polit. Violence 21 (3), 412–427.

Iyer, S., Shrivastava, A., 2018. Religious riots and electoral politics in India. J. Dev. Econ. 131 (March), 104–122.

Jensenius, F., 2017. Social Justice through Inclusion: the Consequences of Electoral Quotas in India. Oxford University Press, Oxford.

Jensenius, F.R., Verniers, G., 2017. Studying Indian politics with large-scale data: Indian election data 1961–today. Studies in Indian Politics 5 (2), 269–275.

Mahmood, Z., 2020. Governance and electoral integrity: evidence from subnational India. Studies in Indian Politics 8 (2), 230–246.

Matanock, A., Staniland, P., 2018. How and why armed groups participate in elections. Perspect. Polit. 16 (3), 710–727.

Mitra, A., Ray, D., 2014. Implications of an economic theory of conflict: Hindu-Muslim violence in India. J. Polit. Econ. 122 (4), 719–765.

Mudgal, V., 2011. Rural coverage in the Hindi and English dailies. Econ. Polit. Wkly. 46 (35), 92–97.

Norris, P., Frank, R., Martínez i Coma, F., 2014. Measuring electoral integrity around the world: a new dataset. PS Political Sci. Polit. 47 (4), 789–798.

Raleigh, C., Linke, A., Hegre, H., Karlsen, J., 2010. Introducing ACLED: an armed conflict location and event dataset: special data feature. J. Peace Res. 47 (5), 651–660.

Rozenas, A., 2017. Detecting election fraud from irregularities in vote-share distributions. Polit. Anal. 25 (1), 41–56. https://doi.org/10.1017/pan.2016.9.

Staniland, P., 2016. Insurgencies in India. In: Kohli, A., Singh, P. (Eds.), Routledge Handbook of Indian Politics. Routledge, London, pp. 167–177.

Sundberg, R., Melander, E., 2013. Introducing the UCDP georeferenced event dataset. J. Peace Res. 50 (4), 523–532.

Varshney, A., Wilkinson, S., 2006. Varshney–Wilkinson Dataset on Hindu–Muslim Violence in India, 1950–1995. Inter-university Consortium for Political and Social Research. Version 2. https://www.icpsr.umich.edu/web/ICPSR/studies/4342/versions/V1.

Von Borzyskowski, I., Wahman, M., 2019. Systematic measurement error in election violence data: causes and consequences. Br. J. Polit. Sci. 51 (1), 230–252.