The Political Boundaries of Ethnic Divisions

Samuel Bazzi *is an Assistant Professor of Economics at Boston University and Core Faculty at the Human Capital Initiative at the Global Development Policy Center.*

Matthew Gudgeon *is a Ph.D candidate in Economics at Boston University, focusing on labor economics, public economics and political economy.*

SAMUEL BAZZI, MATTHEW GUDGEON

ABSTRACT

This paper argues that redrawing subnational political boundaries can transform ethnic divisions. We use a policy experiment in Indonesia to show how the effects of ethnic diversity on conflict depend on the political units within which groups are organized. Redistricting along group lines can reduce conflict, but these gains are undone or even reversed when the new borders introduce greater polarization. These adverse effects of polarization are further amplified around majoritarian elections, consistent with strong incentives to capture new local governments in settings with ethnic favoritism. Overall, our findings illustrate the promise and pitfalls of redistricting in diverse countries.

JEL Classifications: D72, D74, H41, H77, O13, Q34

Keywords: Conflict, Decentralization, Ethnic Divisions, Polarization, Political Boundaries

1 Introduction

Political boundaries can shape group identity and behavior in fundamental ways. For example, individuals from the same two ethnic backgrounds may face incentives for conflict in one political unit and incentives for cooperation in another (Michalopoulos and Papaioannou, 2016; Posner, 2004). A vast social science literature takes boundaries as given and links diversity therein to weak public goods provision, underdevelopment, and conflict (e.g., Alesina et al., 1999; Montalvo and Reynal-Querol, 2005). Would redrawing political boundaries overhaul the ethnic divisions that fuel such adverse outcomes? If political control is a key source of intergroup contestation, then changes in the composition of the electorate could affect violence. Pioneering work on ethnicity and conflict recognized this possibility (Horowitz, 1985). We provide the first attempt to investigate this important policy question.

This paper focuses on what is perhaps the largest shift in the locus of politics since the creation of new nation states after World War II. The proliferation of subnational administrative units over the last 30 years of decentralization has profoundly changed political boundaries, as Grossman and Lewis (2014) document across numerous, often diverse, countries.[1] This massive wave of redistricting offers a unique opportunity to examine how changes in political boundaries affect ethnic divisions by changing the incentives for group mobilization. We study the implications for interethnic contestation and violence in Indonesia, the world's fourth largest country and among its most ethnically diverse.

The key contribution of this paper is to show that political boundaries endogenously determine whether ethnic diversity causes instability. We exploit a policy experiment in Indonesia to estimate the effects of redistricting on social conflict. Where feasible, creating homogeneous new political units reduces violence. Yet, in practice, redistricting along group lines is rarely so simple and may instead foster new divisions.[2] We show that the potential gains in stability are often undone and even reversed in areas where new group divisions become salient and polarizing. These boundary-induced changes in ethnic divisions necessarily change the way different groups interact and compete for political relevance. In newly polarized settings, we find that the ensuing majoritarian electoral contests to control valuable public resources further amplify incentives for violence and may lead to fresh cycles of conflict beyond the first few years after redistricting.

Indonesia offers a rich setting to study changes in ethnic divisions. First, like many other new democracies, political violence remains a major policy concern (Butcher and Goldsmith, 2017). After the fall of a highly centralized, authoritarian regime, many feared the diverse country would break apart as local conflicts tore at ethnic, religious, and regional seams. These large-scale internecine conflicts have subsided, replaced by more sporadic outbursts of low-intensity social conflict and recurring electoral violence. Amplified by decentralization, ethnic mobilization around elections and political patronage are pervasive.[3] Meanwhile, institutional constraints on violence remain relatively weak. While distinct

[1]For example, from 1990 to 2010, Nigerian states increased from 22 to 37, Ugandan districts from 34 to 112, and Kenyan districts from 47 to 70. Czechoslovakia and Hungary increased their municipalities by 50 percent from 1989 to 1993. Brazilian municipalities increased from 3,974 in 1980 to 5,560 in 2000. Vietnam increased its provinces from 40 to 64 from 1996 to 2003.

[2]Consider a stylized case where two minority groups break away to form their own district. Previously pitted together against a larger majority, these two groups now face fresh incentives for conflict in the newly polarized district. Related arguments have been made in the context of the newly created country of South Sudan where the North–South ethnic divide between Arab and non-Arab in pre-partition Sudan gave way to interethnic divisions among the non-Arab Nuer and Dinka ethnic groups in South Sudan (Posner, 2017).

[3]Aspinall (2011, p. 298) notes that, "Since the introduction of direct local government head elections, there has been a shift in

from civil wars, these outbreaks of local conflict may accumulate over time and also exert lasting adverse effects by eroding trust in the political process (Dercon and Gutiérrez-Romero, 2012), by undermining local institutions (Michalopoulos and Papaioannou, 2013), by weakening social capital and deepening ethnic cleavages (Dower et al., 2017; Rohner et al., 2013a), and by planting the seeds for more serious, future conflict (Rohner et al., 2013b).

Second, the Indonesian context allows us to overcome a host of endogeneity and measurement challenges typically associated with identifying the effects of changing political boundaries. Decentralization reforms beginning in the late 1990s led to a dramatic increase in the number of district governments across the archipelago from 302 in 1999 to 514 in 2014 (see Figures 1 and 2). Motivated by a desire to reduce sociopolitical tensions, the central government created a very favorable environment for such redistricting. This allows us to rule out first order concerns about strategic violence aimed at achieving or preventing certain types of partitions. Moreover, we show that an abrupt, centrally-imposed moratorium between 2004 and 2006 generates plausibly exogenous variation in the timing of redistricting across locations.[4] We exploit this staggered process in a generalized difference-in-difference (DiD) framework focused on districts that split around the moratorium. This approach prioritizes internal over external validity and constrains the effective sample size, two limitations we explore at length. But, the combination of exogenous timing and granular data on diversity and conflict make Indonesia arguably the most compelling context to study the effects of political boundary changes on ethnic divisions.

Using universal Population Census data from 2000, we measure how homogenized the new, smaller districts are relative to the original district based on the same, initial populations. Indonesia is home to over 1,000 self-identified ethnic groups, allowing us to distinguish two key measures in the literature: fractionalization and polarization. Ethnic fractionalization (F), which captures the likelihood of meeting someone outside your group, declines substantially on average at the original district level. However, some of the newly drawn borders encompass fewer, large groups, thereby increasing ethnic polarization (P).[5] While homogenization was an objective of redistricting in many areas, policy constraints on the number of required contiguous subdistricts for a new district limit the set of feasible partitions. We exploit these constraints to address endogeneity concerns relating to the specific choice of new borders.

We estimate the effects of these changes in ethnopolitical boundaries on conflict using new data developed by the Indonesian National Violence Monitoring System. Based on systematic coding of print newspaper archives, the data capture over 230,000 violent incidents at a high spatial and temporal frequency from 2000 to 2014. While limited initially to high-conflict regions, these data offer more comprehensive coverage than other sources for Indonesia and rival some of the best geospatial event-based sources in sub-Saharan Africa and South Asia (e.g., ACLED) in terms of depth and detail. Reported violent events include, among others, attempts to influence the allocation of resources and express popular dissatisfaction with local governance (Barron et al., 2014). Vigilantism and public mobilization along

favor of alignment between the ethnic identity of local government heads and the local populations they govern. Electoral winners tend to be drawn from the largest ethnic group in the district or province." We provide further qualitative context in Section 6.1, which also shows that greater diversity within newly created district boundaries is associated with closer elections and stronger individual preferences for politicians' ethnicities and patronage.

[4]We build upon Burgess et al. (2012) who use similar variation in their study of deforestation externalities due to redistricting.

[5]Formally, ethnic *fractionalization* in district d is $F = \sum_{j=1}^{\mathcal{N}_e} g_j(1-g_j)$, where $\mathcal{N}_e$ is the number of ethnic groups in the district, and g_j is the population share of group j. Adopting the Esteban and Ray (1994) metric, ethnic *polarization* is given by $P = \sum_{j=1}^{\mathcal{N}_e}\sum_{k=1}^{\mathcal{N}_e} g_j^2 g_k \eta_{jk}$, where η_{jk} is the distance between groups j and k, for which we use linguistic distance.

ethnic lines are common (Wilson, 2015) as is strategic violence around elections (Harish and Toha, 2017).

Despite policymakers' goals, redistricting does not change the average incidence of social conflict. This somewhat surprising finding can be explained by heterogeneous changes in ethnic divisions. Reductions in conflict in newly homogeneous units are offset by increases in newly polarized units. Our estimates suggest that political boundaries can explain around half of the endogenous long-run correlation between ethnic polarization and conflict in the cross-section. These changes arise in both the *parent* district, which retains the original seat of government, and the *child* district, which acquires a new local government and capital (see Figure 3 for an example). Robustness checks, including a machine learning approach to selecting controls, rule out numerous other confounding effects of redistricting and changes in media attention. We also rule out small-sample biases to estimation and inference. Overall, the abrupt reconfiguration of the diverse electorate gives rise to fresh incentives for violence that are amplified in polarized settings.

These border-induced changes in ethnic divisions may have persistent effects on conflict as a result of political cycles. Violence ratchets up around the first mayoral elections, particularly where they are closely contested. New ethnic divisions, and polarization in particular, exacerbate violence around these majoritarian elections, especially in child districts where the stakes are high given the fundamental role of the initial administration in establishing the new government and its first budget. We see no such amplification around parliamentary elections, which is consistent with incentives for violence differing under proportional representation. Moreover, this amplification effect persists into the next mayoral election five years later and also drives violence in intervening periods, pointing to new cycles of political conflict. The more pronounced cycles in child districts are consistent with differences in the scale of changes in contestable public resources but may also point to nascent differences in local state capacity. Ethnic favoritism helps rationalize electoral violence and perpetuate grievances, a mechanism we validate using nighttime light intensity to capture village-level changes in access to publicly-provided electricity after redistricting.

Our results line up nicely with case studies in Indonesia. In Section 6.4, we discuss a case of redistricting in West Kalimantan that illustrates how border-induced changes in ethnic divisions affect conflict. Diprose (2009) offers similar insights on Central Sulawesi as do Nolan et al. (2014) on Papua, noting that "local elections in relatively new districts can exacerbate existing social fault lines." With a bottom-up approach to redistricting, it is sometimes difficult for local parties to foresee all the complex new divisions that might arise with new borders. Wilson (2015, p. 33) succinctly summarizes the Indonesian context, arguing that redistricting "created confusion regarding lines of authority" and "involved renegotiating the boundaries of collective identities" as "local government bureaucratic precincts were altered, and some networks of resource distribution that relied upon state agents were disrupted or excluded by these shifting boundaries." As a result, "ethnic identities have been politicised as clientilistic networks. . . mobilising support along communal lines." Of course, conflict is by no means a foregone conclusion. However, by localizing political contestation, redistricting may make it easier to solve collective action problems that might otherwise forestall violence.

Similar findings may be at play outside Indonesia as well. In Uganda, for example, Green (2008) argues descriptively that "the huge expansion in the number of new districts has led to local-level conflict by altering relations between local ethnic groups." Although dramatic, the wave of redistricting we study is comparable in scale and purpose to efforts across sub-Saharan Africa, Asia, and Eastern Europe (see Grossman and Lewis, 2014; Grossman et al., 2017). A small but growing political science literature examines this process, documenting similar institutional underpinnings (see Pierskalla, 2016a, for a survey). In diverse countries, as in Indonesia, the resulting shift in the locus of politics can fundamentally change relevant group boundaries. How these changes affect incentives for conflict is very much an open question to which the Indonesian setting is uniquely well suited to answer.

Related Literature. This paper makes several contributions to the political economy and development literatures. Our central contribution is to show that subnational borders can reshape the ethnic divisions underlying conflict. We argue that the ethnic divisions underlying widely-used diversity measures are neither fixed nor exogenous and instead depend on the political boundaries within which groups are organized. Case studies as well as lab and survey experiments across Africa find that interethnic preferences are context-dependent (Berge et al., 2015; Eifert et al., 2010; Habyarimana et al., 2007; Lowes et al., 2015; Posner, 2004), and Blouin and Mukand (forthcoming) and Yanagizawa-Drott (2014) show that public media can manipulate these preferences. We show that subnational boundaries are a policy choice that can change ethnic cleavages by reorienting which groups are relevant to one's own.[6] While prior literature has recognized this political endogeneity of ethnic divisions (see Fearon, 2006; Posner, 2017), we provide the first empirical test using border changes. Our findings highlight important policy tradeoffs and demonstrate how majoritarian elections may exacerbate the new ethnic divisions.

More generally, we offer policy lessons that go beyond the otherwise pessimistic conclusions from the literature on diversity and conflict. Alesina et al. (2011) and Michalopoulos and Papaioannou (2016) identify adverse long-run effects of colonial partitioning of ethnic groups across national borders. Although infeasible to undo most national borders, within-country redistricting may change the equilibrium relationship between diversity and conflict. Amodio and Chiovelli (2017) also study equilibrium changes but focused on geographic mobility, showing that abrupt, migration-induced changes in local polarization exacerbate conflict in post-Apartheid South Africa. Our effect sizes are smaller. This is intuitive because their results are based on changes in the underlying population, which introduces new groups and non-ethnic divisions (e.g., immigrant–native). By comparison, redistricting holds the local population fixed and instead reorients the relevant group divisions within new boundaries. This is important from a policy perspective given that redistricting is typically more feasible than resettlement.

Furthermore, we provide a rich context to identify which configurations of diversity stoke political conflict. Esteban and Ray (2011a) show that in conflict over public goods (e.g., political power), polarization should be relatively more important than fractionalization as polarization captures differences in intergroup preferences and the strength of in-group ties.[7] Esteban et al. (2012) provide supportive

[6]In this respect, our identification strategy is somewhat akin to Hjort (2014) who uses the random assignment of workers to ethnically mixed teams to understand how diversity shapes productivity in a flower plant in Kenya around a period of national interethnic strife. Hence, our notion of time-varying diversity is distinct from variation due to migration or differential mortality and fertility, which may be confounded with a host of other factors whose effects on conflict are of separate interest.

[7]Spolaore and Wacziarg (2017) make similar points in a related setup with the additional insight that contests over rival goods may be more likely between groups with similar preferences, a prediction borne out empirically when examining international conflict, which typically involves contestation of private resources such as oil (Caselli et al., 2015; Spolaore and Wacziarg, 2016). Other models of conflict emphasize the conditions giving rise to mobilization along ethnic as opposed to other group lines (e.g., Caselli and Coleman, 2013; Esteban and Ray, 2008a, 2011b). Consistent with some of these predictions, our results highlight the important role of ethnic markers in mobilizing around newly contestable local politics.

cross-country evidence using proxies for the relative publicness of resources within countries. Local government proliferation offers a more precise public resource shock. We find that polarization matters relatively more than fractionalization, particularly in child districts, where new government institutions are being formed and elections are being held for the first time. We explicitly link polarization to violence around majoritarian elections but not proportional-representation elections, a distinction in line with predictions in Esteban and Ray (2008b). These results help clarify both the type of conflict and the type of diversity that might matter in new democracies undergoing similar decentralization processes.

Finally, our results suggest that the growing number of studies exploiting subnational variation in diversity should use spatial boundaries at levels of aggregation that are outcome-relevant.[8] This echoes recent work arguing that local diversity can have different effects than diversity at the aggregate, country level given the different nature of interaction within and between jurisdictions (Alesina and Zhuravskaya, 2011; Desmet et al., 2016; Montalvo and Reynal-Querol, 2017).

Roadmap. Section 2 details the context of district proliferation and provides motivating evidence on diversity and conflict in Indonesia. Section 3 presents a conceptual framework showing how redistricting can reshape ethnic divisions. Section 4 proposes measures of border-induced changes in diversity and presents the new conflict data. Section 5 develops the empirical strategy and presents core results. Section 6 identifies political mechanisms linking the changing ethnic divisions to new conflict dynamics. Section 7 concludes with a discussion of policy implications and future research.

2 Background: Local Government Proliferation and Ethnic Divisions

The number of districts in Indonesia increased by nearly 70 percent over a 15 year period. This background section details the central role districts play in government, key features of the redistricting process, and the fundamental changes in the scope of government that follow. For reference, Appendix Figure A.1 provides a timeline of key events discussed below. In a final section, we provide motivating evidence on the role of ethnicity in shaping politics and conflict.

2.1 Decentralization and the Transfer of Power to District Governments

Indonesia has four main tiers of government. The largest tier is the province, of which there were 34 in 2014. Provinces are divided into districts known as *kabupaten* in rural areas and *kota* in urban areas. In 2014, there were 514 districts. Districts are in turn divided into 7,094 subdistricts (*kecamatan*), which are further subdivided into more than 80,000 villages, the smallest unit of government.

The resignation of President Suharto in May 1998 ushered in a wave of decentralization reforms that shifted the balance of power away from the central government and provinces and towards the districts. Effective January 2001, districts assumed responsibility for nearly all public policy with the exception

[8] As better data on subnational diversity become available (e.g., Gershman and Rivera, 2016), it is tempting to take the analysis to increasingly granular geography. This might be appropriate for questions about intergroup contact and preferences but perhaps less appropriate for studying conflict if the resources being contested are determined and allocated at higher levels of administration. This is in line with the warning by Michalopoulos and Papaioannou (2017) regarding the modifiable areal unit problem (MAUP) in the use of spatial data where results depend on the level of aggregation.

of the few areas naturally reserved for the central government (e.g., foreign affairs, fiscal and monetary policy). Accordingly, districts assumed extensive decision-making power over local expenditures.

District heads or mayors, known as *bupati* and *walikota*, supervise the budget process, amalgamating spending requests from lower levels of government and submitting the final budget to local parliament for approval. The district executive plays a significantly larger role than local parliament in determining the composition and allocation of public goods (see Lewis, 2017b; Martinez-Bravo et al., 2017).[9] While parliamentarians remain important, they tend to be less individually accountable than mayors and more beholden to national party politics than to the local population (Lewis, 2017a).

Major electoral reforms also accompanied decentralization. Previously appointed by the central government, district heads were now elected via majority vote by members of the local parliament, who were in turn popularly elected according to a closed-list proportional representation system. Democracy deepened further in 2005 as district heads and their running mates were now directly elected via plurality/majority voting. Given the power of the district executive, these quinquennial mayoral elections, which occur at different times across districts, are a focal point of local political contestation.[10]

2.2 Creating New Districts

Until the late 1990s, district boundaries were relatively stable (Booth, 2011). Many boundaries originated under colonial rule when the Dutch used local leaders for indirect rule. In practice, these administrative divisions spanned large swathes of territory with many different groups (Cribb, 2013). Post-independence, new districts arose with the goal of uniform population (Charras, 2005). However, due to imbalances in population density across islands, many districts continued to span large areas and multiple ethnic homelands. These expansive, arbitrary (post-)colonial boundaries remained largely in place until the fall of Suharto, when subsequent governments facilitated dramatic redrawing of the district map. The number of districts ballooned from 302 in 1999 to 514 in 2014 (see Figure 1) through a process known colloquially as *pemekaran* or blossoming.

Redistricting Process. Subdistricts break off from their original district to create new districts. After a split, the *original district* is divided in (at least) two: The single *parent district* contains the original district name and capital with pre-existing institutions. The *child* district(s) receives a new name, capital, district head, parliament, and government apparatus. Figure 3 provides an example of this distinction with the original district of Aceh Tenggara splitting into the parent of Aceh Tenggara and the child of Gayo Lues. This also highlights the fact that original districts split up along contiguous, pre-existing subdistrict lines.

Local elites and interest groups initiate the action to split in accordance with a redistricting law passed in 2000. The regulation stipulates that new districts must have: (i) at least three subdistricts, (ii) support among the original district mayor and parliamentarians, and (iii) demonstrated socioeco-

[9]Mayors also appoint all subdistrict heads beginning in 2004 and have long appointed village heads in urban areas (Martinez-Bravo, 2014). These local officials play important roles in local public goods provision by navigating relations with the district government where the resources initially flow from the center (see, e.g., Section 6.3 on electricity). They also play an important role in mobilizing voters around elections, effectively greasing the patronage politics discussed in Section 3.

[10]As Aspinall (2011, p. 305) notes, decentralization "shifted state resources and hence the focus of political contestation down toward the base of the political system." Booth (2011, p. 46) argues, "Certainly the devolution of resources to the district level, where it has occurred has made the job of district head. . . very attractive to those who in the past had only managed to achieve lower-level positions in provincial or regional bureaucracies, or who had been largely excluded from official positions."

nomic capacity in terms of basic public goods and economic infrastructure.[11] We discuss these policy constraints in Section 4.1 as they determined the scope of feasible changes in ethnic divisions.

For identification purposes, we focus on the wave of redistricting from 2001–3 and 2007–8.[12] The central government twice stopped the redistricting process via national moratoria, the first of which occurred from 2004–6 and the second from 2009–2012 as clearly seen in Figure 1.[13] In both cases, the duration and enforcement of the ban was uncertain. Indeed, applications for new districts continued to arrive at the national parliament with more than 100 proposals at various stages of completion awaiting consideration in 2005–6 (BAPPENAS, 2007). Our main empirical strategy, which builds on Burgess et al. (2012), exploits this first moratorium by comparing districts that split around this policy shock. What is crucial for our identification is that the timing of redistricting is not driven by trends in conflict or correlated factors therein. Section 4 provides supportive evidence.

Changes in Local Government. To understand how redistricting reshapes local government, a brief timeline is instructive. After a new child district is ratified, an interim executive along with local parliamentarians oversee the transition process with a focus on the first mayoral election. On average, those quinquennial elections first take place within 21 months after redistricting. The mayor is then tasked with operationalizing and staffing up to 30 new local government agencies in the new capital. By this time, roughly two years after redistricting, central government transfers have begun flowing into child district coffers. Local public expenditures begin increasing shortly thereafter, taking similar shape as in other districts with around 40–50 percent of spending on personnel. In the parent district, elections take place on the same local, five-yearly timeline as they would have absent redistricting, but in the meantime, local institutions undergo restructuring as the governed area and populace change.

Taking the original district as a whole, there are considerable gains to redistricting in terms of political representation and public resources as shown in Appendix A. First, most child district residents experience a significant reduction in distance to government representatives and institutions in the district capital. Second, the number of legislators per capita always (weakly) increases with redistricting due to apportionment rules. Third, fiscal allocation rules imply a significant, roughly 20 percent long-run increase in annual transfers, relative to a base of around 200 USD per capita. While the scale of changes in local government may differ between parent and child districts, the significant shock to politically-relevant ethnic divisions is shared by residents of both.

2.3 Motivating Evidence on Diversity and Conflict

Like other countries, ethnic diversity in Indonesia has long been associated with adverse social consequences ranging from weaker social capital to greater conflict. In the late 1990s, ethnic divisions were a factor in major conflicts across the archipelago, including, among others, separatist movements in Aceh

[11] Given the favorable returns to parent and child districts (see below), splitting proposals were generally widely supported by original district parliamentarians who represent constituencies in both areas prior to redistricting (Pierskalla, 2016b).

[12] The redistricting in 1999 occurred before the new government regulation on *pemekaran* and substantively differs from later rounds of redistricting. Several were long-standing requests from the Suharto era, and others were initiated by the central government (Fitrani et al., 2005). These areas only enter our main analysis if they later split again.

[13] The primary stated reason for the moratoria was the drain on fiscal resources and lack of capacity to meet the staffing needs of new child district governments. Upon lifting the first moratorium, the government tightened the law on redistricting by increasing the minimum number of subdistricts to five and requiring original districts to have existed for at least seven years.

and Papua, communal violence in Maluku and Central Sulawesi, and political violence in West and Central Kalimantan (Barron et al., 2014).[14] Consistent with these pervasive divisions, Mavridis (2015) shows that residents of more ethnically diverse districts exhibit lower generalized trust and less community participation. Even as many of the most intense conflicts subsided by the mid-2000s, ethnicity remained a key vehicle for political mobilization across the country (see Allen, 2014; Aspinall, 2011; Fox and Menchik, 2011). The introduction of direct majority/plurality elections for mayors reinforced this tendency.[15] Indeed, electoral competition is much stiffer in diverse districts, where fractionalization and polarization are associated with narrower victory margins, especially in newly created districts (see Appendix E.1). Relatedly, clientelism and patronage pervade local politics, with patronage networks often based on local ethnic identities and intermediated by lower-level officials such as village heads (Aspinall, 2013; Aspinall and Sukmajati, eds, 2016).

These ethnic group dynamics occasionally manifest in conflict. Ideally, group-based contests over distributive goods occur peacefully through the political process. However, in settings with weak institutions (e.g., newly democratic countries), violence may be an effective means of influencing elections or the allocation of public resources. These low-intensity, localized bouts of violence are prevalent in modern Indonesia.[16] While distinct from full-blown civil war, these episodes of violence are a major policy concern given their potential to undermine efforts to build local state capacity and to snowball into more systematic, large-scale conflict. Using new data detailed in Section 4.2, we document a strong positive correlation between ethnic diversity—as measured by either fractionalization or polarization—and the incidence of conflict in Indonesian districts since 2000. Simple regressions suggest that a one standard deviation increase in fractionalization (polarization) is associated with a 15 (8) percent increase in the likelihood of conflict.

3 Conceptual Framework: The Political Boundaries of Ethnic Divisions

Implicit in much of the literature on diversity is the notion that ethnic divisions are shaped by political boundaries. The scope for ethnic mobilization and patronage networks depends on the size and cohesion of groups *within the electorate*. This suggests that by changing political boundaries, redistricting has the potential to reshape these intergroup divisions and, in turn, social tensions and conflict. But whether this actually happens, and how quickly it happens in practice, is unclear and yet crucial for informing policy. In the remainder of this section, we discuss a simple conceptual framework relating changes in conflict to border-induced changes in local diversity.

Several models have linked ethnic divisions to *between-group* or *social conflict* (Ray and Esteban, 2016). For example, Esteban and Ray (2011a) show in a contest-based model that equilibrium conflict can be

[14] Like Fearon (2006), we view ethnicity as determined by descent but subject to politics and history and not merely biological.

[15] Ethnic or regional political parties are effectively banned in Indonesia due to a host of legal requirements implemented with democratization in 1999, mandating that political parties must have widespread geographic coverage—in terms of institutional presence—in order to be eligible to contest elections (see Hillman, 2012, for details). Nevertheless, mobilization along identity lines in local elections is widespread (Aspinall, 2011). At the same time, class cleavages and family dynasties are less important in local politics than elsewhere in Southeast Asia (see Aspinall and Asad, 2016).

[16] For example, Tadjoeddin (2012) finds that violence occurred in 23 percent of the first direct mayoral elections between 2005–2007. Harish and Toha (2017) show that conflict around local elections is a persistent problem. They classify over 1,000 electoral violence episodes into voter-targeted, candidate-targeted, and government-targeted. We provide examples using our data in Section 4.2 and discuss their findings in Appendix C.2.

approximated by a function of diversity measures, including polarization and fractionalization, the size and relative publicness of contestable resources, and within-group cohesion. These models take political boundaries to be fixed. However, if the contest depends on the groups within a given boundary (as it may for conflict relating to local government), then changing boundaries has direct implications for conflict. This section discusses these implications for the purposes of guiding our eventual empirical specification. We keep the discussion general here, avoid taking a stance on the exact underlying model, and refer the interested reader to Appendix B for two specific examples of possible models.

Effects of Redistricting. Suppose that per-capita violence in the original district ($\mathcal{O}$) before redistricting is a function $h(\cdot) : \mathbb{R}^N \to \mathbb{R}^1$ of the original district's underlying diversity $\mathbf{D}_\mathcal{O} \in \mathbb{R}^N$, measured using one or more diversity indices: $\mathbf{D}_\mathcal{O} = (D_{\mathcal{O},1}, D_{\mathcal{O},2}, ..., D_{\mathcal{O},N})$ for $N \in \mathbb{Z}^+$. Under the assumption that splitting creates new, separate contests in parent ($\mathcal{P}$) and child districts ($\mathcal{C}$), per-capita violence in these districts post-split is determined by $h_\mathcal{P}(\mathbf{D}_\mathcal{P})$ and $h_\mathcal{C}(\mathbf{D}_\mathcal{C})$. If the functional relationship between violence and diversity remains fixed, which corresponds to keeping per-capita resources and group cohesion constant in Esteban and Ray (2011a), then $h_\mathcal{P} \equiv h_\mathcal{C} \equiv h$. Splitting then implies that the change in total per-capita violence at the original district level will be given by the difference between population weighted functions of diversity in the newly created and original districts, $\frac{\Delta V}{G_\mathcal{O}} = \frac{G_\mathcal{P}}{G_\mathcal{O}} h(\mathbf{D}_\mathcal{P}) + \frac{G_\mathcal{C}}{G_\mathcal{O}} h(\mathbf{D}_\mathcal{C}) - h(\mathbf{D}_\mathcal{O})$, where $G_\mathcal{O}$, $G_\mathcal{P}$, and $G_\mathcal{C}$ correspond to original, parent, and child district population, respectively. For typical specifications of $h(\cdot)$, this suggests that when groups separate into perfectly homogeneous child and parent districts, all violence in the original district ceases. In the special case of $h(\cdot)$ being linear and separable in diversity measures, the change in per-capita violence at the original district level resulting from redistricting is: $\frac{\Delta V}{G_\mathcal{O}} = h\left(\left(\frac{G_\mathcal{P}}{G_\mathcal{O}} \mathbf{D}_\mathcal{P} + \frac{G_\mathcal{C}}{G_\mathcal{O}} \mathbf{D}_\mathcal{C}\right) - \mathbf{D}_\mathcal{O}\right)$. This relationship informs our empirical specification relating conflict to border-induced changes in diversity at the original district level.

When examining violence at lower levels of aggregation, such as within the (eventual) parent district, we need to take a stance on how violence is initially distributed across parent and child. Letting σ be the share of total violence falling in the parent district, the change in violence within the parent district boundary is given by: $\frac{\Delta V}{G_\mathcal{P}} = h\left(\mathbf{D}_\mathcal{P} - \sigma \frac{G_\mathcal{O}}{G_\mathcal{P}} \mathbf{D}_\mathcal{O}\right)$. If violence is initially distributed according to population $\left(\sigma = \frac{G_\mathcal{P}}{G_\mathcal{O}}\right)$, the change in violence within the parent boundary is given by the difference in the diversity within that new unit and the overall diversity in the original district.[17] The same holds for child districts.

Discussion. Following the conflict literature, we focus empirically on (changes in) fractionalization and polarization. Fractionalization, a measure of the probability that two randomly selected individuals within a given boundary belong to different groups, always declines with splitting (in population-weighted terms). This is not the case for polarization. An original district with four equally-sized, different groups could split into two districts, each with two similarly sized groups contesting power. In this example, polarization, which is largest with two equally sized, competing groups, increases. Esteban and Ray (2011a) show that this specific type of diversity is relevant to conflict over public goods, which can be tailored to the winning group's preferences but not fully excluded from losers. Fractionalization meanwhile is more relevant for private goods that can be fully captured and divided among the winners. We will let the data speak by considering both measures.

[17]We validate this assumption in the data using the pre-split distribution of violence across parent and child districts.

The Esteban and Ray setup also implies that changes in the value of the contestable resource (or its relative publicness) and changes in the costs of violence will affect conflict. Redistricting is accompanied by an influx of government resources, which plausibly increase the value of the public prize, and changes in the distance to the capital, which could alter the costs of violence. These changes amplify the effect of splitting on per-capita violence, particularly in newly polarized areas.

With this in mind, we pursue a reduced form empirical approach that focuses on border-induced changes in diversity as a baseline while incorporating changes in the value of resources and costs of conflict for robustness. We also examine differences between parent and child districts, as child districts experience relatively larger changes in both resources and distance to the capital (see Appendix A).

4 Data: Measuring Changing Ethnic Divisions and Conflict

4.1 Border-Induced Changes in Diversity

Indonesia is the fourth most populous country in the world and among its most diverse. More than 1,000 self-identified ethnic groups speaking more than 400 languages span the archipelago. Indonesians are predominantly Muslim (87 percent) with minority Christian, Hindu, and Buddhist groups. From a policy perspective, diversity manifests at different levels of governance but became especially salient at the district level with decentralization. This section shows how we measure the changes in ethnic divisions that accompany redistricting.

Initial Diversity. We measure diversity using microdata from the universal 2000 Population Census. This data allow us to link the initial subdistricts in 2000 to their final 2010 district boundaries, providing us with measures of diversity at the original, parent, and child district levels. These are all based on the initial population in 2000 and hence not subject to concerns about endogenous sorting in response to redistricting. The 2000 Census was the first since 1930 to record ethnicity, allowing respondents to report a single affiliation. This led to remarkable cultural distinction with over 1,000 self-identified ethnicities (Ananta et al., 2015).[18] The sub-ethnic variation within broader ethnic groups may be relevant for conflict (Desmet et al., 2017). We also capture deeper interethnic cleavages using linguistic differences by mapping Indonesian ethnic groups to languages in the *Ethnologue* data.

We focus on two widely-used measures of ethnic diversity. Ethnic *fractionalization* in district d is given by $F = \sum_{j=1}^{\mathcal{N}_e} g_j(1-g_j)$, where $\mathcal{N}_e$ is the number of ethnic groups in the district, and g_j is the population share of group j. Given the large number of self-reported ethnic identities, fractionalization is quite high and indeed above 0.5 in many districts we study. Adopting the Esteban and Ray (1994) metric, ethnic *polarization* is given by $P = \sum_{j=1}^{\mathcal{N}_e}\sum_{k=1}^{\mathcal{N}_e} g_j^2 g_k \eta_{jk}$, where η_{jk} is the distance between groups j and k.

Polarization aims to capture the deeper cleavages in society and differs from fractionalization in two key respects. First, the squaring of the own-group term emphasizes that stronger within-group identification coincides with greater out-group alienation, which together exacerbate intergroup tensions. As such, polarization is maximized when there are two distinct, equally sized groups. Second, it formally incorporates distances between groups while the standard measure of fractionalization (F) does not.

[18] In our average original district, there are 549 distinct ethnic groups with 21 having more than 0.1 percent of the population. Consolidating subgroups based on language reduces these numbers to 271 groups, 18 with more than 0.1 percent.

We specify η using the same parametrization of linguistic distances as in Desmet et al. (2009, 2012) and Esteban et al. (2012) (see Appendix F for details).

Although less pervasive, religious divisions may be important in some locations. In what follows, we focus our discussion around ethnic diversity but nevertheless incorporate religious diversity in our empirical analysis. We account for religious polarization, but lacking an obvious notion of distance, set $\eta_{jk} = 1 \; \forall j \neq k$. Given that religious diversity typically means one sizable non-Muslim group, polarization is nearly identical to fractionalization (rank correlation≈0.99).[19]

Changes in Diversity. Prior to redistricting, these measures of diversity at the original district level demarcated the boundaries of politically-relevant ethnic divisions. After redistricting, diversity within the new parent and child district boundaries becomes salient. Motivated by our conceptual framework, we propose simple but generalizable measures to capture these changes in ethnic divisions.

As suggested by Section 3, we compute the difference between the population-weighted average diversity in the new units and diversity in the original 2000 district. For example, if original district $\mathcal{O}$ splits into parent $\mathcal{P}$ and child $\mathcal{C}$, the change in ethnic fractionalization $\Delta F = \left(\frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}F_{\mathcal{P}} + \frac{G_{\mathcal{C}}}{G_{\mathcal{O}}}F_{\mathcal{C}}\right) - F_{\mathcal{O}}$, where the first term captures the implied F within the new borders. By definition, ΔF (weakly) declines. However, the sign of ΔP is less clear, and sometimes, the new borders increase polarization.

On average, $\Delta F = -0.059$ (std. dev.$= 0.083$) while $\Delta P = -0.0002$ (0.005). Figure 5 compares the original district diversity to the implied diversity after redistricting with the distance to the 45 degree line capturing the Δ. Importantly, there is variation in $\Delta diversity$ across districts with similar initial diversity. Later, we link this variation to constraints imposed by redistricting regulations. Note also that Indonesia's remarkable diversity implies scope for differentiating between the two measures (the correlation between ΔF and ΔP is 0.38). Moreover, these ΔP reflect sizable shifts in ethnic divisions.

To make these numbers concrete, consider two examples of redistricting in our setting. First, some areas, like the original district of Aceh Tenggara, were able to leverage the geographic distribution of groups across subdistricts to split along ethnic lines, and create homogeneous governing bodies. Aceh Tenggara split into one child with 93 percent ethnic Gayo while the parent comprised 47 percent Alas, 17 percent Batak, and 15 percent Gayo. This implied a significant reduction in diversity relative to the original district, which had 39 percent Gayo, 33 percent Alas, and 12 percent Batak. These changes led to $\Delta F = -0.180$ and $\Delta P = -0.003$ and can be seen in Figure 4(a), which plots the boundaries of villages (colored by ethnic majority), subdistricts, and districts, with the latter shown pre- and post-redistricting.

Other districts split in ways that thrust hitherto less salient divisions into the limelight. One interesting example comes from Kotawaringin Timur, once the largest district in Central Kalimantan province and a legacy of Dutch administration in the 1930s, which was comprised of six relatively large groups spanning 26 subdistricts. It was not feasible to homogenize in the way that Aceh Tenggara did (see Figure 4(b)). Instead, the original district split into two child districts and one parent. This reduced fractionalization $\Delta F = -0.068$ but increased polarization $\Delta P = 0.004$ as the new districts comprised similar or fewer groups in more equal proportions than the original district. Section 6.4 discusses at length another interesting case from Bengkayang in West Kalimantan that further clarifies why these border-induced changes in diversity matter politically.

[19] By comparison, ethnic polarization and fractionalization exhibit lower correlation (<0.2) and are statistically independent.

Together, these examples point to the institutional constraints on feasible redistricting schemes. With multiple groups spanning the same subdistricts, creating completely homogeneous new districts would have been unworkable given the policy regulations on economic viability, which required sufficient scale. Using these policy constraints, we show in Section 5.5 how the set of feasible partitions shaped the resulting changes in ethnic diversity and conflict.

4.2 New Conflict Data

We estimate the effects of changes in ethnic divisions using new monthly data on conflict from the National Violence Monitoring System, referred to hereafter by its Indonesian acronym, SNPK (*Sistem Nasional Pemantauan Kekerasan*). Coverage begins in 1998 for nine conflict-prone provinces and increases to 15 provinces plus greater Jakarta beginning in 2005.[20] Crucially, conflict locations are recorded at administrative levels that allow us to link event locations to parent and child districts prior to redistricting.

Like other geospatial conflict databases such as the Armed Conflict Location & Event Data (ACLED), media reports of violence are the key input to the SNPK. Over a four-year period, project architects collected over 2 million images from the print archives of around 120 local newspapers, including multiple outlets for each province and excluding those with clear biases or no fact-checking (see Barron et al., 2009, 2014). Despite this rigor, as with all event-based conflict data, one may still worry about bias from selective reporting. We systematically address these concerns in two ways. First, we flexibly control for the number of papers available to coders for each province–month. Second, we use auxiliary Google Trends data to rule out confounding media attention due to redistricting.

Coders used a standard template to assign incidents to 10 mutually exclusive categories based on the underlying trigger. They first code incidents as domestic violence, violent crime, violence during law enforcement, or conflict. Within conflict, coders further categorize based on what is being violently contested: elections and appointments, governance, resource, identity, popular justice, separatist, and other (could not be classified). When reported, coders also indicate the number of injuries or deaths and details on property damage. Further background can be found in Appendix C.1, which also highlights important advantages of the SNPK relative to other violence data.

Outcome: Social Conflict Incidence. Our core outcome is the likelihood of *any social conflict* in a given district–month. We define this measure as any incident excluding domestic violence and crime. From 2000 to 2014, these events occurred in around 63 (36) percent of district–months based on the original district (parent/child) borders. Appendix C.3 demonstrates robustness to category misclassification, and Appendix C.4 shows that our results are driven by events with injuries and property damage rather than deaths per se. Such events are still costly and broadly reflect the sort of social instability that concerns policymakers in newly democratic countries. This is also consistent with our conceptual framework, motivated by Esteban and Ray (2011a) who note that "social conflict need not manifest itself in civil war alone, and there are various other measures . . . for instance, strikes, demonstrations, riots, ..."

Before proceeding to results, we offer a few illustrative examples of these types of conflict that characterize our outcome measure. The following translated incident descriptions in Maluku Utara district

[20]While the data is not representative of Indonesia, it spans all major island groups and covers a majority of the Indonesian population. We omit districts in Papua due to problems with the underlying administrative and census data. Data coverage is less reliable in the earliest years, and hence we exclude 1998 and 1999.

help fix ideas: "(July 13, 2010): In Galela Selatan, supporters of Djasa (a mayoral candidate) destroyed the office of Galela subdistrict, 2 official cars, and billboards of other candidates." "(August 18, 2011): Office of Morotai District Legislature was bombed; it is suspected as terrorism to prevent the inauguration of the elected mayor." SNPK records point to various forms of political violence—protests over voter eligibility, clashes between supporters, direct targeting of candidates and government offices overseeing elections—often related to local, mayoral elections (see Harish and Toha, 2017, for a rich accounting). Such violence often involves building damage and injuries rather than deaths. Nevertheless, such incidents can and do escalate.[21] Moreover, violence is not limited to election periods. Many events capture groups violently expressing grievances over policy and resource allocation issues (see Appendix C).

5 Results: The Effects of Political Boundaries on Diversity and Conflict

This section presents our main empirical results. We first provide motivating evidence that boundary-induced changes in diversity are associated with conflict. We then estimate a rigorous, generalized difference-in-difference (DiD) that recovers the causal effects of redistricting and links changes in ethnopolitical boundaries to violence. We show that despite policymakers' goals, redistricting causes no change in the average incidence of conflict. We argue that this somewhat surprising finding can be explained by heterogeneous changes in ethnic divisions. In areas able to create homogeneous new districts, conflict falls. However, conflict increases in areas where redistricting led to greater polarization among the newly defined electorate.

5.1 Sample Construction

Our main analysis comprises 52 original districts (d) in 2000 that split into 133 districts by 2014 and are covered by the SNPK. Nearly all redistricting occurs in the two years before and after the moratorium on splitting from 2004–6.[22] By focusing on districts that split around the moratorium, our empirical strategy prioritizes internal over external validity. Prior studies identify incentives for creating new districts in Indonesia, including efficiency gains in the provision of public goods, ethnic homogenization, and rent seeking (see Fitrani et al., 2005; Nolan et al., 2014; Pierskalla, 2016b). We find similar evidence, but what's important for causal identification is that these underlying incentives do not predict early versus later timing of redistricting.[23]

While our effective sample size is constrained, the Indonesian setting is plausibly the most compelling available to study the effects of changes in political boundaries. Granular data on diversity and

[21] Consider these incidents from the districts of Kota Subulussalam and Maluku Tenggara Barat: "(November 2, 2013): Demonstrations involving hundreds of supporters of candidates for mayor and vice mayor. The masses demanded an explanation from the Independent Election Commissioner. [7 injured]." and "(May 30, 2002): The chaos of the mayoral election of West Southeast Maluku district is bad. Supporters of Heri Kadubun who were riding in boats were attacked by supporters of the Taher Hanubun group [3 killed, 8 injured]."

[22] We observe 38 original districts with conflict data spanning pre- and post-split years with the remaining 14 only observed in the SNPK after splitting. This is due to the expansion of SNPK coverage to include additional provinces beginning in 2005. Only one district in our study splits again after 2008 (in January 2013), and for simplicity we drop observations in 2013 and 2014 for this district. Results are unchanged under other treatments. Four other areas split for the first time in late 2012–13. However, we exclude these from the analysis in order to focus on areas that were credibly affected by the moratoria.

[23] In Section 5.6, we discuss a reweighting exercise that is consistent with our results not being driven by selection into splitting.

conflict as well as plausibly exogenous timing of redistricting are simply not available in other countries undergoing similar decentralization reforms. Furthermore, we take several steps to address small-sample biases in Section 5.4.

5.2 Simple Difference-in-Difference: Changes in Ethnic Divisions and Conflict

We begin with motivating evidence that changes in ethnic divisions are associated with changes in conflict incidence. We regress the change in the average monthly likelihood of social conflict after redistricting on the change in ethnic diversity implied by the new borders. Figure 6 presents results in graphical form, normalizing $\Delta diversity$ and including regression lines with robust 95 percent confidence intervals. In graph (b), a one standard deviation increase in $\Delta polarization\ (P)$ is associated with a significant 6.8 percentage point (p.p.) increase in group conflict after redistricting. As a benchmark, consider two districts with roughly one standard deviation difference in ΔP. In Kupang, $\Delta P = -0.001$, and conflict fell by 22.2 p.p., whereas in Kotawaringin Barat, $\Delta P = 0.002$, and conflict increased by 12.4 p.p.[24] Meanwhile, graph (a) shows that conflict is less responsive to border-induced changes in fractionalization, despite the positive long-run correlation between initial fractionalization and conflict.

The data in Figure 6 suggests a strong relationship between the political boundaries of ethnic polarization and conflict.[25] Of course, there are many reasons why these simple DiD estimates might not reflect causal relationships. The rest of the paper aims to rule out these concerns, to understand the dynamics of conflict after redistricting, and to offer a deeper interpretation of the underlying changes in group divisions and social tensions.

5.3 Average Effects of Redistricting

We use a standard generalized DiD specification to identify the effects of redistricting on conflict:

$$conflict_{dt} = \nu + \beta post\text{-}split_{dt} + \theta_t + \theta_d + \theta_d \times t + \varepsilon_{dt}, \tag{1}$$

where $post\text{-}split_{dt}$ is an indicator equal to one for all months after the original district's first redistricting was passed into law.[26] The month fixed effects, θ_t, sweep out shocks to conflict incidence that are common across all districts. The district fixed effects, θ_d, take out time-invariant level differences in conflict incidence across districts. Meanwhile, the district-specific time trends, $\theta_d \times t$, are important given differential long-run trends in violence and consistent with fixed effects specifications in the conflict literature

[24] Recall that these small changes in polarization often imply large changes in ethnic divisions. In Kupang, for example, $\Delta P = -0.001$ captures the split of an original district with three fairly large groups (38 percent Atoni Metto, 32 percent Rote, and 18 percent Sabu) into three homogeneous new districts for each (parent with 63 percent Atoni Metto, one child with 93 percent Rote, and another child with 98 percent Sabu).

[25] These descriptive results exclude one extreme outlier with ΔP six standard deviations below the mean, are robust to controlling for all three $\Delta diversity$ regressors simultaneously, estimating robust Huber (1973) regressions rather than OLS, and conducting small-sample inference. Note, however, that our main generalized DiD estimates below retain the full sample of districts and are robust to small-sample inference and principled outlier removal procedures detailed in Appendix D.4.

[26] Districts that split into three or more all at once pose no particular difficulty. Out of 52 original districts, 11 split at multiple points in time. Consider, for example, Manggarai district, which first created one child, Manggarai Barat in 2003, and then later the parent district was further subdivided to create Manggarai Timur in 2007. In our baseline, we code these using the first date of the split. Results are robust to dropping these multi-split areas or to assigning the date of the split to the month in which the most splits took place for the given original district. However, in all cases, $\Delta diversity$ is computed over the full period, taking the original district and final parent and child districts.

(e.g., Dube et al., 2016; Dube and Vargas, 2013; Miguel et al., 2004). As noted earlier, we define $conflict_{dt}$ as a binary indicator for any reported incident of social conflict and show robustness to intensive margin specifications. We estimate all specifications using linear probability models (LPM). The coefficient β identifies the average post-redistricting deviation from district-specific conflict trends. We further estimate versions of equation (1) that disaggregate the original districts d and identify separate β for parent and child districts, allowing each to have its own fixed effect and trend.

The monthly panel specification leverages the granularity of both the conflict data and the policy changes (as split approvals vary at the month level). This allows us to capture episodic as well as recurring violence associated with discontent. This is especially useful for exploring political cycles of violence in Section 6, which vary at a sub-annual frequency. In addition, it may offer power benefits relative to a coarser annual frequency given the considerable within-year variation and relatively weak autocorrelation of conflict (see McKenzie, 2012).[27] We cluster standard errors at the original district level but demonstrate robustness to a battery of alternative approaches to inference.

Identifying Assumptions. Two key assumptions underlie the generalized DiD strategy: (i) the timing of redistricting is orthogonal to conflict and its determinants, and (ii) districts would have exhibited parallel departures from their conflict trends in the post-split period absent splitting. We provide supportive evidence here and present further robustness checks in Section 5.6.

First, we show that the timing of redistricting is plausibly exogenous. Cross-sectional regressions in Table 1 relate a normalized x variable to the timing of the initial split in original district d—measured either as the number of months since January 2000 or an indicator for whether splitting occurred after the moratorium. In Panel A, there are no statistically or economically significant effects of initial diversity within the original district borders in 2000, the eventual parent/child district borders in 2010, or the border-induced change, $\Delta diversity$. Although diverse districts are more likely to split, they are no more likely to do so earlier. Moreover, as seen in Panel B, we find similarly insignificant timing effects for a large set of 65 time-invariant or predetermined confounding variables considered in robustness checks. To ensure that more violent districts did not split earlier, we estimate a Cox (1972) proportional hazards (survival) model for time to split. The results in Panel C of Table 1 provide no evidence of a systematic relationship between changes in any social conflict and the timing of redistricting, yielding an insignificant hazard ratio of 1.18 with a p-value of 0.8.[28] Our results in Table 1 are consistent with Burgess et al. (2012) who present complementary evidence on exogenous timing.

Second, we provide evidence of parallel pre-trends. While the estimates in Table 1 are reassuring, one might still worry about spikes or dips in conflict prior to redistricting, particularly if such trends were differential with respect to (changes in) diversity. Event study specifications discussed below rule out these concerns. Overall, this evidence is consistent with the favorable environment for redistricting and the arbitrariness of approval timing due to the moratorium and other factors.[29]

[27]To be sure, all results are robust to controlling for up to twelve monthly lags of conflict. However, we do not use the lagged dependent variable specification as a baseline because it is subject to dynamic panel bias in the presence of fixed effects.

[28]The survival model framework is appropriate given that, unlike the covariates in Panels A and B, conflict is time-varying, and the Cox (1972) estimator allows, for example, for conflict incidence closer to the time of splitting to matter differentially from incidence in earlier periods.

[29] That redistricting is a largely peaceful process here is likely a consequence of the limited and often favorable fiscal and legislative consequences for the parent, combined with these obvious benefits for child districts. While there are cases of violence

Null Average Effects. Estimation results for equation (1) can be found in Table 2, which shows that redistricting does not change the local incidence of conflict. Column 1 shows a null effect of splitting on the likelihood of conflict incidents at the original district level. The point estimate and standard error are very small relative to the pre-split mean of around 57 percent of original district–months with any incidents. Nor are the null effects explained by differential pre-trends or countervailing ups and downs in conflict after redistricting. This can be seen in the event study specification in Appendix Figure D.1.1(a). The average likelihood of conflict is relatively flat pre- and post-split.

While informative about overall changes in violence, these original district level results may obscure the different implications of redistricting for child and parent districts. Nevertheless, estimating at this more granular level in column 2 leads to similarly null effects. Moreover, this is not due to differential, offsetting effects in parent or child districts as seen in columns 3 and 4.

5.4 Main Results: New Ethnic Divisions Amplify Conflict after Redistricting

Next, we identify a much richer set of post-split conflict dynamics by explicitly considering how redistricting transforms ethnic divisions. At the original district level, we therefore augment equation (1):

$$conflict_{dt} = \nu + \beta post\text{-}split_{dt} + \phi\,(post\text{-}split_{dt} \times \Delta\; diversity_d) + \theta_t + \theta_d + \theta_d \times t + \varepsilon_{dt}. \qquad (2)$$

The ϕ coefficients identify whether areas that split into more homogeneous and less polarized units experience a differential reduction in conflict. As outlined in Section 3, if original district diversity shapes conflict pre-split while diversity within parent and child districts shapes conflict post-split, then post-split changes in conflict incidence should be a function of $\Delta diversity_d$ ($\phi > 0$).

Importantly, the fixed effects, θ_d, absorb the time-invariant effects of initial diversity on conflict. Unlike the simpler DiD in Figure 6, equation (2) accounts for common shocks across districts, as well as differences in district-specific conflict trends. Recall that the timing of redistricting is uncorrelated with $\Delta diversity$, which also does not exhibit differential pre-trends in conflict (see Appendix Figure D.1.2). Moreover, we show that $\Delta diversity$ is not simply proxying for other initial district characteristics that amplify conflict after splitting. These robustness checks detailed in Section 5.5 further clarify that the identifying variation in $\Delta diversity$ is orthogonal to the timing of redistricting and driven largely by the policy-induced constraints on partitioning original districts.

Baseline Results: Original District. Table 3 reports estimates of equation (2) at different administrative levels, normalizing the $\Delta diversity$ measures. Column 1 demonstrates the conflict-enhancing effects of creating more polarized districts. A one standard deviation increase in ΔP implies a 3.6 percentage point (p.p.) increase in social conflict that is significant at the 5 percent level. This is smaller than the simple difference-in-difference estimate (Figure 6) and implies a 6.3 percent increase in the likelihood of conflict relative to the mean before redistricting. Furthermore, while Δ religious polarization also enters positively, it is less precisely estimated, perhaps due to the more limited range of districts with significant

perpetrated for and against redistricting, such episodes are limited and do not explain the timing of splitting. Indeed, the leading conflict watchdog group in Indonesia highlights a case of violent pressure for splitting in a district of West Sulawesi, which is not in our sample, but notes that "In most cases, this fragmentation [redistricting] has taken part without violence and indeed without much opposition" (International Crisis Group, 2005).

religious divisions. Meanwhile, ΔF has a weaker effect close to zero, albeit less precisely estimated and hence not statistically different from ΔP (p-value=0.17).

We address upfront two important concerns about the small effective sample size, leaving further identification and robustness checks to Sections 5.5 and 5.6. First, we remove outliers in $post\text{-}split \times \Delta diversity$ following the residual-influence approach of Belsley et al. (2005). Second, we demonstrate that inference is generally robust to accounting for the small number of clusters (Cameron et al., 2008; MacKinnon and Webb, 2018a), "effective degrees of freedom" (Young, 2016), and spatial correlation in unobservables (Conley, 1999), among other concerns about overstating precision.[30] We detail these important robustness checks in Appendix D.4 but demonstrate the main takeaway in Table 4. In short, in Panel B, the simultaneous removal of outliers in $post\text{-}split \times \Delta P$ and robust inference adjustments—including a quasi-randomization inference procedure that permutes $\Delta diversity$ across districts—leads to the strongest evidence that border-induced increases in ethnic polarization exert a significant positive effect on social conflict. It is clear that outliers work against our core result in column 1 in terms of both precision and effect sizes.[31]

We consider several benchmarks for this baseline result to shed deeper light on the causal component of the diversity–conflict relationship that can be attributed to political borders. First, the estimated effects of ΔP and ΔF are smaller than the large positive correlations between conflict incidence and diversity across all districts as discussed in Section 3. In other words, the causal component due to political boundaries is slightly smaller for polarization and significantly smaller if not null for fractionalization. Second, the effect of ΔP on conflict is around one-half of the cross-country correlation between ethnic polarization levels (P)—defined similarly based on *Ethnologue* definitions of ethnolinguistic groups—and low-intensity civil conflict in Esteban et al. (2012).[32] Third, the effect of border-induced ΔP in Indonesia is roughly one-quarter as large as the effects of migration-induced ΔP within the black population of South Africa (Amodio and Chiovelli, 2017).[33] This large difference may be due to the fact that changes in local diversity due to migration capture additional conflict-relevant group divisions besides ethnic ones, e.g., between native "sons of the soil" and immigrants (Fearon and Laitin, 2011; Weiner, 1978).

The strong effects of ΔP in column 1 of Table 3 highlight a potentially important unintended consequence of using redistricting to create homogeneous political entities. In diverse settings with many groups, while it is possible to reduce the number of groups and hence F, it may not be feasible to simultaneously ensure that the new boundaries do not engender new polarizing divisions. The two districts of Maluku Utara (MU) and Bolaang Mengondow (BW) illustrate the importance of ΔP. While both sig-

[30] MacKinnon and Webb (2018a,b,c) suggest that randomization inference may work better than the widely-used wild cluster bootstrap of Cameron et al. (2008) given the imbalance in cluster sizes. We do not take a strong stand on which is the correct approach among the five considered in Table 4 but believe that the weight of the evidence is in favor of a statistically meaningful result based on the 52 original districts that comprise our core sample. For brevity, we stick to the baseline approach of clustering by original district in subsequent tables that go beyond the core baseline results.

[31] It could of course have gone the other way, but the residual plots in Appendix Figure D.4.1 provide clear visual evidence for why the outsized influence of a few outlying observations works against us. We restrict attention to $post\text{-}split \times \Delta P$ but find similar results when also removing outliers in $post\text{-}split \times \Delta F$.

[32] This is based on re-estimating column 1 of Table 4 in their paper and comparing normalized effect sizes. We omit the Greenberg-Gini index of diversity to maintain a stricter comparison to our results, but the effect sizes are similar when including it. Their estimates are based on pooled OLS panel regressions at the five-year frequency, and the dependent variable equals one if the country experiences more than 25 conflict-related deaths with a mean of 0.28. By comparison, this same threshold and five-year frequency in our data implies a mean of 0.29. Interestingly, they too find weaker effects of ethnic fractionalization on conflict when using the full breadth of groups in *Ethnologue*.

[33] This is based on rescaling their estimate in column 5 of Table 3 by the standard deviation of ΔP, 0.09, in Table A.2.

nificantly homogenized based on F ($\Delta F^{MU} = -0.125$ and $\Delta F^{BW} = -0.097$), the new borders generated new ethnic divisions and a larger increase in polarization in MU ($\Delta P^{MU} = 0.005$ and $\Delta P^{BW} = 0.002$). We observe a 36 p.p. increase in conflict in MU compared to a 7 p.p. decline in BW after redistricting.

Baseline Results: Parent and Child Districts. To better understand the tradeoffs associated with redistricting, we estimate equation (2) at the smaller, child and parent district boundaries. Changes in violence within these smaller units should also be a function of $\Delta diversity$, specified here as the difference in diversity between the given child or parent district and the original district. As with specification (2), the goal is to identify how changes in the salience of local diversity affect conflict.

There are several advantages to running our specification at this lower geographic level. First, it leverages greater variation in $\Delta diversity$. Second, it allows us to analyze violence around post-split elections, which occur at different times in the parent and child. Finally, it provides a natural way to investigate whether the effects of changing ethnic divisions vary with the scale of changes in local government. As documented in Appendix A, children experience larger increases in fiscal transfers and a reduction in distance to the capital. In the framework of Section 3, this arguably implies that the effects of $\Delta diversity$ should be amplified in child districts.

Columns 2–4 of Table 3 estimate these relationships within the given child and parent boundaries. And, as before, Table 4 demonstrates robustness to outliers and small-sample inference. Together, the results reveal similar effects as those at the aggregate original district level, but also point to potential changes in the geography of violence. Pooling parent and child districts in column 2 yields effects that are statistically indistinguishable from those in column 1. Turning to columns 3 and 4, ΔP has similar effects when splitting the sample into parent and child districts, respectively. Our baseline hints at potentially larger effects in the child, although this difference is imprecise and not robust to outliers (see Panel B of Table 4). When turning to mechanisms in Section 6, we will see clearer evidence of differences between parent and child districts around local political cycles.

Discussion. While recent models clarify how conflict responds to diversity in long-run equilibrium, our results identify how political boundaries can shock that equilibrium by changing the salience of different group cleavages in society. The estimates in Table 3 identify the magnitude of this shift over a 5–10 year period. That redistricting can alter the effects of diversity so quickly is perhaps surprising. In Section 6, we link the resulting conflict dynamics to the nature of politics and majoritarian elections. While these effects may yet die out over time, the new ethnic divisions may also deepen, bringing us closer to the long-run correlation between diversity and conflict (within stable boundaries).

Our findings offer two policy-relevant, methodological innovations. First, prior work shows how the colonial partitioning of ethnic groups across international borders contributes to modern conflict (e.g., Michalopoulos and Papaioannou, 2016). Our strategy moves beyond static differences to show how changes in political boundaries reshape the ethnic divisions underlying conflict. While infeasible to study these dynamic changes in national boundaries, within-country boundary changes are pervasive and informative about general mechanisms linking diversity to conflict. Second, by isolating the contribution of political boundaries, we rule out a host of other confounding changes in local ethnic diversity often associated with migration flows, which are of independent interest in the study of conflict.

Our focus on boundary-induced changes in ethnic divisions further contributes to an ongoing debate over which type of diversity matters for conflict. To the extent that redistricting changes the incidence of local public goods, one expects greater heterogeneity in preferences among the governed population to lead to greater conflict. Moreover, if private resources are changing more slowly than public ones, one would expect changes in fractionalization to be less important than changes in polarization (Esteban and Ray, 2011a). Further probing the differences between polarization and fractionalization shows that border-induced changes in deeper ethnic divisions—as captured by linguistic distances (η) between groups—are driving the changes in violence around redistricting.[34]

5.5 Identification Checks: Isolating Policy-Induced Changes in Ethnic Divisions

This section addresses three key threats to our interpretation thus far: (i) that the identifying variation in $\Delta diversity$ reflects endogenous boundary choices, (ii) that $\Delta diversity$ is a proxy for other features of Indonesian districts correlated with future conflict, and (iii) that changes in contestable public resources confound the effects of $\Delta diversity$.

Feasible Redistricting. Although initial diversity in 2000 is predetermined, the particular way in which the borders are drawn, and hence $\Delta diversity$, may be endogenous. One concern is that districts that chose particularly unfavorable borders with high ΔP were the ones where future conflict would have risen anyway, say because of bad governance. We show here that the effects of $\Delta diversity$ on conflict are not explained by the particular way that districts chose to draw new borders but rather by institutional constraints on redistricting and ethnic geography that did not allow for more homogeneous new districts.

To do so, we consider all the possible ways an original district could split along subdistrict lines into k new districts, given that regulations require each new district has at least three contiguous subdistricts. This provides us with a distribution of feasible partitions, and associated $\Delta diversity$, for each district. Appendix D.3 provides full details on this NP-hard combinatorial problem.

We use this set of feasible partitions to clarify that our identifying variation in $\Delta diversity$ comes from cross-district variation in the set of feasible partitions, as opposed to similar districts choosing very different ways of redrawing their borders. First, we re-estimate our baseline specifications from Table 3, replacing the realized changes in diversity with the mean, feasible ΔP, ΔF, and $\Delta Relig$ for each district. Table 5 shows that this produces very similar results to our baseline, providing initial evidence that strategic redistricting is unlikely to explain our core effects.

To further clarify this point, we look not at the average but rather at random combinations of extreme feasible $\Delta diversity$. Specifically, we re-estimate our baseline specification randomly assigning ΔP (and associated ΔF and $\Delta Relig$) for each district to be either the maximum or minimum from the feasible set.[35] We repeat this randomized procedure 50,000 times at the original district level and plot the resulting estimates in Appendix Figure D.3.2. All the estimated effects of min or max feasible ΔP are greater than zero and smoothly distributed around the baseline estimated effect size. This is consistent with the

[34] Additionally, the key results for ΔF and ΔP hold up to inclusion of $\Delta G/N$ where G/N is the Greenberg-Gini diversity index scaled by population as in the structural equation of Esteban and Ray (2011a) tested by Esteban et al. (2012).

[35] That is, on any given draw, one district would receive its minimum ΔP while another district would receive its maximum ΔP , thereby shuffling randomly across all districts.

fact that within-district variation in feasible splits is small relative to the between-district variation.[36]

Together, these results suggest that unobserved heterogeneity in boundary choice does not explain the effects of $\Delta diversity$ on conflict. This further highlights the importance of the next two exercises that rule out confounding interactions with $post\text{-}split$. More substantively, these results demonstrate the importance of designing redistricting schemes that account for constraints on strategic border formation.

Accounting for Confounders. To rule out omitted factors correlated with $\Delta diversity$ and conflict, we follow a standard approach in heterogeneous DiD specifications and augment equation (2) with interactions of post-split and an array of potential confounders. Our approach is twofold. First, we separately consider groups of initial predetermined variables chosen based on intuition and prior work (see, e.g., Fearon and Laitin, 2003; Esteban et al., 2012). These include, among others: security presence; public goods access; remoteness, transportation infrastructure and access to markets; population size and age distributions; natural resource intensity; educational and occupational distributions; and topography, soil quality, and water access. Many are indeed highly correlated with diversity.[37] As shown in Appendix Tables D.2.1–D.2.4, some of these factors also mediate the effects of redistricting on conflict. However, the key coefficients of interest on ΔP and ΔF remain mostly unchanged across these different specifications. There are of course hundreds of potential confounding variables that one could combine in various ways in this type of exercise. With limited degrees of freedom, this leaves the door open to cherry-picking (Gelman and Loken, 2014).

Therefore, we adopt a second strategy that leverages machine learning to take a more disciplined approach to ruling out omitted variable bias. We utilize the Belloni et al. (2014) double-selection post-Lasso method to select a parsimonious set of influential controls from the large number of variables potentially confounding the relationship between $\Delta diversity$ and conflict. In practice, we expand upon the broad set of covariates noted above and use this approach to select additional interactions with $post\text{-}split$ while penalizing the tendency towards overfitting through a penalty parameter λ.[38] Acknowledging the limits of our natural policy experiment, we choose λ to ensure that the number of variables selected remain sufficiently smaller than the effective degrees of freedom. We set $\lambda = 3,000$ as a baseline and consider alternative values in Appendix D.2.

Panel A of Table 6 presents results based on this principled approach to variable selection. The main qualitative and quantitative findings remain unchanged. The point estimates on ΔP with this rich set of Lasso-selected controls are statistically indistinguishable from the baseline without controls except in column 3 where the estimates here are larger. Note that the fixed λ selects a different number of highly relevant controls across columns, which is due to variation in both the sample size as well as the relevant confounders of $\Delta diversity$. Overall, the conclusions remain unchanged for other reasonable values of λ.

Changes in Local Public Resources. While the confounding effects of other initial district characteristics is limited, it is also important to account for other factors associated with conflict that change with

[36] Appendix Figure D.3.1 further illustrates this point via examples showing relatively little overlap in the distributions of feasible ΔP across original districts. Indeed, original district fixed effects explain 89 percent of the variation in feasible ΔP.

[37] In a cross-section of 310 Indonesian districts, the full set of 65 covariates used in the Lasso procedure below explain 80 (51) percent of the variance in fractionalization (polarization).

[38] Appendix F details the full set of 65 potential confounding variables, and Appendix D.2 details our application of this method, which is particularly effective at dealing with the problem of overfitting in a setting where one aims to learn about a particular causal effect of interest rather than simply develop a good prediction of the outcome.

redistricting. We consider here two significant changes motivated by the conceptual framework, namely, changes in local government transfer revenue, a proxy for the per-capita prize, and changes in the proximity to government institutions stemming from the new capitals in child districts, a proxy for conflict costs. Panel B of Table 6 augments the baseline specification in Table 3 with these two measures (detailed in Appendix A).

We draw a few important takeaways from this exercise. First, the effects of $\Delta diversity$ remain similar, with changes in polarization significantly amplifying conflict. Second, both $\Delta transfers$ and $\Delta distance$ generally enter as expected. Greater transfers and greater proximity to the district capital are associated with more conflict after redistricting, particularly in child districts where these changes are much more pronounced. These findings are suggestive, but we do not push the interpretation too far given concerns about endogeneity and limited degrees of freedom.[39] Importantly, though, these results further establish that border-induced changes in ethnic divisions matter per se.

5.6 Additional Robustness Checks

This section describes additional robustness checks fully detailed in the Appendix. Overall, these tests lend further support to a causal interpretation of the main findings.

First, in Appendix D.5, we address the concern that SNPK coverage might be changing systematically with redistricting. The main concern is that newspapers differentially report on events in locations with greater changes in ethnic divisions, leading to overestimates of the actual effects. The comprehensive coverage of SNPK from many different outlets provides some reassurance. We provide further supportive evidence using auxiliary Google Trends data capturing the relative monthly frequency of search for the original, parent and child district names. While imperfect, this proxy reflects the frequency of general interest in the given location, some of which may be orthogonal to media reporting incentives. To the extent that Google trends are less prone to differential underreporting than the SNPK data, controlling for such trends should dampen the overall effects we estimate if reporting bias is substantial. Nevertheless, doing so leaves our key results unchanged.

In the Appendix, we address other concerns about the baseline specification. Appendices C.3 and C.4 demonstrate robustness across alternative groupings of violence categories and show that our effects are driven by events with injuries and property damage but not deaths per se. Appendix D.6 details a battery of additional checks. First, we consider the number of conflict events rather than a binary indicator, leaving the key takeaways mostly unchanged, though introducing more noise. Second, we omit districts that enter the SNPK in 2005, which is important given that these later entrants were selected on account of policy concerns about recent violence. Third, we separately exclude the regions of Aceh and Maluku, which experienced intense conflict in the late 1990s and early 2000s at the onset of decentralization. Fourth, we consider an alternative identification strategy that includes as additional control areas those nearby districts that have not undergone redistricting. Finally, we offer evidence on the external validity of our findings across Indonesia by reweighting the sample of districts to account for the likelihood of redistricting following the standard Horvitz and Thompson (1952) approach.

[39] Further exogeneity may be possible using transfer rules (Cassidy, 2017) and geographic constraints on capital locations (Campante et al., 2017). However, our goal here is to rule out confounders of $\Delta diversity$, and the OLS results should assuage first-order concerns.

6 Mechanisms: Ethnic Reconfiguration and Political Cycles of Violence

Our core results show that changes in ethnic divisions—arising purely from a reshuffling of *political* boundaries—can affect conflict within a matter of years. This suggests that the boundaries of political contests fundamentally shape the way that diverse groups interact and compete with one another. This section bolsters this interpretation by linking changes in violence to district politics and local elections. District governments play a large role in the local polity with mayors in control of vast public resources. Given the salience of ethnicity during mayoral elections, we investigate how border-induced changes in ethnic divisions affect violence around these political contests. We show that violence surges around new, closely contested mayoral elections, and that these political conflict cycles are amplified in high $\Delta diversity$ areas. Further, we provide evidence of ethnic favoritism in the allocation of public resources as a potential mechanism for generating grievances and amplifying incentives for ethnic mobilization around these majoritarian contests. We conclude by discussing a case of redistricting that nicely captures our broad empirical findings and clarifies the way that political boundaries shape ethnic divisions.

6.1 Diversity and Majoritarian Elections

We begin by establishing a few stylized facts, detailed in Appendix E, that help to characterize the context of mayoral elections. First, ethnic diversity, and particularly polarization, is associated with closer mayoral elections (i.e., lower victory margins). This is especially the case for newly created child districts and for both the first and the second quinquennial mayoral election (see Table E.1.1). Second, closer mayoral elections are associated with greater violence during the period around the election (see Table E.2.1). Third, according to the Indonesian Family Life Survey (IFLS) from 2014, individuals in districts that experienced greater border-induced changes in ethnic polarization are significantly more likely to report mayoral ethnicity and patronage as important factors influencing their vote (see Table E.1.2).[40] Consistent with the background in Section 2.3, these patterns suggest that new ethnic divisions may be particularly salient during mayoral elections, a hypothesis that we verify next.

6.2 New Ethnic Divisions and Electoral Violence

The empirical evidence thus far suggests redistricting may create strong incentives for group mobilization and violence. In settings with weak institutions, such violence may help shape the degree of control that one's group exerts over new institutions responsible for public goods or may help influence resource allocation after another group assumes control. Here, we provide empirical evidence of these mechanisms in the context of mayoral elections.

In Table 7, we augment the parent/child specifications in Table 3 with indicators for direct mayoral election periods. Since these elections are specific to the new parent or child district, we focus on specifications at that administrative level. These quinquennial elections, which began in 2005, vary in their timing (i) in parent districts, due to predetermined path dependence from Suharto-era election schedules (see Martinez-Bravo et al., 2017; Skoufias et al., 2014), and (ii) in child districts, due to the timing

[40]This is consistent with Tanasaldy (2012, p. 263) who notes that in several areas of West Kalimantan (in our study), "Due to ethnic polarization introduced in previous elections, masses from each ethnic group tended to rally for candidates from their own ethnic group."

of redistricting (with elections typically occurring 1.5–2.5 years after the split).[41] Following Harish and Toha (2017) and others cited therein, we define the election period as a six month window centered on the month of the election, but results are similar for other bandwidths. The coefficient on *post-split×first (second) election* identifies whether the incidence of conflict during the first (second) election deviates from the average incidence after redistricting. To see whether electoral violence cycles are more likely in newly polarized areas, we further interact post-redistricting election periods with $\Delta diversity$.

Looking across columns of Table 7, the likelihood of social conflict after redistricting is generally higher around the initial mayoral elections, particularly where redistricting sharply changed ethnic divisions. Column 1 shows this result pooling parent and child districts. Columns 2 and 3 show results separately for parent and child districts, pointing to important differences between pre-existing and new seats of government.

Child districts with ΔP one standard deviation above the mean exhibit nearly 70 percent more violence around the first election (comparing 0.027 to 0.041). This result lines up with the fact that polarization is associated with closer elections, and violence is significantly more pronounced during those close elections. Moreover, the amplifying effect of ΔP on political violence persists and may even be larger during the second election period 5–8 years after redistricting.

Meanwhile, in parent districts, initial elections are generally less violent. Border-induced changes in ethnic polarization seem less important than changes in fractionalization, which exerts a large albeit noisy positive amplification effect. The estimates may be consistent with the stronger correlation of F and $Relig$ than P with narrow victory margins in parent districts (see Appendix Table E.1.1).[42] Second election periods appear differentially less violent in high ΔP areas, which is unexpected and helps explain the null ΔP in that bottom section of column 1.

The differential effects of changing ethnic divisions on electoral violence in parent and child districts admit several possible interpretations. The relatively stronger effects of changes in ethnic polarization in child districts might be explained by the greater stakes of winning initial elections to control a completely new local government as opposed to winning control of an existing one. Another explanation for these diverging results could be differential institutional capacity. Parent districts may not only run more effective and safer initial elections than child districts but also learn more quickly how to manage the changing ethnopolitical divisions.

Overall, the results in Table 7 suggest that ethnic mobilization around mayoral elections are an important feature of conflict dynamics after redistricting. As in the baseline, these results are robust to controlling for confounding media attention using Google trends and to specifying the outcome as the intensive margin number of conflict events.

Furthermore, we show in Table 8 that the amplifying effects of $\Delta diversity$ on conflict around mayoral elections are a distinctive feature of high stakes majoritarian contests for political control. First, in columns 1–3, we find that the effects of $\Delta diversity$ on conflict do not systematically differ during the quinquennial parliamentary (and presidential) elections taking place in April 2004, 2009, and 2014.

[41] All direct elections in child districts follow splitting. Some original districts have a first direct election that precedes splitting, and hence we include a term for first direct elections in addition to the term *post-split×first election*. We observe a second quinquennial election for the three-quarters of new districts in existence long enough to hold a second round by 2014.

[42] And like child districts, more closely contested elections are associated with greater violence in parents (along the intensive margin, see Appendix Table E.2.1).

Although violence is generally more likely during such periods, $\Delta diversity$ exerts no significant amplifying effects as it does during the (staggered) mayoral elections. This difference is consistent with the fact that proportional representation legislative elections impart very different group-based incentives than the majoritarian mayoral elections (see Esteban and Ray, 2008b).[43] As a further validation test in columns 4–6, we show that there are no systematic differential effects of $\Delta diversity$ during the (lunar) holy month of Ramadan when conflict is generally lower but not differential with respect to $\Delta diversity$. Across columns, $\Delta diversity$ exhibits small, insignificant heterogeneous effects that are statistically different from the amplifying effects of $\Delta diversity$ around mayoral elections.

Finally, the changes in violence due to changing ethnic divisions are not merely a transitory phenomenon around elections. Across all results in Tables 7 and 8, the coefficient on $post\text{-}split \times \Delta P$ is statistically indistinguishable from that in Table 3, meaning that the post-split differences in the incidence of social conflict extend beyond election periods. This is what one would expect if grievances among losing group continued to manifest in violent acts protesting governance- and resource distribution-related issues in the future.[44] Next, we provide evidence for one potential reason for such grievances, namely ethnic favoritism in the allocation of public resources.

6.3 Ethnic Favoritism

Given hotly contested mayoral elections and their interplay with ethnic divisions, one would expect political favoritism towards co-ethnics. Recent studies document favoritism in resource allocation towards newly elected leaders' ethnic homelands in sub-Saharan Africa (see, e.g., Burgess et al., 2015; Hodler and Raschky, 2014). While a full accounting of this phenomenon in Indonesia is beyond the scope of this study, we present here evidence consistent with ethnic favoritism as a potential factor contributing to the patterns of discontent underlying the link between changes in ethnic divisions and conflict.

Following prior literature, we use nighttime light intensity as a proxy for local economic development and targeted public resources.[45] Importantly, electricity cannot be provided solely by villages and often requires support from higher levels of government. Mayors have been responsible for setting electricity policy within their districts since 2005 (Jayawardena, 2005). Moreover, mayors have significant discretionary power in their choice of how to allocate resources across villages (Aspinall and Asad, 2015; Pal and Wahhaj, 2016). Altogether, this suggests scope for rivalry across villages in access to this public resource, and indeed, publicly-provided electricity is a factor underlying some of the violent incidents reported in the SNPK data.[46]

[43] Fjelde and Höglund (2016) provide supportive evidence across African countries with different electoral systems.

[44] More fully disaggregating the timing around elections is instructive. Although somewhat imprecise, we can show that during the first election after redistricting, the amplifying effects of ΔP are stronger in the few months just after the actual election with weaker effects just before and during the election. This would be consistent with more limited use of group-based violence leading up to the election—perhaps due to lack of clarity about the newly salient and relevant group divisions—but an uptick associated with grievances thereafter. However, by the time of the second election, the amplifying effects of ΔP show up before, during, and immediately after the election, consistent with the various electoral violence typologies highlighted in Harish and Toha (2017) and discussed at length in Appendix C.2.

[45] Olivia and Gibson (2015) validate a strong correlation of light intensity with district-level output and expenditure data. In the predominantly rural areas of our study, nighttime lights tend to disproportionately capture public street lights. Moreover, electricity provision in Indonesia is almost exclusively concentrated in a single public utility company, which is often subject to the same sort of political manipulation identified in South Asia (see Baskaran et al., 2015; Min, 2015).

[46] For example, in Aceh Selatan district on 17 August 2007, "Hundreds of residents of Meukek subdistrict damaged PLN (Public Electricity Company) office and head of Subranting PLN's house by throwing stones. Citizens were induced to action by

We investigate ethnic favoritism by exploiting the fact that redistricting often changes either the identity or strength of the dominant group in the district. We examine changes in resources flowing to village v after redistricting as a function of that village's initial ethnic composition. Specifically, we ask whether the new boundaries imply that the village's initial population (N) share of ethnics from the new district's largest group is (i) larger, (ii) smaller, or (iii) the same as the village's initial share of ethnics from the original district's largest ethnic group. Formally, we compare $\frac{N_{e_{\mathcal{O}}v}}{N_v}$ to $\frac{N_{e_i v}}{N_v}$ where $e_{\mathcal{O}}$ is the largest ethnic group in original district $\mathcal{O}$, and e_i is the largest group in child or parent district i.[47]

For example, appealing to Figure 4, those Gayo majority villages in the new parent district will fall under (ii) since the Gayo are no longer the largest group. Meanwhile, the Gayo majority villages in the new child district of Gayo Lues will fall under (iii) as the largest group has not changed with redistricting. Note that, like most villages in category (iii), the size of their majority in Gayo Lues has increased, consistent with the general reduction in fractionalization that comes with redistricting.

In Table 9, we show how nighttime light intensity evolves across these three different types of villages after redistricting. We consider interactions of post-split with indicators for the change in alignment status, conditional on year and village fixed effects. Villages that lose their alignment with the largest ethnic group (ii) exhibit differentially lower light intensity after redistricting compared to those that either remain (iii) or become newly (i) aligned with the largest group. The results are consistent at both the parent and child level albeit slightly larger and more precisely estimated for the latter. Column 1 suggests that villages that become newly (remain) aligned with the largest group have 1.1 p.p. (2.8 p.p.) more village area with light coverage post-redistricting relative to those that lose their alignment. These are large differences relative to the mean of 16.3 percent of village area covered with any lights.

These results are suggestive of ethnic favoritism, which may be one vehicle for generating the sort of grievances that lead to persistently higher violence in areas where new boundaries create fresh ethnic divisions, even outside of election periods. We turn now to a case study highlighting some of the particular mechanisms underlying our empirical results.

6.4 Case Study

We discuss here an illustrative case study in West Kalimantan. This region has a significant history of ethnic strife including Dayak violence against the Chinese, and repeated clashes between Dayak and Madurese as well as Malay and Madurese. As in the rest of Indonesia, these large-scale open conflicts largely subsided in the early 2000s, replaced by more sporadic violence that remains a serious concern to policymakers (see Barron et al., 2016). We briefly present this interesting case below.

We focus on the original district of Sambas, which split into three separate districts: Sambas, Bengkayang (in 1999), and Kota Singkawang (from Bengkayang in 2001). Before splitting, Sambas was comprised of 52 percent Malay ethnics, 15 percent Dayak, and 15 percent Chinese, with other smaller groups including Javanese and Madurese. Redistricting significantly altered these group shares.

irritation due to irregular schedules of power outages. Other examples are provided in Appendix C.1. More recent examples, not (yet) in SNPK, provide further insight into the politicization of electricity provision with demonstrations against both PLN and the mayor (KabarNias, 2015) or being led by village heads against PLN (ProKal, 2016).

[47] Note that if the largest group does not change with splitting, the village will fall under category (iii). Nearly 70 percent of these villages are in districts where the share of the largest group increased. There are 1,764 villages in 32 districts in category (i), 558 villages in 28 districts in category (ii), and 12,182 villages in 119 districts in category (iii).

After the fall of Suharto, demands for decentralization and local empowerment spread across Indonesia, and West Kalimantan was no exception. As Tanasaldy (2012, p. 269) notes "In West Kalimantan such [native empowerment] movements were initially led by Dayaks who demanded more top jobs in the government and competed zealously against the Malays, for those political positions. Held in check during the authoritarian New Order [Suharto era], political polarization between the two ethnic groups was now unavoidable."

With these mounting tensions came a push for redistricting, motivated in part by a desire to reduce ethnic divisions. Tanasaldy notes that "government officials thought that separating conflict-prone areas and allowing the Dayaks to govern their own areas was a solution to chronic ethnic conflicts there." After the two splits, the now-parent district of Sambas was about 82 percent Malay, with the next largest group being ethnic Chinese at 8 percent. Bengkayang became 54 percent Dayak, 19 percent Malay, and 6 percent Chinese, while Singkawang was 37 percent Chinese, 20 percent Malay, 17 percent Dayak, and 10 percent Jawa. This split is often depicted as having cleanly separate the three groups: "the Malays in Sambas, the Chinese in Singkawang, and the Dayaks in Bengkayang" (Kobayashi, 2011, p. 374). However, as evident from the group shares, Singkawang remains far more polarized than Benkayang.

In Sambas and Bengkayang, the split was successful in reducing violence as the changing ethnic divisions translated into more amicable politics. Kobayashi nicely summarizes, noting that "district head elections became less tense because the Dayaks and Malays understood each others' rights to lead districts where they were dominant," and more generally that "interethnic strife to obtain political positions has declined since 2000." Government officials often argue that redistricting "contributed to prevention of ethnic violence," drawing connections, for example, between increased Dayak representation in the civil service and the end of violent street demonstrations.

Meanwhile, Singkawang presents an interesting contrast. Despite being a significant 40 percent plurality, the Chinese had generally stayed out of politics, with the "sons of the soil" Malay typically dominating. However, in 2007, the first direct mayoral election after redistricting brought their numeric advantage to the fore amidst a growing "desire within the Chinese community to increase the number of Chinese in the government and to elect a Chinese mayor" (Kobayashi, 2011, p. 295). In the 2007 election, much to the surprise of the Malay candidates, Hasan Karman, an ethnic Chinese beat the three Malay candidates (whose votes were split). Once in power, though, Karman "fumbled the delicate issue of ethnicity", as he "irked Malays by building a [Chinese] dragon statue... in the heart of the city" and "disparaged the Malay community... by [publicly] linking them to pirates" (Sukarsono, 2012).[48] In the 2012 election, tensions mounted amid Chinese accusations of intimidation and vote-tampering after the Malay candidate won. Violent clashes erupted between Malay security personnel and Chinese protestors outside the election commission office.[49]

Overall, West Kalimantan highlights both the promise and pitfalls of redistricting as a vehicle for reshaping ethnic divisions in society. Greater homogeneity in Sambas and Bengkayang may have helped to resolve some of the longstanding interethnic grievances. However, the new district of Singkawang

[48] These violent incidents are reported in the SNPK data with event details such as "series of arson cases by unknown perpetrators believed to be related to ethnical issues" in May 2010, and "there was a clash in the parade of Singkawang Parliament, the village of Pasiran, the city of Singkawang. Clash involving two groups of the pro and contra period of the construction of a dragon statue that will be built at the crossroads (2 injured)" in July 2010.

[49] This event is reported in the SNPK data on October 1st, 2012 as "a clash between the masses and the police when the Mayoral Candidate campaign handed over evidence of more than 3,000 people being denied the vote. [2 injured]"

gave rise to fresh grievances and cycles of violence by invigorating hitherto less salient ethnic divisions. These cases highlight the tradeoffs of redistricting in diverse societies where creating purely homogeneous political units is not feasible everywhere.

7 Discussion

This paper identified the causal effect of political boundaries on ethnic divisions. We showed how redrawing subnational boundaries can alter the salience of different ethnic cleavages in society, and, in turn, affect conflict. By bringing the government closer to the governed, redistricting holds promise for increasing social stability. However, this common policy reform is not without pitfalls. Our natural experiment showed in particular that fresh cycles of violence may erupt when new borders increase ethnic polarization. We argued further that electoral democracy may amplify the underlying incentives for group mobilization that often lead to violence. Overall, our findings provide novel evidence on the interlinkages between ethnic and political divisions. These results help inform ongoing debates on the causes of violence as well as policy efforts aimed at curbing it.

Some of our findings suggest that border-induced changes in ethnic divisions may have persistent effects on conflict. This persistence can be interpreted through the lens of models like Rohner et al. (2013b), which feature vicious cycles of inter-group violence and erosion of trust. Political violence in newly created districts may be particularly prone to such dynamics as seen, for example, in the case of Singkawang. While over the long-run redistricting may foster new interethnic interactions, learning and cooperation, it is important to understand and prepare for the scope for violence during the transition.

Local government proliferation is a pervasive feature of decentralization today. The widespread prevalence of ethnic mobilization (Fearon, 2006) and favoritism (De Luca et al., 2017) suggest that similar conflict dynamics could play out in other diverse countries. For example, Green (2010) discusses some of the same unintended consequences of redistricting in Uganda that we identify empirically in Indonesia. Nevertheless, we acknowledge that ours is only a partial analysis of the vast political and economic implications of redistricting.

We see four important directions for future research on redistricting in Indonesia and elsewhere. First, a small but growing literature highlights the importance of *within*-ethnolinguistic or -religious group heterogeneity in culture (Desmet et al., 2017), genes (Arbatli et al., 2015), or income (Esteban and Ray, 2008a; Mitra and Ray, 2014) in shaping conflict. This is an interesting question in the context of decentralization and one that can be explored using heterogeneity in responses to household survey questions on preferences, variation in vote shares for different parties of the same religion, and within-group educational or occupational inequality.

A second question is whether redistricting can be a vehicle for a central government to constrain secessionist tendencies. Coming on the heels of East Timor's independence and concerns about breakaway regions in Aceh and Papua, Indonesian policymakers in the late 1990s strategically chose districts rather than provinces as the primary administrative units allowed to proliferate. According to observers like Booth (2011), their goal, among others, was to fracture the strength of broader regional identities. It would be interesting to explore whether this policy of "breaking up to stay together" stifles secessionist sentiments and ultimately shifts violence from higher to lower levels.

Third and relatedly, redistricting has the potential to activate more granular cultural distinctions. Although many areas created ethnically homogeneous districts, sub-ethnic distinctions may have emerged over time as groups sought new vehicles for political mobilization. Redistricting may have contributed to the dramatic growth in the number of self-reported ethnic identities from 1,087 in the 2000 Census round to 1,331 in the 2010 round. As recounted in Fearon (2006), Horowitz (1985, p. 66) provides a telling example of this burgeoning of local identity in the Indian context of state proliferation in the 1950s: In Madras state, "... with large Tamil and Telugu populations, cleavages within the Telugu group were not very important. As soon as a separate Telugu-speaking state was carved out of Madras, however, Telugu subgroups—caste, regional, and religious—quickly formed the bases of political action." In the Indonesian context, one could explore empirically how political boundaries lead to new forms of identity related not only to sub-ethnic distinctions but also to shared national identity (Bazzi et al., 2018).

Finally, there are several open questions about the public goods and welfare consequences of redistricting. Recent studies identify environmental externalities (Burgess et al., 2012; Lipscomb and Mobarak, forthcoming). There are other interesting implications of reduced government scale and changes in the network of administrative responsibilities; not to mention increased proximity to service providers in the new district capitals. A full account of the welfare implications of redistricting clearly extends beyond the effects of changing ethnic divisions.

References

Alesina, A. and E. Zhuravskaya, "Segregation and the Quality of Government in a Cross Section of Countries," *American Economic Review*, 2011, *101* (5), 1872–1911.

— , R. Baqir, and W. Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 1999, *114* (4), 1243–1284.

— , W. Easterly, and J. Matuszeski, "Artificial states," *Journal of the European Economic Association*, 2011, *9* (2), 246–277.

Allen, N. W., "From patronage machine to partisan melee: Subnational corruption and the evolution of the Indonesian party system," *Pacific Affairs*, 2014, *87* (2), 221–245.

Amodio, Francesco and Giorgio Chiovelli, "Ethnicity and violence during democratic transitions: Evidence from south africa," *Journal of the European Economic Association*, 2017, *16* (4), 1234–1280.

Ananta, A., E. N. Arifin, M. S. Hasbullah, N. B. Handayani, and A. Pramono, *Demography of Indonesia's ethnicity*, Institute of Southeast Asian Studies, 2015.

Arbatli, C. E., Q. H. Ashraf, and O. Galor, "The Nature of Conflict,", NBER Working Paper No. 21079 2015.

Aspinall, E., "Democratization and Ethnic Politics in Indonesia: Nine Theses," *Journal of East Asian Studies*, 2011, *11* (2), 289–319.

— , "A nation in fragments: Patronage and neoliberalism in contemporary Indonesia," *Critical Asian Studies*, 2013, *45* (1), 27–54.

— and M. Sukmajati, eds, *Electoral Dynamics in Indonesia: Money Politics, Patronage and Clientelism at the Grassroots*, NUS Press, 2016.

— and M. U. Asad, "The Patronage Patchwork," *Bijdragen tot de taal-, land-en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia*, 2015, *171* (2-3), 165–195.

— and — , "Understanding family politics: Successes and failures of political dynasties in regional Indonesia," *South East Asia Research*, 2016, *24* (3), 420–435.

BAPPENAS, *Evaluasi Pemekaran Daerah*, Bappenas, 2007.

Barron, P., A. Engvall, and A. Morel, "Understanding Violence In Southeast Asia: The Contribution of Violent Incidents Monitoring Systems," *The Asia Foundation Working Paper*, 2016.

— , S. Jaffrey, and A. Varshney, "Understanding violent conflict in Indonesia: a mixed methods approach,", World Bank Social Development Working Papers No. 117 2009.

— , — , and — , "How Large Conflicts Subside: Evidence From Indonesia,", World Bank Indonesia Social Development Paper 18 2014.

Baskaran, T., B. Min, and Y. Uppal, "Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections," *Journal of Public Economics*, 2015, *126*, 64–73.

Bazzi, S., A. Gaduh, A. Rothenberg, and M. Wong, "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, September 2016, *106* (9), 2658–98.

— , — , — , and — , "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *CEPR Discussion Paper No. 12377*, 2018.

— and C. Blattman, "Economic shocks and conflict: Evidence from commodity prices," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 1–38.

Belloni, A., V. Chernozhukov, and C. Hansen, "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 2014, *81* (2), 608–650.

Belsley, David A, Edwin Kuh, and Roy E Welsch, *Regression diagnostics: Identifying influential data and sources of collinearity*, Vol. 571, John Wiley & Sons, 2005.

Berge, L. I. O., K. Bjorvatn, S. Galle, E. Miguel, D. N. Posner, B. Tungodden, and K. Zhang, "How Strong are Ethnic Preferences?,", NBER Working Paper No. 21715 2015.

Bertrand, M., E. Duflo, and S. Mullainathan, "How much should we trust differences-in-differences estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Blouin, A. and S. Mukand, "Erasing Ethnicity? Nation Building, (Mis)Trust and the Salience of Identity in Rwanda," *Journal of Political Economy*, forthcoming.

Bollen, K. A. and R. W. Jackman, "Regression diagnostics: An expository treatment of outliers and influential cases," *Modern methods of data analysis*, 1990, pp. 257–291.

Booth, A., "Splitting, splitting and splitting again: a brief history of the development of regional government in Indonesia since independence," *Bijdragen tot de Taal-, Land-en Volkenkunde*, 2011, pp. 31–59.

Burbidge, J. B., L. Magee, and A. L. Robb, "Alternative transformations to handle extreme values of the dependent variable," *Journal of the American Statistical Association*, 1988, *83* (401), 123–127.

Burgess, R., M. Hansen, B. A. Olken, P. Potapov, and S. Sieber, "The Political Economy of Deforestation in the Tropics," *The Quarterly Journal of Economics*, 2012, *127* (4), 1707–1754.

—, R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel, "The Value of Democracy: Evidence from Road Building in Kenya," *American Economic Review*, 2015, *105* (6), 1817–1851.

Butcher, C. and B. E. Goldsmith, "Elections, Ethnicity, and Political Instability," *Comparative Political Studies*, 2017, *50* (10), 1390–1419.

Cameron, A. C., J. B. Gelbach, and D. L. Miller, "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Campante, F., Q.A. Do, and B. Guimaraes, "Capital Cities, Conflict, and Misgovernance," Working Paper, Harvard Kennedy School 2017.

Caselli, F. and W. J. Coleman, "On the theory of ethnic conflict," *Journal of the European Economic Association*, 2013, *11* (s1), 161–192.

—, M. Morelli, and D. Rohner, "The Geography of Interstate Resource Wars," *The Quarterly Journal of Economics*, 2015, *130* (1), 267–315.

Cassidy, T., "Local Government Responses to Permanent vs. Transitory Grant Revenue: Evidence from Indonesia," Unpublished Manuscript, University of Michigan 2017.

Charras, M., "The reshaping of the Indonesian archipelago after 50 years of regional imbalance," in M. Erb, C. Faucher, and P. Sulistiyanto, eds., *Regionalism in Post-Suharto Indonesia*, Vol. 4, Routledge, 2005, pp. 91–114.

Conley, T. G., "GMM estimation with cross sectional dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

Cox, D. R., "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1972, *34* (2), 87–22.

Cribb, R., *Historical atlas of Indonesia*, Routledge, 2013.

Dercon, S. and R. Gutiérrez-Romero, "Triggers and characteristics of the 2007 Kenyan electoral violence," *World Development*, 2012, *40* (4), 731–744.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg, "The political economy of linguistic cleavages," *Journal of Development Economics*, 2012, *97* (2), 322–338.

—, —, and —, "Culture, ethnicity, and diversity," *American Economic Review*, 2017, *107* (9), 2479–2513.

Desmet, K, I. Ortuño-Ortín, and S. Weber, "Linguistic Diversity and Redistribution," *Journal of the European Economic Association*, 2009, *7* (6), 1291–1318.

Desmet, K., J. F. Gomes, and I. Ortuño-Ortín, "The geography of linguistic diversity and the provision of public goods," Technical Report, CESifo Working Paper Series 6238 2016.

Diprose, R., "Decentralization, horizontal inequalities and conflict management in Indonesia," *Ethnopolitics*, 2009, *8* (1), 107–134.

Dower, P. C., V. Ginsburgh, and S. Weber, "Colonial legacy, polarization and linguistic disenfranchisement: The case of the Sri Lankan War," *Journal of Development Economics*, 2017, *127*, 440–448.

Dube, O. and J. F. Vargas, "Commodity price shocks and civil conflict: Evidence from Colombia," *The Review of Economic Studies*, 2013, *80* (4), 1384–1421.

—, O. García-Ponce, and K. Thom, "From maize to haze: Agricultural shocks and the growth of the mexican drug sector," *Journal of the European Economic Association*, 2016, *14* (5), 1181–1224.

Eifert, B., E. Miguel, and D. N. Posner, "Political competition and ethnic identification in Africa," *American Journal of Political Science*, 2010, *54* (2), 494–510.

Esteban, J. and D. Ray, "On the salience of ethnic conflict," *American Economic Review*, 2008, *98* (5), 2185–2202.

_ and _ , "Polarization, Fractionalization and Conflict," *Journal of Peace Research*, 2008, *45* (2), 163–182.

_ and _ , "Linking conflict to inequality and polarization," *American Economic Review*, 2011a, *101* (4), 1345–1374.

_ and _ , "A model of ethnic conflict," *Journal of the European Economic Association*, 2011b, *9* (3), 496–521.

_ , L. Mayoral, and D. Ray, "Ethnicity and conflict: An empirical Study," *American Economic Review*, 2012, *102* (4), 1310–1342.

Esteban, J.M. and D. Ray, "On the measurement of polarization," *Econometrica*, 1994, pp. 819–851.

Fearon, J. and A. Hoeffler, "Benefits and Costs of the Conflict and Violence Targets for the Post-2015 Development Agenda," *Conflict and violence assessment paper, Copenhagen Consensus Center*, 2014.

Fearon, J. D., "Ethnic mobilization and ethnic violence," *The Oxford Handbook of Political Economy*, 2006, pp. 852–868.

_ and D. D. Laitin, "Ethnicity, insurgency, and civil war," *American Political Science Review*, 2003, *97* (01), 75–90.

_ and _ , "Sons of the soil, migrants, and civil war," *World Development*, 2011, *39* (2), 199–211.

Fitrani, F., B. Hofman, and K. Kaiser, "Unity in diversity? The creation of new local governments in a decentralising Indonesia," *Bulletin of Indonesian Economic Studies*, 2005, *41* (1), 57–79.

Fjelde, H. and K. Höglund, "Electoral institutions and electoral violence in Sub-Saharan Africa," *British Journal of Political Science*, 2016, *46* (2), 297–320.

Fox, C. and J. Menchik, "The Politics of Identity in Indonesia: Results from Campaign Advertisements," *APSA 2011 Annual Meeting Paper*, 2011.

Gelman, A. and E. Loken, "The Statistical Crisis in Science Data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up.," *American Scientist*, 2014, *102* (6), 460.

Gershman, B. and D. Rivera, "Subnational Diversity in Sub-Saharan Africa: Insights from a New Dataset," *Journal of Development Economics*, 2016, *133*, 231–263.

Green, E., "Patronage, district creation, and reform in Uganda," *Studies in Comparative International Development*, 2010, *45* (1), 83–103.

Green, E. D., "Decentralisation and conflict in Uganda," *Conflict, Security & Development*, 2008, *8* (4), 427–450.

Grossman, G. and J. Lewis, "Administrative Unit Proliferation," *American Political Science Review*, 2014, *108* (01), 196–217.

_ , J. H. Pierskalla, and E. B. Dean, "Government fragmentation and public goods provision," *The Journal of Politics*, 2017, *79* (3), 823–840.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein, "Why does ethnic diversity undermine public goods provision?," *American Political Science Review*, 2007, *101* (04), 709–725.

Harish, S. P. and R. Toha, "A New Typology of Electoral Violence: Insights from Indonesia," *Terrorism and Political Violence*, 2017, pp. 1–25.

Henderson, J. V., A. Storeygard, and D. N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 2012, *102* (2), 994–1028.

Hillman, B., "Ethnic politics and local political parties in Indonesia," *Asian Ethnicity*, 2012, *13* (4), 419–440.

Hjort, J., "Ethnic divisions and production in firms," *Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.

Hodler, R. and P. A. Raschky, "Regional Favoritism," *Quarterly Journal of Economics*, 2014, *129* (2), 995–1033.

Horowitz, D. L., *Ethnic groups in conflict*, University of California Press, 1985.

Horvitz, D. G. and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *Journal of the American statistical Association*, 1952, *47* (260), 663–685.

Huber, P. J., "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *The Annals of Statistics*, 1973, *1* (5), 799–821.

International Crisis Group, "Decentralisation and Conflict in Indonesia: The Mamasa Case," *Asia Briefing No. 37*, 2005.

Jayawardena, M., "Electricity for all: options for increasing access in Indonesia," Working Paper, Asia Sustainable and Alternative Energy Program, The World Bank 2005.

Jr., R. G. Fryer and R. Holden, "Measuring the Compactness of Political Districting Plans," *Journal of Law and Economics*, 2011, *54* (3), 493–535.

KabarNias, "Rekening Listrik Melonjak, Masyarakat Demo PLN," http://kabarnias.com/kabar-dari-nias/nias/rekening-listrik-melonjak-masyarakat-demo-pln-1979 2015. Accessed: 2017-11-30.

Kobayashi, H., "Civic Associations, Local Governance and Conflict Prevention in Indonesia." PhD dissertation, University of Southern California 2011.

Lewis, B., "Political Fragmentation in Indonesia: Pork Barrel Spending, Service Delivery, and Corruption," Working Paper, Crawford School of Public Policy Australian National University 2017.

Lewis, B. D., "Does local government proliferation improve public service delivery? Evidence from Indonesia," *Journal of Urban Affairs*, 2017, *39* (8), 1047–1065.

Lipscomb, M. and A. M. Mobarak, "Decentralization and the Political Economy of Water Pollution: Evidence from the Re-drawing of County Borders in Brazil," *Review of Economics Studies*, forthcoming.

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel, "Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test (IAT)," *American Economic Review*, 2015, *105* (5), 340–345.

Luca, Giacomo De, Roland Hodler, Paul A Raschky, and Michele Valsecchi, "Ethnic Favoritism: An Axiom of Politics?," *Journal of Development Economics*, 2017, *132*, 115–129.

MacKinnon, J. G. and M. D. Webb, "Randomization inference for difference-in-differences for few treated clusters," Technical Report 2018.

_ and _ , "The wild bootstrap for few (treated) clusters," *The Econometrics Journal*, 2018, *21* (2), 114–135.

_ and _ , "Wild bootstrap randomization inference for few treated clusters," Technical Report 2018.

Martinez-Bravo, M., P. Mukherjee, and A. Stegmann, "An Empirical Investigation of the Legacies of Non-Democratic Regimes: The Case of Soeharto's Mayors in Indonesia," *Econometrica*, 2017, *85* (6), 1991–2010.

Martinez-Bravo, Monica, "The role of local officials in new democracies: Evidence from Indonesia," *American Economic Review*, 2014, *104* (4), 1244–87.

Mavridis, D., "Ethnic diversity and social capital in Indonesia," *World Development*, 2015, *67*, 376–395.

McKenzie, David, "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 2012, *99* (2), 210–221.

Michalopoulos, S. and E. Papaioannou, "National institutions and subnational development in Africa," *Quarterly Journal of Economics*, 2013, *129* (1), 151–213.

_ and _ , "The Long-Run Effects of the Scramble for Africa," *American Economic Review*, 2016, *106* (7), 1802–48.

_ and _ , "Spatial Patterns of Development: A Meso Approach," *Annual Review of Economics*, 2017.

Miguel, E., S. Satyanath, and E. Sergenti, "Economic shocks and civil conflict: An instrumental variables approach," *Journal of Political Economy*, 2004, *112* (4), 725–753.

Min, B., *Power and the Vote: Elections and Electricity in the Developing World*, Cambridge University Press, 2015.

Mitra, A. and D. Ray, "Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India," *Journal of Political Economy*, 2014, *122* (4), 719–765.

Montalvo, J. G. and M. Reynal-Querol, "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 2005, *95* (3), 796–816.

— **and** — , "Ethnic diversity and growth: Revisiting the evidence," Unpublished Manuscript, Universitat Pompeu Fabra 2017.

Nolan, C., S. Jones, and Solahudin, "The Political Impact of Carving Up Papua," in H. Hill, ed., *Regional dynamics in a decentralized Indonesia*, Singapore: Institute of Southeast Asian Studies, 2014, chapter 17, pp. 409–433.

Olivia, S. and J. Gibson, "Value Added from Space? Empirical Evidence from Indonesia," Unpublished Manuscript, University of Waikato 2015.

Pal, S. and Z. Wahhaj, "Fiscal decentralisation, local institutions and public good provision: evidence from Indonesia," *Journal of Comparative Economics*, 2016.

Pierskalla, J. H., "The Proliferation of Decentralized Governing Units," Unpublished Manuscript, Ohio State University 2016.

— , "Splitting the Difference? The Politics of District Creation in Indonesia," *Comparative Politics*, 2016, *48* (2), 249–268.

Posner, D. N., "The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi," *American Political Science Review*, 2004, *98* (04), 529–545.

— , "When and why do some social cleavages become politically salient rather than others?," *Ethnic and Racial Studies*, 2017, *40* (12), 2001–2019.

ProKal, "Kepala Desa se Kabupaten Tanah Bumbu Demo PLN," http://kalsel.prokal.co/read/news/1911-kepala-desa-se-kabupaten-tanah-bumbu-demo-pln 2016. Accessed: 2017-11-30.

Ray, D. and J. Esteban, "Development and Conflict," *Annual Review of Economics*, 2016, *8* (1).

Rohner, D., M. Thoenig, and F. Zilibotti, "Seeds of distrust: Conflict in Uganda," *Journal of Economic Growth*, 2013, *18* (3), 217–252.

— **,** — **, and** — , "War signals: A theory of trade, trust, and conflict," *Review of Economic Studies*, 2013, *80* (3), 1114–1147.

Skoufias, E, A. Narayan, B. Dasgupta, and K. Kaiser, "Electoral Accountability and Local Government Spending in Indonesia," *World Bank Policy Research Working Paper 6782*, 2014.

Spolaore, E. and R. Wacziarg, "War and Relatedness," *Review of Economics and Statistics*, 2016, *98* (5).

Spolaore, Enrico and Romain Wacziarg, "The Political Economy of Heterogeneity and Conflict," Technical Report, NBER Working Paper No. 23278 2017.

Sukarsono, A., "Showing True Colours: Identity Politics in Indonesia's Local Elections," https://www.crisisgroup.org/asia/south-east-asia/indonesia/showing-true-colours-identity-politics-indonesias-local-elections 2012. Accessed: 2017-09-30.

Sundberg, R. and E. Melander, "Introducing the UCDP georeferenced event dataset," *Journal of Peace Research*, 2013, *50* (4), 523–532.

Tadjoeddin, M. Z., "Electoral conflict and the maturity of local democracy in Indonesia: testing the modernisation hypothesis," *Journal of the Asia Pacific Economy*, 2012, *17* (3), 476–497.

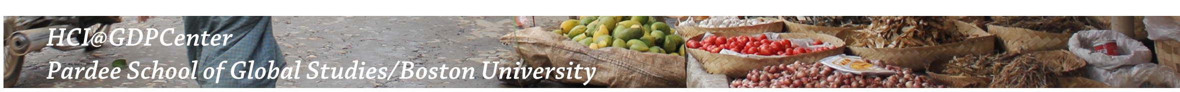
Tanasaldy, T., *Regime change and ethnic politics in Indonesia; Dayak politics of West Kalimantan*, Brill, 2012.

Weiner, M., *Sons of the soil: Migration and ethnic conflict in India*, Princeton University Press, 1978.

Wilson, I. D., *The politics of protection rackets in post-New Order Indonesia: Coercive capital, authority and street politics*, Vol. 47, Routledge, 2015.

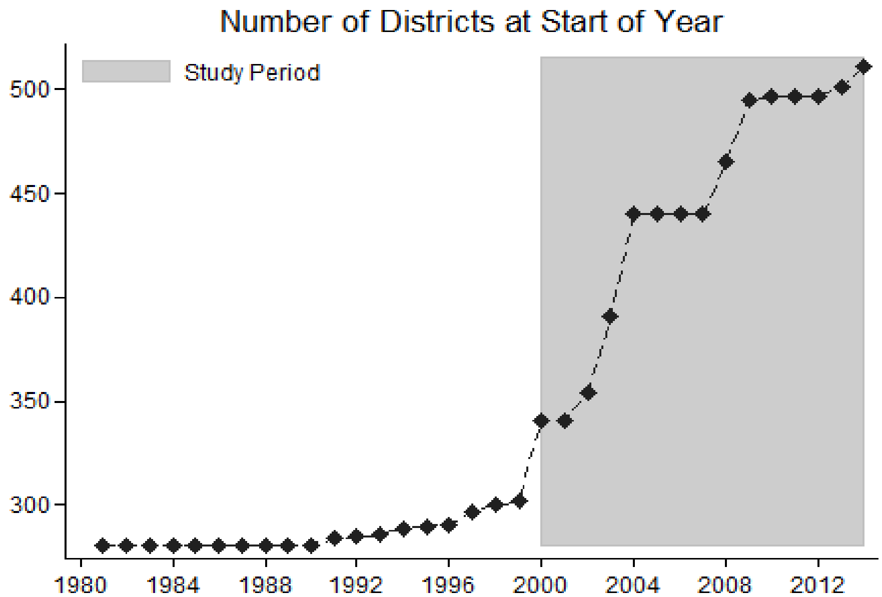
Yanagizawa-Drott, D., "Propaganda and conflict: Evidence from the Rwandan genocide," *Quarterly Journal of Economics*, 2014, *129* (4), 1947–1994.

Young, A., "Improved, nearly exact, statistical inference with robust and clustered covariance matrices using effective degrees of freedom corrections," Unpublished Manuscript, London School of Economics 2016.



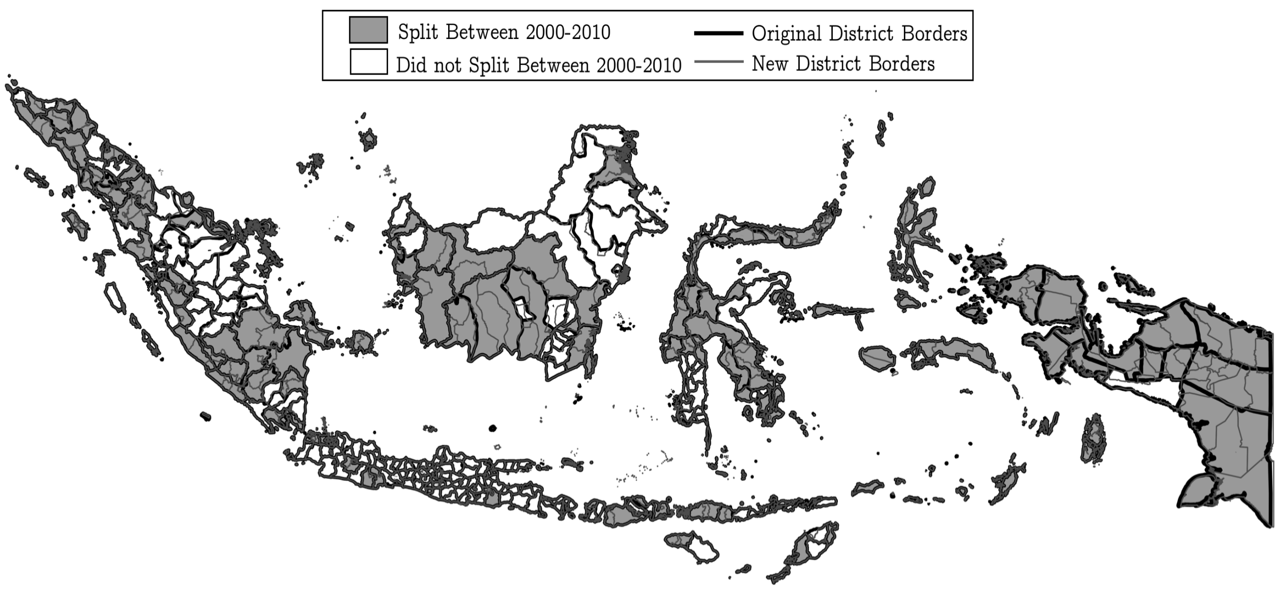
Figures

Figure 1: Indonesia's Remarkable Wave of Redistricting



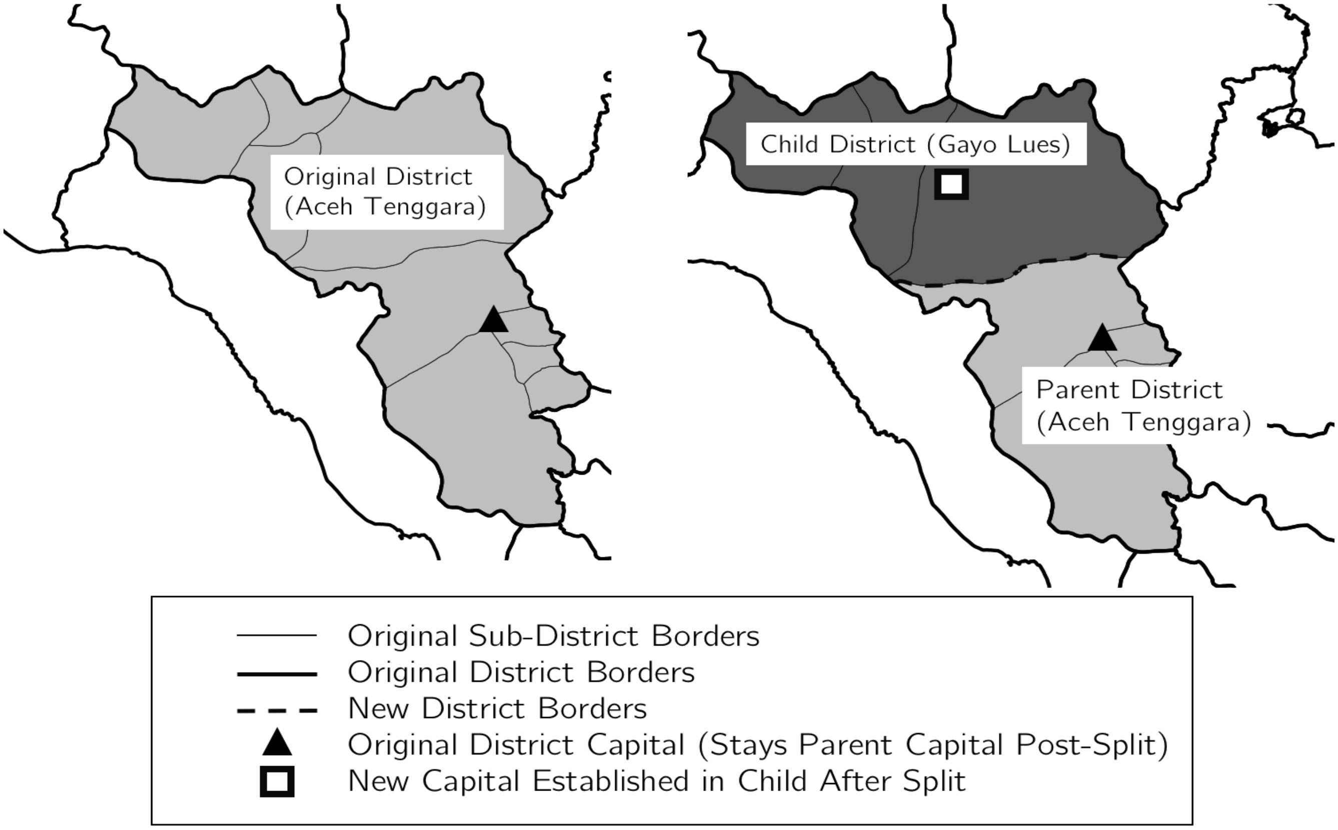
Notes: This figure captures the evolution of new districts across Indonesia from 1980–2014 based on the month each district was passed into law.

Figure 2: Redistricting across the Country



Notes: This map plots the original and new district borders based on district-level shapefiles for 2000 and 2010.

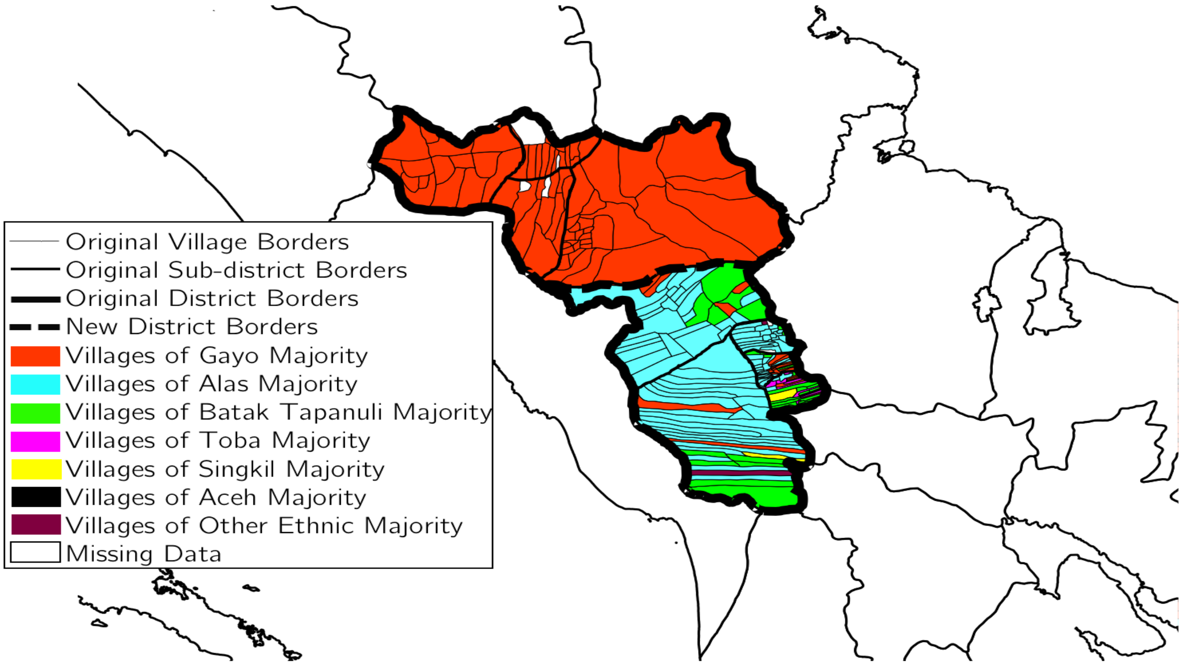
Figure 3: Example of Redistricting into Parent and Child Districts



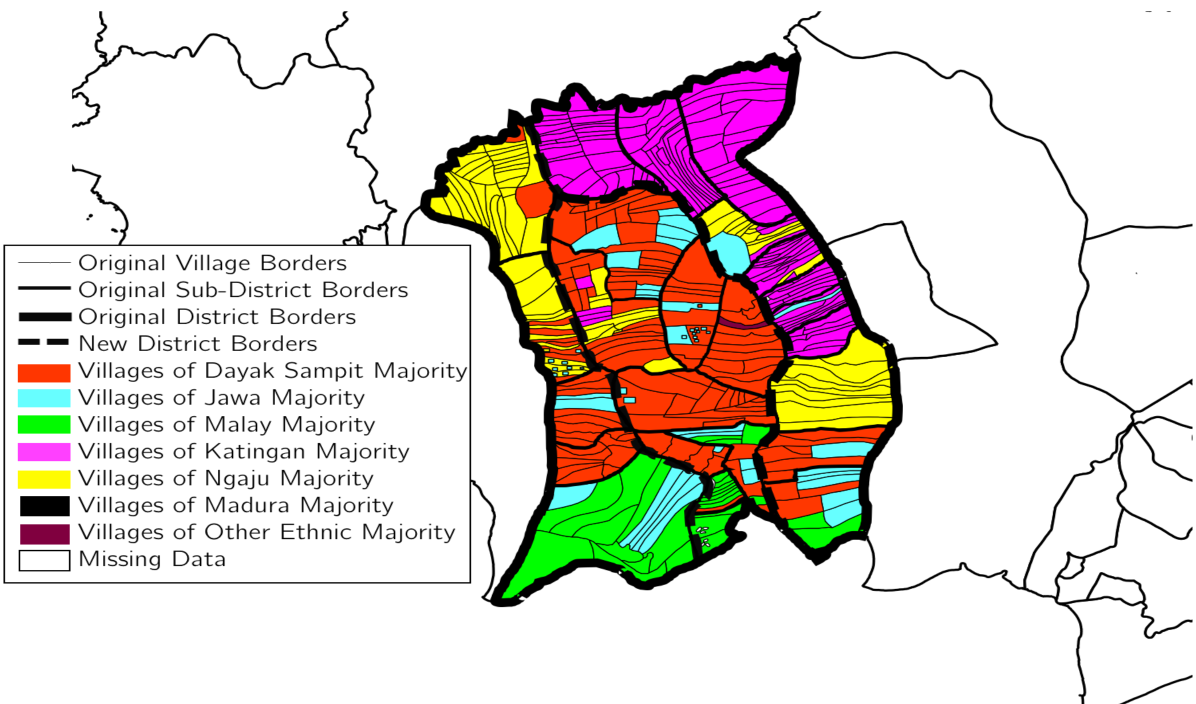
Notes: This figure provides an example of the redistricting process and our nomenclature for the different administrative divisions.

Figure 4: Examples of Border-Induced Δ Diversity

(a) *Homogenizing Case*: Aceh Tenggara District



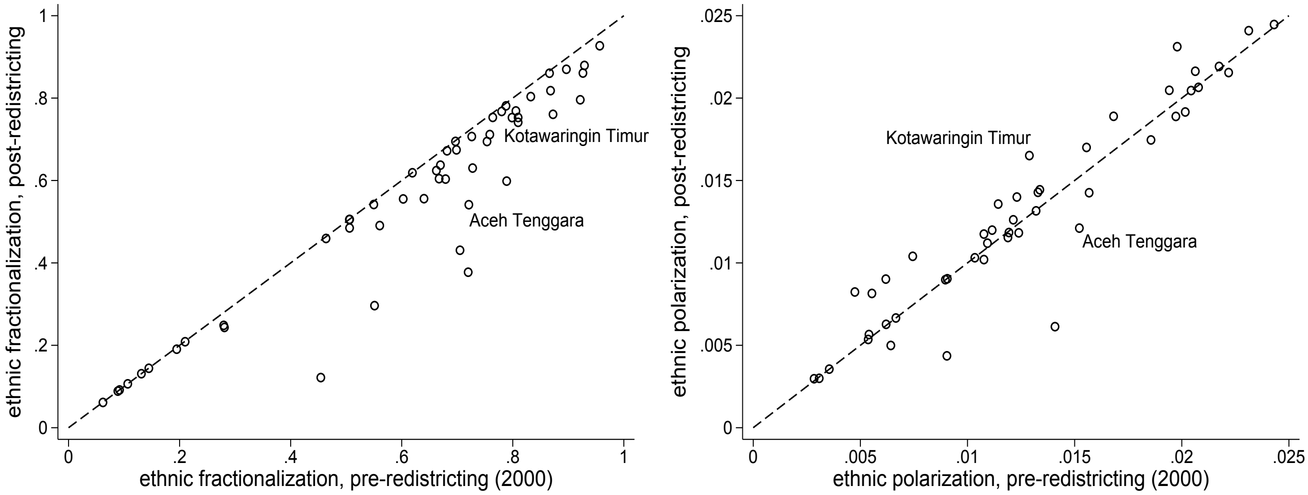
(b) *Newly Salient Divisions*: Kotawaringin Timur District



Notes: This figure provides two examples of the types of redistricting patterns that we see in our data in terms of changes in policy-relevant ethnic diversity. Figure (a) shows the original district of Aceh Tenggara as in Figure 3, and Figure (b) shows the original district of Kotawaringin Timur, which splits into two child districts, Seruyan on the left and Katingan on the right with the parent district in the middle. In both figures, we color code the villages based on the majority ethnic group in the village in the 2000 Population Census with red indicating the largest group in the original district as a whole, aqua the second largest, and so on, with a few villages in white with missing data or inability to match with shapefiles.

Figure 5: Diversity Before and After Redistricting

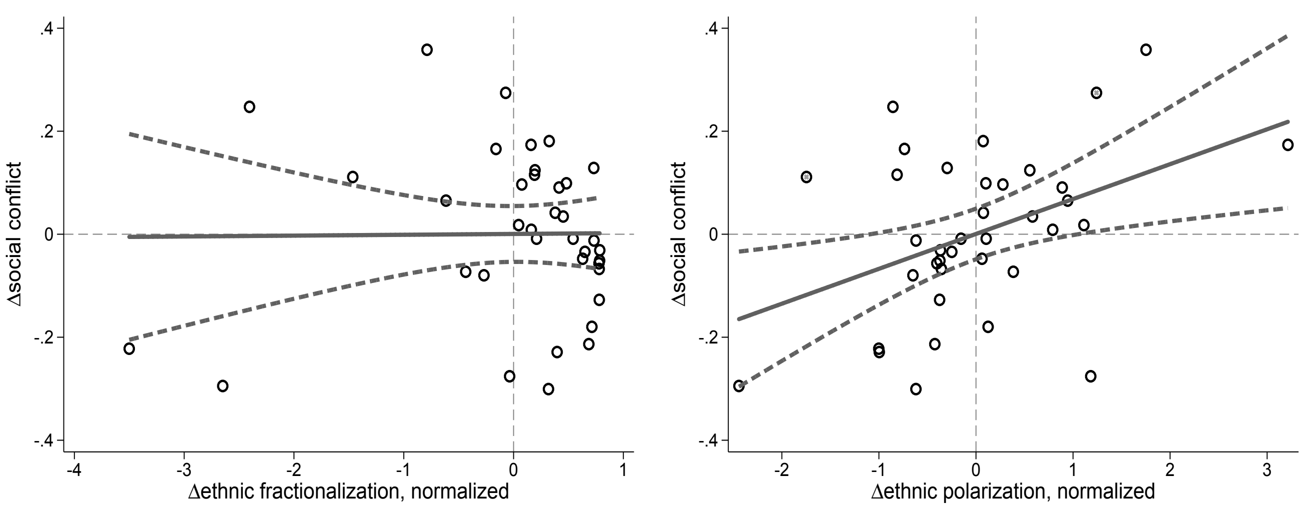
(a) Fractionalization (b) Polarization



Notes: These figures plot diversity at the original district level baseline boundaries in 2000 (x-axis) against the 2010 boundaries after redistricting (y-axis). The latter measure is the population-weighted average of diversity in the new parent and child districts, but again based on the population residing in those areas at baseline in 2000. The dashed 45 degree line indicates the locus of points along which the new borders imply no change in diversity, and the vertical distance between each point and the line captures our $\Delta diversity$ measure. For presentational purposes, we omit the top 4 districts with baseline polarization >0.025, though they are of course included in all regression analysis. In both graphs, we identify the two example districts seen in Figure 4.

Figure 6: Simple Difference-in-Difference: Δ Conflict against Δ Diversity

(a) ΔFractionalization (b) ΔPolarization



Notes: This figure presents a simple difference-in-difference regression relating $\Delta diversity$ to changes in social conflict captured by the difference in the mean monthly likelihood of any social violence before versus after redistricting. The results are restricted to those 38 original districts with conflict data pre- and post-redistricting. Graph (a) shows results for Δfractionalization and (b) for Δpolarization. The thick dashed lines are robust 95 percent confidence intervals. The point estimate in (a) is 0.002 (0.050) and in (b) is 0.068 (0.029)** with robust standard and the HC3 degrees-of-freedom adjustment. We omit one extreme outlier (with ΔP six standard deviations below the mean) based on standard outlier detection methods. All districts, including this outlier, are retained in the main generalized DiD estimates in the paper, and Appendix D.4 presents a battery of small-sample robustness and inference procedures.

Tables

Table 1: Plausibly Exogenous Timing of Redistricting

	Dependent Variable:	
	no. months until split	1(post-moratorium split)
	mean: 53	mean: 0.31
Panel A: Diversity	Standardized Coefficient	
original district ethnic fractionalization	2.183	0.011
	(4.267)	(0.071)
original district Δ ethnic fractionalization	2.328	0.033
	(3.417)	(0.061)
child district ethnic fractionalization	2.641	0.009
	(3.984)	(0.069)
parent district ethnic fractionalization	4.416	0.051
	(3.962)	(0.066)
original district ethnic polarization	-1.829	-0.006
	(2.882)	(0.048)
original district Δ ethnic polarization	2.168	0.039
	(1.880)	(0.026)
child district ethnic polarization	-1.545	-0.002
	(3.122)	(0.045)
parent district ethnic polarization	1.412	0.045
	(4.002)	(0.071)
original district religious polarization	1.530	-0.024
	(3.343)	(0.060)
original district Δ religious polarization	-1.461	0.022
	(2.276)	(0.033)
child district religious polarization	-1.107	-0.063
	(3.821)	(0.060)
parent district religious polarization	2.653	0.013
	(3.749)	(0.063)
Panel B: 65 Potential Confounders (see Appendix D.2)		
mean standardized coefficient	0.096	-0.005
actual number of significant predictors at 5% level	4	3
expected number of significant predictors at 5% level by chance	3.3	3.3
Panel C: Changes in Prior Conflict	Survival Analysis Months-to-Split Cox Proportional Hazard Ratio [p-value]	
any social conflict	1.180 [0.789]	

Notes: Each cell is a different bivariate OLS regression of the timing of the first split on initial district characteristics, each of which is measured in 2000 before the onset of redistricting. The dependent variable in column (1) counts the number of months between January 2000 and the month in which each original district split, and in column (2) is an indicator for whether the split happened after the moratorium from 2004–6. Coefficients are based on standardized variables. Panel A looks at ethnolinguistic and religious diversity, including the Δ measure capturing differences between parent/child and original district diversity levels. Panel B looks at the 65 controls capturing a broad array of confounders associated with proximity to security forces, economic development, public goods, demographics, natural resource intensity, political factors, economic structure, geography/topography, and remoteness. See Appendix D.2 for a discussion of the variables and Appendix F for further details. The mean effect size is the average standardized coefficient. The sample size is the 52 original districts in our main analysis, and all regressions include an indicator if the district entered the SNPK data in 2005. Robust standard errors are in parentheses. Panel C estimates a Cox Proportional Hazards model relating the time-varying incidence of any social conflict to the likelihood of splitting in a survival regression framework.

Table 2: Average Effects of Redistricting on Social Conflict

Administrative Unit:	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
post-split	-0.008	0.001	0.002	0.001
	(0.026)	(0.022)	(0.028)	(0.025)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.57	0.33	0.47	0.25

Notes: The dependent variable in all columns is an indicator equal to one if there is any social conflict in that district–month. *post-split* is an indicator equal to one for all months after which the original or parent district experiences its first redistricting and a child district is officially passed into law. Columns 1–4 are estimated at the respective administrative unit level listed at the top of the column. There are 52 original districts in column 1, 133 parent/child districts in column 2, 52 parent districts in column 3, and 81 child districts in column 4. All specifications include month FE, district FE, district-specific time trends, and dummies for the number of papers used by coders for the given province–month. Standard errors are clustered by original district. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 3: Redistricting, Changing Ethnic Divisions, and Conflict

Administrative Unit	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
post-split	-0.012	-0.003	0.001	-0.005
	(0.025)	(0.021)	(0.026)	(0.025)
post-split×Δ ethnic polarization	0.036**	0.032	0.027**	0.043*
	(0.018)	(0.019)	(0.013)	(0.025)
post-split×Δ ethnic fractionalization	-0.003	0.000	0.035	-0.011
	(0.019)	(0.012)	(0.026)	(0.019)
post-split×Δ religious diversity	0.014	-0.009	-0.031	-0.005
	(0.013)	(0.011)	(0.021)	(0.014)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.57	0.33	0.47	0.25

Notes: The dependent variable in all columns is an indicator equal to one if there was any social conflict in that district–month. *post-split* is an indicator equal to one for all months after which the original or parent district experiences its first redistricting and the child district is officially passed into law. $\Delta diversity$ measures are normalized to mean zero, standard deviation one. Columns 1–4 are estimated at the respective administrative unit level listed at the top of the column. Δ of the given diversity measure captures the difference in diversity between pre-redistricting (2000) and post-redistricting (2010) boundaries, based on the initial population in 2000. For the original district, this is a weighted average of the parent and child districts. For the parent and child districts, this is a simple difference of their diversity and the original district's diversity. All specifications include month FE, district FE, district-specific time trends, and dummies for the number of papers used by coders for the given province–month. Standard errors are clustered by original district. See the notes to Table 2. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 4: Robust Inference and Outlier Removal

Administrative Unit	Original District **(1)**	Parent & Child **(2)**	Parent **(3)**	Child **(4)**
	Panel A: Baseline			
post-split × Δ ethnic polarization	0.036	0.032	0.027	0.043
baseline: clustering on original district (OD)	(0.018)**	(0.019)	(0.013)**	(0.026)*
spatial HAC, 500 km uniform bandwidth	(0.018)**	(0.010)***	(0.014)*	(0.015)***
effective degrees of freedom adjustment	(0.029)	(0.023)	(0.019)	(0.030)
wild bootstrap, clustering on OD [p-value]	[0.117]	[0.180]	[0.166]	[0.472]
randomization inference [p-value]	[0.090]*	[0.017]**	[0.032]**	[0.003]***
post-split × Δ ethnic fractionalization	-0.003	0.000	0.035	-0.011
baseline: clustering on original district	(0.019)	(0.012)	(0.046)	(0.017)
spatial HAC, 500 km uniform bandwidth	(0.021)	(0.011)	(0.018)***	(0.016)
effective degrees of freedom adjustment	(0.022)	(0.013)	(0.030)	(0.021)
wild bootstrap, clustering on OD [p-value]	[0.880]	[0.982]	[0.214]	[0.587]
randomization inference [p-value]	[0.536]	[0.480]	[0.012]**	[0.773]
Number of District-Months	7,956	20,220	7,956	12,264
Dep. Var. Mean Pre-Split	0.57	0.33	0.47	0.25
	Panel B: Residual Outlier Removal			
post-split × Δ ethnic polarization	0.081	0.025	0.041	0.028
baseline: clustering on original district (OD)	(0.016)***	(0.008)***	(0.014)***	(0.013)**
spatial HAC, 500 km uniform bandwidth	(0.015)***	(0.008)***	(0.010)***	(0.011)**
effective degrees of freedom adjustment	(0.024)***	(0.009)**	(0.021)*	(0.015)
wild bootstrap, clustering on OD [p-value]	[0.013]**	[0.033]**	[0.069]*	[0.134]
randomization inference [p-value]	[0.051]*	[0.100]*	[0.027]**	[0.082]*
post-split × Δ ethnic fractionalization	-0.027	0.007	0.045	-0.007
baseline: clustering on original district (OD)	(0.018)	(0.013)	(0.025)*	(0.018)
spatial HAC, 500 km uniform bandwidth	(0.020)	(0.012)	(0.018)**	(0.015)
effective degrees of freedom adjustment	(0.020)	(0.013)	(0.028)	(0.020)
wild bootstrap, clustering on OD [p-value]	[0.255]	[0.551]	[0.067]*	[0.733]
randomization inference [p-value]	[0.881]	[0.296]	[0.005]***	[0.605]
Number of District-Months	7,696	19,753	7,788	11,927
Dep. Var. Mean Pre-Split	0.59	0.30	0.47	0.21

Notes: This table demonstrates robustness of the baseline results for ΔP and ΔF in Table 3 to alternative inference and outlier removal procedures detailed at length in Appendix D.4. The results for $post\text{-}split$ and $\Delta religious\ diversity$ are suppressed for presentational purposes. Panel A reports several alternative approaches to inference besides our baseline of clustering by original district: (i) the Conley (1999) spatial HAC estimator that allows for contemporaneous correlation in unobservables between all districts within 500 km in addition to the usual within-district correlation over time; (ii) a new "effective degrees of freedom adjustment" due to Young (2016), who adjusts standard errors by the effective sample size implied by the influence of each observation; (iii) a cluster wild bootstrap procedure due to Cameron et al. (2008); and (iv) a quasi-randomization inference (RI) procedure that randomly permutes the $\Delta diversity$ vector across each of the districts in the given regression before estimation, repeating 10,000 times to recover the implied p-values. Panel B additionally removes outliers in $post\text{-}split \times \Delta P$ following the residual-influence approach in Belsley et al. (2005). See the notes to Table 2 for further details on the baseline specification used in both panels. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 5: Further Isolating the Effects of Changes in Ethnic Divisions (I)

Mean Feasible Δ Diversity Based on Potential Redistricting Schemes

Administrative Unit	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
post-split	-0.015	-0.010	-0.005	-0.011
	(0.027)	(0.022)	(0.027)	(0.026)
post-split× mean feasible Δ ethnic polarization	0.053**	0.032**	0.046***	0.031**
	(0.023)	(0.014)	(0.017)	(0.014)
post-split× mean feasible Δ ethnic fractionalization	-0.021	-0.009	0.016	-0.019
	(0.017)	(0.011)	(0.021)	(0.016)
post-split× mean feasible Δ religious diversity	0.009	-0.009	-0.012	-0.001
	(0.022)	(0.013)	(0.026)	(0.014)
Number of District-Months	7,680	18,540	7,680	10,860
Number of Districts	50	121	50	71
Dep. Var. Mean, Pre-Split	0.57	0.34	0.47	0.25

Notes: This table presents a robustness check on the main results in Table 3. Instead of actual $\Delta diversity$, this table uses the mean of feasible $\Delta diversity$ based on the simulation of potential legal redistricting schemes (see Section 5.6 and Appendix D.3, which also considers other moments besides the mean). $\Delta diversity$ measures are normalized to mean zero, standard deviation one. These regressions omit two original districts for which a large number of feasible partitions (over which to compute reliable moments) was computationally intractable. The dependent variable in all columns is an indicator equal to one if there was any social conflict in that district–month. All specifications include month FE, district FE, district-specific time trends, and dummies for the number of papers used by coders for the given province–month. Standard errors are clustered by original district. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 6: Further Isolating the Effects of Changes in Ethnic Divisions (II)

Administrative Unit	Original District	Parent & Child	Parent	Child
	(1)	**(2)**	**(3)**	**(4)**
	Panel A: Other Controls×Post-Split Selected via Double Lasso ($\lambda = 3,000$)			
post-split	-0.013	-0.001	-0.017	-0.003
	(0.026)	(0.017)	(0.024)	(0.022)
post-split×Δ ethnic polarization	0.028**	0.032**	0.079***	0.040
	(0.013)	(0.016)	(0.016)	(0.026)
post-split×Δ ethnic fractionalization	-0.024	-0.003	0.017	0.016
	(0.020)	(0.012)	(0.040)	(0.019)
post-split×Δ religious diversity	0.019	-0.005	-0.026	-0.010
	(0.013)	(0.012)	(0.022)	(0.016)
number of post-split × Lasso-selected controls	6	19	14	12
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.57	0.33	0.47	0.25
	Panel B: Accounting for Changes in Local Public Resources After Redistricting			
post-split	-0.006	-0.001	0.002	-0.002
	(0.027)	(0.021)	(0.027)	(0.024)
post-split×Δ ethnic polarization	0.038**	0.032*	0.023	0.034*
	(0.014)	(0.017)	(0.015)	(0.020)
post-split×Δ ethnic fractionalization	-0.006	0.006	0.035	0.005
	(0.022)	(0.012)	(0.027)	(0.017)
post-split×Δ religious diversity	0.013	-0.013	-0.031	-0.014
	(0.013)	(0.012)	(0.021)	(0.016)
post-split×Δ transfer revenue	0.037	0.027**	-0.004	0.038***
	(0.028)	(0.012)	(0.021)	(0.012)
post-split×Δ distance to district capital	-0.002	-0.012	0.010	-0.033
	(0.019)	(0.013)	(0.016)	(0.020)
Number of District-Months	7,836	19,980	7,836	12,144
Number of Districts	51	131	51	80
Dep. Var. Mean, Pre-Split	0.57	0.33	0.47	0.25

Notes: This table presents two robustness checks on the main results in Table 3. Panel A introduces additional interactions of *post-split* and initial district characteristics. We rely on the Belloni et al. (2014) double-selection post-Lasso approach to select a parsimonious set of influential confounders from the large set of potential covariates we marshal from various data (see Appendix D.2 for details). Subject to a penalty parameter, λ, that helps control overfitting, each column includes a given set of additional *post-split* interactions with the number varying with the specification. Alternative values of λ are explored in Appendix D.2. Panel B augments the baseline specification in Table 3 with two measures capturing changes in public resources and proximity to government institutions due to redistricting. Δ *transfer revenue* is the difference in log average annual transfer revenue post-split and average annual transfer revenue pre-split under the assumptions of allocations proportional to population pre-split in the parent and child district specifications. Δ *distance to district capital* is the population-weighted average village-level difference in log reported travel distance to the district capital in 2011 (post-split) and 2000 (pre-split) as reported by village officials. All variables are normalized to mean zero and standard deviation one. These regressions omit one original district in Jakarta on account of it being in the national capital and not receiving the same stream of general district transfer revenue. The dependent variable in all columns is an indicator equal to one if there was any social conflict in that district–month. All specifications include month FE, district FE, district-specific time trends, and dummies for the number of papers used by coders for the given province–month. Standard errors are clustered by original district. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 7: Changes in Ethnic Divisions, Mayoral Elections and Conflict

Administrative Unit	Parent & Child (1)	Parent (2)	Child (3)
post-split	-0.003	-0.002	-0.008
	(0.022)	(0.027)	(0.025)
post-split$\times\Delta$ ethnic polarization	0.030	0.025*	0.042
	(0.019)	(0.014)	(0.025)
post-split $\times\Delta$ ethnic fractionalization	-0.000	0.034	-0.011
	(0.012)	(0.027)	(0.019)
post-split$\times\Delta$ religious diversity	-0.007	-0.033	-0.002
	(0.012)	(0.022)	(0.015)
post-split$\times$ first election period	0.044	-0.006	0.041*
	(0.044)	(0.051)	(0.021)
post-split$\times$ first election period $\times\Delta$ ethnic polarization	0.026**	0.014	0.027***
	(0.011)	(0.022)	(0.010)
post-split$\times$ first election period $\times\Delta$ ethnic fractionalization	-0.004	0.043	-0.010
	(0.018)	(0.034)	(0.018)
post-split$\times$ first election period $\times\Delta$ religious diversity	-0.009	0.016	-0.017
	(0.016)	(0.035)	(0.021)
post-split$\times$ second election period	0.049**	0.042	0.055*
	(0.022)	(0.027)	(0.032)
post-split$\times$ second election period $\times\Delta$ ethnic polarization	0.009	-0.039***	0.056**
	(0.010)	(0.014)	(0.022)
post-split$\times$ second election period $\times\Delta$ ethnic fractionalization	0.001	0.026	-0.012
	(0.018)	(0.030)	(0.019)
post-split$\times$ second election period $\times\Delta$ religious diversity	0.025	0.052**	0.015
	(0.027)	(0.024)	(0.031)
Number of District-Months	19,980	7,836	12,144
Num Dist	131	51	80
Pre-Split Mean	0.325	0.467	0.245

Notes: The dependent variable in all columns is an indicator equal to one if there was any social conflict in that parent or child district–month (see the notes to Table 2). $post\text{-}split$ is an indicator equal to one for all months after which the child district is passed into law. $\Delta diversity$ measures are normalized to mean zero, standard deviation one. Since post-split mayoral elections are specific to the new parent or child district, we focus here on specifications at that administrative level rather than the original district level. The first election period is an indicator capturing the 6 month window around the district-specific date of the first direct election for the district head after splitting into child and parent districts. The parent district elections occur based on the predetermined schedule inherited from the Suharto era while the child district elections typically occur around 1.5–2 years after redistricting. Hence, parent and child district elections occur at different times. The second election period is defined similarly and occurs five years after the initial election. These second election coefficients are only identified for the three-quarters of districts observed for long enough to hold that second round during our study period. We also include controls for the pre-split election periods, which take place in 10 of the original districts. Hence, the reference period in all columns is the pre-redistricting, non-election period. See Appendix F for details. All specifications include month FE, district FE, district-specific time trends, and dummies for the number of papers used by coders for the given province–month. Standard errors are clustered by original district. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 8: Political Mechanisms: Mayoral versus Parliamentary Elections versus Ramadan

Administrative Unit	Parent & Child	Parent	Child	Parent & Child	Parent	Child
	event: Parliamentary Election			**event**: Ramadan		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
post-split	-0.008	-0.009	-0.010	0.007	0.003	0.005
	(0.021)	(0.027)	(0.025)	(0.022)	(0.028)	(0.026)
post-split×Δ ethnic polarization	0.031	0.027**	0.041	0.031	0.029**	0.040
	(0.019)	(0.013)	(0.026)	(0.019)	(0.014)	(0.024)
post-split ×Δ ethnic fractionalization	-0.000	0.034	-0.011	-0.000	0.035	-0.012
	(0.012)	(0.027)	(0.019)	(0.013)	(0.027)	(0.020)
post-split×Δ religious diversity	-0.011	-0.038*	-0.005	-0.011	-0.038*	-0.005
	(0.012)	(0.021)	(0.015)	(0.012)	(0.021)	(0.015)
post-split× mayoral election period	0.065	0.053	0.046***	0.059	0.046	0.046***
	(0.043)	(0.046)	(0.017)	(0.043)	(0.046)	(0.017)
post-split× mayoral election period ×Δ ethnic polarization	0.016**	-0.010	0.039***	0.017**	-0.009	0.040***
	(0.007)	(0.008)	(0.010)	(0.007)	(0.007)	(0.011)
post-split× mayoral election period ×Δ ethnic fractionalization	0.001	0.037**	-0.008	0.000	0.036**	-0.010
	(0.013)	(0.016)	(0.014)	(0.013)	(0.017)	(0.014)
post-split× mayoral election period ×Δ religious diversity	0.007	0.041**	-0.003	0.007	0.040**	-0.004
	(0.016)	(0.016)	(0.018)	(0.016)	(0.016)	(0.017)
post-split× **event** period	0.154**	0.266***	0.091	-0.068**	-0.038	-0.079*
	(0.066)	(0.077)	(0.085)	(0.029)	(0.038)	(0.041)
post-split× **event** period ×Δ ethnic polarization	-0.013	0.002	-0.020	-0.001	-0.008	0.003
	(0.010)	(0.029)	(0.016)	(0.006)	(0.009)	(0.008)
post-split× **event** period ×Δ ethnic fractionalization	0.008	0.025	0.012	0.003	-0.002	0.007
	(0.017)	(0.040)	(0.018)	(0.010)	(0.012)	(0.014)
post-split× **event** period ×Δ religious diversity	0.019	0.063**	0.007	0.011	0.021	0.006
	(0.011)	(0.029)	(0.013)	(0.009)	(0.016)	(0.011)
Number of District-Months	19,980	7,836	12,144	19,980	7,836	12,144
Number of Districts	131	51	80	131	51	80
Pre-Split Mean	0.33	0.47	0.25	0.33	0.47	0.25
Δ polarization × election = Δ polarization× **event**, p-value	0.005	0.654	0.004	0.001	0.915	0.011
Δ fractionalization × election = Δ fractionalization× **event**, p-value	0.689	0.788	0.324	0.883	0.101	0.433

Notes: This table allows the effects of $\Delta diversity$ to vary with mayoral election periods and parliamentary election periods in columns 1–3 and holy month of Ramadan periods in columns 4–6. Mayoral election periods pool the first and second election windows considered separately in Table 7. Parliamentary election periods comprise the 6 month windows centered on the month of April in 2004, 2009, and 2014 when such elections take place simultaneously around the country. Ramadan periods include all Gregorian calendar months during which any of the lunar holy month falls. The own post-split×$\Delta diversity$ terms therefore capture the differential effects of diversity in all other months outside the given periods of interest. The dependent variable in all columns is an indicator equal to one if there was any social conflict in that parent or child district–month (see the notes to Table 2). $post\text{-}split$ is an indicator equal to one for all months after which the child district is passed into law. $\Delta diversity$ measures are normalized to mean zero, standard deviation one. All specifications include month FE, district FE, district-specific time trends, and dummies for the number of papers used by coders for the given province–month. Standard errors are clustered by original district. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table 9: Light Intensity and Changes in Village-Level Alignment with the Largest Ethnic Group in the New Versus Original District

	Parent Child (1)	Parent (2)	Child (3)
post-split	-0.020	-0.006	-0.033**
	(0.012)	(0.022)	(0.013)
post-split $\times$ Δ share of village in largest ethnic group in district > 0	0.011*	0.012	0.021*
	(0.006)	(0.009)	(0.011)
post-split $\times$ Δ share of village in largest ethnic group in district $= 0$	0.028**	0.017	0.040**
	(0.013)	(0.017)	(0.018)
Number of Village–Years	164,594	85,401	79,193
Dep. Var. Mean, Pre-Split	0.163	0.229	0.114

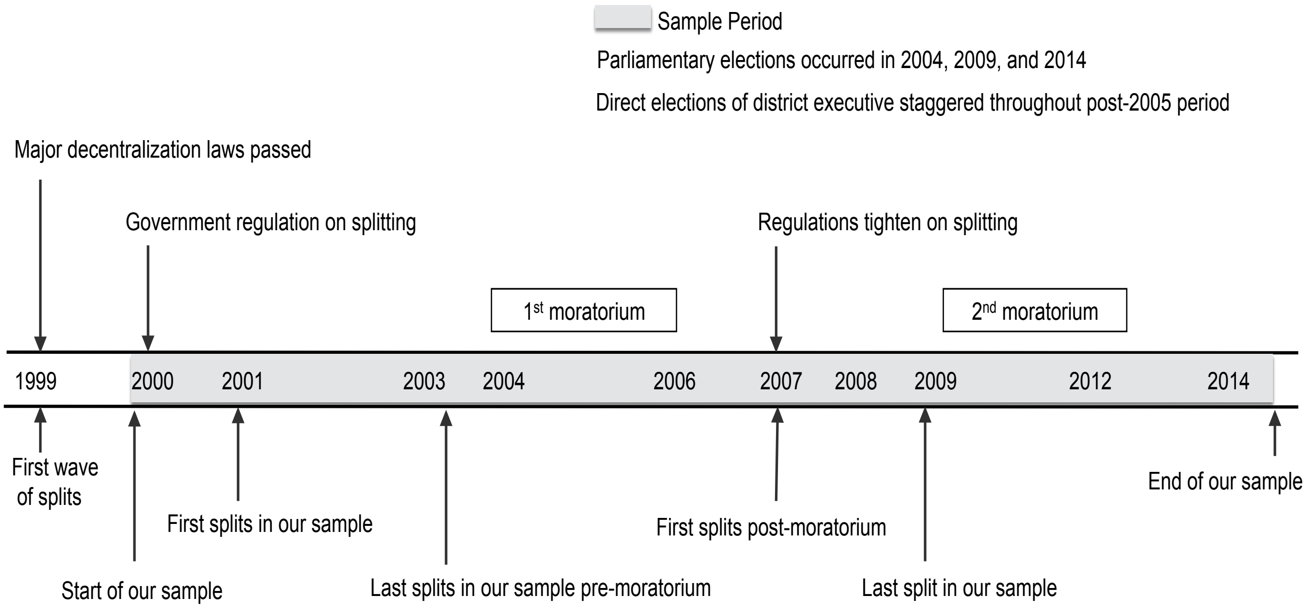
Notes: The dependent variable is share of village area covered with any nighttime lights. The village-level panel spans 2000–2013. The regressions also control for village and year fixed effects. The regressor in row 2 (3) equals one if the share of the given village belonging to the largest ethnic group in the district increased (remained the same) as a result of redistricting. The ethnic shares are, as throughout the paper, defined based on the population in 2000. Standard errors are clustered at the original district level. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Online Appendix

A Further Background on District Proliferation

Figure A.1 provides a timeline of events over our study period, including major decentralization, redistricting, and electoral reforms. Below, we provide further details on the implications of redistricting discussed in Section 2.2.

Figure A.1: Timeline of Events



Size of Government. In the typical district, between 1,200 and 2,000 new jobs are created (according to interviews and province-level yearbooks). We have found no evidence that the total number of offices and jobs decrease in the parent district. Thus, the overall number of civil servants per capita increases substantially, and these newly created jobs are important for setting and executing public policy.

In addition, there are apportionment gains to redistricting due to the step function rule used to determine the seat-to-population ratio. Seats in local parliament always weakly increase with redistricting. For example, an original district with 400,000 people initially would have 40 seats. If it split into two equally sized districts, each would have 30 seats for a total of 60 compared with 40 originally.

Fiscal Resources. Redistricting also leads to an increase in transfers from the central government. We estimate the effects of splitting on total per capita transfers in our sample using the within-district identification strategy detailed in Section 4.[1] Once new funds for the child district start flowing in approximately two years after the split, real transfers at the original district level increase by 18–25 log points off a mean of roughly USD 200 (Table A.1, Panel A, Column 1).[2] These revenue increases pass through to significant increases in local government expenditures in the following year.

We cannot observe how transfers were divided between child and parent areas before redistricting. However, one natural benchmark is to assume that pre-split transfers (T) were allocated according to population with the parent receiving $\left(\frac{N_{parent}}{N}\right) T$ and the child receiving $\left(\frac{N_{child}}{N}\right) T$. We use this benchmark to perform two exercises that clarify the overall fiscal benefits of redistricting and the differential

[1]Initial population is absorbed in the fixed effect, and while including time-varying population does little to change the point estimates, it introduces unnecessary noise as the data is incomplete and requires estimation and imputation.

[2]Note that the decline in transfers in the year after splitting reflects a short adjustment period when child district transfers have only slowly started to flow into the new public coffers while parent district transfers have begun to adjust downward to account for their now smaller population.

gains to child districts. First, we take the original district transfers as given and compare realized transfers post-split to the expected transfers if they had continued to be allocated proportional to population. Second, we assume that parents and children receive their population shares of the original district transfers pre-split (and in the year of the split when nothing yet changes). Then, we continue this time-series post-split using the actual, observed transfers at these lower administrative units. This allows us to re-estimate regressions like that in column 1 of Panel A in Table A.1 at the smaller units.

First, we simply compare realized transfers in post-split years at the parent/child level to expected transfers based on population shares of the realized original district level transfers in all post-split years. We plot the distribution of these differences between actual and expected transfers (based on population shares of the realized original district transfers) in Figure A.2. This shows the difference (in USD) for all post-split years and districts in our sample but looks comparable if plotted year-by-year. It is evident that children receive more than expected based on population shares and, consequently, parents less. In the average post-split year, parents receive USD 7.4 million less than expected (USD 16 per capita) and children receive USD 5.1 million more (USD 58 per capita). This strongly suggests that the gains from redistricting accrue disproportionately towards children. This finding is in line with the upfront costs of establishing new government institutions. For example, around 40–50 percent of expenditures go towards staff, which expanded greatly in the child but not the parent.

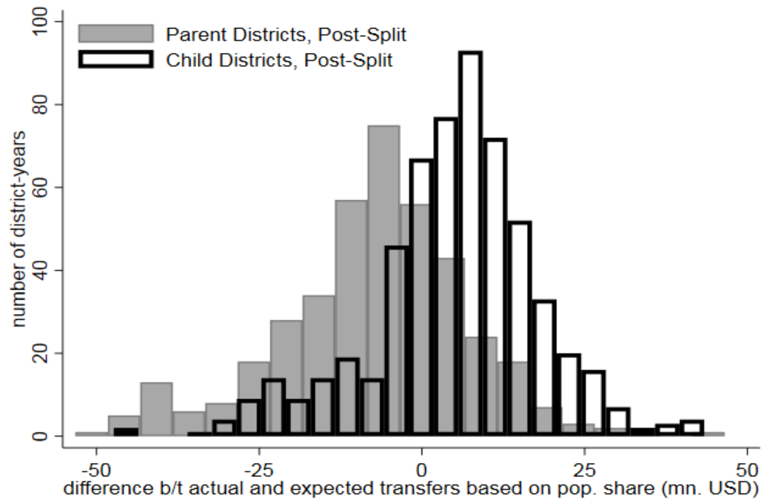
Note that while children gain disproportionately from splitting, parents nevertheless tend to see an increase in transfers as well. To see this, suppose that parents received their population share of original district transfers pre-split. Parents receive a lower share of a larger total transfer 'pie'. In practice, this still results in an increase in transfers per capita at the parent level as made clear in the second exercise.

Second, columns 2–4 of Panel A in Table A.1 make these patterns even clearer by showing that both parents and children benefit from splitting in terms of real transfers, but children clearly benefit more. Parent districts experience roughly a 19 log point increase in long-run transfers relative to the pre-split period (column 3), whereas child districts experience a 59 log point increase (column 4). While these results are subject to strong assumptions about the pre-split allocation of transfers, different assumptions are unlikely to explain away the main takeaways that (i) overall transfers increase in both parent and child districts, and (ii) child districts benefit relatively more than parent districts.

Proximity to Government. In addition to receiving increased transfers, child district residents experience a significant reduction in the average distance to government institutions. Panel B of Table A.1 shows how reported travel distance to the capital (in kilometers) changed after splitting. These estimates are based on reports by the village head in 2000 and 2011 from *Podes*, which we aggregate to 2010 district borders using population weights. While parent districts experienced little change in distance to the capital, child districts register an average reduction of around 55 km off of a pre-split mean of 100 km.

Finally, note that these changes imply a significant reduction in the size of the population governed by any given district. According to Census data from 2000, districts based on 2000 boundaries have a median population of 400,000 whereas the 2010 district boundaries imply a median of 250,000.

Figure A.2: Comparing Fiscal Transfers Between Parent and Child Districts



Notes: This figure plots the density of the difference in actual versus expected fiscal transfers for parent and child districts post-split under the assumption that the expected transfers are allocated proportional to population share of the original district.

Table A.1: Splitting-Induced Changes in Transfer Revenue and Distance to Capital

Panel A: Effects on ln(total fiscal transfers)

Administrative Unit	Original District (1)	Parent Child (2)	Parent (3)	Child (4)
$\leq$ 2 Years Pre-Split	0.073*	-0.005	0.028	-0.017
	(0.041)	(0.032)	(0.030)	(0.038)
1 Year Pre-Split		reference period		
Year of split	-0.029	-0.022	0.002	-0.035
	(0.023)	(0.026)	(0.021)	(0.035)
1 Year after Split	-0.113*	0.073*	0.252***	-0.047
	(0.059)	(0.041)	(0.051)	(0.061)
2 Years after Split	0.093	0.314***	0.211***	0.368***
	(0.057)	(0.047)	(0.051)	(0.060)
3 Years after Split	0.180***	0.474***	0.263***	0.596***
	(0.058)	(0.042)	(0.051)	(0.052)
4 Years after Split	0.246***	0.500***	0.290***	0.620***
	(0.064)	(0.038)	(0.056)	(0.047)
5+ Years after Split	0.207***	0.444***	0.187***	0.593***
	(0.053)	(0.050)	(0.058)	(0.064)
No. of District–Years	765	1,965	765	1,200
Dep. Var. Mean	26.6	25.6	26.0	25.3

Panel B: Effects on Distance to District Capital (kilometers)

	Pre-Split Mean	Mean Change	Median Change
Parent Districts	48.9	-5.7	-1.14
	[33.3]	[18.2]	
Child Districts	99.8	-55.5	-38.5
	[79.5]	[8.04]	

Notes: Panel A reports a regression of log per capita transfer revenue in real 2010 USD (see Appendix F) on dummies pre- and post-split as well as district fixed effects, year fixed effects, and district-specific time trends. Details on the transfer time series are discussed in the text above. Standard errors are clustered at the original district level. Panel B reports the average change in distance to the capital in kilometers, constructed from the *Podes* 2000 and 2011 administrative censuses, for parent (and child districts separately. We are missing data for a small number of the districts in Aceh in 2003. Standard deviations in brackets.

B Conceptual Framework: Further Details

In this appendix, we make explicit the general points made in Section 3 using two specific models. First, we present a simple two-group model of conflict over a public good, which shows that changes in violence can be derived from a simple, linear function of border-induced changes in diversity. Second, we present the implications of border changes for a special case of Esteban and Ray's (2011a) approximation, which generates richer predictions.

B.1 A Simple Two-Group Framework[1]

Baseline Setup. Suppose an original district ($\mathcal{O}$) is composed of two groups with population $G_{\mathcal{O},1}$ and $G_{\mathcal{O},2}$ and total population $G_{\mathcal{O}} = G_{\mathcal{O},1} + G_{\mathcal{O},2}$. Denote by $g_{\mathcal{O},i}$ the share of group i in the population. These groups compete over a public prize. The prize, being non-excludable and non-rival, is not diminished by group size, but the winner of the contest gets to choose the mix of public goods that their group prefers. We assume the winner chooses a level that provides their own group with value R per person and the other group with 0. The leader of each group i chooses total violence $V_{\mathcal{O},i}$ given its per unit cost γ. The probability of group i winning control over the public prize is given by the contest function $\frac{V_{\mathcal{O},i}}{V_{\mathcal{O},1}+V_{\mathcal{O},2}}$. Each group leader chooses $V_{\mathcal{O},i}$ to maximize per capita payoffs. That is, each group i, taking as given the other's choices $V_{\mathcal{O},-i}$, maximizes $\left(\frac{V_{\mathcal{O},i}}{V_{\mathcal{O},1}+V_{\mathcal{O},2}} R - V_{\mathcal{O},i}\frac{\gamma}{G_{\mathcal{O},i}}\right)$.

The Nash equilibrium level of conflict per capita, $\frac{V_{\mathcal{O}}}{G_{\mathcal{O}}} = \frac{V_{\mathcal{O},1}+V_{\mathcal{O},2}}{G_{\mathcal{O},1}+G_{\mathcal{O},2}}$, is given by $\frac{V_{\mathcal{O}}}{G_{\mathcal{O}}} = g_{\mathcal{O},1} g_{\mathcal{O},2}\left(\frac{R}{\gamma}\right)$. Thus, total violence per-capita in the district depends on the value of the public prize, R, the costs of conflict, γ, and a measure of diversity, $g_{\mathcal{O},1} g_{\mathcal{O},2}$, which is greatest when both groups are of equal size. For two groups, this measure equals $0.5\times$ a standard fractionalization measure.

Effects of Redistricting. We trace out the implications of changing borders on conflict under the assumption that splitting creates new, separate contests in parent and child districts. Conflict within each of the new districts will now be a function of the diversity *within* each new area. Assume for now that winning the prize continues to yield R per person within group and 0 otherwise and that the costs of conflict remain unchanged. Using $\mathcal{P}$ to denote the parent and $\mathcal{C}$ the child district, the change in total violence per capita at the original district level is:

$$\frac{\Delta V_{\mathcal{O}}}{G_{\mathcal{O}}} = \left(\frac{G_{\mathcal{P}}}{G_{\mathcal{O}}} g_{\mathcal{P},1} g_{\mathcal{P},2} + \frac{G_{\mathcal{C}}}{G_{\mathcal{O}}} g_{\mathcal{C},1} g_{\mathcal{C},2} - g_{\mathcal{O},1} g_{\mathcal{O},2}\right)\frac{R}{\gamma},$$

where $g_{i,j}$ is the share of group j in new district i, and $G_{\mathcal{O}} = G_{\mathcal{P}} + G_{\mathcal{C}}$. That is, the change in violence per capita is explicitly a function of the difference in the population-weighted average of diversity within the new units relative to the diversity in the original district pre-split. In the event that the groups separate into perfectly homogeneous child and parent districts, all violence in the original district ceases.

It is also interesting to consider changes *within* the new borders. This requires taking a stance on how violence is initially distributed across parent and child. Letting σ be the share of total violence falling in the parent district, the change in conflict within the parent district is given by:

$$\frac{\Delta V_{\mathcal{P}}}{G_{\mathcal{P}}} = \left(g_{\mathcal{P},1} g_{\mathcal{P},2} - \sigma \frac{G_{\mathcal{O}}}{G_{\mathcal{P}}} g_1 g_2\right)\frac{R}{\gamma}.$$

If violence is initially distributed according to population ($\sigma = \frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}$), the change in per-capita violence within the eventual parent border is given by the difference in the diversity within that new unit and the overall diversity in the original district.

[1] We thank Enrico Spolaore for suggesting this framework.

The model also implies that changes in the value of the public prize (R) or the costs of violence (γ) will change conflict. Redistricting is accompanied by an influx of government resources (R) as well as reductions in the distance to the new capital, which could affect γ. If R/γ increases in the newer units, then this will exacerbate Δ conflict, particularly in newly diverse areas. Adapting the model to incorporate these changes, we find a similar link between changes in diversity and changes in conflict but one that puts greater than population-weight on diversity in the newer units. Finally, note that this simple model uses a single metric of diversity (a scalar multiple of fractionalization) and does not have the richer dynamics of a model like Esteban and Ray (2011a) which distinguishes between polarization and fractionalization, which we turn to next.

B.2 An Esteban and Ray (2011a) Framework

Esteban and Ray (2011a) model groups contesting a budget with per-capita value $\pi + \mu$, some of which can be distributed privately and the remaining is public with the winning group choosing their preferred mix of public goods. For large N and for an isoelastic cost function $c(r) = (1/\theta)r^\theta$, per-capita conflict is given by $\frac{V}{G} \approx (\alpha[\pi P + \mu F])^{1/\theta}$, where α is within-group cohesion, π (μ) is the population-normalized public (private) payoff of the conflict prize, P is ethnic polarization, and F is ethnic fractionalization. The paper explains the sense in which this is an approximation.

Effects of Redistricting. We trace out the implications of changing borders on conflict in this model under the assumption that splitting creates new, separate contests in parent and child districts. Conflict within each of the new districts will now be a function of the diversity *within* each new area. Let $\mathcal{O}$ denote the original district boundaries, $\mathcal{P}$ denote the parent district, and $\mathcal{C}$ the child district. Further assume that per-capita payoffs remain unchanged within each new area. Then, the change in total violence per-capita at the original district level is:

$$\frac{\Delta V_{\mathcal{O}}}{G_{\mathcal{O}}} = \alpha^{1/\theta}\left(\frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}([\pi P_{\mathcal{P}} + \mu F_{\mathcal{P}}])^{1/\theta} + \frac{G_{\mathcal{C}}}{G_{\mathcal{O}}}([\pi P_{\mathcal{C}} + \mu F_{\mathcal{C}}])^{1/\theta} - ([\pi P + \mu F])^{1/\theta}\right),$$

for original district population $G_{\mathcal{O}} = G_{\mathcal{P}} + G_{\mathcal{C}}$. That is, the change in violence per capita is explicitly a difference between population-weighted functions of diversity within the new units relative to a function of diversity in the original district pre-split. In the event that the groups separate into perfectly homogeneous child and parent districts, all violence in the original district ceases.

It is also interesting to consider changes *within* the new borders. This requires taking a stance on how violence is initially distributed across parent and child. Letting σ be the share of total violence falling in the parent district, the change in conflict within the parent district is given by:

$$\frac{\Delta V}{G_{\mathcal{P}}} = \alpha^{1/\theta}\left(([\pi P_{\mathcal{P}} + \mu F_{\mathcal{P}}]^{1/\theta} - \sigma\frac{G_{\mathcal{O}}}{G_{\mathcal{P}}}([\pi P + \mu F])^{1/\theta}\right)$$

If violence is initially distributed according to population ($\sigma = \frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}$), the change in per-capita violence within the eventual parent border is given by the difference in the diversity within that new unit and the overall diversity in the original district.

The model also implies that changes in the value of the prize (π, μ), social cohesion (α) or the costs of violence (which vary with θ) will change conflict. Redistricting is accompanied by an influx of government resources as well as reductions in the distance to the new capital, which could affect costs. If the value of the prize increases in the newer units, then this will exacerbate Δ conflict, particularly in newly diverse areas.

C Measuring Conflict: Background and Robustness

C.1 Indonesia's National Violence Monitoring System (SNPK)

Indonesia's National Violence Monitoring System (NVMS) or SNPK by its Indonesian acronym (*Sistem Nasional Pemantauan Kekerasan*) is among the world's largest single-country, geospatial conflict databases. After compiling several million images from over 120 carefully screened local newspapers, data entrants classify the nature of violence underlying each reported event into one of the 10 categories listed below in Table C.1.1.[1] There are further subcategories within each category of conflict. For example, when available, each event also includes information on the number of deaths, injuries and buildings destroyed.

Table C.1.1: Violence Categories in the SNPK

Resource Conflict	Violence triggered by resource disputes (land, mining, access to employment, salary, pollution, etc.).
Governance Conflict	Violence is triggered by government policies or programs (public services, corruption, subsidy, region splitting, etc).
Popular Justice Conflict	Violence perpetrated to respond to/punish actual or perceived wrong (group violence only).
Elections and Appointment Conflict	Conflict Violence triggered by electoral competition or bureaucratic appointments.
Separatist Conflict	Violence triggered by efforts to secede from the Unitary State of the Republic of Indonesia (NKRI).
Identity-Based Conflict	Violence triggered by group identity (religion, ethnicity, tribe, etc).
Other Conflict	Violence triggered by other issue.
Violence During Law Enforcement	Violent action taken by members of formal security forces to perform law-enforcement functions (includes use of violence mandated by law as well as violence that exceeds mandate for example torture or extrajudicial-shooting).
Violent Crime	Criminal violence not triggered by prior dispute or directed towards specific targets.
Domestic Violence	Physical violence perpetrated by family member(s) against other family member(s) living under one roof/same house including against domestic workers and violence between cohabiting couples.

As discussed in Section 4.2, we rely on this rich, human-led classification system to isolate social conflict as opposed to (unorganized) interpersonal violence or crime. Of course, the lines between categories are often fuzzy.[2] Nevertheless, in a robustness check in Appendix C.3, we effectively show that our core results are not driven by the particular measure of social conflict. Moreover, as proof of concept, it is reassuring that the differential social conflict around mayoral elections in Table 7 is indeed driven in large part by violence categorized as "elections and appointments conflict."

[1] The data report other information about each event such as the actors involved, the organizational form of violence (e.g., riot, kidnapping), weapons used, and outcome of external intervention. While potentially useful, this information is much less systematic and comprehensive than the categorization into types of violence, which is the most directly related to the conceptual framework and broader interest in the paper.

[2] This description from the data manual provides further background that may be illustrative: "According to NVMS system, violent crime comprises acts of violence that occur without any prior dispute between parties. The motivation behind a criminal act can be monetary, for example, robbery or abduction; or personal pleasure, for example, rape or serial killings. In contrast, violence in the context of conflict occurs due to pre-existing disputes between those involved such as dispute over land, election, religion or other such matters. As such, in the NVMS system, an act of killing can be coded as 'Conflict' if there is a dispute behind it, e.g., in a killing of a certain group figure by other groups, or can be coded as 'Crime' if there is no pre-existing dispute between parties, for example, serial killings."

Event Descriptions. The following Appendix C.2 provides several examples of events in the "elections and appointment" conflict category. Below, we provide examples from a few of the other categories beginning with "governance", which, like elections/appointments, is plausibly responsive to a similar sort of sociopolitical changes associated with redistricting.

1. Pontianak City, 24 July 2006: *Hundreds of residents from 6 villages came to the office of Sungai Kunyit Subdistrict. They protested the perceived unfair distribution of the unconditional cash transfer (BLT) funds. They then threw a chair at the sight of a BPS (Central Statistical Agency) representative. Some community leaders and the subdistrict head calmed the masses.*

2. Kotawaringin Timur District, 21 June 2012: *People burnt a temporary bridge in Seruyan Hilir subdistrict because they argued that the government took too long to build the main permanent bridge.*

3. Singkawang District, 5 December 2008: *Protests led by Front Pembela Islam (FPI), Front Pembela Melayu (FPM), and Aliansi LSM Perintis Singkawang. They argued that dragon statue is a religious symbol, and hence a public road is not the proper place to build that symbol. In addition, the dragon statue is perceived as Chinese symbol. FPI claimed that symbols for particular ethnic groups cannot be placed in public places.*

Note that the last example above could clearly have also been classified as 'Identity-Based Conflict', pointing to the fuzziness across categories as noted earlier. As noted in Section 6.3, there are also numerous governance incidents involving violence directed at the public electricity monopoly (PLN) centered on frustration with electricity allocation. A few examples follow:

1. Tapanuli Selatan District, 23 November 2011: *In the office of PLN . . . , about 200 people demonstrated and damaged the office. The action was triggered by anger over electricity being out for three months.*

2. Sumbawa District, 14 December 2014: *In the office of PLN . . . , there was an attack perpetrated by local residents on the head of Human Resources. The incident started when hundreds of residents went to the PLN office to protest the frequent power outages in the last 4 months and the recent total power outage for two days without notice. Angry demonstrators hit the HR's head, leaving him injured. Fortunately, the action did not escalate further because the military, police and village heads intervened to quell emotions.*

A few other illustrative examples come from the "resource conflict" category:

1. Aceh Singkil District, 30 May 2011: *Two hundred people demonstrated in front of the mayor's office of Aceh Singkil in relation to land disputes with companies of Malaysian origin. They also demanded a fair and fixed land [compensation].*

2. Halmahera Tengah, 30 Jan 2012: *Hundreds of East Halmahera residents burned tires and blocked roads at the PT Kemakmuran Pertiwi Tambang (PT Harita Grup) nickel mining site in Loleba village.*

Comparison to Other Conflict Data. The SNPK data offer several advantages over two alternative sources of information on violence in Indonesia. First, it offers more comprehensive temporal coverage than the triennial *Potensi Desa* (or *Podes*) data, which records information on the violent events at the village-level over the prior three-year period. This coarse coverage would not allow for the systematic generalized difference-in-difference identification strategy we deploy here. Moreover, *Podes* accounts are based on the self-reports of village leaders as opposed to the plausibly more objective, cross-validated newspapers reports in the SNPK.

Second, the SNPK offers significantly more comprehensive coverage compared to a widely used, cross-country, subnational data source. The Uppsala Conflict Data Program (UCDP) Georeferenced Event Data (GED) (Sundberg and Melander, 2013) has been fruitfully deployed in a range of subnational conflict studies and with particular success in sub-Saharan Africa alongside the widely used Armed Conflict Location & Event Data Project (ACLED) data. The UCDP-GED is available for Indonesia whereas the ACLED is not (yet). Mapping the UCDP-GED events to our original district monthly panel, we find very limited coverage of social conflict events in Indonesia. While SNPK covers 223 of the 230 original district–month incidents in the UCDP-GED data, there are 4,795 additional district–months with social conflict incidents in the SNPK. Together, these violent events involve nearly 5,000 deaths over a 15 year period. The more limited coverage by UCDP-GED is explained by both its more narrow focus on large-scale conflict and by its reliance on international news sources and or English-based ones in Jakarta. The SNPK offers much deeper coverage precisely because it digitized millions of old newspapers from outlying regions of the country that allowed for coverage of violence that may have otherwise missed the attention of international reporters. Barron et al. (2016) offer a more systematic comparison (for all of Indonesia) by applying particular restrictions in the SNPK that more closely match those applied in the UCDP-GED. Their conclusion is similar to ours; the UCDP-GED cover around one-third of the events and and deaths reported in the SNPK.

Costs of Conflict. The violent episodes in SNPK can be costly. Even if we examine the least violent years and restrict to social conflict, we observe around 500 annual deaths, 7,000 annual injuries, and 1,500 annual buildings damaged. Including crime and domestic violence more than doubles these numbers. Using a methodology due to Fearon and Hoeffler (2014), we estimate that the direct costs of social conflict in the post-2005 period range from 0.2–0.5% of GDP.

C.2 Electoral Violence in the SNPK

As discussed in Sections 2 and 3, Harish and Toha (2017) use the SNPK data to identify three salient types of electoral violence in Indonesia: (1) *voter-targeting* is “any kind of election-related violence that affects voters’ preferences participation in elections”, (2) *candidate-targeting* directs violence towards “candidates themselves and those around them by intimidating them into withdrawing and/or physically and forcefully removing them from the race”, and (3) *government-aimed* is “violence mounted against a government agency responsible for monitoring and enforcing rules of elections.” The authors use SNPK data combined with supplementary reporting to categorize over 1,000 episodes of local election violence in Indonesia since 2005. Attacks targeting candidates are the most common, occurring on 35 percent of the days in a six month window centered on the election. Voter-targeting occurred in 25 percent of those days, and agency-targeting on 17 percent of days. Not surprisingly, most candidate-targeting is concentrated in the lead-up to the election with attacks on election-related government agencies occurring thereafter.

Drawing upon the same SNPK data, we provide some concrete examples of incident reports that clarify the types of electoral violence underlying these patterns. The following are district-specific examples that we translate from the SNPK:

1. Aceh Singkil District, 2 November 2013: *Protest at Komisi Independen Pemilihan (KIP, Independent Commission for Elections) by supporters of Affan Alfian-Pianti Mala (Walikota-Wakil Walikota [mayor–vice mayor] candidate) regarding fraud in mayoral election.* Seven people were reported seriously injured. The election took place on 29 October.

2. Aceh Barat Daya District, 28 June 2012: *Supporters of FD (mayoral candidate for Aceh Barat Daya) were attacked by their competitors in Kuala Terubu Village and Alue Sungai Pinang village.* The election took place on 9 April 2012.

3. Halmahera Utara District, 16 April 2005: *Komisi Pemilihan Umum Daerah (KPUD, Local General Elections Commission) office and the house of the Partai Demokrasi Kebansaan (PDK) chairman were destroyed by people because one of the candidate was not selected in mayor–vice mayor ticket.* Two buildings were damaged and one destroyed. The election took place on 27 June 2005.

4. Kepulauan Sula District, 12 May 2005: *Molotov bombing of the local Electoral Commission office due to anger with the decision about four mayoral candidates.* The election took place on 27 June 2005.

5. Pulau Morotai District, 21 May 2011: *Mass supporters of RS and WP [mayoral candidate and running mate] who did not accept the decision of the Morotai Electoral Commission in the election took action in the Morotai air force base, South Morotai, northern Maluku, by trying to break...* Four people were injured, and one building was damaged. Subsequent violent incidents were reported on May 26 and 27. The election took place on 16 May 2011.

6. Kotawaringin Timur, 6 June 2005: *Incident between supporters of mayoral candidates Wahyu-Amrullah and Thamrina-Mullan Safri because one of them established billboard in the other candidates' area (Seruyan) K Timur: On Jalan Mayjen Suprapto, Seruyan Hilir subdistrict, billboard of mayoral candidate was destroyed, occurred around mayoral election time. In Danau Sembuluh subdistrict, AS (legislative member candidate for Dapil [electoral region] II) was attacked by people (one of them was legislative member candidate for Dapil [electoral districdt] II).* Two people were seriously injured. The election took place on 23 June 2005.

7. Bengakayang, 21 May 2010: *In the Local Electoral Commission office, demonstrations took place with rioters throwing stones at the building and officials out of anger over the election outcome.* One building was damaged. The election took place on 19 May 2010.

C.3 Robustness of Main Results to Alternative Categorizations of Conflict

Our main analysis considered a set of violence categories in the SNPK that aimed to capture group-based conflict. This appendix rules out two potential concerns with the measure of social conflict we use based on the SNPK groupings.

First, some of the crime-related categories of violence may be shaped by similar (changes in) ethnic divisions as other categories deemed to fall under conflict.[3] Hence, their omission may be deemed arbitrary at best and biasing at worst. Table C.3.1 shows that the main results in Table 3 are robust to not restricting the definition of violence. Indeed, the estimated effects of $\Delta diversity$ are very similar. The increase in precision may be due to the fact that the broader grouping reduces classical measurement error of the sort discussed in Appendix C.1.

[3] Echoing this interpretation, one of the architects of the SNPK notes in a later reappraisal that "What may appear to be local violence (crime, interpersonal clashes over land) is often linked in complicated ways to the broader conflict" (Barron et al., 2016, p. 25). This would be consistent with the ethnic-related criminal gangs documented at length in the Wilson (2015) book that we cite in the paper. Indeed, many of these gangs are often mobilized for conflict by political actors during times of instability around elections. Another, broader interpretation of this concern would be that changes in ethnic divisions further undermine local state capacity that helps to forestall a breakdown in social order and prevent various types of crime.

Table C.3.1: Effects are Similar When Not Restricting to Social Conflict

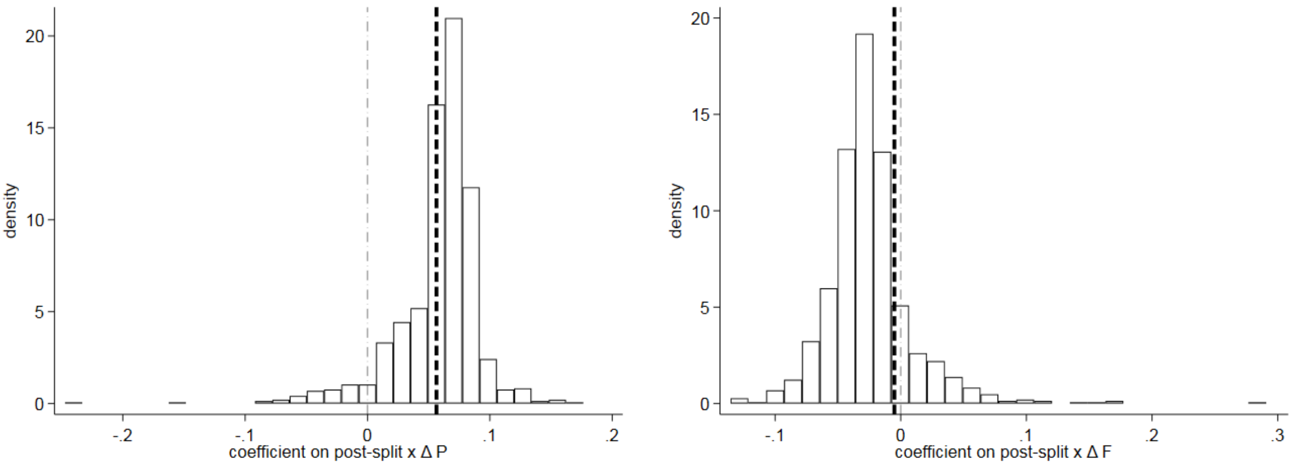
Administrative Unit	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
	Dep. Var.: Any Violence			
post-split	-0.007	0.018	-0.008	0.036
	(0.021)	(0.021)	(0.021)	(0.027)
post-split $\times$ Δ ethnic polarization	0.049***	0.043***	0.054***	0.028**
	(0.016)	(0.015)	(0.020)	(0.011)
post-split $\times$ Δ ethnic fractionalization	-0.013	-0.008	-0.002	0.001
	(0.028)	(0.019)	(0.025)	(0.027)
post-split $\times$ Δ religious diversity	0.006	-0.008	-0.055**	0.000
	(0.012)	(0.019)	(0.026)	(0.023)
Number of District-Months	7,956	20,220	7,956	12,264
Dep. Var. Mean, Pre-Split	0.76	0.52	0.69	0.42

Notes: This table re-estimates our baseline specification but for all violence reported in the SNPK. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

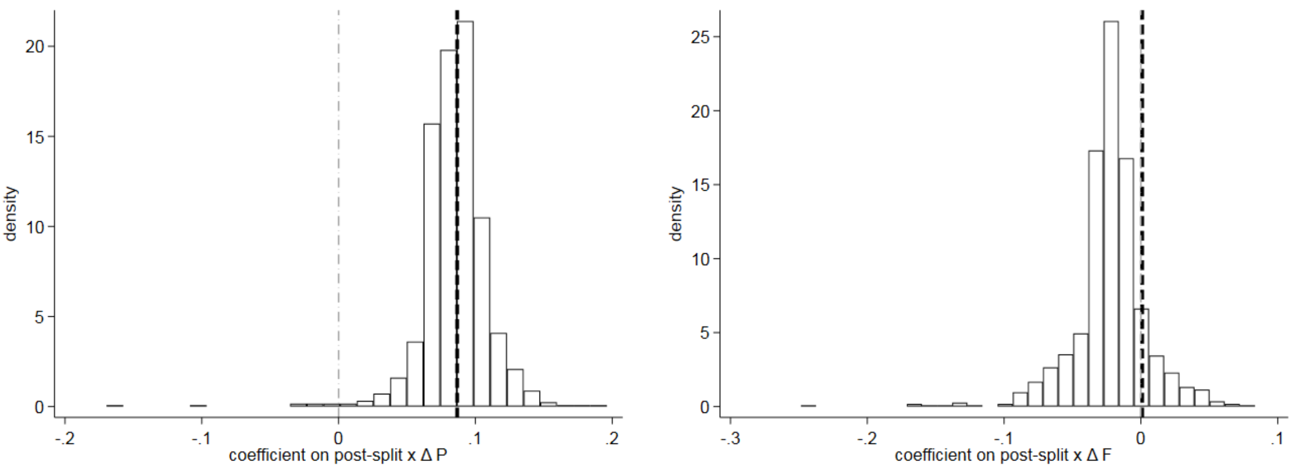
Second, we further gauge robustness to event misclassification by re-estimating our regressions for all possible combinations of the ten main categories of violence in the SNPK. Figure C.3.1 presents the distribution of the estimated coefficients on $post\text{-}split \times \Delta P$ and $post\text{-}split \times \Delta F$ for these 1,023 regressions with the given baseline estimate for social conflict indicated by the dashed, vertical black line. For both our baseline and each separate regression, we scale the reported coefficient by the mean of the given dependent variable, which varies across groups of categories. The magnitudes are therefore standard deviation $\Delta diversity$ effect sizes relative to the mean outcome over the sample period. Note that we are not using this data mining approach for inference purposes but rather to address concerns that our particular designation of categories as conflict was somehow spuriously generating our results. Figure C.3.1 helps to dispel such concerns and shows that our core estimated effect of ΔP on social conflict appears to be around the middle of the distribution of effect sizes across all possible combinations of violence categories. Moreover, the distribution of these coefficients seems to lie mostly above zero, which again points to the fact that changing ethnic divisions shifts most types of violence in the same direction. The takeaways are similar for ΔF.

Figure C.3.1: Distribution of Estimated Effects of $\Delta\ diversity$ across All Possible Groupings of Violence Categories in SNPK

(a) Original Districts



(b) Parent and Child Districts



Notes: These graphs present the distribution of estimated effects of $\Delta diversity$ across all possible groupings of the violence categories reported in the SNPK. The estimates are rescaled by the mean of the dependent variable such that the effects are standard deviations relative to the mean violence in the given grouping. The dashed line is our baseline estimate from Table 3.

C.4 Main Results Broken Down by Injuries, Deaths, and Damages

SNPK records injuries, deaths, and property damage. We show in Table C.4.1 that our results are robust to redefining any social conflict to include only the roughly 90% of incidents that record at least one of these outcomes. Further, we show in Table C.4.2 that our results are driven by events with injuries and are strongest for 'any social incident with an injury or property damage'. Our results are not identifying changes in violence resulting in deaths.

Table C.4.1: Effects are Similar When Restricting to Social Conflict Events with an Injury, Death, or Property Damage

Administrative Unit:	Original District	Parent & Child	Parent	Child
	(1)	**(2)**	**(3)**	**(4)**
	Panel A: Only events with an injury, death, or property damage			
post-split	-0.005	0.001	-0.007	0.004
	(0.026)	(0.021)	(0.028)	(0.023)
post-split$\times\Delta$ ethnic polarization	0.029*	0.032*	0.031**	0.044*
	(0.017)	(0.017)	(0.012)	(0.022)
post-split$\times\Delta$ ethnic fractionalization	0.003	-0.008	0.050**	-0.025
	(0.015)	(0.013)	(0.025)	(0.021)
post-split$\times\Delta$ religious diversity	0.018	-0.008	-0.003	-0.010
	(0.014)	(0.011)	(0.022)	(0.013)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.53	0.30	0.43	0.22
	Panel B: Only events with an injury or property damage			
post-split	-0.003	-0.002	0.004	-0.005
	(0.029)	(0.022)	(0.029)	(0.023)
post-split$\times\Delta$ ethnic polarization	0.034*	0.037**	0.029***	0.051**
	(0.019)	(0.016)	(0.010)	(0.022)
post-split$\times\Delta$ ethnic fractionalization	-0.002	-0.006	0.033	-0.020
	(0.024)	(0.013)	(0.024)	(0.021)
post-split$\times\Delta$ religious diversity	-0.004	-0.007	-0.018	-0.006
	(0.015)	(0.010)	(0.025)	(0.013)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.49	0.26	0.40	0.18

Notes: Panel A re-estimates our baseline specification only counting social conflict events with at least one recorded injury, death, or property damage. Panel B re-estimates our baseline specification only counting social conflict events with at least one recorded injury or property damage. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Table C.4.2: Effects Are Driven by Social Conflict Events with Injuries and Not Deaths

Administrative Unit:	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
	Panel A: Only events with an injury			
post-split	0.023	0.005	0.018	-0.002
	(0.028)	(0.021)	(0.027)	(0.022)
post-split×Δ ethnic polarization	0.014	0.030*	0.016*	0.048*
	(0.023)	(0.016)	(0.009)	(0.024)
post-split×Δ ethnic fractionalization	0.003	-0.007	0.020	-0.023
	(0.023)	(0.013)	(0.024)	(0.021)
post-split×Δ religious diversity	-0.006	-0.013	-0.031	-0.011
	(0.020)	(0.010)	(0.025)	(0.012)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.43	0.23	0.34	0.16
	Panel B: Only events with a death			
post-split	-0.019	-0.005	-0.021	0.002
	(0.025)	(0.013)	(0.021)	(0.014)
post-split×Δ ethnic polarization	-0.003	-0.003	0.002	-0.002
	(0.014)	(0.005)	(0.013)	(0.007)
post-split×Δ ethnic fractionalization	0.021	-0.006	0.020	-0.013
	(0.023)	(0.009)	(0.025)	(0.013)
post-split×Δ religious diversity	0.061***	-0.003	0.020	-0.005
	(0.010)	(0.007)	(0.028)	(0.007)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.26	0.13	0.19	0.10
	Panel C: Only events with prop. damage			
post-split	-0.005	-0.004	-0.004	-0.002
	(0.026)	(0.012)	(0.025)	(0.011)
post-split×Δ ethnic polarization	0.001	0.008	0.009	0.005
	(0.009)	(0.006)	(0.009)	(0.006)
post-split×Δ ethnic fractionalization	0.005	-0.006	0.020	-0.011
	(0.021)	(0.010)	(0.030)	(0.010)
post-split×Δ religious diversity	0.015	0.007	0.020	0.004
	(0.011)	(0.007)	(0.015)	(0.008)
Number of District-Months	7,956	20,220	7,956	12,264
Number of Districts	52	133	52	81
Dep. Var. Mean, Pre-Split	0.22	0.10	0.17	0.06

Notes: Panel A re-estimates our baseline specification only counting social conflict events with at least one recorded injury. Panel B re-estimates our baseline specification only counting social conflict events with at least one recorded death. Panel C re-estimates our baseline specification only counting social conflict events with at least some property damage. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

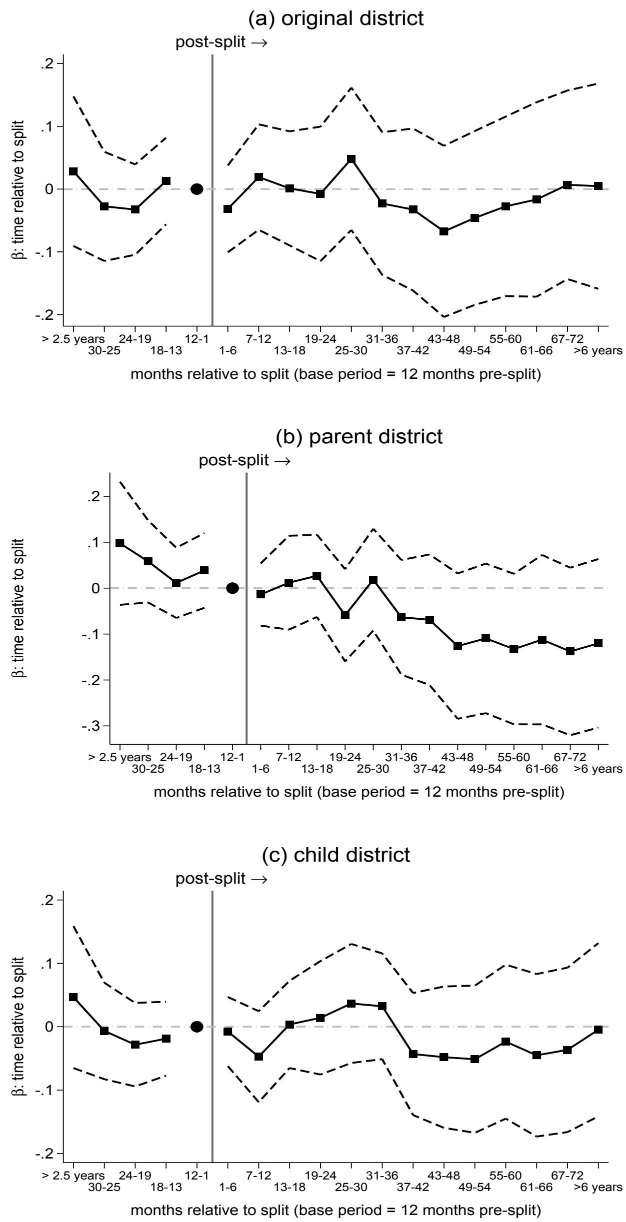
D Robustness Checks on the Main Results in Section 5

This Appendix discusses the main robustness checks discussed throughout the paper.

D.1 Event Study Specifications

We present here the event study generalization of the main equations (1) and (2) as discussed in the paper. These figures highlight both the lack of worrying pre-trends before redistricting as well as provide some insight into the post-redistricting conflict dynamics.

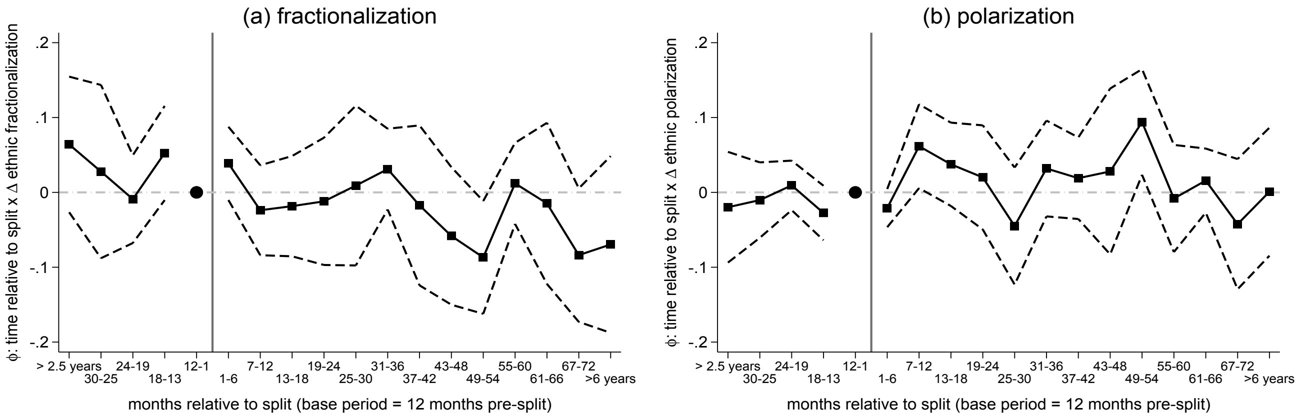
Figure D.1.1: Event Study: Average Effects of Redistricting on Social Conflict (Table 2)



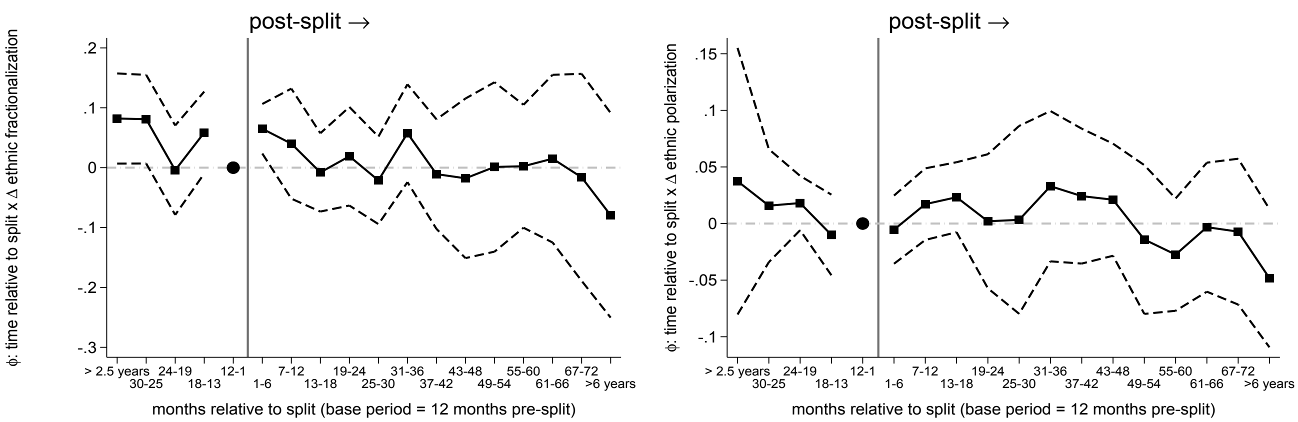
Notes: These figures report coefficient estimates and 95% confidence intervals from event study versions of the regression specification in equation (1) and given by: $conflict_{dt} = \nu + \sum_{j=-5}^{13} \beta_j post\text{-}split_{d,t-j} + \theta_t + \theta_d + \theta_d \times t + \varepsilon_{dt}$, where j denotes 6 month bins beginning 30 months prior to splitting (i.e., $j = -4$ for months 30–24 before splitting) and ending 72 months after (i.e., $j = 12$ for months 67–72 after splitting) with an additional $j = -5$ for greater than 30 months before splitting (where defined) and $j = 13$ for all months after 72. The reference period is the 12 months just prior to splitting. The graph shows the β_j coefficients.

Figure D.1.2: Event Study: Redistricting, Polarization, and Social Conflict (Table 3)

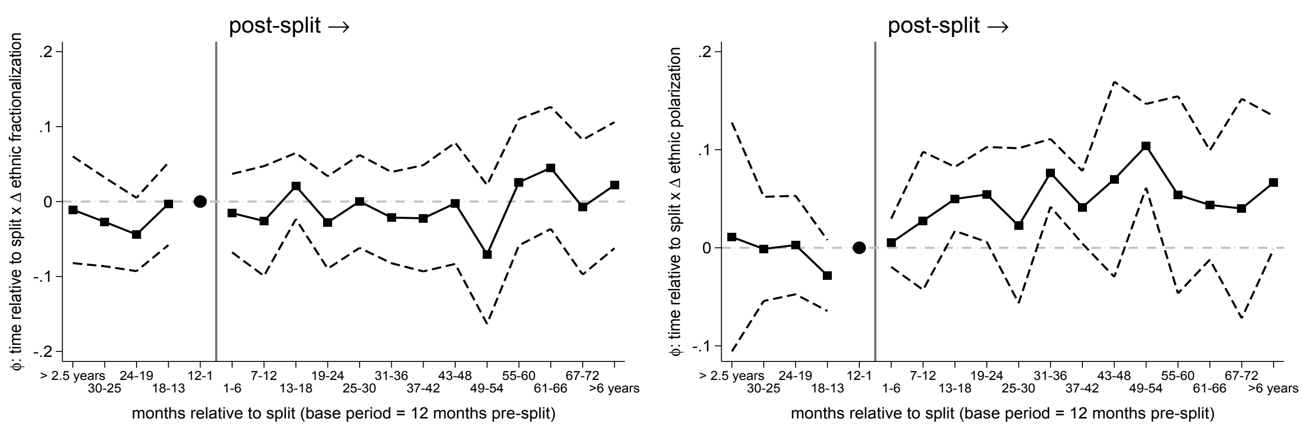
(a) Original District



(b) Parent District



(c) Child District



Notes: These figures report coefficient estimates and 95% confidence intervals from event study versions of the regression specification in equation (2) and given by: $conflict_{dt} = \nu + \sum_{j=-5}^{13}\left[\beta_j post\text{-}split_{d,t-j} + \sum_{k\in\{F,P\}} \phi_j^k \left(post\text{-}split_{d,t-j} \times \Delta k_d\right)\right] + \theta_t + \theta_d + \theta_d \times t + \varepsilon_{dt}$, where j denotes 6 month bins beginning 30 months prior to splitting (i.e., $j = -4$ for months 30–24 before splitting) and ending 72 months after (i.e., $j = 12$ for months 67–72 after splitting) with an additional $j = -5$ for greater than 30 months before splitting (where defined) and $j = 13$ for all months after 72. The reference period is the 12 months just prior to splitting. The graph shows the ϕ_j coefficients for ethnic fractionalization on the left and polarization on the right.

D.2 Ruling Out Confounding Effects of Other Initial District Characteristics

As discussed in Section 5.5, we take several steps to address the concern that the effects of $\Delta diversity$ on conflict are confounded by omitted district-specific characteristics that also differentially matter after redistricting. We follow the standard method of assessing omitted variable bias in heterogeneous effects DiD specifications, namely interacting treatment (post-split) with other factors besides the primary one(s) of interest ($\Delta diversity$) and assessing coefficient stability. The key question is how to select those variables. We consider two approaches: one, subjective and researcher-driven, and a second, more objective and machine-led. In both cases, we marshal a large set of variables across Census, administrative, and GIS-based data sources, mapping each measure to the district level of analysis in the given specification. All variables are time-invariant or predetermined as measured in 1999 or 2000.

First, we consider groups of variables plausibly correlated with diversity and conflict based on prior literature and intuition. After reproducing our baseline estimate in column (1), Tables D.2.1–D.2.4 present results based on variables broadly capturing: (2) proximity to security forces, (3) economic development, (4) public goods, (5) demographics, (6) natural resource intensity, (7) political factors, (8) economic structure, (9) geography/topography, and (10) remoteness. Across all specifications at different administrative levels, the estimated effects of $\Delta diversity$ are statistically indistinguishable from the baseline in column 1. While reassuring, these tables are nevertheless subject to researcher degrees of freedom in which variables we include and how we combine them across different columns.

Therefore, we address such concerns by taking a second, more agnostic approach to variable selection based on the double-selection post-Lasso method of Belloni et al. (2014) to identify covariates that are particularly important in explaining both diversity and social conflict. We elaborate briefly on this method here.

We assume that $post\text{-}split \times \Delta P$ and $post\text{-}split \times \Delta F$ can be taken as exogenous, once one controls linearly for a relatively small number of variables–a simple sparsity assumption. The method uses a three-step approach to help the researcher determine which controls to include. First, we select, from the set of $post\text{-}split \times control$ variables, the covariates that predict $post\text{-}split \times \Delta P$, and separately, $post\text{-}split \times \Delta F$, conditioning on the usual baseline fixed effects and $post\text{-}split \times \Delta Religious\ diversity$. This first step accounts for important confounding factors that are related to ΔP and ΔF. We use 65 $post\text{-}split \times control$ variables (detailed in Appendix F), drawn from key Indonesian data sources that cover 1999/2000 and are granular enough to construct controls at the eventual 2010 boundaries. Selection is accomplished using Lasso. The Lasso penalty parameter λ is a choice parameter, so we consider a range of values that yield a reasonable number of controls in the final step. In the second step, we select variables that predict the incidence of social conflict from the same set of $post\text{-}split \times control$ variables, again conditioning on the baseline specification. This step, also operationalized using Lasso, helps capture any important predictors of changes in violence intensity, which keeps residual variance small and can identify additional confounds. Finally, we estimate our baseline OLS equation including the union of selected controls from these two prior stages (hence post-lasso). Inference is uniformly valid for a large class of models under the assumed sparsity condition.

Table 6 showed that our main results are unchanged when including these machine-selected covariate interactions with $post\text{-}split$. The fact that these machine-chosen covariates do not alter our results provides some reassurance that the relationship between post-split changes in the incidence of violence are driven by cross-district variation in ΔP and not other observable, cross-district variation. Figure D.2.1 below shows further that these results are robust to varying the penalty parameter, λ, allowing for the inclusion of more or fewer additional covariates.[1] We see that the estimated effects of $\Delta diversity$ are fairly stable across λ despite large changes in the number of controls selected. In some cases, estimated effects drop and become noisier as we drop λ and grow the number of controls, which is to be expected.

[1] In practice, the variable selection tends to pick variables that predict $post\text{-}split \times \Delta P$ and $post\text{-}split \times \Delta F$, rather than social conflict. The full listing of included covariates in each specification, including the baseline, are available upon request.

Table D.2.1: Robustness to Additional Controls × Post-Split, Original District Level

+ controls for:	–	sec. forces	development	pub. goods	demog.	nat. res.	politics	occup.	geog.	remoteness
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
post-split	-0.012	-0.017	-0.017	-0.011	-0.014	-0.014	-0.014	-0.013	-0.012	-0.018
	(0.025)	(0.025)	(0.027)	(0.027)	(0.024)	(0.025)	(0.026)	(0.027)	(0.026)	(0.025)
× Δ ethnic polarization	0.036**	0.037*	0.034*	0.036**	0.028	0.039**	0.036*	0.036*	0.030**	0.034*
	(0.018)	(0.018)	(0.017)	(0.016)	(0.018)	(0.018)	(0.018)	(0.019)	(0.014)	(0.018)
× Δ ethnic fractionalization	-0.003	0.005	0.000	-0.003	-0.021	-0.007	-0.010	-0.013	-0.022	0.004
	(0.019)	(0.018)	(0.021)	(0.021)	(0.018)	(0.020)	(0.019)	(0.021)	(0.017)	(0.017)
× Δ religious polarization	0.014	0.021	0.015	0.030**	-0.000	0.017	0.009	0.014	0.026	0.023
	(0.013)	(0.017)	(0.013)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.019)	(0.018)
× log distance to security post		-0.009								
		(0.019)								
× log distance to police station		0.037								
		(0.024)								
× nighttime light intensity			-0.011							
			(0.018)							
× share with > primary education			-0.014							
			(0.024)							
× distance to public market				-0.007						
				(0.026)						
× share villages with electricity				-0.025						
				(0.025)						
× share villages with safe water				0.018						
				(0.026)						
× share villages with street light				0.019						
				(0.032)						
× share villages with transport center				0.055***						
				(0.014)						
× health centers per capita				-0.015						
				(0.028)						
× high schools per capita				0.020						
				(0.019)						
× log initial population					0.027					
					(0.020)					
× population share, 5–14					0.063*					
					(0.034)					
× population share, 15-49					0.052**					
					(0.026)					
× nat. resource revenue per capita						0.020*				
						(0.010)				
× cash crop share of total ag. output						0.025				
						(0.022)				
× share of land area with forest						-0.013				
						(0.016)				
× parliamentary vote polarization							-0.019			
							(0.019)			
× fiscal transfers per capita							-0.014			
							(0.016)			
× share in agriculture								-0.012		
								(0.044)		
× share in forestry/fishing								0.019		
								(0.044)		
× share in other								-0.007		
								(0.041)		
× land area									0.031**	
									(0.015)	
× share villages on coast									-0.278**	
									(0.131)	
× share villages in valley									-0.156**	
									(0.076)	
× share villages on hill									-0.199*	
									(0.104)	
× share villages on flatland									-0.239**	
									(0.115)	
× shares villages in highlands									0.026	
									(0.043)	
× log elevation									-0.004	
									(0.026)	
× log distance to coast									0.020	
									(0.041)	
× log distance to river									0.025	
									(0.033)	
× log distance to subdistrict capital										0.020
										(0.030)
× log distance to district capital										0.003
										(0.037)
× log distance to major roads										0.019
										(0.028)
Num. of Observations	7,956	7,956	7,956	7,956	7,956	7,956	7,956	7,956	7,956	7,956

Notes: This table augments our baseline specification from column 1 of Table 3 with additional interactions of *post-split* and potentially confounding initial district characteristics.

Table D.2.2: Robustness to Additional Controls × Post-Split, Parent/Child District Level

+ controls for:	–	sec. forces	development	pub. goods	demog.	nat. res.	politics	occup.	geog.	remoteness
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
post-split	-0.003	-0.007	-0.004	-0.001	-0.002	-0.003	-0.002	-0.004	-0.005	-0.009
	(0.021)	(0.021)	(0.022)	(0.019)	(0.022)	(0.020)	(0.021)	(0.023)	(0.020)	(0.021)
× Δ ethnic polarization	0.032	0.033*	0.031	0.031**	0.032	0.029	0.032	0.033	0.033	0.032*
	(0.019)	(0.019)	(0.019)	(0.015)	(0.020)	(0.021)	(0.019)	(0.020)	(0.021)	(0.019)
× Δ ethnic fractionalization	0.000	-0.002	0.001	-0.001	0.001	-0.001	0.000	-0.000	-0.003	0.005
	(0.012)	(0.011)	(0.013)	(0.013)	(0.013)	(0.011)	(0.011)	(0.012)	(0.012)	(0.013)
× Δ religious polarization	-0.009	-0.004	-0.009	-0.006	-0.009	-0.007	-0.010	-0.008	-0.010	-0.013
	(0.011)	(0.010)	(0.011)	(0.013)	(0.012)	(0.011)	(0.011)	(0.011)	(0.012)	(0.011)
× log distance to security post		-0.031*								
		(0.016)								
× log distance to police station		0.030*								
		(0.017)								
× nighttime light intensity			0.003							
			(0.015)							
× share with > primary education			-0.011							
			(0.017)							
× distance to public market				-0.004						
				(0.014)						
× share villages with electricity				-0.017						
				(0.015)						
× share villages with safe water				0.023						
				(0.016)						
× share villages with street light				0.001						
				(0.018)						
× share villages with transport center				0.053***						
				(0.014)						
× health centers per capita				0.007						
				(0.015)						
× high schools per capita				-0.018						
				(0.016)						
× log initial population					0.003					
					(0.015)					
× population share, 5–14					-0.001					
					(0.016)					
× population share, 15-49					-0.005					
					(0.020)					
× nat. resource revenue per capita						-0.016				
						(0.011)				
× cash crop share of total ag. output						0.011				
						(0.016)				
× share of land area with forest						0.006				
						(0.012)				
× parliamentary vote polarization							0.002			
							(0.013)			
× fiscal transfers per capita							-0.007			
							(0.010)			
× share in agriculture								-0.008		
								(0.022)		
× share in forestry/fishing								0.013		
								(0.023)		
× share in other								-0.011		
								(0.024)		
× land area									0.008	
									(0.024)	
× share villages on coast									0.021	
									(0.076)	
× share villages in valley									0.019	
									(0.046)	
× share villages on hill									-0.038	
									(0.072)	
× share villages on flatland									0.009	
									(0.070)	
× shares villages in highlands									0.036	
									(0.031)	
× log elevation									0.005	
									(0.016)	
× log distance to coast									0.006	
									(0.030)	
× log distance to river									-0.013	
									(0.019)	
× log distance to subdistrict capital										0.018
										(0.020)
× log distance to district capital										0.030
										(0.024)
× log distance to major roads										-0.031*
										(0.017)
Num. of Observations	20,220	20,220	20,220	20,220	20,220	20,220	20,220	20,220	20,220	20,220

Notes: This table augments our baseline specification from column 2 of Table 3 with additional interactions of *post-split* and potentially confounding initial district characteristics.

Table D.2.3: Robustness to Additional Controls × Post-Split, Parent District Level

+ controls for:	–	sec. forces	development	pub. goods	demog.	nat. res.	politics	occup.	geog.	remoteness
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
post-split	0.001	-0.006	-0.007	-0.018	-0.001	0.001	-0.004	-0.002	-0.006	-0.003
	(0.026)	(0.024)	(0.028)	(0.026)	(0.026)	(0.027)	(0.027)	(0.028)	(0.024)	(0.026)
× Δ ethnic polarization	0.027**	0.029**	0.030**	0.063***	0.036**	0.029 **	0.028**	0.031**	0.030*	0.037**
	(0.013)	(0.014)	(0.014)	(0.011)	(0.016)	(0.012)	(0.013)	(0.014)	(0.017)	(0.014)
× Δ ethnic fractionalization	0.035	0.044*	0.031	0.056***	0.032	0.041	0.040	0.033	0.045*	0.038
	(0.026)	(0.024)	(0.027)	(0.020)	(0.024)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
× Δ religious polarization	-0.031	-0.021	-0.027	-0.012	-0.040*	-0.029	-0.040 *	-0.035	-0.034	-0.032
	(0.021)	(0.022)	(0.021)	(0.018)	(0.020)	(0.019)	(0.022)	(0.023)	(0.022)	(0.023)
× log distance to security post		-0.028								
		(0.021)								
× log distance to police station		0.053**								
		(0.023)								
× nighttime light intensity			-0.020							
			(0.021)							
× share with > primary education			-0.024							
			(0.030)							
× distance to public market				-0.060**						
				(0.028)						
× share villages with electricity				-0.067**						
				(0.025)						
× share villages with safe water				-0.010						
				(0.027)						
× share villages with street light				-0.012						
				(0.022)						
× share villages with transport center				0.076***						
				(0.017)						
× health centers per capita				0.049***						
				(0.017)						
× high schools per capita				0.019						
				(0.017)						
× log initial population					0.017					
					(0.026)					
× population share, 5–14					0.058**					
					(0.027)					
× population share, 15-49					0.011					
					(0.033)					
× nat. resource revenue per capita						0.010				
						(0.009)				
× cash crop share of total ag. output						-0.005				
						(0.027)				
× share of land area with forest						-0.022				
						(0.024)				
× parliamentary vote polarization							-0.027			
							(0.019)			
× fiscal transfers per capita							0.017			
							(0.022)			
× share in agriculture								-0.019		
								(0.036)		
× share in forestry/fishing								0.008		
								(0.041)		
× share in other								-0.036		
								(0.038)		
× land area									-0.018	
									(0.025)	
× share villages on coast									0.006	
									(0.124)	
× share villages in valley									0.041	
									(0.083)	
× share villages on hill									-0.024	
									(0.128)	
× share villages on flatland									0.003	
									(0.105)	
× shares villages in highlands									0.032	
									(0.047)	
× log elevation									-0.029	
									(0.024)	
× log distance to coast									0.013	
									(0.033)	
× log distance to river									0.041	
									(0.039)	
× log distance to subdistrict capital										0.027
										(0.033)
× log distance to district capital										-0.012
										(0.039)
× log distance to major roads										0.030
										(0.029)
Num. of Observations	7,956	7,956	7,956	7,956	7,956	7,956	7,956	7,956	7,956	7,956

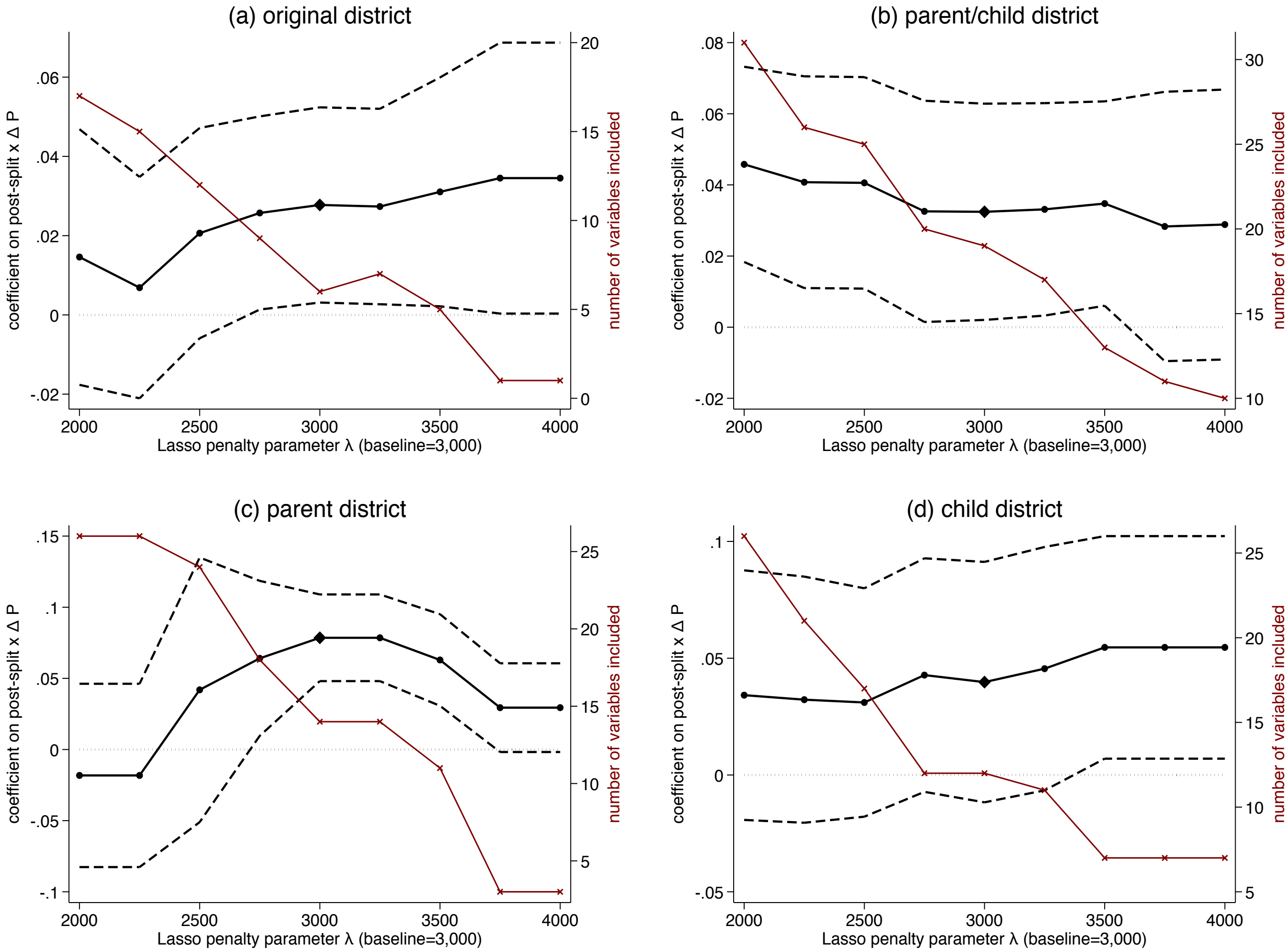
Notes: This table augments our baseline specification from column 3 of Table 3 with additional interactions of *post-split* and potentially confounding initial district characteristics.

Table D.2.4: Robustness to Additional Controls × Post-Split, Child District Level

+ controls for:	–	sec. forces	development	pub. goods	demog.	nat. res.	politics	occup.	geog.	remoteness
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
post-split	-0.005	-0.010	-0.004	-0.000	-0.004	-0.004	-0.005	-0.006	-0.010	-0.011
	(0.025)	(0.027)	(0.026)	(0.023)	(0.025)	(0.022)	(0.025)	(0.027)	(0.024)	(0.024)
× Δ ethnic polarization	0.043*	0.048*	0.042*	0.027	0.045*	0.035	0.046*	0.044*	0.055*	0.049**
	(0.025)	(0.024)	(0.023)	(0.026)	(0.025)	(0.027)	(0.023)	(0.026)	(0.028)	(0.019)
× Δ ethnic fractionalization	-0.011	-0.018	-0.012	-0.007	-0.007	-0.015	-0.006	-0.012	-0.016	-0.006
	(0.019)	(0.019)	(0.020)	(0.020)	(0.024)	(0.018)	(0.019)	(0.020)	(0.021)	(0.021)
× Δ religious polarization	-0.005	0.000	-0.006	-0.006	-0.004	0.002	-0.011	-0.004	-0.008	-0.020
	(0.014)	(0.012)	(0.014)	(0.016)	(0.016)	(0.014)	(0.015)	(0.014)	(0.017)	(0.015)
× log distance to security post		-0.034								
		(0.026)								
× log distance to police station		0.027								
		(0.022)								
× nighttime light intensity			0.013							
			(0.019)							
× share with > primary education			-0.004							
			(0.026)							
× distance to public market				0.007						
				(0.020)						
× share villages with electricity				-0.008						
				(0.024)						
× share villages with safe water				0.026						
				(0.024)						
× share villages with street light				0.011						
				(0.031)						
× share villages with transport center				0.044**						
				(0.020)						
× health centers per capita				-0.002						
				(0.026)						
× high schools per capita				-0.019						
				(0.023)						
× log initial population					-0.005					
					(0.018)					
× population share, 5–14					-0.019					
					(0.019)					
× population share, 15-49					-0.012					
					(0.027)					
× nat. resource revenue per capita						-0.034*				
						(0.018)				
× cash crop share of total ag. output						0.029				
						(0.021)				
× share of land area with forest						0.012				
						(0.015)				
× parliamentary vote polarization							0.019			
							(0.017)			
× fiscal transfers per capita							-0.012			
							(0.013)			
× share in agriculture								-0.006		
								(0.026)		
× share in forestry/fishing								0.013		
								(0.022)		
× share in other								-0.007		
								(0.037)		
× land area									0.006	
									(0.035)	
× share villages on coast									-0.001	
									(0.097)	
× share villages in valley									-0.025	
									(0.051)	
× share villages on hill									-0.072	
									(0.092)	
× share villages on flatland									-0.020	
									(0.087)	
× shares villages in highlands									0.061	
									(0.047)	
× log elevation									0.007	
									(0.024)	
× log distance to coast									0.011	
									(0.042)	
× log distance to river									-0.036	
									(0.029)	
× log distance to subdistrict capital										0.032
										(0.032)
× log distance to district capital										0.042
										(0.030)
× log distance to major roads										-0.066**
										(0.027)
Num. of Observations	12,264	12,264	12,264	12,264	12,264	12,264	12,264	12,264	12,264	12,264

Notes: This table augments our baseline specification from column 4 of Table 3 with additional interactions of *post-split* and potentially confounding initial district characteristics.

Figure D.2.1: Varying the Penalty Parameter in Lasso Robustness Procedure



Notes: This figure reports alternative estimated effects of $post\text{-}split \times \Delta P$ based on varying the penalty parameter λ used to discipline variable selection in the double Lasso procedure. Table 6 in the paper reported results for $\lambda = 3,000$ as a baseline. These figures vary that value from 2,000 to 4,000, leading to a range of variables included as seen in the red line and "x" points plotted on the right y-axis. The dashed lines are 95 percent confidence intervals on the point estimates from each individual regression.

D.3 Constraints on Redistricting and Changes in Ethnic Divisions

Section 5.5 presented a policy exercise aimed at clarifying the sources of identifying variation in $\Delta diversity$. This appendix provides further background on that exercise and also demonstrates further results consistent with the takeaway in the paper: The redistricting policy itself constrained the possible changes in ethnic divisions, and the associated effects on conflict can be explained by these policy constraints and the underlying ethnic geography rather than the particular way in which the boundaries were drawn. The concern is that better or worse borders were chosen by districts in manner correlated with latent conflict. We argue here that this is not consistent with the data.

We construct the distribution of feasible $\Delta diversity$ based on redistricting schemes that satisfied the legal restrictions in terms of the minimum number of subdistricts (3) and basic viability proxied by contiguity. This "NP-hard" problem is challenging given the large number of possible splits.[2] In order to make headway, we use a heuristic, randomized approach. Specifically, we randomly partition the district and then check to ensure the partition satisfies the contiguity requirements.[3] We repeat this process until we get 1,000 valid partitions for each original district, which we achieve for all but two original districts. Within each of the valid partitions, we then compute the corresponding ΔP and ΔF, creating a distribution of feasible ΔP and ΔF for each split. When constructing $\Delta diversity$ for parent and child districts separately, we simply assign the simulated partition with the original district capital to the parent and the residual partition(s) to the child(ren).[4] This procedure should provide a reasonably unbiased estimate of various moments of the distribution of $\Delta diversity$, taking the number of splits as given.

While some districts have relatively few feasible options, or many that result in very similar $\Delta diversity$, others have a range of feasible $\Delta diversity$. It is not obvious, in such cases, which moment of the feasible $\Delta diversity$ distribution is most appropriate. Table 6 used the mean. Results hold with the minimum or maximum.

More generally, though, the key insight we derive from this exercise is that the variation *across* districts in feasible $\Delta diversity$ swamps variation *within* districts. Indeed, stacking all random draws r for each district and regressing ΔP_{rd} on district fixed effects delivers a R^2 of nearly 0.9. While some districts certainly had choices that would result in different $\Delta diversity$, in general, regardless of their choice, their $\Delta diversity$ would differ from feasible changes in other districts. This can be seen graphically in Appendix Figure D.3.1, which plots the distribution of feasible ΔP for six districts across several major regions of Indonesia.

To formally develop this intuition, we re-estimate our baseline regressions randomly assigning each of the 50 original districts to either the minimum or the maximum of their simulated feasible $\Delta diversity$. We then repeat this a large number of times (50,000 in practice) and plot the distribution of resulting estimates for ΔP and ΔF.[5] If strategic border formation is driving our results, then the baseline estimates in Table 3 should look very different for at least some of these permutations.

Figure D.3.2 shows that this is not the case. In fact, the entire distribution of estimated effects of ΔP lies above zero and is roughly centered on our baseline estimate. This suggests that regardless of how local policymakers drew the borders, the constraints on redistricting and underlying geography limited the extent to which redistricting could reshape ethnic divisions.[6]

[2] The number of possible splits of n subdistricts (of a given original district) into k new districts given by the Stirling number of the second kind (see Fryer Jr. and Holden, 2011). For example, although Aceh Tenggara only has 255 possible partitions of its 9 subdistricts into the two new districts, Kotawaringin Timur has 4.236×10^{11} possible partitions into its three new districts (see Figure 4).

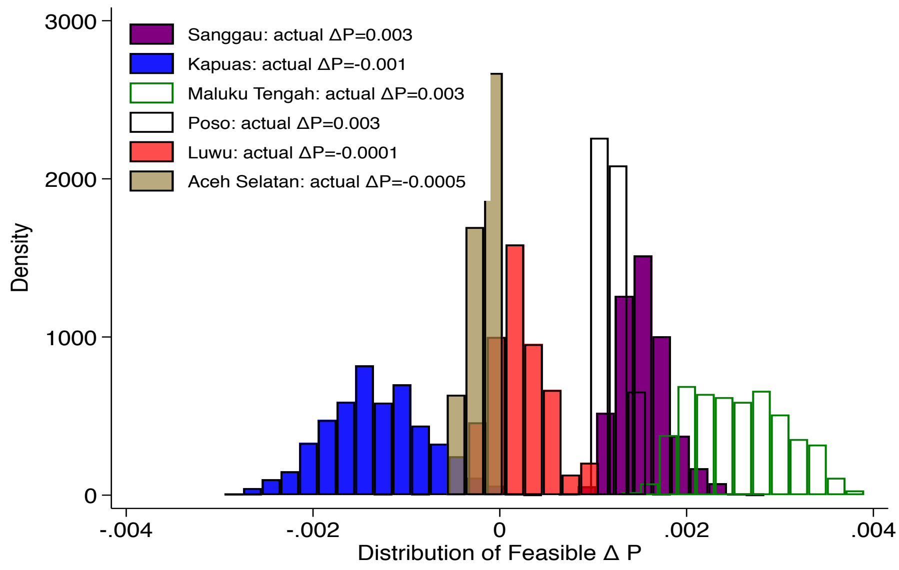
[3] Contiguity matrices are computed from shapefiles. We connect islands to the closest non-island.

[4] If there are multiple children we use the location of the eventual capital to distinguish among them.

[5] There are 2^{50} possible ways to permute min and max $\Delta diversity$ across the districts in our regressions. Given computational constraints, we randomize this 50,000 times and appeal to the law of large numbers.

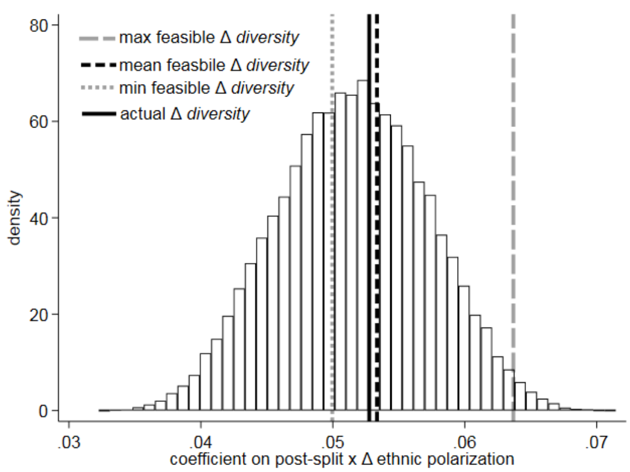
[6] These graphs look similar when including the Lasso-selected control variables used in the other robustness checks.

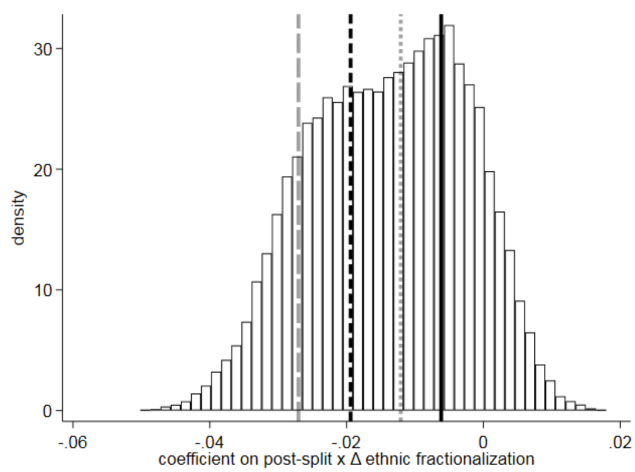
Figure D.3.1: Comparing Distribution of Feasible ΔP Across Districts



Notes: These figures plot the distribution of randomly drawn feasible ΔP for six original districts in our data.

Figure D.3.2: Distribution of Estimated Effects of Randomized Min or Max $\Delta diversity$





Notes: These figures plot the distribution of estimated effect sizes on $post\text{-}split \times \Delta diversity$ based on randomly assigning each district either its minimum or maximum feasible $\Delta diversity$ from the set of feasible partitions. We repeat this exercise 50,000 times and the bars reflect the density of each effect size (standard deviation change relative to mean outcome). The black solid line is our baseline effect size with actual $\Delta diversity$, the dashed line is based on the mean $\Delta diversity$ as reported in Table 6, and the dashed lines are based on the observed min and max $\Delta diversity$.

D.4 Outliers and Influential Observations: Robust Estimation and Inference

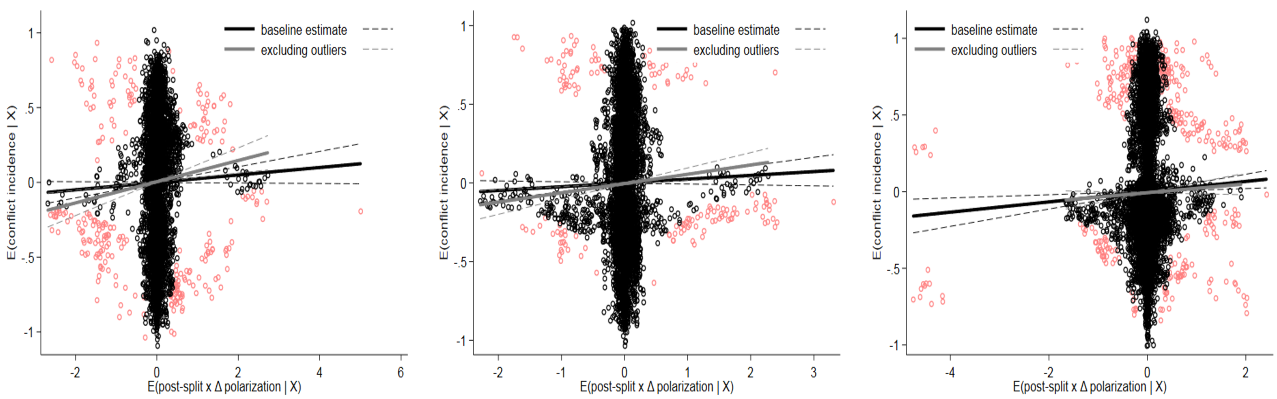
This section shows details our findings in Table 4 on outliers and robust inference. The use of district fixed effects and time trends in our main, generalized DiD specification can help rule out some of the high leverage observations that might otherwise drive the results. Nevertheless, we elaborate here why a principled approach to their removal does not change our findings and, if anything, strengthens them.

Point Estimates. We begin by demonstrating graphically how outliers affect the main results in Table 3. We adopt the widely used approach of Belsley et al. (2005) to identify observations with high influence as captured by a $dfbeta_i^k$ measure, which captures the difference between the regression coefficient θ for variable k when the ith observation is included versus excluded, with the difference being further scaled by the estimated standard error on the regressor coefficient, θ_k. Belsley et al. (2005, p. 28) recommend as a rule-of-thumb to remove all observations for which $|dfbeta_i^k| > 2/\sqrt{N}$ where N is the number of observations. Other authors recommend weaker cutoffs of 1 (Bollen and Jackman, 1990).

To visualize outliers detected using this method, Figure D.4.1 plots the baseline partial regression coefficients and scatterplot of residuals for the original district, parent and child specifications in columns 1, 3, and 4 of Table 3. The red circles identify those residuals with high $|dfbeta_i|$ for ΔP. The black lines correspond to our baseline estimate, and the gray lines are estimates based on removing the influential observations. The only regression line that seems significantly affected by the inclusion of outliers is $post\text{-}split \times \Delta P$ at the original district level, which becomes more starkly positive when removing the high-influence observations. Panel B of Table 4 presented the corresponding regression results alongside our baseline estimates for reference in Panel A.

Figure D.4.1: Principled Removal of Outliers from Baseline Estimates of Table 3

(a) Original District (b) Parent District (c) Child District



Notes: These figures present the partial regression plots for $post\text{-}split \times \Delta P$ in our baseline regressions. The black regression line and 95 percent confidence interval are the results from columns 1 (a), 3 (b), and 4 (c) of Table 3. The red observations are district-months identified by the Belsley et al. (2005) method for removing outliers described earlier. The gray regression line and 95 percent confidence interval are based on removing those observations and re-running the baseline regressions.

Inference. Besides influencing point estimates and implied effect sizes, outliers and small sample sizes more generally can also affect inference. We considered in Table 4 several alternative approaches to inference in the generalized DiD panel setup. The baseline point estimates and standard errors clustered at the original district level are as suggested by the usual Bertrand et al. (2004) motivation for clustering in fixed effects DiD designs. Below those, we presented a series of standard errors or p-values. First, we considered the Conley (1999) spatial HAC estimator that allows for contemporaneous correlation in unobservables between all districts within 500 km in addition to the usual within-district correlation over time. Results are similar using other distance bandwidth. Second, we adopt the new "effective degrees of freedom" adjustment due to Young (2016), who adjusts standard errors by the effective sample size implied by the influence of each observation.[7] Third, we implemented a cluster wild bootstrap procedure with Webb weights and 9,999 draws (Cameron et al., 2008). Finally, we take seriously the quasi-random timing of redistricting seen in Table 1 and implement a randomization inference procedure that randomly reassigns the set of three $\Delta diversity$ vector components across each of the districts in the given regression before estimation. We repeat this 10,000 times and recover the implied nearly exact p-values on our baseline estimates.[8]

Overall, the main qualitative takeaway of significant effects of ΔP remain fairly robust with the exception of the wild cluster bootstrap and the "effective degrees of freedom" adjustment. Nevertheless, as shown in Panel B of Table 4, both of these inference procedures are sensitive to outliers. Indeed, the simultaneous removal of outliers and adjustment of inference to account for remaining high influence observations delivers the most consistent evidence that ΔP exerts a significant positive effect on social conflict.

[7]This novel approach to inference delivers coefficient-specific degrees-of-freedom (DoF). For example, for ΔP, the DoF across columns 1–4 are 11.3, 4.7, 6.2, and 5.6.

[8]These are *nearly* exact as they do not recover the entire distribution of possible estimates as there 2^D possible ways to reassign $\Delta diversity$ across D districts and with a relatively large number of $D > 50$ across all specifications, this would require far longer than necessary to identify the general shape of the distribution (and size of the tails) of estimated coefficient sizes.

D.5 Validating the Conflict Measures and Ruling Out Systematic Reporting Bias

Recall that the SNPK data is based on an exhaustive and carefully vetted set of local media sources across Indonesia. However, like other conflict event sdata, the SNPK still has the potential concern that it systematically underreports violence in certain areas of the country. While we control for the number of sources being used by coders in any given province–month, we can still not completely rule out the possibility that media outlets differentially report on events in (and hence reallocate resources and reporters to) more interesting locations. If "interesting" coincides with redistricting and changes in ethnic divisions, then one might worry that we are over-estimating the effects of $\Delta diversity$ on conflict. Subjective reporting is a basic fact facing all conflict research.[9] We offer here one important robustness check on our own results that might also be fruitfully applied to others using similar data.

In particular, we draw upon Google Trends data in an attempt to rule out confounding effects of time-varying media intensity. The idea here is that the events taking place in any given district–month in our data should attract a baseline level of interest from the (internet-using) population among whom are media actors trying to follow that interest. Once we partial out that general location-specific interest in that period, the SNPK conflict report is more likely to reflect the true likelihood of any incidents rather than just a general uptick in popular (media) attention. These Google Trends, which capture the relative frequency of searches for the given district name (original, parent, or child), are indeed highly correlated with major local events such as mayoral elections.[10]

More importantly, though, our core results remain qualitatively and quantitatively unchanged when controlling for these Google Trends, which we measure on a $[0, 1]$ continuum. Table D.5.1 below shows this for our baseline results from Table 3. Results look similar for the intensive margin specifications in Table D.6.1.

Table D.5.1: Table 3 Robust to Controlling for Google Trends

Administrative Unit:	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
post-split	-0.011	-0.004	0.003	-0.007
	(0.025)	(0.021)	(0.026)	(0.025)
post-split × Δ ethnic polarization	0.040**	0.033*	0.032**	0.043*
	(0.018)	(0.019)	(0.013)	(0.025)
post-split × Δ ethnic fractionalization	-0.005	0.001	0.035	-0.011
	(0.019)	(0.012)	(0.026)	(0.019)
post-split × Δ religious diversity	0.015	-0.009	-0.030	-0.005
	(0.013)	(0.011)	(0.021)	(0.014)
Google trends	0.125*	0.071*	0.145*	0.042
	(0.067)	(0.041)	(0.073)	(0.039)
Number of District-Months	7,956	20,220	7,956	12,264
Dep. Var. Mean, Pre-Split	0.57	0.33	0.47	0.25

Notes: This table re-estimates the baseline specification controlling for monthly Google Trends in searches for each district's name. This measure takes on a value ranging from 0 to 1 indicating for each district–month the relative frequency of searches for its name when compared to other benchmark searches. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

[9] Similar concerns apply to nearly every study of conflict based on media reports, e.g. regions facing weather or commodity price shocks might draw media resources and reporters away from other areas of a given country. Studies at the country level suffer from similar concerns insomuch as they rely on either media reporting of deaths to define civil conflict/war or subjective assessments of conflict scholars as to the timing of conflict outbreaks and cessation (see Bazzi and Blattman, 2014).

[10] A fixed effects specification suggests that parent/child district names are around 10 percent more likely to be searched for during the six month window around the direct mayoral elections.

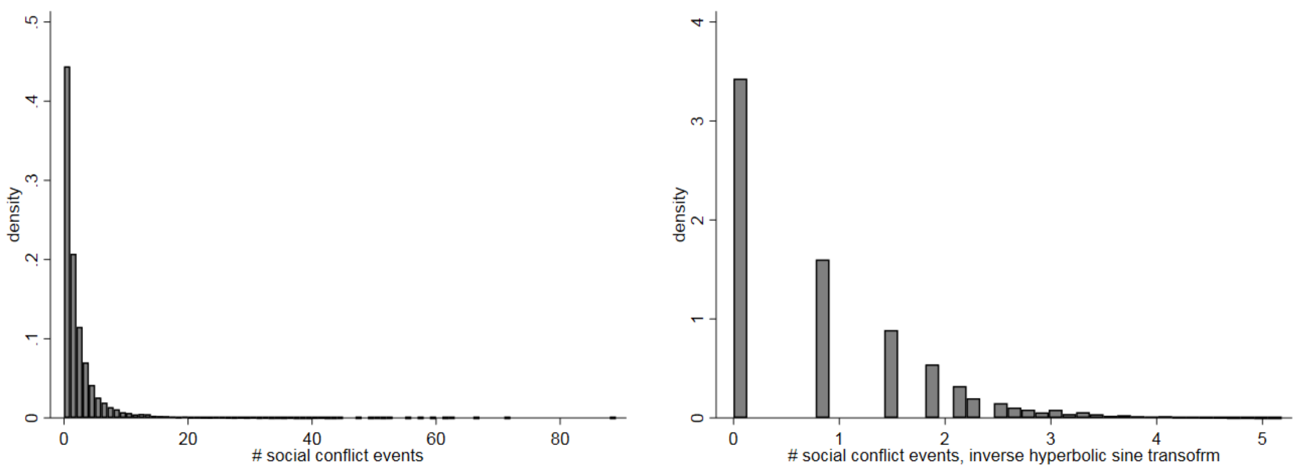
D.6 Alternative Econometric Specifications

As discussed in Section 5.6, we present here several robustness checks on the econometric specifications in Table 3.

D.6.1 Intensive Margin of Violence

Our baseline specification focuses on the extensive margin of whether there are any social conflict events in the given district–month. This is a sensible baseline given that most district-months with any events have one or two events (see the left graphs in Figure D.6.1 below). Even at the aggregate original district level—where 63 percent of district–months have any social conflict in column 1 of Table 3—80 percent of observations with any conflict have 5 or fewer events with a very long tail up to 89 events. The skewness is even starker at the more granular parent–child district level. While each of these separate event records is meant to capture a different incident, it is of course possible that they are part of the same underlying conflict, which means that the intensive margin specification might simply introduce noise. Nevertheless, there may be substantive empirical content in the intensive margin variation of incidents.

Figure D.6.1: Number of Social Conflict Incidents by Original District-Month



Notes: This figure plots the distribution of the number of social conflict events by month at the original district level. The left figure is the raw data. The right figure is the inverse hyperbolic since transformation used in the regressions.

Table D.6.1 presents intensive margin specifications based on the widely used hyperbolic inverse sine transformation, $\log\big(\#events_{dt} + (\#events_{dt}^2 + 1)^{1/2}\big)$, due to Burbidge et al. (1988). This approach to dealing with zeros has much better properties than the usual method of adding a small constant inside the log and similarly can help mitigate the effect of skewness in the outcome distribution. It also allows us to maintain the basic fixed effects OLS specification. While interpreting magnitudes is less straightforward,[11] the main takeaway from Table D.6.1 is that the results look very similar to the baseline extensive margin specification albeit slightly less precise. We increase precision by winsorizing the top 5th percentile of $\#events$ to further deal with the extreme skew (see the right graphs in Figure D.6.1).

[11] Except for very small outcome values, the transformation can be interpreted in approximately the same way as a log dependent variable.

Table D.6.1: Intensive Margin Specification: Number of Conflict Events

Administrative Unit:	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
post-split	-0.038	-0.021	-0.045	-0.008
	(0.060)	(0.033)	(0.051)	(0.032)
post-split × Δ ethnic polarization	0.021	0.041	0.030*	0.060*
	(0.031)	(0.025)	(0.018)	(0.036)
post-split × Δ ethnic fractionalization	0.010	-0.018	0.046	-0.037
	(0.046)	(0.019)	(0.049)	(0.030)
post-split × Δ religious diversity	0.031	-0.006	-0.005	-0.005
	(0.026)	(0.014)	(0.032)	(0.016)
Number of District-Months	7,956	20,220	7,956	12,264
Dep. Var. Mean, Pre-Split	0.96	0.44	0.68	0.31

Notes: The dependent variable is the hyperbolic sine transformation of the number of social conflict incidents in the given month. We winsorize at the 95th percentile of the outcome distribution. Otherwise, the specification is the same as in the baseline Table 3 with time and district FE, district-specific time trends, and standard errors clustered at the original district level. Significance levels: ∗ : 10% ∗∗ : 5% ∗ ∗ ∗ : 1%.

D.6.2 Omitting Later Entrants to SNPK Data

Table D.6.2 omits districts that enter the SNPK data in 2005, thereby ensuring a balanced panel. The similarity in results is reassuring insomuch as these later entrants were selected on account of policy concerns about recent violence.

Table D.6.2: Alternative Time Restriction: Excluding 2005 Entrants to SNPK

Administrative Unit:	Original District (1)	Parent & Child (2)	Parent (3)	Child (4)
post-split	-0.015	-0.018	-0.013	-0.022
	(0.032)	(0.027)	(0.031)	(0.031)
post-split × Δ ethnic polarization	0.046*	0.041*	0.034**	0.060*
	(0.023)	(0.024)	(0.016)	(0.031)
post-split × Δ ethnic fractionalization	-0.006	-0.008	0.025	-0.024
	(0.022)	(0.015)	(0.026)	(0.025)
post-split × Δ religious diversity	0.017	-0.005	-0.025	-0.001
	(0.016)	(0.013)	(0.021)	(0.016)
Number of District-Months	5,196	13,020	5,196	7,824
Dep. Var. Mean, Pre-Split	0.52	0.31	0.43	0.24

Notes: This table drops all districts that entered the SNPK conflict data starting in 2005, thereby imposing a balanced panel. The specification is otherwise the same as in the baseline Table 3 with time and district FE, district-specific time trends, and standard errors clustered at the original district level. Significance levels: ∗ : 10% ∗∗ : 5% ∗ ∗ ∗ : 1%.

D.6.3 Omitting Historic Conflict Zones

Table D.6.3 below excludes the two regions of Indonesia with the most intense civil conflict in the late 1990s and early 2000s at the onset of democratization. Panel A excludes districts in the Maluku islands, which saw fierce interreligious warfare from early 1999 through early 2002. Panel B excludes the entire province of Aceh, which was home to a longstanding guerilla movement to secede from Indonesia. The violent campaigned ended with a peace agreement in mid-2005. Omitting either of these two regions leaves the main takeaways intact, which is reassuring from the standpoint of generalization outside historic conflict zones per se.

Table D.6.3: Excluding Civil War Regions

Administrative Unit:	Original District	Parent & Child	Parent	Child
	(1)	**(2)**	**(3)**	**(4)**
	Panel A: Excluding Maluku			
post-split	-0.025	-0.020	-0.014	-0.026
	(0.029)	(0.023)	(0.028)	(0.029)
post-split × Δ ethnic polarization	0.067***	0.044***	0.043***	0.051***
	(0.021)	(0.008)	(0.012)	(0.014)
post-split × Δ ethnic fractionalization	-0.016	0.009	0.071**	-0.002
	(0.021)	(0.012)	(0.035)	(0.017)
post-split × Δ religious diversity	0.021	-0.012	-0.026	-0.015
	(0.013)	(0.016)	(0.026)	(0.018)
Number of District-Months	6,900	17,100	6,900	10,200
Dep. Var. Mean, Pre-Split	0.64	0.38	0.53	0.29
	Panel B: Excluding Aceh			
post-split	0.027	0.019	0.035	0.009
	(0.023)	(0.020)	(0.029)	(0.024)
post-split × Δ ethnic polarization	0.026*	0.032	0.024	0.048*
	(0.014)	(0.022)	(0.015)	(0.026)
post-split × Δ ethnic fractionalization	0.009	0.009	0.037	0.001
	(0.021)	(0.013)	(0.030)	(0.019)
post-split × Δ religious diversity	0.032**	-0.001	-0.017	0.001
	(0.015)	(0.012)	(0.027)	(0.015)
Number of District-Months	6,696	17,340	6,696	10,644
Dep. Var. Mean, Pre-Split	0.55	0.30	0.45	0.21

Notes: Panel A excludes the districts in the Maluku islands, and B excludes districts in Aceh. The specification in both panels is otherwise the same as in the baseline Table 3 with time and district FE, district-specific time trends, and standard errors clustered at the original district level. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

D.6.4 Alternative Identification Strategy

Table D.6.4 considers an alternative strategy that removes the district fixed effects altogether and incorporates an additional control group in those districts that never split from 2000 to 2010. In this case, the coefficients identify a mixture of differential effects of redistricting (i) among early versus late splitters and never-splitters, and (ii) splitters versus never-splitters. We include province×month fixed effects to ensure that each of these comparisons takes place between nearby districts. Note that the removal of district fixed effects means that we not longer identify within-district changes in conflict but rather cross-sectional differences (i) and (ii) within a given province akin to a nearest (geographic) neighbor matching-type design. In other words, all estimates below are with reference to the mean social conflict among never-splitters within the same province in a given month. Interestingly, as seen in Table D.6.4, this alternative identification strategy delivers estimated effects of changes in ethnic divisions among those that undergo redistricting that looks very similar to the baseline.

Table D.6.4: Expanding the Counterfactual to Include Nearby Never-Splitters

Administrative Unit:	Original District	Parent & Child	Parent	Child
	(1)	**(2)**	**(3)**	**(4)**
post-split	0.067	-0.146***	-0.102*	-0.249***
	(0.051)	(0.044)	(0.052)	(0.050)
post-split × Δ ethnic polarization	0.038***	0.030**	0.057**	0.047*
	(0.011)	(0.012)	(0.023)	(0.024)
post-split × Δ ethnic fractionalization	0.012	0.031**	-0.006	0.028
	(0.029)	(0.015)	(0.037)	(0.023)
post-split × Δ religious diversity	0.015	0.011	-0.039	0.030
	(0.016)	(0.015)	(0.040)	(0.022)
Number of District-Months	22,896	35,160	22,896	27,204
Dep. Var. Mean, Pre-Split	0.62	0.55	0.61	0.56

Notes: This table adds districts in the SNPK data that never split over the sample period. Relative to the baseline Table 3, the specification therefore omits district FE and district-specific time trends but includes province× month FE. Standard errors clustered at the original district level. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

D.6.5 Reweighting for External Validity (within Indonesia)

As discussed in Section 5.6, we consider a reweighting approach to account for the two dimensions of selection in our generalized DiD sample of districts. First, Table D.6.4 above notwithstanding, our results are based solely on districts that split between 2000 and 2010. We know from prior work and our analysis that this sample of districts look different along a number of observable dimensions. One way to assess whether these differences are important for our results is to reweight the estimates by the inverse probability of redistricting, assigning greater weight to those original districts that look more like the ones that never split. We do this using a standard inverse probability weighting (IPW) approach in which we first predict the cross-sectional probability of redistricting and then apply IPW to our main estimates from Table 3. We estimate that probability based on a logit specification, using Lasso to select relevant variables from a set of 67 baseline district characteristics used for other exercises in Section 5.5. Applying the IPW in Table D.6.5 leaves our key results unchanged.

Table D.6.5: Inverse Probability Weighting (IPW) for External Validity

Administrative Unit:	Original District	Parent & Child	Parent	Child
	(1)	**(2)**	**(3)**	**(4)**
post-split	-0.019	-0.016	-0.001	-0.028
	(0.027)	(0.019)	(0.026)	(0.024)
post-split × Δ ethnic polarization	0.036	0.037*	0.030*	0.057**
	(0.026)	(0.021)	(0.015)	(0.027)
post-split × Δ ethnic fractionalization	-0.021	0.003	0.053**	-0.019
	(0.037)	(0.014)	(0.024)	(0.023)
post-split × Δ religious diversity	0.001	-0.014	-0.033	-0.013
	(0.014)	(0.013)	(0.022)	(0.017)
Number of District-Months	7,956	20,220	7,956	12,264
Dep. Var. Mean, Pre-Split	0.57	0.33	0.47	0.25

Notes: This table reweights each observation in the baseline specification in Table 3 by its inverse probability of redistricting. The IPW are estimated in an initial step based on a logit specification with a battery of Lasso-selected controls. Standard errors are clustered at the original district level Significance levels: $*$: 10% $**$: 5% $***$: 1%.

E Additional Evidence Supporting Political Violence Results in Section 6

E.1 Further Background on Ethnicity in Indonesian Politics

Section 2.3 offered background on the changing role of ethnicity in Indonesian politics that could be broadly summarized in the following four takeaways: (i) ethnicity is an important organizing principle for political mobilization, (ii) ethnic-based clientilism and patronage networks are pervasive, (iii) decentralization and direct, majoritarian mayoral elections deepen (i) and (ii), and (iv) redistricting further amplifies all of these forces. Here, we provide additional background from the political science literature as well as fresh empirical evidence consistent with this context.

Political Science Literature. Wilson (2015, p. 92) offers a helpful summary of views on ethnicity and patronage in the context of redistricting: *"As local government and administrative boundaries were altered, 'local selfishness' was reinforced, resulting in conflicts and tensions at the local level (Firman 2013, 180). Just like national politics, local-level politics was an intense 'arena of contestation between competing coalitions of social interests' as networks that had relied upon central state patronage or been regime middlemen moved to establish new means to access resources (Hadiz, 2011a, 171). This contestation involved renegotiating the boundaries of collective identities, in doing so defining a social economy of who had to access to what, and under what circumstances. According to Klinken, from 1998 local elites throughout the country attempted to build 'an exclusive discourse of ethnicity', one that in its construction of group identity formed a 'language with which elites compete for power by mobilising supporters' (Klinken 2002, 68)."*

In the context of our case study in Section 6.4, Kobayashi (2011) notes from personal interviews that *"A Dayak politician, a strong supporter of the creation of Bengkayang district, clearly explained that increase of Dayak government employees was one objective of* pemekaran *[redistricting]. A Dayak department head admitted that* pemekaran *increased job opportunities for Dayaks in government by commenting that he himself would not have been promoted to the position of department head without creation of Bengkayang.*

Diversity and Close Elections. Table E.1.1 demonstrates that ethnic diversity is associated with significantly closer mayoral elections. In particular, we regress the victory margin for the winning candidate on ethnic and religious diversity within the newly created parent and child districts. We consider both the first and second (when possible) quinquennial direct election after redistricting.[1]

The main takeaway is that greater diversity is associated with closer elections, consistent with the importance of ethnic mobilization highlighted in recent literature. Column 1 shows this when pooling across both the first and second elections taking place in the new parent and child districts. Both polarization (P) and fractionalization (F) matter, though the former is more precisely estimated. The effect sizes, though, are not trivial. A one standard deviation increase in P or F is associated with 10 percent lower victory margin relative to a mean of around 0.14 across all elections from 2005–2014 in these new districts. Results look similar if not slightly more pronounced for second elections.

Looking separately at parent and child contests, however, reveals a difference between P and F. In parent districts, fractionalization matters much more than polarization whereas the opposite holds for child districts. These patterns line up nicely with the results in Table 7 where P does more than F to exacerbate violence around elections in child districts whereas the opposite holds for parent districts. Though perhaps consistent with differences in the stakes of political control (as noted in Section 6), these differences between P and F in Table E.1.1 are not statistically significant. The patterns for religious polarization are similarly in line with the differential effects of $\Delta religious\ diversity$ for parent relative to child districts in Table 7. Together, the results are suggestive about the importance of group configuration in driving electoral competition.[2]

[1] As discussed in Appendix F, several newly created districts had not yet had their second election by the end of our study period, while others have missing data on election outcomes.

[2] It is also worth noting that we can estimate the relationship between $\Delta diversity$ and $\Delta victory\ margins$ for 22 parent and child districts with a direct election at the original district level prior to redistricting. In particular, we find that a one standard

Table E.1.1: Diversity and Close Elections After Redistricting

	Dependent Variable: Victory Margin for Winning Mayoral Candidate in								
Administrative Unit	Parent/Child			Parent			Child		
Which Election?	All	1st	2nd	All	1st	2nd	All	1st	2nd
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
ethnic polarization, new district	-0.013***	-0.012*	-0.031***	-0.012	-0.003	-0.014	-0.015***	-0.016*	-0.029**
	(0.004)	(0.007)	(0.009)	(0.014)	(0.016)	(0.021)	(0.005)	(0.008)	(0.013)
ethnic fractionalization, new district	-0.006	-0.006	-0.006	-0.008	-0.029	-0.031	0.001	0.004	0.026
	(0.010)	(0.012)	(0.015)	(0.015)	(0.030)	(0.021)	(0.013)	(0.015)	(0.018)
religious polarization, new district	-0.015*	-0.019*	-0.028**	-0.042**	-0.015	-0.071***	-0.002	-0.016	-0.001
	(0.009)	(0.010)	(0.013)	(0.018)	(0.023)	(0.026)	(0.009)	(0.012)	(0.015)
Number of Districts	113	103	68	44	34	32	69	69	36
Mean Victory Margin	0.14	0.14	0.13	0.14	0.14	0.14	0.14	0.14	0.12

Notes: This table presents simple regressions relating ethnic diversity in the newly created parent/child districts to the victory margin in the first and second direct mayoral elections post-redistricting. Columns 1–3 pool parent/child districts, and columns 4–9 examine each separately. The $\Delta diversity$ measures are normalized, and standard errors are clustered at the original district level. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Ethnic Divisions and Preferences for Mayoral Candidates. We draw upon the *Indonesia Family Life Survey* (IFLS) to provide some evidence in line with these claims as they relate to border-induced changes in ethnic and religious divisions. In particular, we draw upon the 2014 round of data, which asks individuals "What factors do you consider in electing a mayor?". We observe individuals in 40 of the parent and child districts in our main sample. In Table E.1.2 below, we control for basic demographics and relate $\Delta diversity$ to preferences over a large set of mayoral qualities. The results suggest that changes in ethnic divisions as a result of redistricting are strongly associated with preferences for mayor's ethnicity as well as their provision of patronage. We find weaker correlations with mayoral experience, political affiliation and proposed program quality, among others. Note that this observation is at the end of the study period by which time many of these districts have had multiple mayoral elections, some of which may have been among those that witnessed violence of the sort identified in Section 6 of the paper.

Table E.1.2: Changes in Ethnic Divisions and Preferences for Mayoral Candidates

	Dep. Var. (binary): Respondent in 2014 Believes that the Mayor's ... Is Important							
	Appearance	Popularity	Program Quality	Political Affiliation	Religion	Ethnicity	Experience	Patronage
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Δ ethnic polarization	0.029**	-0.010	-0.002	0.032	0.087**	0.086***	-0.012	0.045***
	(0.013)	(0.014)	(0.010)	(0.020)	(0.037)	(0.023)	(0.009)	(0.012)
Δ ethnic fractionalization	-0.002	-0.029**	0.013	0.014	0.056	0.044**	-0.006	0.025***
	(0.012)	(0.013)	(0.010)	(0.016)	(0.045)	(0.017)	(0.009)	(0.006)
Δ religious polarization	0.011	0.008	-0.003	-0.001	0.031	0.004	-0.013***	0.002
	(0.009)	(0.017)	(0.009)	(0.014)	(0.026)	(0.012)	(0.005)	(0.004)
Number of Districts	1,887	1,887	1,887	1,887	1,887	1,887	1,887	1,887
Number of Districts	40	40	40	40	40	40	40	40
Dep. Var. Mean	0.75	0.76	0.93	0.70	0.77	0.60	0.93	0.46

Notes: The dependent variable in each column is a binary indicator that equals one if the respondent in the 2014 IFLS agrees that the mayoral candidates' given trait is an important factor in determining his/her vote. The regressions control for age, age squared, education level fixed effects, and gender. The $\Delta diversity$ measures are normalized, and standard errors are clustered at the district level. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

deviation increase in ΔP (ΔF) is associated with a 2.2 p.p. (3.3 p.p.) reduction in $\Delta victory\ margin$ relative to its mean of 5.3 percent. There are only 10 prior elections and hence it is not meaningful to conduct inference, but the patterns are nevertheless supportive of the level results in Table E.1.1.

E.2 Close Elections and Conflict

The first set of results in Table E.2.1 demonstrates that violence is more likely around new elections after redistricting when those contests are closely contested. In particular, we interact the post-split×election period indicator with the victory margin (ranging from 0.004 to 0.55). Panel A examines the baseline outcome of any social conflict, and Panel B examine the intensive margin number of conflict incidents transformed via the inverse hyperbolic sine used in baseline robustness checks in Appendix D.6.1. This latter specification allows for the possibility that the intensive margin may be differentially more important around election periods, which may be generally more intense periods of violence. Together, these results are broadly consistent with the fact that victory margins are significantly lower in more diverse, newly created districts as seen in Appendix Table E.1.1.

Table E.2.1: Differential Conflict Around Close Elections After Redistricting

Administrative Unit	Parent Child	Parent	Child
	(1)	**(2)**	**(3)**
	Panel A: Any Social Conflict		
post-split	0.020	-0.002	0.025
	(0.029)	(0.038)	(0.036)
post-split × 1st election period	0.069	-0.030	0.091**
	(0.049)	(0.063)	(0.035)
post-split × 1st election period × victory margin	-0.259	0.077	-0.427**
	(0.155)	(0.201)	(0.206)
Number of District-Months	18,120	7,176	10,944
Dep. Var. Mean	0.35	0.51	0.26
	Panel B: # Social Conflict Events Hyperbolic Inverse Sine		
post-split	0.016	-0.050	0.053
	(0.059)	(0.105)	(0.050)
post-split × 1st election period	0.204**	0.236*	0.106*
	(0.097)	(0.126)	(0.053)
post-split × 1st election period × victory margin	-0.652*	-1.465***	-0.306
	(0.341)	(0.515)	(0.355)
Number of District-Months	18,120	7,176	10,944
Dep. Var. Mean	0.52	0.84	0.34

Notes: This table examines interactions of the first mayoral election period with the victory margin in that election. The interaction of post-split and that victory margin is included but not shown. The specification is otherwise similar to the one in Table 7. Significance levels: $*$: 10% $**$: 5% $***$: 1%.

Looking across specifications, the evidence in Table E.2.1 suggests that after redistricting, violence is significantly more pronounced around close mayoral elections. These patterns are consistent with both (i) the qualitative background on election violence and incident descriptions discussed in Appendix C.2, and (ii) the conflict-amplifying effects of $\Delta diversity$ around elections seen in Table 7. While victory margins are potentially endogenous with respect to contemporaneous electoral violence, these results provide an important validation check on our interpretation. Together with the results linking ethnic diversity to closer elections , these findings paint a rich picture of how (changes in) ethnic divisions reshape conflict dynamics in settings with high returns to local political control.

F Data and Variable Construction

We describe here the key variables and data sources used in the paper.

Administrative Divisions

Indonesia's administrative divisions proceed down from the province to the district to the subdistrict to the village. These different levels of administration and our terminology for original, child and parent districts as defined below can be seen in Figure 3, which shows one of the districts in our study.

Original District: This administrative unit defines all areas based on the 2000 boundaries.

Child District: This represents the subdistricts that eventually become their own new district with an accompanying capital.

Parent District: This represents the subdistricts that stay with the original district capital after other subdistricts split off.

Post-Split: This is an indicator that turns on in the month that national parliamentary legislation first established a new district within the original district boundaries. In our main results, post-split equals one for the original district and parent district once the first child district splits off from 2000 onward. For child districts, the indicator equals one once it is ratified into law.

Conflict

The conflict data comes from the Indonesian National Violence Monitoring System (known by its Indonesian acronym SNPK).[1] The data are reported at or below the 2011 district level, and hence we can calculate conflict within both the 2010 and 2000 borders over the years 2000–2014. Our main conflict measures are binary indicators for any conflict in a given district–month, but we also consider the number of incidents as a robustness check. Coders read articles and then assign the incident to mutually exclusive categories based on the underlying trigger. The incidents are first coded as domestic violence, violent crime, violence during law enforcement, or conflict. Eighty-two percent of incidents record some property damage, injuries, or deaths.

Any Social Conflict: A dummy for whether SNPK recorded any non-crime and non-domestic violence incidents in the given month.

Active Media: Using data obtained directly from SNPK managers on newspaper availability and usage by province and month, we calculate the number of papers used in any given province-month. All conflict specifications control flexibly for media availability by including dummies for the number of active papers in any given province-month.

Entered 2005: SNPK coverage begins in 1998 for nine conflict-prone provinces and increases to 15 provinces plus parts of 3 provinces in greater Jakarta beginning in 2005. The data coverage is less complete and reliable for 1998 and 1999, and hence we focus on 2000–2014 for most results in the paper.

[1] We downloaded the data from http://www.snpk-indonesia.com, which is no longer active due to a recent contracting change. However, as of June 2016, the data hosted on and available through the World Bank website. A search in their Central Microdata Catalog for "Sistem Nasional Pemantauan Kekerasan" yields data from 1998 to 2014.

Diversity

All measures are computed using the universal 2000 Population Census. Since this contains data at the village level, metrics can be constructed at both the 2000 and 2010 borders.

Ethnic Fractionalization: Ethnic fractionalization in district d is given by $F = \sum_{j=1}^{\mathcal{N}_e} g_j(1-g_j)$, where $\mathcal{N}_e$ is the number of ethnic groups in the district, and g_j is the population share of group j as reported in the 2000 Census. We observe over 1000 ethnicities and sub-ethnicities speaking over 400 languages. We also consider the related Greenberg-Gini version, which allows for non-binary distances between groups: $G = \sum_{j=1}^{\mathcal{N}_e}\sum_{k=1}^{\mathcal{N}_e} g_j g_k \eta_{jk}$, where η_{jk} captures the linguistic distance between groups j and k as detailed below.

Ethnic Polarization: $P = \sum_{j=1}^{\mathcal{N}_e}\sum_{k=1}^{\mathcal{N}_e} g_j^2 g_k \eta_{jk}$, where $\mathcal{N}_e$, g_j, and g_k are as defined before, and η_{jk} is the distance between groups j and k. We map each ethnic group in the 2000 Census to a language in *Ethnologue*, which provides a full classification of the linguistic origins of each language (see the Online Appendix Section A.3 in Bazzi et al., 2016, for details). We set $\eta_{gh} = 1 - s_{gh}^{\delta}$, where s_{gh} is the degree of similarity between the languages spoken by g and h as given by the ratio of common branches on the language classification tree to the maximum possible (14), and δ is a parameter that selects the level of linguistic dissimilarity to be emphasized. We set $\delta = 0.05$ in our baseline, but consider alternate values. Ethnicities with missing languages are given province-specific average pairwise distances (η's) between all other languages. Missing ethnic groups are necessarily grouped together, but separately from the "other" category, and also given province-specific average distances. We drop foreigners as they represent a minute fraction of the population, but we retain the ethnic Chinese.

Religious Polarization: Religious polarization, $Relig = \sum_{j=1}^{\mathcal{N}_r}\sum_{k=1}^{\mathcal{N}_r} g_j^2 g_k$, where $\mathcal{N}_r$ is the number of religious groups, and g_j (g_k) is the population share of group j (k). There are seven religions recorded in the Census, but in most districts, there is a single cleavage between a Muslim and a non-Muslim group. As a result religious polarization is effectively identical to religious fractionalization in our data (with a correlation of 0.96).

Δ **Ethnic Polarization**: To examine changes in diversity at the original district level, we compute the population-weighted average polarization in the new units (children and parent district) and subtract the polarization in the original district. If original district $\mathcal{O}$ splits into parent $\mathcal{P}$ and child(ren) $\mathcal{C}_1$ ($\mathcal{C}_2$ if multiple), with populations $G_{\mathcal{O}} = G_{\mathcal{P}} + G_{\mathcal{C}_1}(+G_{\mathcal{C}_2})$ the change in ethnic polarization is $\Delta P = \left(\frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}P_{\mathcal{P}} + \frac{G_{\mathcal{C}_1}}{G_{\mathcal{O}}}P_{\mathcal{C}_1} + \frac{G_{\mathcal{C}_2}}{G_{\mathcal{O}}}P_{\mathcal{C}_2}\right) - P_{\mathcal{O}}$. We construct changes in ethnic polarization at the child/parent level analogously as: $\Delta P = P_{\mathcal{P}} - P_{\mathcal{O}}$ for the parent and $\Delta P = P_{\mathcal{C}} - P_{\mathcal{O}}$ for each child.

Δ **Ethnic Fractionalization**: For original district $\mathcal{O}$ splitting into parent $\mathcal{P}$ and child(ren) $\mathcal{C}_1$ ($\mathcal{C}_2$ if multiple), with populations $G_{\mathcal{O}} = G_{\mathcal{P}} + G_{\mathcal{C}_1}(+G_{\mathcal{C}_2})$ the change in ethnic fictionalization is given by $\Delta F = \left(\frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}F_{\mathcal{P}} + \frac{G_{\mathcal{C}_1}}{G_{\mathcal{O}}}F_{\mathcal{C}_1} + \frac{G_{\mathcal{C}_2}}{G_{\mathcal{O}}}F_{\mathcal{C}_2}\right) - F_{\mathcal{O}}$. We construct changes in ethnic fractionalization at the child/parent level analogously as: $\Delta F = F_{\mathcal{P}} - F_{\mathcal{O}}$ for the parent and $\Delta F = F_{\mathcal{C}} - F_{\mathcal{O}}$ for each child.

Δ **Religious Polarization**: For original district $\mathcal{O}$ splitting into parent $\mathcal{P}$ and child(ren) $\mathcal{C}_1$ ($\mathcal{C}_2$ if multiple), with populations $G_{\mathcal{O}} = G_{\mathcal{P}} + G_{\mathcal{C}_1}(+G_{\mathcal{C}_2})$ the change in religious polarization is given by $\Delta Relig = \left(\frac{G_{\mathcal{P}}}{G_{\mathcal{O}}}Relig_{\mathcal{P}} + \frac{G_{\mathcal{C}_1}}{G_{\mathcal{O}}}Relig_{\mathcal{C}_1} + \frac{G_{\mathcal{C}_2}}{G_{\mathcal{O}}}Relig_{\mathcal{C}_2}\right) - Relig_{\mathcal{O}}$. We construct changes in ethnic fractionalization at the child/parent level analogously as: $\Delta Relig = Relig_{\mathcal{P}} - Relig_{\mathcal{O}}$ for the parent and $\Delta Relig = Relig_{\mathcal{C}} - Relig_{\mathcal{O}}$ for each child.

Table F.1.1: Summary Statistics for Baseline Variables

	Mean	Std. Dev.	Min.	Median	Max.
	2000 Borders: 52 Original Districts				
any social conflict incidents	0.631	0.483	0.000	1.000	1.000
number of social conflict incidents	2.631	5.185	0.000	1.000	89.000
post-split	0.787	0.409	0.000	1.000	1.000
ethnic polarization	0.017	0.016	0.003	0.013	0.095
ethnic fractionalization	0.612	0.256	0.062	0.689	0.957
religious polarization	0.119	0.070	0.001	0.130	0.233
Δ ethnic polariation	-0.000	0.011	-0.062	0.000	0.061
Δ ethnic fractionalization	-0.078	0.153	-0.677	-0.034	0.193
Δ religious polarization	-0.008	0.049	-0.192	-0.000	0.109
	2010 Borders: 133 Parent and Child Districts				
any social conflict incidents	0.364	0.481	0.000	0.000	1.000
number of social conflict incidents	1.035	2.941	0.000	0.000	76.000
post-split	0.768	0.422	0.000	1.000	1.000
ethnic polarization	0.017	0.016	0.003	0.013	0.095
ethnic fractionalization	0.609	0.258	0.062	0.682	0.957
religious polarization	0.122	0.067	0.001	0.131	0.233
Δ ethnic polarization	-0.000	0.005	-0.035	0.000	0.008
Δ ethnic fractionalization	-0.059	0.083	-0.342	-0.032	-0.000
Δ religious polarization	-0.008	0.020	-0.129	-0.001	0.017

Notes: At the 2000 level, there are 52 districts and 7,956 monthly observations. At the 2010 level, there are 133 Districts (52 parents and 81 children) and 20,220 monthly observations. See Appendix F for variable definitions.

Voting and Elections

District Head Elections: District elections occur every 5 years. Prior to 2005, district head elections were conducted by parliament and varied across districts in terms of timing. From 2005 onward, district and vice-district heads were directly elected by plurality vote contingent on that vote being at least 30 percent. If not, a second round between the top two candidates takes place. District heads directly appoint subdistrict heads. We collect data on the date of and vote shares in all direct elections from documents published by the General Election Commissions, many of which were graciously provided by Monica Martinez-Bravo, Andreas Stegmann, and Audrey Sacks. Elections in child districts typically occur 1.5–2.5 years after the split. Elections in parent districts are determined by the pre-Suharto election cycles carried over into the democratic era (see Martinez-Bravo et al., 2017).

1st Direct Election Period: Using the exact date of all direct elections, we construct an indicator that equals one in the 6 month window around the parent/child's first direct election date. In the case of the latest splits, this can occur pre-split.

2nd Election Period: We construct an indicator that equals one in the 6 month window around the parent/child's second direct election date. There are some children (the latest splits) for which we do not observe a second election post-split.

District Head Election Victory Margins: Using the General Election Commissions records, we compute victory margins in the district head elections conducted after redistricting. This continuous measure is simply equal to the vote share for the winner minus the vote share for the loser.

Control Variables

We list here the rich set of 65 variables from 1999 and 2000 that we interact with $post - split$ and use as controls to ensure that the cross-district variation picked up by $post - split \times Diversity$ is not picking up other observable differences across districts. These are carefully constructed from a variety of data sources, and are generally non-missing. Several variables are missing for at most one original district, and are imputed simply using the average across districts.

PODES Variables

We use the 2000 administrative village census (*Potensi Desa* or *Podes*) to construct a number of control variables relating to education, public goods provision, security, and development. Each of these measures are aggregated to the district level at both the original district level, and eventual, 2010 boundaries.

Health Variables: We construct a variable for the number of health care facilities (polyclinics and PHCs) per capita in 2000 at the 2000/2010 district levels. We construct the (population weighted) share of villages that say they have a midwife available. Further, we construct the (population weighted) share of villages that say they have a doctor or access to a PHC.

Education Variables: We construct the number of high schools per capita in 2000 at the 2000/2010 district levels. We also construct the number of Islamic schools per capita.

Public Goods: We construct the (population weighted) share of villages that have access to water from a pump or a water company; have a trash disposal system (bin/hole); have most households using gas/kerosene or electricity; and have road lighting. We also use the number of households per capita with electricity, with a telephone, and with a television.

Economy: We construct the number of permanent markets per-capita and the (population weighted) average distance to the nearest market. In addition we calculate the (population weighted) share of villages with a transportation hub (airport, seaport, or bus terminal). We also construct the (population weighted) share of villages reporting good or great economic conditions and the share of villages for which agriculture is the main source of income. Finally, we construct the (population weighted) average number of natural disasters in the past 3 years.

Security: We construct the (population weighted) mean distance to the nearest police post and office. We construct two variables: the logarithm of (one plus) the distance to the nearest police outfit and the logarithm of (one plus) the distance to the nearest police office (which is always larger).

Geography: We construct the (population weighted) share of villages on the shore, on the coast, in a valley, on a hill, on flat land, and at high altitude. We also construct the logarithm of total land area. Importantly, we also include the logarithm of (one plus) the (population weighted) mean distance from the village to the 2000 capital and the logarithm of (one plus) the (population weighted) mean distance from the village to the sub-district capital.

Census Variables

Using the 2000 population micro census we construct a number of additional demographic variables. We construct each of the below at both the original district and the eventual 2010 boundaries.

Population Shares: We use the Population Census in 2000 to compute the share of the population that is aged 5–14 and 15–29 at the original, child, and parent district levels. We also include the logarithm of total population and mean household size.

Education Shares: We compute the share of the population whose highest educational attainment is primary school, as well as the share of the population whose highest educational attainment is post-primary.

Migration: We compute the share of the population who arrived from a different province in the last five years and the share arrived from a different district in the last five years.

Geography: We include an indicator for the share of the population living in rural areas.

Sectors of the Economy: We compute the fraction of workers in agriculture, the fraction of workers in forestry, fishing and livestock, and the fraction of workers in other sectors (industry, trade, service, and transport).

Government Transfers

District Revenues: District revenue figures come from the World Bank's Indonesia Database for Policy and Economic Research (DAPOER), which in turn obtains data from the Indonesia Ministry of Finance. They are given for each district at the time of existence up to 2013. We add in the 2014 revenue data directly from the Ministry of Finance. Population data is taken from the same dataset. We construct all revenue and population variables at the original district level by aggregating up to the 2000 borders. Both the population and revenue data are missing in some cases. In our baseline, we impute these missing observations as described below, but our results are very similar if either or both variables are left as missing. Population data is missing in 2014 for all districts and in 2000 for 6 original districts. We impute population using the preceding/following year and the median growth rate of 1.5 percent. Revenue data is missing in 2000 for 4 of our original districts, and thereafter there are occasional within-district gaps in the data. These gaps occur between 2001–2005 and to a lesser extent between 2012–2013, never exceeding 8 missing districts. We impute missing revenues using annual median revenue growth rates. All revenue figures are adjusted for inflation using 2010 as the base year.

Total district revenue comes from the general allocation grant (Dana Alokasi Umum, DAU), the special allocation grant (Dana Alokasi Khusus, DAK), shared taxes, shared natural resource rents, as well as limited own revenue, and limited revenue from other sources. We construct 5 control variables, all using the information from year 2000, that account for all of district revenues while keeping information disaggregated: grants (DAU + DAK) per capita, shared taxes per capita, shared natural resource rents per capita, own revenue per-capita, and other revenue per capita. This allows natural resources, for example, to enter separately. These are necessarily only computed at the original district level, and are included at that level in the child/parent regressions.

When we examine how transfers evolve over time in Appendix A, we use the full time series of total revenues less own revenue, to capture total transfers from the central government. At the Original District level we simply use the logarithm of real total transfers.

At the parent and child level, we have to make an additional assumption, since we do not observe how parent and child districts shared transfers pre-split. Specifically, we assume that parent and child districts get their initial 2000 population share of the original district transfers and use these values up to and including the year of the split. For all subsequent years, we use actual realized transfers at the lower level, imputing any missing values using the prior years value and median growth rates.

Light Intensity

Fraction of District Area Covered by Lights: We use night lights in 2000 as a proxy for initial GDP (Henderson et al., 2012). We have data on the coverage of each village by any lights in 2000, and take the average percentage coverage across villages at the original district and eventual, 2010, borders.

Village Level Light Data: For our ethnic favoritism results, we also use the village level light directly. When looking at how nighttime light intensity varies by share of residents in 2000 belonging to the largest ethnic group in the eventual child district, we use the fraction of the village area covered with any lights in each year 2000–2013.

Other Variables

Climate: We compute the population weighted average rainfall and temperature from 1948 to 1978 using village level information from NOAA-GPCP.

GIS Data: We compute the logarithm of the population weighted average distance to the nearest road, to the coast, and to the nearest river. We also compute the logarithm of elevation (30 as), and the ruggedness of the terrain (RUGGED3). We include the population weighted average forest coverage in 2000. Finally we include detailed indicators for the slope of the terrain (slope 1– 8). See Bazzi et al. (2016) for details on the underlying sources and construction.

Cash Crop Share: We use the 2003 administrative village census (*Potensi Desa* or *Podes*) to calculate the value (price $\times$ quantity) of each crop produced within the 2000 and 2010 district borders. To proxy for agricultural resources, we compute the fraction of district agricultural output that is composed of nearly 30 cash crops, the most important among which include palm oil, rubber, coffee, and cocoa.

Party Vote Share Polarization: We use the 1999 parliamentary (proportional system) vote shares for all 48 political parties at the subdistrict level to construct a measure of party polarization at the original district and eventual 2010 borders level. The measure for a given district is given by $\sum_i \sum_j \text{share}_i^2 \text{share}_j$ over each party i and j. The underlying data was graciously shared by Audrey Sacks.

Time Varying Transfers and Distance

Δ **Distance**: Using PODES 2000 and PODES 2011 we calculate the population-weighted average distance (in km) to the district capital across villages within the eventual parent and child units. At the child and parent level we construct Δ Distance as the difference in the natural logarithm of reported distance to the capital in 2011 less that in 2000. At the original district we take the average of these measures across parent and children, weighted by district population.

Δ **Transfers**: We use the information from DAPOER on total transfers less own revenue (which encompasses the general and specific allocation grants and all tax and natural resource sharing). As discussed above, we impute missing values using median annual growth rates and we adjust for inflation. At the original district level, we compute Δ Transfers as the change in the logarithm of real transfers post-split. We compare the average post-split to the average pre-split (including the year of the split).

We do not observe how parent and child districts shared transfers pre-split. So for the child and parent level we assume original district transfers were divided according to the child/parents population share in all pre-split years and in the year of the split. Thereafter, we use actual realized transfers at the lower level, imputing any missing values using the prior years value and median growth rates. Similar to the original district level, we then construct Δ Transfers as the change in the average logarithm of real transfers post-split to that pre-split.

BOSTON UNIVERSITY Global Development Policy Center

Boston University
53 Bay State Road
Boston, MA 02215

gdp@bu.edu
www.twitter.com/gdpc_bu
www.bu.edu/gdp

The Human Capital Initiative (HCI) is a research inititiative at Boston University's Global Development Policy Center. The GDP Center is a University wide center in partnership with the Frederick S. Pardee School for Global Studies. The Center's mission is to advance policy-oriented research for financial stability, human wellbeing, and environmental sustainability.

www.bu.edu/gdp