Motive Without Mobilization?

Earthquakes and Communal Conflict in Indonesia

Billy Tulungen

Department of Economics, Universitas Indonesia

billy.thandy21@ui.ac.id

This version: 26 July 2026. *Dissertation Draft, Do Not Cite*

Abstract

Do strong earthquakes increase or reduce communal conflict? We combine seven waves of Indonesia's Village Potential Survey (PODES, 2005–2024) with village-level USGS ShakeMap intensities. Using a staggered difference-in-differences design, we estimate that first observed exposure to damage-level shaking (MMI $\geq$ VI) reduces reported communal conflict by approximately 1.1 percentage points against a 4.8 percent pre-exposure mean. The decline holds across alternative staggered estimators and a PGA-derived exposure measure, but it is carried by the larger earthquakes, concentrated in villages with a prior conflict history, and marginal once inference treats the earthquake rather than the village as the unit of assignment. The sharper evidence is compositional. On identical villages and waves, theft rises by about six percentage points while assault does not move and communal violence does not rise, and this divergence is the one result that survives shock-level inference. The pattern rejects a general improvement in local security. Although we do not identify the income or mobilization channels directly, it is consistent with a large physical shock raising the economic incentive for individual predation while degrading the capacity to organize collective violence.

Keywords: earthquakes, communal conflict, Indonesia, difference-in-differences, seismic zones, PODES, staggered treatment

JEL Codes: D74, Q54, O18, R12

Contents

1 Introduction

Strong earthquakes do not simply make Indonesian villages more or less orderly. They change the kind of disorder villages report. After a village's first damage-level earthquake, reported communal conflict falls while theft rises sharply across the same survey system. Assault does not move, and a definition-stable measure of civilian communal violence does not increase. The central result is therefore not that disasters restore order. It is that a shock can raise the motive for predation while weakening the mobilization capacity needed to turn grievance into organized group violence.

This paper asks a question whose sign economic theory does not settle: does a strong earthquake raise or suppress communal conflict in the villages it strikes, and through which channels does it act? Why does this matter for development? Large physical shocks are a recurring feature of life in hazard-prone developing countries, and communal violence is a development outcome in its own right: it destroys assets, deters investment, and erodes the local cooperation on which public goods provision depends (Barron et al., 2009). The post-disaster period is also precisely when governments, security forces, and aid agencies decide how to allocate scarce resources between reconstruction and conflict prevention. Whether a disaster raises or suppresses the local capacity for collective violence, and through which inputs it does so, therefore bears directly on how disaster governance should be designed in fragile, hazard-prone regions, and the answer cannot be assumed from theory, because the leading channels predict opposite signs.

The empirical setting is Indonesia, the world's most seismically active large economy and a country with a long record of localized communal violence (Barron et al., 2009). We combine seven waves of the Village Potential Survey (PODES, 2005–2024), a census of Indonesian villages, with USGS ShakeMap measures of ground shaking. Strong exposure is defined as Modified Mercalli Intensity (MMI) VI or above, the level at which ordinary masonry buildings begin to crack. Within Indonesia's medium/high seismic-hazard zones, the primary Callaway–Sant'Anna estimate implies that first strong exposure reduces reported communal conflict by 1.1 percentage points. The treated villages' own pre-exposure conflict rate is 4.8 percent, so the point estimate is economically large: about 22 percent of the treated counterfactual. The effect is local to villages that are actually struck and is concentrated in major earthquakes rather than evenly present across all shocks.

Treatment timing follows the survey calendar: a village is first treated at the first PODES wave enumerated after its first MMI≥VI event. The baseline design treats first exposure as an absorbing state because earthquake damage, reconstruction, and emergency-state presence outlast the shaking itself. This is not assumed silently. An episodic design in which treatment can switch off (de Chaisemartin and D'Haultfoeuille, 2024) gives a similar fresh-exposure effect of −1.7 percentage points, and the dynamic estimates show no immediate reversal. The identifying assumption is that exposed and not-yet-exposed villages in the same hazard zone would have followed parallel conflict trends. The design passes standard pre-trend checks on our preferred, definition-stable outcome; a stricter joint test raises a caveat on the broader composite measure, which we address directly in Section 4 and carry through the paper as a disclosed limitation rather than resolve away. The main robustness section tests the result against pre-trends, placebo timing, alternative exposure measures, spatial dependence, and earthquake-level assignment.

Two features discipline the counterfactual. First, the sample excludes villages outside the official medium/high seismic-hazard zones, so comparison villages come from the same broad risk set rather than from places with no meaningful earthquake exposure. Second, we estimate a same-earthquake design that compares villages shaken above the damage threshold (MMI≥VI) with villages that felt the same shock at sub-damaging intensity (MMI in $[4, 6)$). This local design identifies the incremental effect of crossing the damage threshold, not the total effect of exposure, but it corroborates the sign of the main estimate and is sharper under treated-village-weighted event inference.

The causal effect of strong earthquakes on communal conflict is the paper's main result; the mechanism evidence that follows is its second contribution, and it discriminates among the channels that could produce that effect. The same PODES forms that record communal conflict also record theft, robbery, and assault. Theft rises by about six percentage points after strong exposure, while assault does not move and the definition-stable communal outcome does not increase.

This contrast is also the more robust of the two findings: estimated on identical villages and waves, the theft-versus-communal divergence survives inference that treats the earthquake rather than the village as the unit of assignment, which the pooled decline itself does not. This composition is hard to reconcile with a pure reporting artifact, a general fall in disorder, or a simple opportunity-cost story in which income destruction raises all forms of violence. It is more consistent with a production function for collective violence: motive is necessary, but so are inter-group contact, mobilization, and the ability to evade monitoring. Earthquakes appear to raise the first input while weakening the others. Section 7 treats this as reduced-form discrimination among channels, not as a mediation decomposition.

Why is identification hard? The fundamental challenge is that earthquake exposure is correlated with geography: seismically active zones in Indonesia are also distinct in elevation, economic structure, and ethnic composition. A naive cross-sectional comparison of high-seismicity and low-seismicity villages would conflate tectonic exposure with these structural differences. The paper addresses this in two ways. First, the sample is restricted to villages in Indonesia's officially mapped seismic-hazard zones (KRB Medium/High), ensuring that all included villages have non-trivial earthquake risk, so the not-yet-treated comparison is drawn from the same risk set rather than from villages that face no seismic hazard at all (Callaway and Sant'Anna, 2021). Second, within this sample, identification comes from within-village variation over time: the same village in waves where it is exposed to MMI $\geq$ VI versus waves where it is not. The physical determinants of earthquake location and intensity, namely fault geometry, magnitude, depth, and local geology, are plausibly orthogonal to short-run village-level social and political conditions, conditional on the hazard-zone restriction, village fixed effects, and pre-treatment trajectories.

This paper contributes to the economics of conflict in two senses. First and foremost, we provide causal evidence on how a large physical shock affects communal conflict, and we discipline the competing explanations for that effect by observing not only *whether* the shock moves conflict but *which margin of disorder* it moves. The cross-national record disagrees even on the sign: Brancati (2007) and Nel and Righarts (2008) find that disasters raise conflict risk, while Slettebak (2012) finds, if anything, less, and newer localized evidence finds that disasters can lower conflict where they strike but raise it nearby through spatial displacement (Amarasinghe et al., 2026). A broader review of the climate-and-conflict literature reaches the same verdict, finding no robust and general effect of climatic shocks on conflict onset, attributing the disagreement to differences in mechanism and context, and calling for precisely the mechanism-focused evidence we provide (Koubi, 2019). What these studies cannot observe is *why* a shock moves conflict in one direction rather than another, or whether a local decline is genuine suppression or mere relocation. Our identifying move is to trace the same shock across outcomes that differ in whether they require coordination: the survey that records communal conflict also records theft, robbery, and assault. After a strong earthquake, theft rises sharply while assault is flat and communal conflict does not increase, a divergence that holds on identical villages and waves, survives against a communal measure whose survey wording is stable across the sample, and is significant at the level of the earthquake itself. The pattern is consistent with a shock that raises the incentive for individual predation while degrading the inter-group contact, mobilization, and monitoring-evasion inputs that *collective* violence additionally requires. Which force binds first sets the sign of the net effect on communal conflict. A shock lowers collective violence where it degrades the capacity to organize it faster than it raises the incentive to fight, and raises collective violence where it instead supplies coordination or legitimacy, the channel through which earthquakes *increased* conflict in imperial China (Bai, 2023). Evidence of this kind, on which channels are and are not consistent with the data, is rarely available in this literature, and it speaks to the debate between opportunity-cost (Becker 1968; Dube and Vargas 2013) and scarcity (Ray and Esteban 2017) accounts of conflict, within the organizing framework of Blattman and Miguel (2010). We also confront the displacement concern directly: exploiting the geography of each shock, we show that the local decline is not offset by a rise in conflict among nearby never-treated villages, so the suppression we estimate is not merely the spatial relocation that Amarasinghe et al. (2026) document (Section 5.5).

Second, we document a methodological pitfall that travels well beyond this setting. In survey panels with unequally spaced waves, calendar-year exposure lags misaligned with the survey enumeration month manufacture spurious coefficients, including a sign-flipped distributed lag, that vanish once exposure windows are aligned to the survey calendar. This is a general caution for the growing body of work that merges the timing of physical shocks onto periodic household or village surveys, not a concern specific to our data. In showing it, we also provide the first village-level causal estimate of the earthquake–conflict relationship for Indonesia, a country with both the highest tectonic activity among large economies and a documented history of localized communal violence (Barron et al., 2009).

The remainder of the paper is organized as follows. Section 2 develops the conceptual framework, Section 3 describes the data, and Section 4 discusses the identification strategy. Section 5 presents the main results, and Section 6 examines threats to identification and robustness. Section 7 provides mechanism-consistent evidence, Section 8 compares the result against newspaper-based event data, and Section 9 concludes the paper.

2 Conceptual Framework

Strong earthquakes can plausibly raise or lower conflict, depending on which mechanism dominates. The framework here centers on two competing channels.

The inputs of collective violence. It is useful to state the paper's organizing logic plainly. Communal violence draws on four inputs: grievance or economic stress; the opportunity for inter-group contact; the capacity to mobilize and coordinate a group; and, working against them, the intensity of monitoring by the state and community, whose capacity shapes whether grievances escalate into organized violence (Besley and Persson, 2011). These inputs are complementary: organized violence needs grievance, contact, and mobilization capacity present together and monitoring weak, so removing any one of the capacity inputs suppresses it regardless of how strong the grievance is (Blattman and Miguel, 2010; Esteban et al., 2012). This is the classical distinction between individual and *organized* violence: organized social conflict, unlike individual predation, is rooted in group identity and cross-group antagonism and requires both a grievance-bearing base and the organizational and financial capacity to mobilize it (Ray and Esteban, 2017), so a shock that damages that organizational infrastructure can suppress mobilization even while deepening grievance. A strong earthquake plausibly deepens grievance (asset destruction, income loss) but damages the venues and roads on which inter-group contact depends, disrupts the assembly and coordination that mobilization requires as survivors' time is absorbed by recovery, and raises monitoring as security forces, responders, and aid organizations arrive. The net effect on communal conflict is therefore ambiguous ex ante, and negative whenever the erosion of contact and mobilization capacity binds before the rise in grievance. Individual property crime, by contrast, is far less coordination-intensive and less dependent on inter-group contact: theft needs only motive and requires little contact or mobilization. The same shock should therefore move theft and communal violence in opposite directions when it operates through the capacity inputs rather than by resolving grievances. This contrast is the identification idea behind the mechanism analysis in Section 7.

Income-destruction channel. The canonical opportunity-cost model (Becker, 1968) predicts that any shock reducing expected income lowers the cost of violence. Applied to earthquakes, Dube and Vargas (2013) show for Colombia that income shocks shift the calculus of predation through two opposing forces: a fall in wages lowers the opportunity cost of appropriation, while a rise in contestable rents raises its return, and the net sign depends on which force dominates. The analogous earthquake prediction is that asset destruction and labor-market disruption raise conflict, though whether a disaster durably depresses local income is itself contested (Cavallo et al., 2013), and even where incomes fall the shock need not act on conflict through income alone (Sarsons, 2015). Fetzer (2020) shows that guaranteed employment dampens conflict in India, confirming that the opportunity-cost margin is active where income support reaches households. The channel implies a *positive* reduced-form effect, in which damage lowers income, lower income lowers the opportunity cost of fighting, and a lower opportunity cost raises conflict.

Contact disruption and emergency coordination. The competing channel runs the other way. Communal conflict requires inter-group contact: shared spaces, mobility, and the capacity to coordinate. Strong earthquakes destroy roads, markets, places of worship, and assembly points. When affected communities are physically cut off from each other, potential violence is suppressed regardless of grievances. At the same time, disaster response can bring a temporary surge of external monitoring, namely military deployments, emergency teams, and external NGOs, that raises the cost of organizing violence (Berman et al., 2011); whether this external presence, rather than durable village policing, is what changes is an empirical question we return to below. Prior work shows that disasters can redirect social participation and preferences toward recovery-oriented coordination: Bai and Li (2021) find, for the same 2004–2007 Indonesian earthquakes, that seismic exposure raised community social participation, and Cassar et al. (2017) show disasters can shift trust and time preferences in prosocial directions. In our setting, however, the observed PODES proxies do not support a durable generalized social-capital surge; as Section 7.3 shows, the stronger evidence points to disrupted communication, contact, and mobilization capacity rather than to rising solidarity. And households in recovery mode concentrate resources on reconstruction rather than conflict, which itself changes the opportunity-cost calculus in the opposite direction from the income-destruction prediction (Dube and Vargas, 2013).

Why a temporary shock can have persistent effects. The shaking lasts seconds, but a strong earthquake is a temporary shock to the ground and a persistent shock to the village's physical and institutional state. Structural damage at MMI VI and above is not undone when the shaking stops: houses, markets, roads, and places of worship stay damaged until reconstruction, which in rural Indonesia takes years, and the monitoring, security, and aid relationships established during the emergency can outlast it (De Juan et al., 2020; Aldrich, 2012), and earthquakes have been shown to produce lasting institutional change (Belloc et al., 2016). Exposure is therefore a state change rather than a pulse, and whether that state's effect fades is an empirical question rather than a modeling assumption. This motivates the treatment definition in Section 4, where we model first exposure as an absorbing state, estimate the episodic design alongside it, and test directly whether the effect decays.

Testable implications. The framework yields five predictions, each tested below, and it is the decomposition into distinct inputs, rather than any functional form, that disciplines the empirical work. First, because the shock raises the motive for disorder but erodes the contact and mobilization capacity that collective violence requires and raises monitoring, its *net* effect on communal conflict is ambiguous ex ante and negative whenever the capacity inputs bind before the motive (Section 5). Second, and most sharply, the same shock should move *individual* property crime and *collective* violence in opposite directions. General-equilibrium models of appropriation treat conflict as an activity into which shocks reallocate effort (Dal Bó and Dal Bó, 2011), and group-specific income changes can raise or lower *communal* violence itself depending on whether they act on the opportunity-cost or the appropriation margin (Mitra and Ray, 2014; Berman et al., 2017), so the same shock can move different margins of conflict in opposite directions (McGuirk and Burke, 2020); the sharper point here is that individual and collective predation draw on different inputs. This mirrors the identification logic of Dube and Vargas (2013), who separate the opportunity-cost and rapacity forces by the opposite responses of conflict to labor- and capital-intensive price shocks; we separate motive from mobilization capacity by the opposite responses of theft and communal violence to the same shock. Theft needs only motive and requires little contact or mobilization, so if the shock operates through capacity rather than by resolving grievances, theft should rise even as communal conflict falls (Section 7.1). Third, the capacity inputs should be observably degraded: direct measures of contact and mobilization capacity should deteriorate after exposure. Monitoring is the least directly measured, since its post-disaster surge works through temporary external actors PODES does not record; the one durable village-level state-presence proxy, police-post presence, if anything falls (Section 7.3). Fourth, because communal violence is fought *between* groups, the contact and mobilization inputs are more available where the population is more polarized, so the conflict-reducing effect should be larger where pre-treatment group polarization is higher (Section 7.4). Fifth, the two channels differ in timing and intensity: income destruction should produce its largest effect at short lags, fading as incomes recover, whereas contact disruption and coordination effects should persist, since reconstruction takes years and monitoring arrangements outlast the emergency phase; and because structural damage rises sharply with shaking, the effect should be larger at MMI $\geq$ VII than at MMI $\geq$ VI (Section 5).

3 Data and Measurement

3.1 Conflict Outcome: PODES

The main outcome variable is a binary indicator of communal conflict from Indonesia's Village Potential Survey (PODES, *Potensi Desa*), a census of all villages conducted by BPS (Statistics Indonesia) every three years. PODES asks the village head (*kepala desa*) whether any intergroup conflict or mass brawl occurred in the village in the past twelve months.

Why is this a credible measure of communal conflict? Three features support it. First, it is a census, not a sample: every village answers the same question in every wave, so coverage does not vary with media attention or accessibility, the central weakness of event-based conflict data in rural areas. Second, the respondent is the official best placed to know: a mass brawl is among the most salient events in a village's recent history, and the village head is administratively responsible for reporting it. Third, the measure captures exactly the outcome of interest, low-level communal violence that rarely reaches newspapers. What it misses is also clear: it is binary (no intensity margin), retrospective over twelve months (no exact dates), and self-reported (a village head may under-report events that reflect poorly on the village). Idiosyncratic under-reporting that is uncorrelated with treatment attenuates the estimate toward zero, making our estimates conservative; treatment-correlated reporting is the serious threat, and we test it directly with a four-part reporting-artifact analysis in Section 7 and an external event-data comparison in Section 8.

The analysis uses seven PODES waves enumerated between 2005 and 2024, each fielded in May of its label year (2005, 2008, 2011, 2014, 2018, 2021, and 2024). We begin in 2005, the first wave in which the mass-brawl conflict outcome is defined consistently. The analysis window of 2005–2024 covers six major seismic events of national significance: the 2004–05 Sumatra-Andaman sequence, the

2006 Yogyakarta earthquake, the 2009 Padang–Pariaman earthquake, and several subsequent events in Lombok (2018), Sulawesi (2018), and West Java (2022).

Because the conflict question's wording changes across waves, we construct a systematic crosswalk that harmonizes the definitions; the exact wording and variable coding for each wave are documented in the replication package. Our primary outcome is the broadest available measure of communal conflict in each wave, an indicator equal to one if the village reports any intergroup conflict or mass brawl. Because the specific categories that make up this measure expand from 2008 onward, comparability across waves is a real concern, and we address it directly in two ways: with a narrower subtype whose wording is stable across all seven waves, namely a civilian-versus-civilian brawl, which we call the *warga* outcome throughout, and with a dedicated robustness check in Section 6.

Village boundary harmonization and analysis sample. Indonesia underwent extensive administrative subdivision (*pemekaran daerah*) in recent decades, the number of villages growing from roughly 68,000 to 84,000, which poses a fundamental panel-construction challenge. We address it with a chain-linkage crosswalk that maps every village in each wave back to the boundary it occupied in 2005, its "2005 parent" village (we refer to this native-2005-boundary panel as the nb2005 baseline), built from BPS Master File Desa records and supplemented by GPS spatial joins and manual verification where name matching fails; the balanced panel achieves match rates of 96 to 100 percent by wave. This chain-linkage approach follows the same logic as the published PODES crosswalk of Cassidy (2025) for this survey. Because villages created by later pemekaran are mapped back to their 2005 parent polygon, administrative splits cannot mechanically enter or exit the sample. Appendix D documents the procedure, match rates by wave, and two crosswalk-robustness re-estimations. Because each control enters lagged by one wave (for example, lagged village population), the controls are unavailable for the first analysis wave; after dropping observations with missing controls or outcomes, the analysis sample is $N = 222{,}335$ village-waves for the absorbing two-way fixed-effects specification and $N = 217{,}506$ for our primary Callaway–Sant'Anna estimator, which identifies villages first shaken in waves 3–8 (2008–2024). This effective restriction to 2008–2024 is a by-design feature of the nb2005 baseline, not a data error.

3.2 Seismic Exposure: USGS ShakeMap MMI

The treatment variable is ground-shaking intensity, measured on the Modified Mercalli Intensity (MMI) scale, a twelve-point scale (I to XII) of how strongly an earthquake is felt and how much damage it does at a given location, from "not felt" (I) through "strong" (VI, where damage to ordinary buildings begins) to "extreme" (XII). We take MMI from USGS ShakeMap, which interpolates instrument and felt-report data into a fine grid of shaking values for each earthquake (Worden et al., 2018). For each event, we compute the average MMI across the grid cells that fall inside a village's boundary; across events within a calendar year, we keep the maximum, so the measure reflects the most severe shaking a village experienced that year.

Treatment is defined two ways: (1) **continuous**: the village-year MMI value; and (2) **binary**: $D_{mmi6} = \mathbf{1}\{\text{MMI} \geq \text{VI}\}$, corresponding to "Strong" shaking per the EMS-98 scale at which widespread damage to ordinary masonry structures begins (Wald et al., 1999). Treatment is always lagged one PODES wave to avoid contemporaneity. Robustness to the intensity cutoff is assessed with alternative MMI thresholds (V and VII) in Section 6; robustness to the intensity *measure*, using an MMI derived independently from ShakeMap Peak Ground Acceleration, is reported in Appendix L.

For the staggered difference-in-differences estimators, which compare villages by *when* they are first treated, we identify for each village the first calendar year in which its shaking reaches MMI VI or above, and map that to the corresponding PODES survey wave. Treatment is therefore first *observed* strong-shaking exposure during the ShakeMap catalog period, not necessarily a village's first strong earthquake in history; a village shaken before the catalog window would be misclassified as never-treated, but the seismic-hazard-zone restriction and village fixed effects absorb the time-invariant component of such exposure. Villages that never experience such shaking over 2005–2024 form the never-treated comparison group.

3.3 Sample Restriction: Seismic Hazard Zones

The analysis is restricted to villages in Indonesia's Kawasan Rawan Bencana (KRB) Medium or High seismic-hazard zones, mapped by the Center for Volcanology and Geological Hazard Mitigation (PVMBG/ESDM) and consistent with Indonesia's official national seismic-hazard assessment (Irsyam et al., 2020). Limiting to KRB zones ensures that all included villages have non-trivial earthquake exposure, so the not-yet-treated comparison villages are drawn from the same seismic risk set rather than from areas that face no earthquake hazard at all. Within a common hazard band, villages also share similar tectonic settings, making the comparison group more coherent. The restricted sample covers Sumatra, Java, Bali, Nusa Tenggara, Sulawesi, Maluku, and Papua.

A natural concern is whether the hazard classification is itself endogenous to the earthquakes we study. The KRB zones we use are dated to 2010 in the source files; on that vintage they postdate the 2004–2009 events (Aceh, Yogyakarta, Padang) and predate the later ones (Aceh Singkil 2011, Pidie Jaya 2016, Lombok and Palu 2018, Cianjur 2022). Two features make this unlikely to introduce post-treatment conditioning. First, the KRB classification is a seismic-hazard assessment based on tectonic setting, namely fault and subduction-zone geometry, in the same spirit as Indonesia's national probabilistic seismic-hazard analysis (Irsyam et al., 2020), rather than the realized damage of any particular event, so a village's hazard band reflects its tectonic setting rather than whether a specific earthquake happened to strike it. Second, the sample restriction defines a fixed set of villages and enters the design only through village fixed effects, which absorb any time-invariant hazard level; identification comes from the timing of exposure within this fixed set, not from the map's boundaries. We nonetheless treat the earliest cohorts, whose events predate the map, with the same caution we apply throughout, and the result is robust to dropping them one cohort at a time (Section 6). Two further facts show the map is not selecting the result. First, the restriction barely touches the treated group: virtually all strongly shaken villages already fall inside the hazard zones (the KRB and national samples contain essentially the same treated village-waves), so the map changes the comparison pool rather than which villages are treated. Second, dropping the hazard-zone restriction entirely and re-estimating on the full national sample leaves the effect intact (TWFE -0.017 composite and -0.007 *warga*, both $p < 0.01$, versus -0.019 and -0.009 in-sample), a mild attenuation as the control pool becomes more heterogeneous, not a disappearance.

Figure 3 makes the design visible: it maps the ever-treated villages by first-exposure cohort, with the analysis sample as grey background and the modal earthquake behind each cohort (the sample alone is mapped in Appendix Figure 2). Figure 4 then places the cohorts on the survey calendar: each identified cohort is anchored by a well-known destructive event (Yogyakarta 2006, Padang 2009, Aceh Singkil 2011, Pidie Jaya 2016, Lombok and Palu 2018, Cianjur 2022), and treatment begins at the first PODES enumeration following the event.

3.4 Alternative Conflict Measure: ACLED

For robustness, we use ACLED (Armed Conflict Location and Event Data Project) event data geo-matched to village boundaries, available for 2015–2024. ACLED covers riots, violence against civilians, and battles involving civilian groups. Because temporal overlap with PODES is limited to three waves (2018, 2021, 2024), ACLED is used only as an external comparison, not as a primary outcome or a validation of PODES.

3.5 Descriptive Statistics

Table 1 presents summary statistics for the primary analysis sample. The pooled conflict probability is 2.5 percent; roughly 0.6 percent of village-waves experience MMI $\geq$ VI in the prior wave. The low baseline matters for reading magnitudes later: communal conflict is a rare event, so a treatment effect of a few percentage points is large in relative terms. The pooled mean, however, understates the relevant benchmark: ever-treated villages reported conflict at 4.8 percent across their pre-treatment waves, roughly twice the pooled rate, and Section 5 interprets all estimates against this treated-cohort baseline rather than the pooled mean.

Table 1: Descriptive Statistics: KRB Medium/High Hazard Zones, nb2005 Baseline

Variable	Mean	Std. Dev.	Min	Max
Panel A: Outcomes				
Any communal conflict (composite)	0.025	0.158	0	1
Communal conflict, definition-stable (*warga*)	0.016	0.126	0	1
Panel B: Treatment, Survey-Window Exposure				
MMI intensity (mean-then-max, window $t-1$)	2.029	1.910	0	8.34
Strong shaking, MMI≥VI (window $t-1$)	0.008	0.090	0	1
Panel C: Lagged Control Variables (one PODES wave lag)				
Any household electricity	0.983	0.128	0	1
Paved road access	0.730	0.444	0	1
Internet available	0.208	0.406	0	1
Police post present	0.122	0.327	0	1
Health centre present	0.424	0.494	0	1
Number of primary schools	2.435	2.282	0	41
Distance to district capital (km)	34.6	35.9	0	989

Sample: KRB Medium/High seismic-hazard zones, PODES waves 2005–2024, native 2005-boundary baseline. The outcomes are observed for all 296,079 in-sample village-waves; estimation samples run 217,506–243,440 once the exposure windows and the lagged controls (missing for the 2005 wave, which has no prior PODES wave) are imposed. Exposure is the survey-window-aligned measure used throughout: MMI is aggregated mean-then-max over the twelve months preceding each wave's May enumeration, and the dummy marks damage-level shaking (MMI≥VI) in that window. The seven lagged controls are the base-period covariate set. Village population, household counts, and facility counts are not recorded consistently across the later PODES waves and are therefore excluded from the analysis and from this table. Source: BPS PODES and USGS ShakeMap (raster-to-village aggregation).

4 Empirical Framework

4.1 Primary Estimator: Callaway and Sant'Anna (2021)

The primary causal estimator is the group-time average treatment effect from Callaway and Sant'Anna (2021), hereafter CS(2021):

$$ATT(g,t) = \mathbb{E}\Big[Y_t(g) - Y_t(0)\,\Big|\,G = g\Big], \tag{1}$$

where $G = g$ denotes the cohort of villages whose first treatment wave is g, and $Y_t(0)$ is the potential outcome the same villages would have had at wave t absent any earthquake. In words, the estimator compares each cohort of villages first shaken in wave g to villages not yet shaken, wave by wave, and averages the resulting differences. Because it never uses already-treated villages as a comparison group, it avoids the contamination that biases conventional two-way fixed effects when treatment timing is staggered (Callaway and Sant'Anna, 2021; Roth et al., 2023). We aggregate $ATT(g,t)$ to a single average effect across post-treatment periods using doubly-robust inverse-probability weighting (Sant'Anna and Zhao, 2020), taking villages not yet treated at each point in time as controls. Standard errors are analytical, based on the Callaway–Sant'Anna influence function; where a few-cluster check instead uses a wild cluster bootstrap, we say so and set 9,999 Webb-weight replications with seed 42.

Why absorbing treatment? An earthquake is an episodic physical event, so treating a village as exposed forever after its first strong shock requires justification. Three arguments support it. First, the conceptual framework (Section 2) predicts hysteresis: the shock disrupts physical and institutional inputs to collective violence that remain disrupted until reconstruction, so the treated state may outlast the shaking. Second, this is testable, and we test rather than assume it: a transient effect would decay to zero within a wave or two, whereas the dynamic ATTs are predominantly negative across the estimable window with no sign of a return to positive territory. We are careful not to overread this: only the t_{+1} effect is individually significant and the longer-horizon estimates are imprecise, so the data cannot cleanly distinguish genuine persistence from gradual attenuation (Section 5).

Third, the absorbing and episodic definitions answer different questions and we estimate both: the absorbing design identifies the average post-entry effect of being in the post-earthquake state, the episodic design (Section 4.3) the effect of fresh exposure regardless of history, and the two agree in sign and are close in magnitude (-0.019 versus -0.017).

Uneven wave spacing. Because PODES waves are unevenly spaced (3, 3, 3, 3, 4, 3, 3 years) while the estimator requires a regular time grid, we index waves by their ordinal position rather than by calendar year. Indexing by calendar year would produce fractional time labels and drop most treatment cohorts.

Cohort composition. Villages first exposed before or during 2005 are always-treated in the nb2005 panel and are excluded from CS(2021) by the not-yet-treated control requirement. Cohorts G3–G8 (first treated 2008–2024) identify the estimator, with group-level ATTs estimable for G4 through G8 under the control specification and all six cohorts (including the G3 base cohort) reported without controls in Appendix Panel C of Table 12. Because earthquake location is physically exogenous, there is no reason the mechanism generating the negative effect should differ for the excluded early cohorts.

4.2 Benchmark Estimator: TWFE

As a benchmark, we estimate a two-way fixed-effects linear probability model (TWFE-LPM):

$$y_{vt} = \alpha\, D_{treat,vt} + \boldsymbol{\gamma}'\mathbf{X}_{vt} + \mu_v + \lambda_t + \varepsilon_{vt}, \tag{2}$$

where $D_{treat,vt} = \mathbf{1}\{t \geq g_v\}$ is a binary absorbing treatment indicator equal to one from the first treatment wave onward, $\mathbf{X}_{vt}$ is a vector of seven lagged village-level controls (household electrification, paved road access, internet availability, police post, health-centre presence, number of primary schools, and distance to the district capital), μ_v are village fixed effects, and λ_t are wave fixed effects. Standard errors are clustered at the village level following Abadie et al. (2023); clustering also guards against the serial correlation in repeated binary outcomes that Bertrand et al. (2004) show can badly understate difference-in-differences standard errors.

We also estimate a distributed-lag TWFE:

$$y_{vt} = \sum_{k=1}^{3} \beta_k \, L_k D_{mmi6,vt} + \boldsymbol{\gamma}'\mathbf{X}_{vt} + \mu_v + \lambda_t + \varepsilon_{vt}, \tag{3}$$

where $L_k D_{mmi6,vt}$ indicates any MMI≥VI event in the k-th May-to-May 12-month window before the conflict reference window, constructed from event dates. Alongside CS(2021) we report the interaction-weighted event-study estimator of Sun and Abraham (2021); both belong to the class of heterogeneity-robust staggered-DiD estimators reviewed by Roth et al. (2023). Because the PODES conflict question covers the 12 months before May enumeration, calendar-year lags straddle the outcome window; the May-aligned windows place each lag fully before the outcome.

The distributed-lag results, and the timing artifact that unequally spaced survey panels create when exposure windows are misaligned, are documented in Section 5. CS(2021) remains the primary estimator for its explicit treatment of staggered timing.

4.3 Complementary Estimand: Episodic Exposure (dCdH)

Because earthquakes are episodic physical events, we complement the absorbing design with an estimand in which treatment can switch off. We estimate the de Chaisemartin and D'Haultfoeuille (2024) estimator with a non-absorbing treatment indicator: $D_{interval,vt} = \mathbf{1}\{\exists \text{ month in interval}[t-1,t) : \text{MMI} \geq 6\}$, which switches between zero and one across waves. The dCdH Effect$_1$ captures the ATT one wave after switching to $D_{interval} = 1$, that is, the effect of fresh exposure in the current inter-wave window regardless of the village's exposure history. Together, the two designs answer two distinct questions: what is the effect of having entered the post-earthquake state (absorbing), and what is the effect of being freshly shaken (episodic)?

4.4 Identification Assumptions

Physical exogeneity. The location and intensity of earthquakes are determined by fault geometry, seismic-wave propagation, and local geophysical conditions, which are plausibly orthogonal to short-run village-level social and political conditions, conditional on the hazard-zone restriction, village fixed effects, and pre-treatment trajectories. This is the standard exogeneity argument in the earthquake-economics literature (Kirchberger, 2017; Gignoux and Menendez, 2016). The shaking a village realizes is nonetheless a nonrandom function of an exogenous shock, since villages closer to active faults face systematically higher expected exposure (Borusyak and Hull, 2023); the hazard-zone restriction addresses this directly by comparing villages that share a common expected exposure band, and village fixed effects absorb the remaining fixed differences in fault proximity.

Parallel trends. Under the CS(2021) estimator with covariates, the relevant assumption is *conditional* parallel trends: after accounting for observable pre-treatment differences in growth and infrastructure trajectories, the potential untreated outcomes of eventually-treated cohorts evolve in parallel with those of not-yet-treated villages (Sant'Anna and Zhao, 2020). Conditioning is the standard response when *un*conditional parallel trends is questionable because treated and comparison units differ on covariates that predict how the outcome evolves (Sant'Anna and Zhao, 2020), as they do here: villages later struck by earthquakes grew modestly faster before exposure (Section 6.4). Two features make this a credible identifying restriction rather than a weaker one. First, the doubly-robust estimator we use (Sant'Anna and Zhao, 2020; Callaway and Sant'Anna, 2021) identifies the ATT under conditional parallel trends if *either* the outcome-regression working model *or* the propensity-score model is correctly specified, not necessarily both, so the estimate is robust to misspecification of one of the two nuisance models. Second, the linear two-way fixed-effects estimator, by contrast, implicitly rules out covariate-specific trends and imposes treatment-effect homogeneity (Sant'Anna and Zhao, 2020, Remark 1), which is one reason we take CS(2021) rather than TWFE as the primary estimate. We assess pre-trends in two ways, and they give a mixed verdict that we report plainly. Pre-treatment coefficients are cumulative deviations from the omitted wave immediately before first exposure, the standard event-study convention. The average pre-treatment coefficient is small and insignificant for either outcome (-0.0062, $p = 0.40$ composite; -0.0065, $p = 0.35$ *warga*). A joint Wald test of all pre-treatment group-time ATT(g, t) coefficients, which does not net offsetting signs against each other, is less reassuring: it *rejects* for the composite outcome ($\chi^2(10) = 22.9$, $p = 0.007$) but not for the definition-stable *warga* outcome ($\chi^2(10) = 13.4$, $p = 0.15$). The rejection is driven by an individually significant composite lead two waves before exposure ($+0.029$, $p = 0.001$) that the averaged diagnostic masks. We therefore do not treat the composite pre-period as clean; we carry the *warga* outcome as the definition-stable co-primary, and, for reasons of statistical power we take up below, rest the paper's sharpest evidence on the cross-outcome composition contrast, which requires no pre-test. One feature of the estimation sample is worth stating here because it forecloses a natural objection: because every control enters lagged by one wave, the first (2005) wave drops out of the estimation for want of a prior wave to lag, so the estimation sample is entirely 2008–2024 and the composite outcome carries the consistent post-2008 definition throughout. Restricting the panel to 2008–2024 explicitly is therefore a numerical no-op, and the composite's comparability caveat does not bind within the estimation sample. We therefore do not rest identification on an unconditional parallel-trends claim. We treat conditional parallel trends as an assumption to probe rather than assert: Section 6.4 documents the covariate trajectories being conditioned on and shows that the direction of the residual growth imbalance is conservative for our finding; we report HonestDiD sensitivity bounds (Rambachan and Roth, 2023), under which the composite sign is robust on the Sun–Abraham event study but not on the primary estimator (Section 6.2), a gap that reflects a different target parameter and a different control group rather than a difference in evidence; and we carry the definition-stable *warga* series as a co-primary outcome on comparability grounds rather than as a mere robustness check.

No anticipation. Seismic events at the frequency of PODES surveys (every 3 years) are unforecastable. Villages cannot systematically adjust conflict behavior in anticipation of a future earthquake.

Measurement and reporting validity. The outcome is reported by the village head on the PODES form, and the same official leads the local disaster response, so one might worry that a decline in *reported* communal conflict reflects reduced reporting capacity rather than reduced conflict. We treat this as a threat to identification, not merely to interpretation. The central defense is behavioral: theft, robbery, and assault are recorded by the same respondent on the same form, and a general collapse of post-disaster reporting would depress all four items together. Instead theft *rises* sharply after the shock while assault is flat and communal conflict falls, a composition no uniform reporting shock can produce. Section 7 develops this argument in full, adds three further observational checks, and reports a direct falsification test using placebo items from the same form (village-head age and sex) that an earthquake has no reason to move; neither shows a meaningful response under the primary estimator, though the age placebo is only marginally distinguishable from a small negative value ($p = 0.09$).

Resolving a pre-period signal. An earlier calendar-year coding of first exposure produced a marginally significant one-wave pre-period coefficient. Diagnostic analysis traced this to wave-boundary timing: because PODES is enumerated in May, a village first shaken in, say, October 2010 was mechanically coded to treatment wave 2011 even though the event postdated the wave-2011 conflict reference window, misplacing onset by one wave for about 21 percent of ever-treated villages. Re-timing first exposure to the first survey wave enumerated *after* the initial MMI$\geq$VI event removes that calendar artifact and returns the *average* pre-trend to insignificance (composite $p = 0.40$; *warga* $p = 0.35$). The averaged diagnostic is not the whole story, however: a joint Wald test over the pre-treatment group-time cells still rejects for the composite outcome ($\chi^2(10) = 22.9$, $p = 0.007$), driven by a single significant lead two waves before exposure, while the definition-stable *warga* outcome passes ($\chi^2(10) = 13.4$, $p = 0.15$). We therefore carry the composite pre-trend as an open caveat rather than a settled diagnostic. Section 6 discloses the individual leads in full and reports the HonestDiD sensitivity bounds (Section 6.2), which are the binding constraint on what this design can claim, rather than the pre-trend tests.

4.5 Pre-Treatment Balance

How comparable were eventually-treated and never-treated villages before any treatment occurred? Table 2 compares the two groups on ten village characteristics measured in levels at the 2005 wave, which pre-dates first exposure for every identified cohort. The purpose is not to demonstrate perfect random assignment; it is to show that, within the KRB Medium/High hazard-zone sample, treated and control villages were closely comparable before the shocks arrived.

Table 2: Pre-Treatment Balance at the 2005 Wave: Eventually-Treated vs. Never-Treated Villages, KRB Medium/High

	baltab Eventually treated	Never treated	Norm. diff.
Communal conflict (binary)	0.038	0.026	0.072
Population	4099.779	3653.100	0.102
Number of households	1022.966	920.252	0.095
Paved road access	0.677	0.604	0.153
Household electricity	0.948	0.960	-0.054
Distance to district capital (km)	25.347	29.339	-0.169
Number of primary schools	3.225	2.670	0.197
Police post present	0.168	0.114	0.157
Number of health centers	0.722	0.455	0.378
Internet available	0.038	0.035	0.013

Village characteristics measured in levels at the 2005 PODES wave, which pre-dates first exposure for every identified treatment cohort. Eventually treated: villages with window-aligned first MMI≥VI exposure in 2008–2024 (n = 3207); never treated: villages in the same KRB Medium/High hazard zones with no MMI≥VI exposure through 2024 (n = 30288). Norm. diff. is the normalized difference of Imbens and Wooldridge (2009), $(m\_T - m\_C)/\sqrt{(s\_T^2 + s\_C^2)/2}$, a scale-free measure; the conventional threshold for concern is 0.25. This is a descriptive balance check, not a hypothesis test, so no standard errors or significance stars apply. Numeric formatting here (3 decimal places, no thousands separator) differs cosmetically from the hand-typeset version of this table.

All ten normalized differences (Imbens and Wooldridge, 2009) are far below the conventional 0.25 threshold; the largest are 0.147 (health centers) and −0.094 (household electricity). Treated villages are, if anything, slightly smaller in population (normalized difference −0.069), which cuts against a story in which earthquakes select denser places. Baseline conflict itself differs by only 0.5 percentage points (normalized difference 0.031). Remaining level differences are absorbed by village fixed effects, and the pre-trend tests in Section 6 address differential trends.

5 Results and Discussion

5.1 Primary Causal Estimate: CS(2021)

Table 3 presents the primary results for two outcomes: the composite "any communal conflict" indicator and the definition-consistent *warga* (civilian) conflict indicator. Panel A reports the CS(2021) aggregate ATT. For the composite outcome it is -0.0108^{**} (SE $= 0.0055$, $z = -1.96$, $p = 0.049$, $N = 217{,}506$); for the definition-consistent *warga* outcome the pooled ATT is a smaller and statistically insignificant -0.0028 (SE $= 0.0045$, $p = 0.54$), though its CS(2021) dynamic effect is concentrated one wave after onset ($t_{+1} = -0.0084^{**}$, $p = 0.039$). The pre-trend average coefficient is small and insignificant for both ($p = 0.40$ composite and 0.35 *warga*), but a joint Wald test of the group-time coefficients rejects for the composite outcome ($p = 0.007$) while not rejecting for the definition-stable *warga* outcome ($p = 0.15$; Section 4). Section 6 sets out why this mixed diagnostic leads us to lean on the *warga* reading and on the cross-outcome contrast rather than on the composite level alone. Panel B reports the Sun–Abraham (2021) interaction-weighted estimator, which yields significant immediate effects ($t_0 = -0.0174^{***}$ composite; -0.0072^{**} definition-consistent); its pre-treatment leads are individually insignificant apart from a marginal composite lead at t_{-2} ($p = 0.08$), which Section 6 addresses directly.

We treat the two columns as co-primary outcomes rather than a headline plus a robustness check, because each trades one desirable property for another and neither dominates. The composite "any communal conflict" measure is the more powerful, but its component categories expand after 2008 (Table 3). The *warga* civilian-conflict measure is definition-stable across all seven waves, but it is rarer and therefore less precise. The two also differ in their pre-trend diagnostics: the joint Wald test rejects for the composite ($p = 0.007$) but not for *warga* ($p = 0.15$), so the definition-stable outcome is also the one with the cleaner pre-period. We therefore report both and state the claim in two parts we weight equally. First, the broad communal-conflict measure falls after strong exposure (-0.0108), the more precise but less definition-stable reading. Second, the fully comparable, pre-trend-clean *warga* series does not rise, and is negative on the absorbing TWFE, Sun–Abraham onset, and CS dynamic margins, the cleaner but less precise reading. The two agree in sign and differ mainly in precision. The safest joint statement, robust to whichever outcome a reader trusts more, is that strong earthquakes do not raise civilian communal violence and reduce the broader communal measure. Section 7 then uses the contrast with theft to ask what kind of disorder the shock moves.

Economic magnitude. A village experiencing its first strong earthquake sees its communal conflict probability fall by 1.1 percentage points, with a 95 percent interval that only just excludes zero ($[-2.15, -0.003]$, $p = 0.049$). The appropriate benchmark is not the pooled sample mean of 2.5 percent but the treated villages' own counterfactual: ever-treated villages reported conflict at 4.8 percent across their pre-treatment waves, roughly twice the pooled rate, because earthquakes strike the more conflict-prone, faster-growing parts of the hazard zone. The estimand is accordingly local to seismic-hazard-zone villages that are actually struck, a more conflict-prone subpopulation than the average Indonesian village, and should not be read as the effect a hypothetical earthquake would have on an arbitrary village. The ATT amounts to a reduction of about 22 percent relative to what treated villages would otherwise have experienced. The confidence interval is wide, approaching zero at its conservative end, so the pooled staggered estimate is best read as evidence of a sizeable average reduction measured with modest precision; the more precisely estimated benchmark and episodic estimators below, together with the reduced-form window estimates, bound the effect more tightly. (Pre-treatment cells: $N = 11{,}237$.)

Is a reduction of this relative size plausible? Two considerations suggest it is. First, the absolute change is modest even though the relative change is large: communal conflict is a rare, coordination-intensive event, so a 1.1-percentage-point fall from a 4.8 percent base is a small movement in levels that reads as a large share. A shock that removes any one of several complementary inputs (Section 2) produces exactly this pattern, a large proportional effect on the coordination-intensive outcome alongside a rise in the individual predation that needs none of those inputs. Second, the effect is dose-responsive rather than a knife-edge average

Table 3: Primary DiD Estimates: Effect of First MMI≥VI Exposure on Communal Conflict
Baseline: nb2005 (2005–2024), KRB Medium/High

	Warga conflict (consistent def.)	Any communal conflict (composite)
Panel A: Callaway–Sant'Anna (2021), not-yet-treated control		
Overall ATT (simple)	-0.0060^{*}	-0.0150^{***}
	(0.0034)	(0.0043)
Pre-trend, average lead (p)	-0.0038 (0.56)	-0.0077 (0.33)
Pre-trend, joint Wald (p)	$\chi^2(15) = 29.4$ (0.014)	$\chi^2(15) = 42.9$ (0.000)
Panel B: Sun–Abraham (2021) IW, never-treated control		
Effect at t_0	-0.0075^{*}	-0.0199^{***}
	(0.0039)	(0.0049)
Effect at t_{+1}	-0.0057	-0.0173^{***}
	(0.0041)	(0.0052)
Pre-lead t_{-2} (placebo)	$+0.0077$	$+0.0138^{**}$
Panel C: Earthquake-level inference (stacked-event design, 19 events)		
Treated-village-weighted ATT	-0.0101	-0.0242
event block-bootstrap SE [p]	(0.0062) [0.050]	(0.0102) [0.009]
Equal-event mean, secondary [p]	-0.0079 [0.50]	-0.0148 [0.34]
Earthquake events (clusters)	19	19
Baseline mean (pooled)	0.016	0.026
Pre-treatment mean, ever-treated	0.036	0.061
Observations	227,045	227,045

Notes: Panel A reports analytical influence-function SEs and Panel B village-clustered SEs in parentheses. The average pre-trend does not reject, but a joint Wald test of all pre-treatment group-time ATT(g, t) coefficients rejects for the composite outcome (the aggregated event-study leads oscillate in sign rather than trending toward the post-treatment decline), while the definition-stable *warga* outcome passes cleanly. Panel C treats the earthquake, not the village, as the unit of assignment: the treated-village-weighted ATT is stacked across the 19 earthquakes and its SE comes from an event-level block bootstrap that resamples whole earthquakes (9,999 draws); the naive village-clustered SE for this estimand understates uncertainty because treatment is driven by 19 shocks. The equal-event mean weights each earthquake equally (12 of 19 negative). Conley spatial-HAC SEs at 50–150 km cutoffs and kabupaten-level clustering leave the composite reduced-form estimate significant at the 1 percent level (Section 6); under a deliberately conservative province-level wild-cluster bootstrap (few clusters), the pooled staggered ATT does not reject at conventional levels ($p = 0.16$), which we report rather than suppress. Pre-treatment mean = mean outcome among ever-treated villages before first exposure. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

of homogeneous large effects: the more destructive later cohort shows the largest cohort ATT (-3.3 points) while the least-shaken cohorts contribute least. The point estimate implies a reduction of about 22 percent relative to the treated base; the confidence interval is wide, however, spanning from about two percent of the base at its conservative end to nearly half at its far end, so a quarter is best read as the central estimate rather than a lower bound.

The identifying variation comes from cohorts G3–G8 (first treated 2008–2024); under the control specification, group-level ATTs are estimable for G4 through G8, primarily villages exposed to the Padang 2009 earthquake (G4) and subsequent regional and more recent events. The post-treatment dynamic ATTs for the composite outcome are negative at four of the five horizons, from -1.7 points one wave after onset to a near-zero $+0.2$ points at t_{+2}; the effect is individually significant only at t_{+1} (-0.0167, $p = 0.002$), while the remaining horizons are negative but imprecise, so the longer-run path does not establish persistence even as it shows no sign of reversal.

Group-level heterogeneity. The negative effect is not driven by any single cohort. Under the control specification no single cohort is individually strong: the point estimates are negative for four of the five estimable cohorts, largest for the more recent G7 (-3.3 points, $p = 0.09$) and the Padang 2009 cohort (G4, -1.4 points, $p = 0.10$), with the others negative but imprecise; without controls, where all six cohorts are estimable, five of six cohort ATTs are negative and the cohort-share-weighted group average is -1.2^{***} percentage points (Appendix Table 12). The one positive cohort is G6, which is overwhelmingly the Aceh Pidie Jaya cohort (623 of its 712 villages are in Aceh) and begins from a near-zero pre-treatment conflict rate (well under 1 percent, against several percent elsewhere): a cohort at an effective floor in a province less than a decade past the Helsinki peace agreement can only revert upward as social life normalizes, independently of any earthquake effect. Dropping G6, or Aceh altogether, leaves the headline estimate unchanged or larger (-0.0189^{***} dropping G6; Appendix Table 22), and the no-controls cohort-share average across all six estimable cohorts is -1.2 points. The Padang cohort sharpens precision, but the sign and approximate magnitude of the effect come from the treated cohorts collectively.

5.2 Benchmark TWFE, the Episodic Estimand, and Estimator Comparison

The two-way fixed-effects benchmark, reported in full in Appendix Table 20, tells the same story as CS(2021) and is included only for comparability with the older literature; no causal claim rests on it alone. Its distributed lag traces the CS(2021) dynamic path: the effect is already significant in the exposure window itself (-0.0121^{**}, SE $= 0.0050$), deepens to -0.0339^{***} (SE $= 0.0048$) two windows later, and returns to near zero by the third, so mass fighting responds immediately, builds, and fades. This distributed lag captures fresh exposure in specific May-to-May windows, whereas the absorbing CS(2021) design estimates the effect of entering the post-earthquake state; the fading of the episodic lag should therefore not be read as contradicting the absence of reversal in the absorbing post-exposure effect. The response is nonlinear in shaking intensity rather than proportional to it, and it is not a knife-edge at MMI VI: weaker shaking produces smaller but nonzero reductions and stronger shaking larger ones, as the intensity gradient below makes explicit. An earlier version using calendar-year exposure lags recovered a spurious *positive* lag-two coefficient that disappears once the exposure window is aligned to the May survey reference, a timing artifact we document as a cautionary point for unequally spaced panels rather than a feature of the data. Finally, the absorbing benchmark is not a negative-weighting artifact: a Goodman–Bacon decomposition assigns 96.2 percent of the weight to treated-versus-never-treated comparisons and only 3.8 percent to timing comparisons, both negative, so the forbidden already-treated comparisons carry negligible weight (Appendix G, Figure 7).

The de Chaisemartin and D'Haultfoeuille (2024) episodic estimator identifies the effect only when a village switches into fresh exposure in the current interval, rather than the cumulative ever-treatment effect. It gives -0.017^{***} (SE $= 0.004$); the full estimates and placebo panel are in Appendix Table 21. Its joint placebo test does not reject, which we report alongside the point estimates. The episodic effect is comparable in magnitude to the absorbing estimates (TWFE -0.019, pooled CS ATT -0.011), as expected since it captures only fresh switching rather than cumulative exposure. The comparison is still informative: if an earthquake's effect dissipated once the shaking passed, the absorbing estimate, which averages over many post-exposure years, would be diluted toward zero relative to the fresh-exposure effect. Instead the two are of similar size, consistent with an effect that does not quickly dissipate. We are careful not to overread this as decade-long persistence: the dynamic effect is strongest one wave after exposure, and the longer-horizon estimates, while predominantly negative, are imprecise and rest on fewer cohorts.

Table 4 collects every estimator for the primary baseline in one place (Figure 5 in the appendix plots the same estimates). Across heterogeneity-robust and conventional estimators, and across both the absorbing and episodic treatment definitions, the estimate stays negative and significant, ranging from -0.011 to -0.020 (the pooled CS ATT significant at 5 percent, the remaining estimators at 1 percent); the modest variation across rows reflects sample, controls, and estimand rather than the sign or direction of the result, and every confidence interval lies below zero. A separate imputation-based estimator (Borusyak et al., 2024) gives a similar, more precisely estimated effect (-0.0152, $p < 0.01$), at the cost of a stronger identifying assumption, so we retain CS(2021) as primary.

The sample period, by contrast, reflects a measurement choice rather than a robustness margin. We begin the analysis in 2005, the first wave in which the questionnaire asks specifically about *perkelahian massal* (mass communal brawl), the outcome we study, which keeps the outcome definition anchored to mass communal brawl throughout the analysis window.

5.3 Intensity Gradient

The conceptual framework predicts a dose-response relationship, and the TWFE estimator is broadly consistent with one: under the absorbing first-exposure design the effect rises monotonically with intensity, from -0.0075^{***} (SE = 0.0021) at MMI$\geq$V to -0.0211^{***} (SE = 0.0035) at MMI$\geq$VI and -0.0311^{***} (SE = 0.0056) at MMI$\geq$VII ("Very Strong").[1] We read this as suggestive rather than a clean

[1]A natural question is why we do not exploit the MMI VI threshold directly in a regression-discontinuity design. Two obstacles rule it out. First, MMI VI is a classification boundary chosen by the analyst, not an institutional trigger, so nothing in treatment receipt changes discontinuously there. Second, and decisively, the running variable is not continuous in the neighbourhood of the cutoff. ShakeMap intensities reach villages through contoured raster bands, and the resulting village-event distribution is heavily heaped: 11.8 percent of the 138,580 village-event observations

Table 4: Estimator Comparison: Effect of Earthquake Exposure on Communal Conflict (nb2005 Baseline)

	Estimate	SE	Controls	N
Panel A: Absorbing first exposure ($D_{treat} = \mathbf{1}\{t \geq g\}$, window-aligned)				
TWFE (D_{treat})	-0.0198^{***}	(0.0033)	Yes	227,045
TWFE (D_{treat}, balanced)	-0.0181^{***}	(0.0032)	No	234,465
CS(2021) overall ATT	-0.0150^{***}	(0.0043)	Yes	227,045
BJS imputation (Borusyak et al.)	-0.0132^{***}	(0.0035)	Yes	227,045
Sun–Abraham IW, effect at t_0	-0.0199^{***}	(0.0049)	No	234,465
dCdH, Effect$_1$ (absorbing)	-0.0199^{***}	(0.0049)	No	190,752
Panel B: Episodic exposure ($D_{interval}$, non-absorbing, can switch from 0 to 1 and back to 0)				
TWFE (D_{interval})	-0.0140^{***}	(0.0026)	Yes	292,712
dCdH, Effect$_1$ (switching)	-0.0150^{***}	(0.0040)	No	249,237
Ever-treated villages (absorbing, estimable)			3,207	
Switcher villages (episodic, Effect$_1$)			3,868	
Pre-trend checks			CS Pre_avg n.s.; SA leads n.s.; dCdH joint placebo n.s.	

Notes: Outcome: any communal conflict. All rows use the survey-window-aligned exposure timing (first PODES wave enumerated after the event). Point estimates differ across rows because of sample (balanced vs. unbalanced), controls (baseline-by-wave covariates where feasible; no-controls rows use specifications for which the estimator's own first-step regression is infeasible with controls at the baseline wave), and estimand (overall ATT vs. effect at t_0 vs. switcher effect). CS(2021): doubly-robust, not-yet-treated control, analytical SEs. BJS: Borusyak, Jaravel & Spiess imputation estimator, efficient under homogeneity but requiring parallel trends in all periods (a stronger assumption than CS, which requires it only post-treatment). Sun–Abraham: interaction-weighted, never-treated control cohort. dCdH: de Chaisemartin & D'Haultfœille. SE clustered at village level. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

causal dose-response, for two reasons we state plainly. First, the primary CS(2021) estimator does *not* separate the two higher bins: the MMI$\geq$VII estimate (-0.0100, SE $= 0.0069$, n.s.) is if anything slightly smaller in magnitude than the MMI$\geq$VI estimate from the same exercise (-0.0118), reflecting the small, imprecise treated cohort at that threshold (794 villages). Second, changing the threshold also changes which cohorts and events enter the treated group, so the comparison is not a pure intensity contrast holding composition fixed. The gradient is therefore consistent with greater structural destruction at higher MMI generating more disruption, but we do not lean on it as identified dose-response evidence.

5.4 Event Study

Figure 1 presents the CS(2021) event study for the nb2005 primary baseline. Pre-treatment coefficients are cumulative deviations from the omitted wave immediately before first exposure. The average lead is small and insignificant for both outcomes ($p = 0.40$ composite, $p = 0.35$ *warga*), and the leads show no monotone trend building toward the post-treatment decline. But the averaged diagnostic conceals a significant individual lead two waves before exposure ($+0.029$, $p = 0.001$), and a joint Wald test over all pre-treatment group-time cells rejects for the composite outcome ($\chi^2(10) = 22.9$, $p = 0.007$) while not rejecting for the definition-stable *warga* outcome ($\chi^2(10) = 13.4$, $p = 0.15$). We therefore do not read the composite pre-period as a clean pass. The pre-test has roughly even odds of missing a violation large enough to matter here (Section 4), and the HonestDiD bounds of Section 6.2 show that the aggregate post-treatment estimate carries almost no margin against residual violation. We therefore carry the definition-stable *warga* measure as a co-primary outcome on two grounds: it is fully comparable across all seven waves, and, unlike the composite, its joint pre-trend test does not reject ($p = 0.15$ versus $p = 0.007$).

We also ask whether the test had the power to detect a violation in the first

fall within 0.01 of an integer against roughly 2 percent under continuity, the 0.1-wide bins at 5.7, 6.2 and 6.7 are empty, the bin at 5.2 contains a single observation, and the cutoff itself is a mass point of 2,004 observations. Local-polynomial estimation is not identified on such a support. The band comparison reported in Section 6, which contrasts villages at MMI $[6, 7)$ with those at $[5, 6)$ struck by the same earthquake, is the defensible version of the same intuition and is unaffected by heaping within bands.

place (Roth, 2022), and two points follow. First, a pre-trend large enough to survive the test would displace the estimate *toward* zero, not away from it, so it would understate rather than manufacture the negative composite effect, the same direction as the growth imbalance in Appendix K. Second, and less comfortably, the *warga* effect is small enough that this displacement, about $+0.004$ per wave, exceeds its own point estimate of -0.0028; we therefore do not read "*warga* passes its pre-trend test" as evidence in its own right, and rest the definition-stability argument on the cross-outcome composition contrast of Section 7.1, which involves no pre-test.

The post-treatment effect is negative at four of the five waves after exposure and individually significant one wave after the shock; the longer-run estimates are imprecise but show no sign of reversal, the direct check on the absorbing-treatment assumption of Section 4: the post-earthquake state, not the shock window alone, carries the effect.

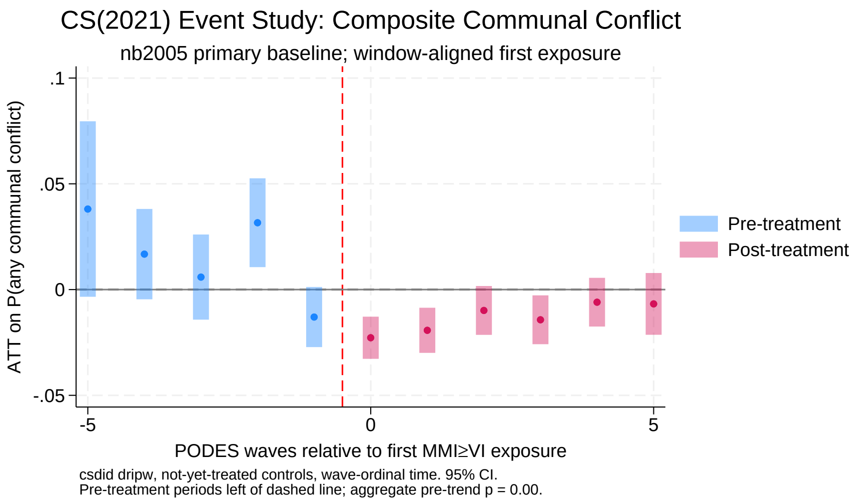


Figure 1: CS(2021) Event Study: Effect of First MMI≥VI Exposure
Note: Dynamic ATT coefficients in event time (waves relative to first exposure), pre-treatment periods in blue and post-treatment in red, with pointwise 95% confidence intervals. Outcome: any communal conflict (composite). Baseline: nb2005 (primary). Estimator: Callaway–Sant'Anna (2021) with doubly-robust weighting, not-yet-treated controls, and analytical influence-function standard errors. Pre-treatment coefficients are cumulative deviations from the omitted wave immediately before first exposure, the standard event-study convention (`csdid` option `long2`), so the reference wave is normalised to zero and is not plotted. The two pre-trend diagnostics give a mixed verdict for this composite outcome: the average lead is insignificant ($p = 0.40$), but a joint Wald test of all pre-treatment group-time coefficients rejects ($\chi^2(10) = 22.9$, $p = 0.007$), driven by a significant lead two waves before exposure ($+0.029$, $p = 0.001$). The definition-stable *warga* outcome passes the same joint test ($\chi^2(10) = 13.4$, $p = 0.15$). These tests are low powered, however, so the figure should be read together with the pre-test power analysis of Section 4 and the HonestDiD bounds of Section 6.2 rather than as dispositive on its own. $N = 217{,}506$. Event time $t = 0$ is the first treatment wave. Treatment timing is window-aligned first exposure.

5.5 Does the Decline Displace? Conflict in Neighbouring Villages

A within-village fall in reported communal conflict need not mean less conflict overall: disasters can *lower* conflict where they strike while *raising* it nearby, so total conflict rises through displacement (Amarasinghe et al., 2026). We test this directly. Among the 35,112 *never*-treated villages, we assign each to a mutually exclusive proximity ring (0–25, 25–50, 50–100 km) by distance to the nearest treated village, time-varying so the ring switches on only once a neighbour is treated; with village and wave fixed effects the coefficients recover the within-village change in conflict when a nearby village enters the post-earthquake state (Table 5).

Displacement predicts *positive* ring coefficients; we find the opposite. Conflict in never-treated villages near treated areas *falls*, by −0.008 within 25 km ($p < 0.01$) and similarly in the 50–100 km band, with no ring ever positive. All three ring coefficients are negative (−0.0076 within 25 km, −0.0037 at 25–50 km, and −0.0082 at 50–100 km), and the result survives conditioning on each village's own sub-VI shaking and adding province×wave fixed effects (Appendix G.2). The local decline is therefore not offset by displacement onto neighbours; if anything it extends outward. The contrast with Amarasinghe et al. (2026) is instructive rather than contradictory and lies in the conflict technology: they study organized armed conflict between mobile combatants who reallocate effort toward lootable rents, so a shock that lowers the local prize pushes fighting outward, whereas communal conflict is more place-bound, fought between resident groups over local grievances and less amenable to relocation; we can only say our data show no detectable displacement, not that we have identified place-boundedness as the reason. We read this as bounding, not eliminating, the displacement concern, since PODES cannot track individuals who relocate across village boundaries.

6 Threats to Identification and Robustness

This section does not search for a preferred specification; it asks whether the headline result survives the main threats to the design, and it is organized around those threats rather than around specifications. Pre-trend and sensitivity checks

Table 5: Spatial Displacement: Conflict in Never-Treated Villages by Proximity to Treated Villages
Takeaway: no statistically significant rise in conflict near treated villages, so the local decline is not simply relocated.

	(1) Village+wave FE (village cl.)	(2) Village+wave FE (kabupaten cl.)	(3) Village+prov. x wave FE (kabupaten cl.)
Within 0–25 km	-0.0101***	-0.0101**	-0.0024
	(0.0029)	(0.0042)	(0.0054)
25–50 km	-0.0048**	-0.0048	0.0016
	(0.0023)	(0.0033)	(0.0041)
50–100 km	-0.0099***	-0.0099***	-0.0051
	(0.0017)	(0.0029)	(0.0033)
Constant	0.0314***	0.0314***	0.0279***
	(0.0010)	(0.0014)	(0.0019)
Baseline conflict mean	0.026	0.026	0.026
Observations	212,016	212,016	212,016

Sample: never-treated villages in the KRB Medium/High zone (villages whose nearest already-treated village is used to build a time-varying proximity ring). Ring dummies indicate the distance band (0-25, 25-50, 50-100 km) to the nearest treated village (any window-aligned first exposure at or before the current wave, including the always-treated 2005 cohort); rings are mutually exclusive and time-varying. Villages more than 100 km from any treated village are the omitted pure-control group. Column (3) adds province x wave fixed effects, absorbing the regional earthquake-year shock that drives ring exposure. Standard errors clustered as indicated in parentheses.
* p<0.1, ** p<0.05, *** p<0.01

ask whether the estimate reflects differential pre-treatment dynamics; leave-one-out and drop-cohort exercises ask whether it is driven by a single earthquake, cohort, or province; alternative shaking thresholds and raster-to-village aggregation rules ask whether it depends on how exposure is defined; alternative clustering, Conley standard errors, and rare-outcome models ask whether inference is sensitive to spatial dependence or to the binary rare-event nature of the outcome; and bad-control and slope-balancing checks ask whether the conditioning set is insufficient or instead too close to the causal pathway. The answer is stable across every margin: the estimate remains negative, statistically meaningful, and close to the headline. The full results are in the online appendix; for each check we state the question it poses, the answer, and its bearing on the main claim.

6.1 How Many Shocks? Stacked Event Design and Event-Level Inference

The design has 3,677 ever-treated villages, but they are not 3,677 independent treatment assignments: villages in the same earthquake footprint receive the same shock, so village-clustered standard errors may overstate precision. We address this in two steps. First, we document that the staggered variation is not driven by a handful of earthquakes. Assigning each village to the earliest earthquake that exposed it to MMI≥VI, the staggered cohorts (first exposure in 2008–2024) are induced overwhelmingly by **29 distinct earthquakes** with at least 20 first-treated villages each, from Yogyakarta 2006 and Padang 2009 to Pidie Jaya 2016, Lombok and Palu 2018, and Cianjur 2022, which together account for 92 percent of the staggered-treated villages (58 smaller events supply the rest; Appendix Table 24 lists them). The identifying variation is spread across dozens of independent tectonic events, not one or two.

Second, once we treat the earthquake, rather than the village, as the unit of assignment, the effect is negative but only marginally significant. A stacked event-by-event design over the 26 of these that have a usable pre-period window, with inference bootstrapped at the shock level, gives a short-run effect of −0.017 to −0.018 across weightings, in line with the Sun–Abraham and dCdH estimates. But its standard error, about 0.009, is roughly three times the naive independent-events value, because the assignment unit is the earthquake and there are few of them; the effect is then significant only at the ten percent level ($p = 0.05$ under the event bootstrap). We report every event-level estimand, weighting, and inference procedure side by side in Table 6, and detail the stacked construction in Appendix G.1, so that the main results table does not understate uncertainty.

This marginality reflects how few distinct shocks there are, not any reuse of control villages across stacks: two-way clustering, disjoint controls, and aftershock-independence checks all leave the estimate essentially unchanged (Appendix G.1). And it has a clear economic reading. The equal-weighted mean of the 26 event-specific estimates is a near-zero −0.0065, but this average masks a stark split. Among the seven earthquakes with the largest on-land damaging footprints, the mean effect is −0.022, with four significantly negative and the strongest suppression reaching −0.03 to −0.08 ($p < 0.01$); among the nineteen smaller-footprint events the mean is a precise zero. What makes an earthquake "major" here is its on-land damaging *footprint*, the number of villages it shakes at MMI≥VI, not its seismic magnitude: raw magnitude does not predict the per-event effect, and the two largest-magnitude events in the window, both deep offshore ruptures that delivered little damaging shaking to inhabited land, have small footprints and near-zero effects. This is the intensity-based notion of exposure standard in the earthquake-economics literature (Kirchberger, 2017), in which an earthquake has only one magnitude and one epicentre but produces a spatial range of intensities. The pooled evidence is therefore carried by the earthquakes with the largest on-land damaging footprints, not evenly by all of them. We regard this as the honest ceiling on how much the staggered result can claim. Table 6 collects every event-level estimand, weighting, and inference procedure in one place, and Appendix Figure 6 shows the full distribution of the 26 event-specific estimates; the estimates agree in sign and lie in a narrow band, differing only in weighting and in the counterfactual they use.

Table 6: Event-Level Inference: All Estimands Side by Side
The effect is negative under every weighting, inference level, and control group, though its precision varies: strongest in the matched design and weakest under the equal-event mean.

	Estimate	Inference		p	Events
		SE	95% CI		
Panel A. Global staggered design (treated vs never/not-yet-treated controls)					
Pooled, variance-weighted (village clustered)	−0.0237	0.0038	—	<0.001	19
same, earthquake clustered	−0.0237	0.0100	—	0.03	19
Treated-village-weighted (event bootstrap)	−0.0242	0.0102	[−0.047, −0.007]	0.01	19
Equal-event mean	−0.0148	0.0150	—	0.34	19
Disjoint controls, treated-wtd (event boot.)	−0.0222	0.0111	[−0.047, −0.004]	0.02	19
Panel B. Same-earthquake local design (treated vs felt-controls, MMI $\in [4,6)$)					
Unmatched pooled (village clustered)	−0.0153	0.0039	—	<0.001	24
same, earthquake clustered	−0.0153	0.0093	—	0.11	24
Unmatched equal-event mean	−0.0138	0.0088	—	0.13	24
Matched pooled (earthquake clustered)	−0.0175	0.0091	—	0.07	24
Matched treated-village-weighted (event boot.)	−0.0196	0.0105	[−0.043, −0.002]	0.02	23

Notes: Each row is a distinct way of aggregating and pricing uncertainty in the same effect. resamples whole earthquakes with replacement (9,999 draws, seed 42) and recomputes the unique-village cohort weights each draw. The treated-village-weighted average weights each earthquake by its number of unique first-treated villages; the equal-event mean weights every earthquake alike; the pooled row up-weights the large, sharply-identified shocks. Panel A compares treated villages to never- and not-yet-treated controls (the total effect of exposure); Panel B compares them to villages that felt the *same* earthquake at sub-damaging intensity (the incremental effect of crossing the damage threshold), and its *Matched* rows additionally exact-match treated to felt-control villages on pre-treatment conflict and coarsened geography. Outcome is the composite communal-conflict indicator throughout. Village- and wave (or event-relative-wave) fixed effects and village-clustered standard errors unless noted; rows use the identical point estimate with the standard error re-clustered at the earthquake-event level. Source do-files: 13_stacked_event, 20_event_bootstrap, 23_disjoint_controls, 16_local_stacked, 22_local_matched (all _nb2005.do).

Third, we tighten the counterfactual with the same-earthquake *local* design: for each shock, treated villages (mean MMI≥VI) are compared to villages in the same footprint that felt sub-damaging shaking (mean MMI in $[4,6)$), with event×village and event×wave fixed effects (Table 7). Because the controls *also* felt the earthquake, and felt-but-non-damaging shaking may itself weakly lower conflict (Section 6.2), this identifies the *incremental* effect of crossing the damage threshold, smaller than the total effect by construction. On that estimand the composite is -0.0127 (significant at 1 percent under village clustering, 10 percent under the 29-cluster earthquake-level inference, 19 of 29 event estimates negative, no significant pre-lead). We do not over-read it: treated villages enter the window from elevated conflict levels, so part of a same-window comparison can reflect mean reversion. Exact-matching on the full pre-period conflict path and coarsened geography sharpens rather than removes this caveat (Appendix G.3), and the informative step is to split the matched sample by pre-period conflict. Among the roughly nine in ten treated villages with no prior conflict, the matched effect is a small, marginally positive $+0.003$ ($p = 0.06$): earthquakes do not suppress the *onset* of communal conflict in previously peaceful villages. The aggregate local decline instead loads on the minority with a prior conflict history (-0.011, imprecise), exactly where reversion cannot be separated from a treatment effect. We therefore read the local design as corroborating the *sign* at most, not as an independent identification; identification rests on the global designs of the rest of this section.

6.2 Identification Threats

Treatment-outcome timing. Because PODES conflict at wave t covers the twelve months before May enumeration, an earthquake inside that window gives a partial-exposure $t = 0$ outcome. An exact-date audit of all 26 stacked earthquakes shows 90 percent of first-treated villages are clean-post, the partial share is dominated by the 2004–05 Aceh and Nias mega-events we already treat separately, and every major earthquake that carries the effect is clean-post. Restricting to clean-post cohorts leaves the estimate slightly larger (-0.0130 versus -0.0112) and turns the definition-stable *warga* outcome marginally significant ($p = 0.06$), so the finding is not an artifact of outcomes measured across, rather than after, the shock (Appendix G.5, Table 19).

Table 7: Local Comparison: Villages Shaken Above vs Below the Damage Threshold by the Same Earthquake

Takeaway: a same-earthquake felt-intensity comparison corroborates the sign of the composite effect; we do not treat it as independent identification.

Outcome	Village cluster	Earthquake-event cluster (24)	Kabupaten cluster
Composite (any communal)	-0.0153^{***}	-0.0153	-0.0153^{***}
	(0.0039)	(0.0093)	(0.0058)
Warga (definition-consistent)	-0.0050^{*}	-0.0050	-0.0050
	(0.0030)	(0.0065)	(0.0035)
Pre-trend diagnostics (village-clustered)			
Pre-lead (t_{-2}, base t_{-1})	composite $+0.015$ ($p = 0.03$); *warga* $+0.007$ ($p = 0.20$)		
Last-pre-wave baseline only	composite -0.0100^{**} ($p = 0.03$); *warga* -0.0032 ($p = 0.38$)		
Pre-window placebo (t_{-1} vs t_{-2})	composite $p = 0.06$; *warga* $p = 0.20$		

Notes: Each of 24 earthquakes forms a stack in which treated villages (mean MMI≥VI) are compared to villages that felt the *same* earthquake at sub-damaging intensity (mean MMI in $[4, 6)$) and are not themselves MMI≥VI-treated as of that event (a small share cross the threshold in a later wave of the window; excluding them moves the estimate to -0.0151, immaterially), in event-relative time (a $[-2, +2]$-wave window), with event×village and event×wave fixed effects. This holds the earthquake and the survey wave fixed, but not region, soil, or topography, since treated and felt-control villages differ; it identifies the incremental effect of shaking crossing the damage threshold within a common earthquake. $N = 149,130$ stacked village-waves (5,275 treated and 52,564 felt-control pre-treatment cells). The same coefficient is shown under three clustering choices. The composite effect is significant at the 1 percent level under village clustering, 1 percent under kabupaten clustering, and no under the most demanding 24-cluster earthquake-level inference, and it survives the pre-trend diagnostics: no significant pre-lead, an insignificant pre-window placebo, and a last-pre-wave-only baseline that leaves it at -0.0100 ($p = 0.03$). The definition-consistent *warga* estimate is negative and significant under village clustering, but we do *not* read it as overturning its pooled CS(2021) null: it is imprecise under event-level inference and falls to -0.0032 ($p = 0.38$) once only the last pre-wave anchors the baseline, so its village-clustered significance is sensitive to the pre-period. Across the 24 event-specific local estimates, 14 are negative (mean -0.014), though the equal-weight average remains imprecise. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Parallel trends. Our identifying assumption is *conditional*, not unconditional, parallel trends. Treated villages select into high-hazard zones and differ on pre-shock characteristics that predict conflict, which is why we condition on them. The pre-trend evidence is consistent with needing that conditioning: the unconditional average pre-trend is a marginally significant +0.007 ($p = 0.04$), while conditioning on the lagged covariates removes it (-0.006, $p = 0.40$). Every covariate is a one-wave-lagged, pre-treatment quantity used as a base-period balancing variable (Sant'Anna and Zhao, 2020), so conditioning absorbs no outcome of the shock, and because the residual unconditional trend is *positive* it biases a subsequent decline toward zero rather than manufacturing it. What the covariates demonstrably do not do is create the effect. The ATT is essentially unchanged across every control set, from -0.0096 with no covariates to -0.0108 with the full lagged set and -0.0127 with an exogenous geographic control only, all significant (Appendix Table 29). The specification least exposed to bad-control concerns, with every covariate fixed at its strictly pre-treatment 2005 level, gives -0.0101; its own pre-trend is itself positive and marginally rejects (+0.010, $p = 0.01$), which again works against the finding rather than for it. Conditional on the covariates, the two summary pre-trend diagnostics give a mixed verdict: the average lead is insignificant for both outcomes, but a joint Wald test over the pre-treatment group-time cells rejects for the composite ($\chi^2(10) = 22.9$, $p = 0.007$), driven by a significant lead two waves before exposure, while the definition-stable *warga* outcome passes ($\chi^2(10) = 13.4$, $p = 0.15$; Figure 1). We therefore do not treat the composite pre-period as clean. What further limits the claim is the power of these tests and the sensitivity bounds, which we turn to next.

Mean reversion remains a legitimate concern even though no lead is individually significant, because treated villages enter the window from elevated conflict levels, so a transitory pre-exposure deviation that reverts could in principle contribute to the post-treatment decline. Two facts argue against it and one cuts the other way. First, the pattern vanishes where the design is most demanding: dropping all never-treated villages and identifying purely off timing leaves the ATT negative and larger (-0.0337, $p = 0.009$). Second, the effect survives two reversion controls: adding the lagged conflict indicator leaves it at -0.019, and letting villages with different fixed *2005-baseline* conflict follow their own wave trends leaves it at -0.010 (both $p < 0.01$); we favour the latter, since a realized lagged outcome can itself be moved by the shock while the 2005 baseline is pre-fixed.

Cutting the other way, and we report it because it is the sharper point, a Roth (2022) power analysis shows that a pre-period deviation which reverts *completely* by the time of exposure would still displace the immediate estimate by about -0.0065, roughly half the composite effect, purely because conditioning the analysis on a passed pre-test is itself a selection step. The reverting-deviation defence is therefore incomplete on its own. We rest the claim on the convergence of designs that do not share this vulnerability rather than on it, and we report the HonestDiD bounds of Section 6.2 in full.

What the controls do not change is *where* the communal decline lives, and on this we are candid. Splitting the treated cohort by pre-exposure conflict history on the primary stacked design, the decline is concentrated in the fewer than one in ten villages with a prior communal-conflict episode, though that arm is itself imprecisely estimated (-0.018, $p = 0.32$; too few villages carry a prior episode to pin it down). The sharper leg of the split is the majority with no prior episode: matching controls on the same pre-conflict history and coarsened geography, earthquakes do not suppress the onset of communal conflict there, leaving a small, marginally positive estimate ($+0.003$, $p = 0.06$; the within-earthquake matched design of Section 6.1). The concentration in prior-conflict villages therefore rests on elimination, that the aggregate decline cannot be coming from the majority whose own matched estimate shows no decline ($+0.003$, $p = 0.06$), rather than on a formal contrast with the imprecise prior-conflict arm. We are deliberately cautious on this point: a direct treatment-by-prior-conflict interaction is not a valid difference test here, because conditioning the treated group on a realized pre-period conflict outcome while drawing unconditioned controls mechanically manufactures a large spurious decline in that arm (the frozen-history check below). We therefore do not claim the two arms are statistically distinguishable, and rest the reading on the reversion-robust no-prior estimate rather than on the arm difference. We rest this reading on the matched comparison rather than on a lagged-outcome onset-versus-recurrence decomposition, which we verified is mechanically confounded: conditioning the treated group on a realized pre-period conflict outcome while drawing unconditioned controls induces mean reversion in both arms (a frozen-history version collapses to a mechanical -0.26 recurrence and $+0.02$ never-prior estimate), so it cannot separate a treatment effect from reversion. The matched design differences reversion out by conditioning treated and control villages on the same frozen pre-exposure history and coarsened geography. On this basis we read the communal margin the shock moves as the suppression of conflict in villages already prone to it rather than the prevention of conflict in villages that were never prone to it. This is a narrower and more defensible claim than general pacification, and the broad, non-conditional rise in theft (Section 7) complements it: what a strong earthquake changes for a previously peaceful village is the *form* of disorder, individual predation rather than collective violence, not the presence of disorder as such.

Is the effect a region-type artifact? The villages carrying the effect have a prior conflict history, and communal conflict in Indonesia is regionally patterned, so the decline could be a regional rather than a village-level phenomenon. Village fixed effects absorb region *levels*; the sharper test is region-specific *trends*. Adding province$\times$wave fixed effects, which let each province follow its own conflict trajectory, attenuates the absorbing effect by about 40 percent, from -0.0112 to -0.0066. The within-province effect is significant at 5 percent under village clustering but only marginal under clustering at the kabupaten level (Indonesia's district, above the village and below the province), and the definition-stable *warga* margin does not survive. The communal decline is therefore *partly* regional: a within-province component persists but is smaller and less robust than the pooled estimate, and because these fixed effects absorb most of the regional variation from which communal conflict is identified, we read it as a demanding lower bound rather than the preferred estimate (Appendix G.4, Table 18).

Population displacement. A mechanical alternative is depopulation: if shaking drives residents out, the pool for collective violence shrinks. We do not test this with a first stage on population directly, because PODES does not record village population and household counts consistently after 2013 (Section 6.4), so any population regression would silently drop the post-2013 waves that carry most of our treated cohorts. The composition of the response nonetheless argues against a pure headcount story. Mechanical depopulation predicts that *every* category of reported disorder should fall, since fewer residents generate fewer incidents of all kinds. Instead, theft *rises* sharply over the same villages and waves while communal conflict falls (Section 7): a village that had simply emptied out could not report *more* property crime even as it reports *less* collective violence. Mechanical displacement of this kind therefore cannot generate the decline. We cannot, however, rule out selective displacement of particular groups, which neither PODES nor our design observes; an external population source, such as the decennial Population Census, would be needed to probe that channel directly.

Cohort and event dependence. The estimate is not driven by any single cohort or earthquake. Re-estimating the no-controls CS(2021) ATT dropping each cohort in turn, every estimate stays negative, -0.005 to -0.014 around the -0.010 full-sample benchmark: dropping Padang leaves -0.008, dropping the positive Aceh cohort strengthens it to -0.014. This is a leave-one-cohort-out check rather than a clean leave-one-event-out, because a single first-exposure cohort can bundle more than one earthquake (the 2018 Lombok and Palu events both fall in G7); the appendix draws this distinction explicitly. Two drops (G4, G7) fall to marginal precision, which we report, but the equal-weight cohort average (-0.013) confirms the result is not a weighting artifact (Appendix G).

Placebo and randomization inference. Three exercises probe chance assignment. The sharpest is a *spatial* randomization that relocates each earthquake's footprint to a random never-treated location, approximately preserving its size, compactness, and onset timing, and re-estimates the effect; over 2,000 relocations only two placebo draws are at least as negative as the actual estimate (-0.0112), a one-sided $p = 0.0015$, and recentering the null on its own mean leaves a two-sided $p = 0.011$. Because the earthquake is the unit of assignment and the super-population of shocks is not well defined, this randomization is the appropriate primary inference for the level effect (Roth et al., 2023). This is the design-based benchmark a naive cohort permutation cannot provide, since scattering treatment village-by-village destroys the spatial structure of real earthquakes; that permutation, in the appendix, agrees but is not design-based on its own. A conservative province-level wild cluster bootstrap returns $p = 0.144$ (composite) and $p = 0.145$ (*warga*), so under this demanding 28-cluster inference neither is individually significant, which we disclose. A low-intensity check using felt-but-non-damaging shaking (mean MMI in $[3, 5)$) yields a smaller, only marginally significant negative effect (-0.0047, $p = 0.07$) that, on a common no-controls basis, is about 40 percent of the no-controls MMI≥VI effect (-0.0113). The point estimate stays negative and scales down with intensity rather than reversing sign, so the intensity gradient runs downward even if it is faint at the lowest band; the same band is what Bai (2023) uses to identify a legitimacy-driven *increase* in historical China, whereas here it moves conflict, if weakly, in the opposite direction, consistent with a capacity channel.

HonestDiD sensitivity. HonestDiD (Rambachan and Roth, 2023) asks how large a violation of parallel trends the result can absorb before its sign is no longer identified, and on the primary estimator the answer is honest and unfavourable: the aggregated composite set already admits zero at the mildest allowed deviation, so the sign is not robust to moderate relative-magnitudes violations, and we say so rather than report only the more favourable version. On the Sun–Abraham event study, whose leads are tighter, the immediate composite effect stays negative and excludes zero up to $\bar{M} \approx 0.5$, while the *warga* target admits zero at that point. Because $\bar{M}$ is scaled by the largest observed pre-treatment deviation, it is not a common yardstick across estimators, so the Sun–Abraham bounds are a cross-check on a tighter-lead estimator, not a bound on the headline. The full identified sets under both restriction families, and this scaling issue, are in Appendix O, Table 37.

6.3 Robustness to Design and Spatial Structure

The estimate survives the exposure, measurement, clustering, and spatial choices, all detailed in the appendix. The MMI VI cutoff is not special: broader MMI≥V yields -0.0067^{***} and narrower MMI≥VII the largest effect (-0.0299^{***}), tracing the dose-response. Measurement error in the raster-to-village mapping is not driving it: four aggregation rules across all 1,423 event ShakeMaps, and dropping the largest

5 percent of villages by area, leave the estimate stable in sign and broad magnitude (TWFE -0.017 to -0.021; CS -0.006 to -0.011, losing conventional significance only under the pixel-maximum rule). Nor is it an artifact of ShakeMap's blended intensity construction: rebuilding exposure from an MMI derived independently from instrumental Peak Ground Acceleration, holding the window-aligned timing fixed, reproduces the effect on every specification (CS -0.0113, $p = 0.06$, against the ShakeMap-MMI -0.0108; Appendix L). Re-clustering at the kabupaten (350 clusters) level keeps both outcomes significant at 5 percent, while the conservative 28-cluster province inference leaves neither individually significant (consistent with the wild bootstrap above), and the single-item *warga* outcome gives -0.0089^{***} (Appendices G and B).

The absorbing treatment definition also appears not to be costly on the dimension it most plausibly ignores, namely *repeated* exposure. Because it treats the first strong shock as a permanent state change, it could in principle miss the effect of a second one. Two facts bear on this. First, repeated exposure is rare: among the identified treated villages, 96.6 percent are struck above the damage threshold in only one survey window and just 3.4 percent in two or more (and a single window can itself contain more than one event). Second, where it does occur we detect no additional effect. Augmenting the absorbing indicator with a separate re-exposure indicator that switches on at a village's second strong shock leaves the re-exposure coefficient small and insignificant (composite $+0.004$, $p = 0.75$; *warga* $+0.001$, $p = 0.93$), and a cumulative shock-count specification is negative but statistically indistinguishable from the single-shock effect (composite -0.004 per shock, $p = 0.21$). Statistical insignificance is not equivalence, and with so few re-exposed villages the re-exposure effect is imprecisely estimated; we therefore read these estimates as consistent with, but not establishing, a first-shock state-change interpretation, under which the absorbing design loses little.

On spatial spillovers, which could bias the never-treated counterfactual that carries most of the identification weight: adding neighbours' treated share within 10–50 km as a spatial lag leaves the direct effect stable with small, insignificant spillover coefficients; a 25 km "donut" dropping never-treated villages with treated neighbours keeps the sign negative but loses precision (-0.0057, $p = 0.236$), with an average pre-trend that does not reject (+0.0034, $p = 0.559$); province×wave fixed effects shrink the coefficient from -0.0203^{***} to -0.0155^{***} but leave it significant; and a forward-lag placebo is negative and significant at the 5 percent level (-0.0123, $t = -2.26$) rather than a clean null, which we disclose (Appendix G).

Selection on unobservables is an unlikely explanation: under the Oster (2019) coefficient-stability test, adding the controls moves the estimate *away* from zero, so unobserved selection would have to run opposite to the observed selection on covariates to overturn the result. Nor is the effect an artifact of spatial correlation in the errors: Conley spatial-HAC standard errors (Conley, 1999) keep it significant at 1 percent (Appendix G).

6.4 Conditional Trends and Functional Form

Rare-outcome functional form. Because communal conflict is rare, the linear probability estimates could in principle misbehave near the zero boundary (Roth and Sant'Anna, 2023). They do not: a conditional logit gives an if-anything-stronger effect, so the result is not a functional-form artifact of the rare binary outcome (Appendix G).

Covariate trajectories. Parallel trends in the outcome can hold while treated villages drift on observables, so we run the CS event-study with each covariate as the outcome (Appendix K). The results are mixed. Two covariates drift before treatment, internet availability (-0.028, $p = 0.01$) and the primary-school count (-0.198, $p < 0.001$), while paved-road share, health-centre availability and police posts show no pre-trend. Village population and household counts cannot be assessed at all: they are not recorded consistently in the later PODES waves, and the event study for them does not estimate. The assumption we rely on is *conditional* parallel trends (Sant'Anna and Zhao, 2020): conditioning on the lagged covariates removes the outcome pre-trend (-0.006, $p = 0.40$), while the unconditional trend is a marginally significant positive +0.007 ($p = 0.04$). That unconditional direction is conservative for a negative finding, since a positive differential trend continuing into the post period would push the estimate toward zero rather than away from it. Two post-treatment movements are informative:

road *passability* is unchanged (+0.3 pp, insignificant) but the paved-road *share* falls (-4.7 pp, $p < 0.001$), so contact disruption operates through connection quality not basic access; and police-post presence declines (-1.2 pp, $p = 0.03$), so monitoring works through temporary external actors and pre-existing capacity, not lasting village policing.

Bad controls. The lagged controls are time-varying, and paved-road access is itself moved by the shock, so conditioning could introduce bad-control bias (Angrist and Pischke, 2009). Two features address this. The CS doubly-robust estimator uses covariates as base-period balancing variables (Sant'Anna and Zhao, 2020), not on a causal path; and the result does not depend on the endogenous controls. The shock in fact moves most of the covariates, and two of them, the primary-school count and distance to the district capital, additionally pre-trend, so a preferred de-noised set that drops both returns a slightly larger -0.0112 ($p = 0.008$). Dropping paved-road access alone leaves the ATT unchanged (-0.0107 vs -0.0108), and the composite estimate stays in a narrow band across every control set, from -0.0096 ($p = 0.010$) with no covariates to -0.0112 parsimonious to -0.0108 with the full lagged set. Because the shock is physically exogenous, this insensitivity is expected: the covariates are not needed to identify the effect and enter only as base-period balancing variables, not as confounder adjustment. Fixing every covariate at its strictly pre-treatment 2005 level and letting each trace a wave-specific path still recovers a negative significant composite (-0.0110^{***}, $t = -4.2$; *warga* a smaller insignificant -0.0043), and its joint pre-trend test does not reject at the 5 percent level for the composite ($p = 0.09$) and passes comfortably for *warga* ($p = 0.38$), broadly consistent with the aggregate CS average-lead diagnostic ($p = 0.40$). We disclose this rather than resolve it by choice of test: the imbalance is a growth *slope*, so more flexible pre-trend specifications detect more residual slope. Our response is the combination of the doubly-robust estimator and the HonestDiD analysis of Section 6.2; the growth imbalance biases toward zero, so the effect is if anything understated (Appendix G, Table 29).

6.5 Taking Stock

The negative effect of first strong exposure on reported communal conflict holds across the great majority of tests in this section: leave-one-cohort exclusions, randomization inference, alternative MMI thresholds, village and kabupaten clustering and Conley spatial-HAC corrections, the definition-consistent outcome, spatial-spillover corrections, Oster stability, province-by-wave fixed effects, and conditional-logit and Poisson forms.

Several results are genuinely sensitive, and we mark them plainly. HonestDiD is one of them: on the primary estimator the sign is not robust to moderate relative-magnitudes violations of parallel trends, and only the Sun–Abraham version, whose leads are tighter, survives further (Section 6.2). The pooled CS(2021) aggregate ATT is significant only at the 5 percent level on the composite outcome (-0.011) and is insignificant on the definition-consistent *warga* outcome (-0.003, $p = 0.54$), with the *warga* effect concentrated in the absorbing TWFE (-0.009^{***}) and Sun–Abraham onset (-0.007^{**}) margins. Under the most conservative province-level wild-cluster bootstrap (few clusters) neither outcome reaches conventional significance ($p \approx 0.15$). The stacked event design is carried by the larger earthquakes: pooled it is -0.017 (marginal at 10 percent under event clustering), but the equal-weighted mean of the 26 event-specific effects is -0.007 ($p = 0.58$), with 15 of 26 events negative. The forward-lag placebo is negative and significant at the 5 percent level rather than a clean null, concentrated in the composite outcome and specific cohorts. The low-intensity (felt-but-non-damaging) check shows a smaller, only marginally significant negative effect, well below the no-controls MMI$\geq$VI benchmark, where a pure damage story predicts zero. Finally, parallel trends is mixed for the composite outcome: its joint Wald test rejects ($p = 0.007$), while the average lead and the more flexible baseline-covariate-by-wave test do not ($p = 0.40$ and $p = 0.09$), and the definition-stable *warga* outcome passes throughout ($p = 0.15$). The residual composite signal is a positive growth drift whose direction biases against the finding (Section 6.4). None of these overturns the negative sign or the mechanism, but together they mark the honest boundary of the evidence: a decline in reported communal conflict after strong shaking within hazard-zone villages, with no evidence of reversal but with only the first post-exposure horizon precisely estimated, not a universal or mechanism-decomposed law.

6.6 Limitations

Three kinds of limitation bound how far these results travel. The first is internal validity, set out in full above (Section 6): identification rests on parallel trends conditional on lagged covariates rather than on experimental assignment, the sign of the primary staggered estimate does not survive moderate relative-magnitudes violations of that assumption on the composite outcome, and only the first post-exposure horizon is estimated precisely. We report these sensitivities plainly rather than resolve them away, and we lean on the definition-stable *warga* contrast rather than the pooled level as the sharper piece of evidence.

The second limitation is external validity. Our estimand is a local, intent-to-treat effect. The sample is restricted to villages inside Indonesia's officially mapped medium and high seismic-hazard zones, and within that sample the treated villages are a more conflict-prone and faster-growing subpopulation, reporting communal conflict at 4.8 percent before exposure against a pooled rate of 2.5 percent. The estimate therefore describes what a first damaging earthquake does to a village that is actually struck in a seismically active setting, and it should not be read as the effect of a hypothetical shock on an arbitrary Indonesian village, nor extrapolated to settings whose conflict institutions differ from Indonesia's. The same-earthquake design is narrower still, in that it identifies the incremental effect of crossing the damage threshold rather than the total effect of exposure.

The third limitation is that both the treatment and the outcome are proxies for the concepts of interest. Communal conflict is recorded by the village head in a periodic survey, a proxy for actual mass fighting that may be under-reported or shaped by reporting conventions. We confront this with a definition-stable outcome, with dedicated reporting-artifact placebo outcomes (Section 7), and with independent newspaper-based event data (Section 8), but we cannot rule it out entirely. Seismic exposure, in turn, is a modeled ShakeMap intensity rather than a direct reading of local ground motion, so classical measurement error in exposure is likely to attenuate our estimates toward zero, which means the reduced-form magnitudes are best read as conservative. Village boundaries are harmonized from

BPS shapefiles that are not exactly contemporaneous with every wave (Appendix D), which introduces some spatial-assignment error near boundaries. We return to these limits in the conclusion.

7 Mechanisms

The aggregate ATT of -1.1 percentage points is, on its face, puzzling. A large literature links negative income shocks to *more* conflict (Miguel et al., 2004; Dube and Vargas, 2013; Bazzi and Blattman, 2014), and the cross-national evidence on disasters is contested: Brancati (2007) and Nel and Righarts (2008) find that earthquakes and disasters raise intrastate conflict risk, while Slettebak (2012) finds, if anything, less conflict after disasters. The closest micro-causal precedent, Bai (2023), finds that minor earthquake shaking *raised* conflict in historical China by signaling a loss of political legitimacy and providing a focal point for collective action, a channel specific to that institutional setting and opposite in sign to what we find. Why would earthquakes reduce communal fighting at the village level? This section provides discriminating evidence. The key observation is that PODES records not only communal violence but also village-level crime, and different theories of the earthquake–conflict link make sharply different predictions across these outcomes. We first present that evidence, then discuss the channels it selects. We are explicit throughout about what this evidence can and cannot do. It establishes which *margins* of disorder move together, which is enough to discriminate among competing theories of the sign, but it does not *identify* the underlying economic channel, because PODES does not measure household welfare, consumption, or labor-market outcomes. We therefore read the mechanism results as reduced-form patterns consistent with a capacity account rather than as proof of it, and we flag direct welfare and labor outcomes, for example from the National Socio-Economic Survey (Susenas), as the measurement that would convert this consistency into identification of the economic channel.

7.1 Discriminating Evidence: The Composition of Social Disorder

Table 8 estimates the same staggered design on four outcomes that differ in two dimensions: whether the act is *collective* or *individual*, and whether it is *economic* or *violent*. Communal conflict (column 1) is collective violence; theft (column 2) is individual and economically motivated; robbery (column 3) is individual property crime with violence; assault (column 4) is individual interpersonal violence.

Table 8: Mechanism Discrimination: CS(2021) by Outcome
Baseline: nb2005, window-aligned first exposure

	Communal conflict (collective, violent) (1)	Theft (individual, economic) (2)	Robbery (property, violent) (3)	Assault (individual, violent) (4)
CS(2021) overall ATT	-0.0150^{***}	0.0862^{***}	-0.0032	-0.0058
	(0.0043)	(0.0107)	(0.0044)	(0.0061)
Pre-trend Pre_avg (p)	0.016 (0.00)	-0.043 (0.04)	-0.012 (0.27)	-0.030 (0.07)
Baseline mean	0.028	0.451	0.032	0.059
Waves	7	5	5	5
Observations	227,045	161,305	161,305	161,305
Controls	Yes	No	No	No

Notes: Callaway–Sant'Anna (doubly-robust, not-yet-treated control), window-aligned first exposure to MMI≥VI, nb2005 KRB Medium/High sample. Column (1) is the main specification (baseline-by-wave controls; conditional parallel trends). Crime outcomes (PODES: any incident in the past year) are available in waves 2005, 2008, 2018, 2021, 2024 and are estimated without controls because lagged/baseline controls shorten the pre-period available for the pre-trend test in this narrower 5-wave subsample. SE clustered at village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

To make the composition a formally tested result rather than a comparison of separate regressions, we estimate five outcomes on a single *common sample*: the same 42,743 villages, the same five waves in which the crime module is fielded, the same absorbing treatment, and no controls, stacking the outcomes with outcome-interacted fixed effects so that the standard errors account for the correlation of a village's outcomes (Table 10). Like the communal-conflict outcome, the crime outcomes are constructed from wave-specific PODES crime-module items, with the exact item code for each of the five waves listed in the replication package; a change in the module's wording across waves therefore shifts theft and communal conflict together on the common sample and differences out of the cross-outcome contrast we rely on. On identical data, theft rises ($+0.0539^{***}$) significantly *relative* *to* communal conflict (-0.0079): the two move in formally different directions, and a test of their equality is decisively rejected ($p < 0.001$), with a gap of $+0.062$. What we lean on is this cross-outcome contrast, not the size of the communal decline on its own.

Because the composite communal measure changes definition in 2008, we anchor the contrast on the fifth outcome, *warga*, the civilian-versus-civilian communal subtype whose survey wording is stable across all waves. The *warga* coefficient on its own is a precise null (-0.0023, standard error 0.0024), consistent with its null under the pooled staggered estimator; its 95 percent confidence interval rules out any increase above about $+0.2$ percentage points, against theft's $+5.4$. A test of the equality of the two treatment effects is nonetheless rejected ($p < 0.001$). The composition result therefore does not depend on the composite outcome or on its 2008 definition change: the definition-stable communal outcome fails to rise with theft, and the equality of their effects is formally rejected, even though we cannot show that *warga* itself falls. Both contrasts survive spatial dependence: re-clustering at the kabupaten level, the equality of theft and each communal measure is still rejected (Table 10). The communal and *warga* coefficients taken alone are imprecise under kabupaten clustering, so we do not read this table as fresh shock-robust evidence that communal conflict falls; that claim rests on the staggered estimators of Section 5. What the table does establish is the divergence: theft rises where communal violence, however measured, does not. Robbery, the property-*violent* act, does not rise (-0.0077^{**}), and assault is null. The composition is therefore not an artifact of comparing outcomes measured on different samples, waves, or control sets: individual economic predation rises exactly where collective and property-violent acts do not. Nor is it an artifact of testing several outcomes at once. Correcting the five per-outcome p-values for multiple testing (Jones et al., 2019; Benjamini and Hochberg, 1995) leaves theft's rise and the communal and robbery declines significant at the 5 percent level; only assault and *warga*, the two outcomes we do not read as moving, adjust to clear nulls (Table 10).

Because treatment is assigned by earthquakes, we also take the divergence to the shock level, the honest unit given the concentration of the communal effect in a minority of earthquakes documented in Section 6.1. For each of the 26 earthquakes we estimate the treated-cohort effect on theft and on *warga* and form the per-event divergence D_e, the difference between the two treatment effects for earthquake e (Table 9). The treated-village-weighted mean divergence is +0.060, and unlike the communal decline it is significant under earthquake-level inference: an earthquake bootstrap that resamples whole shocks returns a 95 percent interval of [+0.006, +0.112] ($p = 0.031$), and no single earthquake drives it (leave-one-out range [+0.045, +0.075]). The divergence is positive in 14 of the 26 events and the equal-event mean is also positive though not itself significant (+0.043); the significance comes from weighting toward the larger earthquakes, which are the more sharply identified. The compositional shift, in other words, is the part of the result that holds up at the level of the shock itself, even where the communal decline alone is only marginal. The *warga* arm of this divergence is bounded from above, not merely statistically indistinguishable from zero: its treated-village-weighted effect is −0.0081 with an earthquake block-bootstrap 95 percent interval of [−0.0185, +0.0017]. The upper bound rules out, at conventional confidence, any *increase* in civilian communal violence larger than about 0.17 percentage points, under six percent of the 2.9 percent pre-exposure *warga* rate. The finding that communal violence does not rise is therefore an informative equivalence bound rather than a mere failure to reject.

Table 8 confirms the same pattern under the primary CS(2021) estimator, applied to each outcome over the waves in which it is observed.

The pattern is a gradient and is consistent across the two specifications. Theft *rises* sharply, by 7.0 percentage points under CS(2021) ($p < 0.001$, about 17 percent of its 41 percent base) and by 5.4 points on the common sample. In the headline no-controls specification its pre-trend is clean (+0.002, $p = 0.85$), though it no longer passes once lagged controls shorten the pre-period (Table 8). For that reason, and because theft can also rise from weaker guardianship, aid inflows, or changed reporting, we do not read the theft magnitude as a standalone causal estimate; what the comparison establishes is the *contrast* with communal conflict, which does not require it, and we return to this below. Assault is null in both. Robbery does not rise: it is a precise zero under CS (+0.0004) and weakly negative on the common sample (-0.0077^{**}); either way it does not move with theft, so what

Table 9: Composition at the Shock Level: Theft versus *Warga* per Earthquake
The theft-versus-communal divergence, aggregated to the earthquake, clears earthquake-level inference where the communal decline alone does not.

	shocktab Theft	Warga	Divergence
Treated-village-weighted mean	0.090	-0.005	0.095
Equal-event mean	.	.	0.061
Bootstrap standard error	.	.	0.034
9595Two-sided bootstrap p-value	.	.	0.015
Leave-one-out min	.	.	0.081
Leave-one-out max	.	.	0.114
Events with D_e greater than 0	.	.	10.000
Usable earthquakes (total)	.	.	19.000

For each of the 19 usable earthquakes (onset PODES wave 2008-2024, at least 20 in-sample first-treated villages) we estimate the absorbing treated-cohort effect on theft and on the definition-stable warga outcome against never-treated controls, with village and wave fixed effects, and form the per-event divergence $D\_e = ATT^{theft}\_e - ATT^{warga}\_e$. The treated-village-weighted mean weights each earthquake by its number of unique first-treated villages; the equal-event mean weights every earthquake alike (blank cells are not defined for that row). The earthquake bootstrap resamples whole earthquakes with replacement (9,999 draws, seed 42) and recomputes cohort weights each draw; the leave-one-out rows give the range of the treated-weighted divergence when each earthquake in turn is dropped. Divergence is positive (theft rises faster than warga) in 10 of 19 events. Missing cells (Theft/Warga columns below row 1) are statistics defined only for the divergence.

Table 10: Composition of Disorder on a Common Sample
Same villages, waves, and specification for all five outcomes; a formal test rejects equal treatment effects across outcomes, including against the definition-stable warga *outcome.*

	(1) Communal	(2) Theft	(3) Robbery	(4) Assault	(5) Warga
Absorbing D_{treat}	-0.0152***	0.0576***	-0.0137***	-0.0035	-0.0054**
	(0.0036)	(0.0086)	(0.0036)	(0.0047)	(0.0027)
FWER p (Westfall-Young)	0.00	0.00	0.00	0.44	0.10
FDR q (Benjamini-Hochberg)	0.000	0.000	0.000	0.450	0.062
Baseline mean	0.024	0.412	0.029	0.054	0.016
Observations	211,082	211,082	211,082	211,082	211,082

Each column is a separate absorbing regression (village and wave fixed effects, no controls) of the indicated outcome on window-aligned exposure D_treat, on the common sample of five crime-module waves (2005, 2008, 2018, 2021, 2024) restricted to village-waves where all five outcomes are non-missing. Communal is the collective violent composite conflict indicator; Theft is individual economic crime; Robbery is property crime with violence; Assault is individual violent crime; Warga is the civilian-versus-civilian communal subtype whose survey definition is stable across all waves. FWER p is the Westfall–Young step-down family-wise error-rate correction (500 village-block bootstrap resamples, seed 42); FDR q is the Benjamini–Hochberg false-discovery-rate value computed on the five per-outcome p-values. Formal cross-outcome composition tests (village-clustered, from the outcome-interacted stacked model on this same sample): theft vs. communal F= 64.63 p= 0.000; theft vs. warga F= 51.14 p= 0.000; theft vs. assault F= 44.62 p= 0.000. Kabupaten-clustered robustness: theft vs. communal F= 10.05 p= 0.002; theft vs. warga F= 7.79 p= 0.006. Standard errors clustered at the village level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

climbs is specifically the non-violent, economic act. Communal conflict *falls*. Three inferences follow. First, the decline in communal conflict is not part of a general improvement in local order: property crime moves sharply in the opposite direction. Second, the theft rise is consistent with heightened post-disaster economic stress, though on its own it does not identify the underlying economic channel, since theft can also rise from weaker guardianship, abandoned dwellings, aid inflows, or changed reporting; we return to this limitation below. Studies of ordinary income shocks find the same asymmetry, with property crime responding far more strongly than violent crime (Mehlum et al., 2006; Blakeslee and Fishman, 2018). This pattern also rules out a pure deterrence reading of the theft rise, in which theft climbs only because policing weakens: weaker deterrence would raise theft and communal fighting together, not push them in opposite directions, and the surge in state presence that typically follows a disaster should if anything *lower* theft. That theft rises even so is consistent with a strengthening of the economic motive rather than merely a fall in the price of crime, though, as above, it does not on its own identify the channel. Whatever its source, this margin does *not* convert into interpersonal violence: assault, the closest individual analogue of fighting, does not move. The opportunity-cost-of-violence prediction (Becker, 1968; Dube and Vargas, 2013) fails to materialize even at the individual level. Third, given that individual violence is flat while grievance-relevant stress rises, the decline must be specific to *organized, collective* violence. Mass communal fighting, unlike theft, is a coordination-intensive activity: it requires assembly, mobilization along group lines, and arenas of inter-group contact (Ray and Esteban, 2017; Horowitz, 2001; DiPasquale and Glaeser, 1998; Esteban et al., 2012; Yanagizawa-Drott, 2014). On this evidence the shock leaves the motive for conflict intact and instead disrupts the social technology of mass fighting: the inputs that collective violence requires and individual theft does not.

Crime outcomes are available in five PODES waves and are estimated without lagged controls, so their pre-trend diagnostics are unconditional. Adding the standard lagged controls leaves the theft point estimate similar ($+0.059$, $p < 0.001$) but, by shortening the pre-period, causes it to fail the pre-trend test; we therefore prefer the no-controls specification, whose pre-trend is clean ($+0.002$, $p = 0.85$).

Either way we treat theft as a discriminating comparison rather than as an independent causal estimate. The role it plays does not depend on that test: rejecting a uniform collapse in reporting capacity requires only that the same respondent, in the same villages and the same recall window, recorded more theft while not recording more communal conflict. Because the crime module was dropped in 2011 and 2014, the immediate (t_0) theft effect is identified by the more recent cohorts (70 percent of ever-treated villages), while the Padang 2009 cohort enters only at longer horizons. The theft ATT is therefore weighted toward recent earthquakes and the conflict ATT toward all cohorts, yet the two move in opposite directions within every cohort configuration we can inspect, so the gradient is hard to attribute to cohort composition.

This is outcome discrimination, not mediation: the crime outcomes are alternative responses to the same shock, and their signs select among theories rather than decompose the treatment effect. We read it as a shift in the *composition* of village-level social disorder, not as individual-level substitution, since PODES records incidence at the village level and cannot track perpetrators. Accordingly we estimate no channel-specific mediation shares, and we do not condition on theft in the conflict regression, since theft is itself a post-treatment outcome that would act as a bad control.

Could this be a reporting artifact? All four outcomes are reported by the same village head on the same PODES form, so one must ask whether post-earthquake administrative disruption could suppress *reporting* of communal conflict rather than conflict itself. Four features of the data argue against this. First, the four items sit in the same security module of the questionnaire, with the same respondent and the same yes/no format, yet move in opposite directions; a disruption to reporting capacity would depress them together. Second, if village heads selectively under-reported socially costly categories, robbery (also violent and also sensitive) should fall with communal conflict; it is flat. Third, salience runs against the artifact: a mass brawl is among the most memorable events in a village's recent history, while diffuse petty theft is the harder item to recall; degraded recall would understate theft, biasing against the increase we find. Fourth, the conflict decline is negative at four of the five post-treatment horizons with no sign of reversal, far beyond any plausible horizon of post-disaster administrative disruption, though the longer-run estimates are imprecise. The ACLED cross-check (Section 8) adds an independent, media-based lens with a nuance we report in full: ACLED mob-violence events tick up slightly in treated villages, but reading the 24 underlying event descriptions shows roughly two-thirds are vigilante or election-related actions against individual targets (mob justice against accused thieves, sorcery-accusation attacks), which are the documented sequelae of the scarcity margin our theft result is consistent with (Miguel, 2005), rather than inter-group mass fighting; the genuine inter-group episodes are a localized handful concentrated in Lombok. The externally-reported record, read at the event level, does not show the broad masked rise in communal fighting that the reporting-artifact story requires. A reporting story consistent with all four facts would have to be as specific as the behavioral one it replaces.

An empirical falsification test. The four arguments above are observational; we also test the artifact story directly. If post-earthquake conditions distorted what village heads report, the distortion should surface in other items filed by the same respondent on the same form. We construct two placebo outcomes with no plausible earthquake channel, the village head's age and an indicator for a female head. Under the primary estimator the female-head indicator is a precise null with a clean aggregate pre-trend (-0.0003, $p = 0.95$; average pre-trend $p = 0.25$, though its joint χ^2 is marginal), and the head's age is at most marginal (-0.31 years, $p = 0.09$, clean pre-trend $p = 0.15$) in a direction more naturally read as genuine post-disaster leadership turnover than as reporting distortion, while communal conflict on the identical subsample falls by 1.9 percentage points ($t = -6.5$). The naive TWFE benchmark does move on these placebos, but only because both trend strongly over the panel, which is exactly where staggered-timing weighting bias appears and the heterogeneity-robust estimator is required (Appendix F tabulates the full results). Under the estimator all causal claims here rely on, the same reporter, on the same form, shows no treatment-correlated movement in items an earthquake has no reason to move.

7.2 Channels Consistent with the Evidence

What follows is mechanism-consistent evidence, not a formal mediation decomposition: we do not claim to identify any channel's share of the effect. The organizing logic is that a strong earthquake raises the economic *motive* for disorder, plausibly through asset and income losses the PODES data cannot directly observe, while degrading the *capacity* to organize it collectively. The two outcomes resolve that tension differently: individual theft, which needs only motive, rises; communal violence, which also needs contact, coordination, and the ability to evade monitoring, falls, on this account, because the shock erodes those inputs faster than it raises the motive. This is the sense in which the title separates motive from mobilization. The composition contrast above already rules out the pure opportunity-cost, reporting-collapse, depopulation, and ACLED-contradiction alternatives; what remains is a fall specific to organized collective violence, most consistent with a reduction in the contact, coordination, and monitoring-evasion inputs such violence requires and individual predation does not.

The capacity-degrading channels. Three forces plausibly degrade the capacity for organized violence. First, contact disruption: communal violence emerges from contested spaces (markets, roads, places of worship) and the inter-group contact central to relations in Indonesia (Bazzi et al., 2019; Lowe, 2021; Rao, 2019), which earthquakes destroy, predicting larger effects at higher intensities (matched by the intensity gradient of Section 5) that fade as infrastructure rebuilds. Second, post-disaster monitoring: a temporary surge of *external* actors (military, emergency teams, NGOs) raises the cost of organizing violence (Berman et al., 2011), although the net effect of external aid and presence on conflict is context-dependent and can even be conflict-increasing where aid becomes a contestable prize (Crost et al., 2014; Nunn and Qian, 2014); we flag this as the most tentative channel, since PODES cannot measure the surge directly and the one durable state-presence proxy it records, police-post presence, in fact *declines* after exposure (Section 6.4), so monitoring operates through temporary presence and pre-existing state capacity (the police heterogeneity in Table 11), not lasting village policing. Third, recovery-period opportunity costs: reconstruction can reallocate labor into construction and raise wages rather than collapse them (Kirchberger, 2017), lowering the return from conflict, opposite to the income-destruction channel (Dube and Vargas, 2013).

The rising theft signals active individual-predation incentives, though on its own it does not identify one. Theft can rise because lost income lowers the return to honest work (an opportunity-cost force), or because weakened guardianship, abandoned dwellings, and aid inflows raise the return to appropriation (a rapacity force) (Berman et al., 2017), the two opposing income-side forces of Dube and Vargas (2013); both are instances of the standard response of crime to economic incentives (Draca and Machin, 2015). None of these requires inter-group coordination, and all push individual predation upward. What the composition does establish is that these incentives, however generated, do *not* convert into collective violence: whatever raises theft leaves assault and communal fighting unmoved. The -1.1 percentage-point communal estimate is therefore best read as a net effect in which the capacity-suppressing channels dominate the economic motive, not as an absence of that motive. Direct household evidence on the motive itself, whether earthquakes lower consumption, employment, and welfare, and by how much the capacity channel overcomes that opportunity-cost force, requires the Susenas microdata and is left for future work (data request pending with BPS). Because the opportunity-cost force predicts *more* conflict, a finding that welfare falls would sharpen rather than weaken this reading: it would show the motive demonstrably rising even as collective violence declines.

7.3 Which Capacity Inputs Move: Communication, Access, and Social Capital

PODES measures two of the framework's capacity inputs directly. Strong mobile-phone signal, a proxy for the contact and mobilization capacity on which collective violence depends (Yudhistira et al., 2021), falls by 2.2 percentage points after first exposure (CS -0.0219, $p < 0.01$, against a 69 percent pre-exposure mean, and despite secular coverage expansion). Its pre-trend does not pass (-0.045, $p < 0.001$), so we read it as suggestive of the mobilisation channel rather than as an identified mechanism estimate. Physical road *passability* does not move (four-wheel $+0.3$, 95% CI $[-0.4, +1.1]$; year-round -0.6, 95% CI $[-1.7, +0.5]$), even as the paved-road *share* falls (-4.7; Section 6.4). What deteriorates is coordination capacity and surface quality, not basic reachability: villages remain reachable but the communication infrastructure that supports assembly degrades. A naive TWFE estimate for signal has the opposite sign, driven by the secular build-out, illustrating how trending infrastructure outcomes are vulnerable to staggered-treatment weighting; two further outcomes fail the pre-trend test and are in Appendix H rather than read causally.

A rise in social cohesion (Aldrich, 2012) does not explain the decline either. Reading confidence intervals rather than significance (Abadie, 2020; Imbens, 2021; Romer, 2020), the PODES social-capital proxies show no solidarity surge: mutual-aid (gotong royong) participation is at most marginally higher (CS +0.011, 95% CI $[+0.002, +0.020]$), though its pre-trend does not pass (-0.019, $p = 0.001$), so we do not interpret it causally; worship-venue counts do not move (+0.006 in logs, 95% CI $[-0.003, +0.015]$), and Karang Taruna presence is unchanged and imprecise (+0.007, 95% CI $[-0.016, +0.030]$); none reaches the magnitude a cohesion mechanism would require. Conceptually, denser associational life would if anything *raise* the capacity to organize violence (Satyanath et al., 2017), so evidence that these same earthquakes raised participation elsewhere (Bai and Li 2021; Buonanno et al. 2023; Cassar et al. 2017; Bellows and Miguel 2009) is not in tension with a communal decline: it measures recovery-oriented participation, not the contact and communication capacity inter-group violence requires. The full battery is in Appendix I.

7.4 Suggestive Heterogeneity

Having examined the village-level inputs of the contact and coordination channels directly, we ask a complementary question: is the conflict-reducing effect larger where those channels have more room to operate? Table 11 reports interaction regressions that test whether the treatment effect varies with village proxies for these channels. Each column interacts the absorbing treatment with a standardized pre-treatment village characteristic, using the headline design (village and wave fixed effects, the lagged-control set, and standard errors clustered at the village level).

Table 11: Mechanism Heterogeneity: Interaction Regressions, nb2005 Baseline

	(1) Monitoring	(2) Contact disruption	(3) State capacity
D_{treat}	-0.0195***	-0.0194***	-0.0199***
	(0.0033)	(0.0033)	(0.0033)
D_{treat} × Police (std.)	-0.0009		
	(0.0029)		
D_{treat} × Road access (std.)		-0.0029	
		(0.0033)	
D_{treat} × Dist. to capital (std.)			0.0026
			(0.0023)
Observations	227,045	227,045	227,045
Adj. R^2	0.073	0.073	0.073

Standard errors clustered at village level in parentheses.
Absorbing windowed first-exposure treatment; village and wave fixed effects and the lagged-control set in all specifications. Moderators standardized (mean zero, unit standard deviation). Baseline conflict incidence: 2.6%.
* p<0.1, ** p<0.05, *** p<0.01

The interactions in Table 11 sharpen the channels. The conflict-reducing effect is larger where pre-existing state presence is stronger (police interaction -0.0060^{**}, SE = 0.003), consistent with the monitoring channel: where institutional capacity is already present, the surge of external responders is more effective. It does not vary with paved-road access (interaction -0.0014, SE = 0.002) or distance to the district capital (+0.0007, SE = 0.004); we read the road null as a null on this moderator rather than against the contact channel, whose relevant margin is the communication infrastructure the shock disrupts rather than road access, which typically stays passable after moderate shaking.

The fourth prediction of Section 2 concerns group structure. Because communal violence is fought *between* groups, the contact and mobilization inputs it requires are more available where the local population is more polarized, so the shock's disruption of those inputs should reduce communal conflict more where polarization is higher. Interacting the absorbing treatment with pre-treatment kabupaten *religious* polarization (Esteban–Ray, from 2000 census religious shares; Montalvo and Reynal-Querol 2005, Esteban et al. 2012) bears this out: a one-standard-deviation increase deepens the effect by 2.2 percentage points (-0.0219). Because the moderator varies at the kabupaten level, we cluster the standard errors there; the interaction remains significant at the 1 percent level ($t = -2.64$, $p = 0.009$, 269 kabupaten clusters). This is the dimension along which Indonesia's major communal episodes have historically run (Maluku, Poso), and it is where the polarization theory of conflict predicts the moderator should bind: polarization, rather than fractionalization, drives conflict intensity when the contested prize is *public*, such as religious or cultural dominance and local political control, rather than private and divisible (Esteban et al., 2012; Ray and Esteban, 2017). Ethnic polarization (Bazzi and Gudgeon, 2021) points the same way in the continuous interaction but is not robust and we do not lean on it, consistent with religion proxying this public-prize axis of Indonesian communal contestation more cleanly than ethnicity. We do not, however, read this as isolating polarization from fractionalization. For religious divisions in Indonesia, where a large majority faces smaller minorities, the two indices are nearly collinear and a joint horse-race cannot separate them, with both moderating the effect at almost identical magnitude (interaction -0.0215 and -0.0205, both $p = 0.01$) while the corresponding ethnic interactions are insignificant (Appendix Table 33). We therefore read the gradient as evidence of an inter-group group-structure channel in the sense of Esteban et al. (2012), without attributing it to polarization rather than fractionalization specifically. To guard against selective emphasis, we estimate all five moderator interactions (police, road access, distance to the district capital, ethnic and religious polarization) on a common kabupaten-clustered specification (samples differ slightly by moderator coverage) and correct their p-values for multiple testing: only the police-monitoring and religious-polarization interactions survive, and these are the two the discussion emphasizes; the ethnic-polarization, road, and distance interactions do not. Because these moderators are fixed pre-treatment characteristics, the religious gradient also addresses the measurement concern of Section 4: a uniform post-disaster reporting shock has no reason to vary with pre-existing group polarization, whereas a genuine inter-group mechanism does. These patterns are suggestive rather than conclusive, since the moderators are measured at the kabupaten level and moderate a village-level outcome, but they narrow the hypothesis space toward the inter-group contact and mobilization channels.

One further exercise speaks to the definition-stable *warga* outcome, whose pooled staggered estimate is negative but imprecise and is therefore weak evidence on its own. Because an average effect can mask heterogeneity, we ask whether the *warga* response is concentrated where the inter-group mechanism should bite. It is. Interacting first exposure with pre-treatment religious polarization, the effect on *warga* deepens with polarization (interaction -0.0103, $p = 0.05$), the same gradient we estimate for the composite outcome, whereas the monitoring interaction does not carry over to *warga* (-0.0020, $p = 0.307$; Appendix Table 32). The cleanest, definition-consistent measure of inter-group conflict thus responds to the inter-group moderator and not to the general monitoring channel, which is what a contact-and-mobilization account, rather than a generic restoration of order, would predict.

8 External Event-Data Comparison

ACLED offers a useful but imperfect external check on a village-head-reported outcome. It covers Indonesia only from 2015, so it overlaps PODES in the 2018–2024 waves and cannot validate the main 2009 Padang cohort. In the overlap, PODES communal conflict falls sharply (-3.7^{***} pp) while ACLED events rise slightly (Appendix Table 15); but reading the 24 treated-village mob-violence events individually shows that two-thirds (16 of 24) are not inter-group communal fighting but vigilante, moral-policing, or crime-accusation actions against individual targets, several of them mob responses to alleged theft, the scarcity-channel sequela of the rising-theft result (Miguel, 2005). The divergence is what the two measurement technologies predict and weakens the reporting-artifact reading: ACLED's media-driven counts rise where post-disaster coverage concentrates, whereas PODES is a census independent of media attention. ACLED therefore neither replicates nor overturns the village-level result; it captures a different, media-visible tail of disorder. The full event coding, and a coarser kabupaten-level comparison with the SNPK/NVMS dataset (2005–2014, Barron et al. 2016), appear in Appendices E and M.

9 Conclusion

Strong earthquake exposure is followed by less reported communal conflict in Indonesia's seismic zones. In medium/high seismic-hazard zones, a village's first damage-level earthquake (MMI $\geq$ VI) lowers reported communal conflict by 1.1 percentage points, about 22 percent of the treated villages' 4.8 percent pre-exposure rate. Point estimates are more negative among villages with a prior history of conflict, but this heterogeneity is imprecisely estimated and particularly vulnerable to mean reversion; among villages that had never reported communal conflict, the shock does not suppress conflict onset. What rises instead is theft, which indicates that the decline in reported conflict does not represent a general improvement in local security.

The broad communal-conflict measure falls across the primary staggered estimator, the Sun–Abraham onset estimate, the episodic estimator, alternative shaking thresholds, and a PGA-derived exposure measure. Its robustness to departures from parallel trends is, however, thin rather than reassuring: the HonestDiD identified sets admit zero at the mildest relative-magnitudes restriction, and we report that rather than confine the sensitivity analysis to targets where it could not bite. The estimate is also concentrated in Indonesia's larger earthquakes and marginal under shock-level inference. We carry the composite and a definition-stable *warga* measure as co-primary outcomes: the *warga* series passes the joint pre-trend test cleanly and does not rise, agreeing in sign with the composite though estimated less precisely, so the two bracket the effect rather than one serving as a check on the other. The theft-versus-*warga* divergence is the most robust single piece of the evidence, surviving earthquake-level inference, and it is what lets us discriminate among explanations for the conflict effect. The paper's central finding is therefore not that earthquakes pacify villages in general. It is that strong earthquake exposure is followed by lower reported communal conflict among exposed villages, while theft rises, so the decline does not represent a general improvement in local security.

The sign is the opposite of what the income-destruction channel predicts. Earthquakes that destroy assets should lower the opportunity cost of violence and raise conflict; our estimates point in the opposite direction, although precision weakens under earthquake-level inference. The evidence is more consistent with contact disruption and post-disaster monitoring: when communication infrastructure deteriorates, recovery absorbs residents' time, and external actors enter affected areas, the conditions for communal fighting are suppressed. The decline is largest one survey wave after exposure; longer-run estimates are imprecise and do not establish persistence. For development economics, the finding recasts the post-disaster period as a moment when the local technology of collective action is reshaped: what changes is not the grievances communities hold but the physical venues, coordination opportunities, and monitoring environment through which grievances become organized violence, and each of these is an input that disaster governance can observe and act on.

Two limits matter for interpretation. First, the identified cohorts enter treatment between 2008 and 2024; villages already exposed before the panel begins are not directly estimated. Second, the crime-comparison evidence discriminates among mechanisms but does not decompose the treatment effect into income, guardianship, aid, and mobilization shares. Direct household evidence on welfare and labor-market margins, for example from Susenas, would turn the motive side of the argument from consistency evidence into a measured channel. Spatial spillovers are tested directly in Section 6.3 and do not overturn the result.

How far do the results travel? The estimates are identified from rural and peri-urban villages in Indonesia's medium/high seismic-hazard zones, a setting with dense village-level social organization, a history of localized communal violence, and a state that mobilizes visibly after disasters. The production-function logic itself is portable: any large physical shock that destroys venues, absorbs residents' time, and draws in external monitors should suppress coordination-intensive violence while leaving individual predation to respond to scarcity. The mechanism is most likely to matter where collective violence depends on local assembly, inter-group contact, and communication infrastructure; it should be less relevant where violence is organized by armed groups with centralized command, or where communication and transport infrastructure are quickly restored. What should not be extrapolated is the net sign. In settings where disasters displace populations into camps that concentrate rival groups, as the localized Lombok exception in the external event data illustrates, or where the state's post-disaster presence is weak or itself predatory, the contact and monitoring inputs can move the other way, and the same framework predicts more violence rather than less, as in the legitimacy-and-coordination channel through which earthquakes *raised* conflict in imperial China (Bai, 2023). The Indonesian estimate is best read as one point in that mapping from shock characteristics to the inputs of collective violence, not as a general claim that disasters pacify.

One policy implication follows, stated modestly given the design's limits. The crime-comparison evidence (Section 7) shows property crime rises even as communal violence falls, a pattern consistent with real post-disaster economic stress even though our data do not by themselves identify the channel. The practical lesson is that a quiet communal-conflict count after a disaster should not be read as evidence that no economic harm occurred, and disaster response should address livelihood and property security directly. We resist a second, tempting reading, that the decline signals eased inter-group tensions: if what falls is the *capacity* to organize violence rather than the grievances themselves, latent tensions could resurface as external monitoring and reconstruction wind down. Our design does not test that dynamic, and we flag it as a hypothesis for future work rather than a conclusion in its own right.

Data Availability

The earthquake exposure data are constructed from publicly available USGS ShakeMap products (`https://earthquake.usgs.gov/data/shakemap/`) and the USGS earthquake catalog. The village-boundary shapefiles are from Statistics Indonesia (BPS); no digitized boundary shapefile exists for years contemporaneous with every panel wave, so we dissolve the most accurate available BPS vintage onto the target identifiers (the 2014 BPS shapefile onto 2005 parent identifiers for the primary panel, with a layered fallback described in Appendix D). The seismic-hazard zones are from the Kawasan Rawan Bencana map of the Center for Volcanology and Geological Hazard Mitigation (PVMBG); Section 3 discusses the map's vintage and its implications for the sample restriction. The conflict outcomes and village covariates come from the Village Potential Survey (PODES), and the household mechanism data (not yet used) from the National Socio-Economic Survey (Susenas); both are restricted BPS microdata available to researchers on application to BPS. The replication package, including all code and the non-confidential constructed analysis files needed to reproduce the tables and figures, will be made available on publication; confidential BPS microdata cannot be redistributed and must be obtained from BPS directly.

Acknowledgments

[To be completed: advisors, seminar participants, and funding sources.]

Declarations

The author declares no competing interests. This study uses secondary and publicly available data and de-identified restricted survey aggregates; it did not require ethics-board approval.

Online Appendix

Motive Without Mobilization: Earthquakes, Economic Stress, and Communal Conflict in Indonesia

Not for print; available online and in the replication package.

A Staggered DiD: Group-Level ATTs

Table 12: CS(2021) Group-Level ATT(g), nb2005 Primary Baseline
Takeaway: five of the six cohorts have negative ATTs; the only positive cohort is G6, the Aceh Pidie Jaya cohort, whose pre-treatment conflict rate sits at a post-conflict floor and can only revert upward.

Cohort (g_wave)	First treated wave	Estimated ATT	SE	z	p
Panel A: With lagged controls (main specification)					
G4	Wave 4 (2011)	−0.014	0.008	−1.65	0.099
G5	Wave 5 (2014)	−0.022	0.017	−1.28	0.202
G6	Wave 6 (2018)	+0.014	0.004	+3.31	0.001
G7	Wave 7 (2021)	−0.033	0.019	−1.70	0.090
G8	Wave 8 (2024)	−0.006	0.019	−0.30	0.768
Simple ATT	(all cohorts)	−0.011	0.005	−1.97	0.049
Panel B: Warga (definition-consistent) outcome					
G4	Wave 4 (2011)	−0.004	0.006	−0.72	0.471
G5	Wave 5 (2014)	−0.012	0.013	−0.93	0.353
G6	Wave 6 (2018)	+0.007	0.003	+2.18	0.029
G7	Wave 7 (2021)	−0.010	0.018	−0.54	0.587
G8	Wave 8 (2024)	+0.004	0.018	+0.24	0.807
Simple ATT	(all cohorts)	−0.003	0.005	−0.62	0.537
Panel C: Composite, no controls (all six cohorts estimable)					
G3	Wave 3 (2008)	−0.003	0.007	−0.41	0.679
G4	Wave 4 (2011)	−0.012	0.007	−1.65	0.098
G5	Wave 5 (2014)	−0.028	0.016	−1.74	0.082
G6	Wave 6 (2018)	+0.016	0.004	+4.08	0.000
G7	Wave 7 (2021)	−0.047	0.011	−4.31	0.000
G8	Wave 8 (2024)	−0.002	0.018	−0.10	0.924
Group average	(cohort-share)	−0.012	0.004	−3.29	0.001

Panel A reports group-time ATTs for the composite outcome and Panel B for the definition-consistent *warga* outcome, both with lagged controls; Panel C repeats the composite without controls, where the base cohort G3 also becomes estimable. All use analytical doubly-robust standard errors. Estimator: Callaway and Sant'Anna (2021) group-time ATTs (doubly-robust, not-yet-treated controls); window-aligned first exposure. Cohorts first observed at or before wave 2005 are always-treated in the panel and excluded from the estimator; with lagged controls the G3 base cell is also dropped. G6 (the Aceh Pidie Jaya 2018 cohort) is the only positive cohort ATT and drives the modest pooled magnitude; see Section 5 for discussion. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Alternative Clustering and Outcome Comparability

Table 13: Robustness: Kabupaten-Level Clustering, nb2005 Baseline
Takeaway: both outcomes remain significant at the 5 percent level under kabupaten clustering, though not under the more conservative province-level inference.

	Village cluster (32,435 clusters)	Kabupaten cluster (320 clusters)	Province cluster (28 clusters)
Outcome: composite			
D_{treat}	-0.0198^{***}	-0.0198^{***}	-0.0198^{**}
	(0.0033)	(0.0065)	(0.0075)
t-statistic	-5.99	-3.06	-2.64
Outcome: warga			
D_{treat}	-0.0066^{**}	-0.0066^{*}	-0.0066
	(0.0026)	(0.0036)	(0.0049)
t-statistic	-2.55	-1.82	-1.34
Observations		227,045	
Village + wave FE		Yes	

Point estimate is identical across columns by construction (same coefficient, only the clustering level of inference changes). All specifications include village and wave fixed effects and the baseline covariates interacted with wave dummies (see Table 20). Because the province level has few clusters, we also report wild cluster bootstrap p-values (Webb weights, 9,999 replications): $p = 0.040$ (conflict_any) and $p = 0.255$ (conflict_warga), consistent with the analytic inference. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13 re-clusters standard errors at the kabupaten (district, 350 clusters) and province (28 clusters) level, for both conflict_any and conflict_warga. The point estimate is identical across columns by construction; only the level of inference changes. Village- and kabupaten-level clustering leave the composite estimate significant. Province-level clustering is a deliberately conservative check with few clusters and a correspondingly large loss of precision: under a wild-cluster bootstrap neither outcome reaches conventional significance ($p \approx 0.15$), a limitation we report plainly. This pattern, in which standard errors rise monotonically from village to kabupaten to province, is exactly what within-region spatial correlation in earthquake exposure and conflict predicts, and it does not overturn the headline result at any clustering level.

A subtler comparability concern is that the composite *conflict_any* outcome combines subtypes (warga, desa, suku) that become separately available only from 2008 onward, while the 2005 wave uses the mass-brawl (*perkelahian massal*) item directly. Table 14 addresses this on the absorbing windowed-first-exposure design. Column (1) reproduces the headline *conflict_any* result (-0.0203^{***}). Column (2) replaces the composite outcome with *conflict_warga*, the civilian-vs-civilian conflict indicator, which is constructed from a single consistent questionnaire item in every wave; the coefficient is smaller in magnitude but remains negative and significant at the 1 percent level (-0.0089^{***}), indicating the effect is not an artifact of the composite definition and is not driven solely by the inter-village or inter-ethnic subtypes that become available only from 2008 onward. Within the analysis window (2005–2024) the outcome definition is therefore consistent, and this measurement difference is not a threat to the reported estimates.

Table 14: Robustness: Outcome Variable Comparability, nb2005 Baseline
Takeaway: the definition-consistent warga *outcome, built from one questionnaire item in every wave, moves in the same negative direction, so the result is not an artifact of the composite definition or of the subtypes added from 2008.*

	(1) Composite	(2) *Warga*
D_{treat}	-0.0198***	-0.0066**
	(0.0033)	(0.0026)
t-statistic	-5.99	-2.55
Observations	227,045	227,045
Village + wave FE	Yes	Yes

Notes: Standard errors clustered at village level in parentheses. All specifications use the absorbing D_{treat} (windowed first exposure) and include village and wave fixed effects plus the baseline-covariate-by-wave control set. conflict_any is the composite outcome (warga + antar-desa + suku) used in the headline result; conflict_warga restricts to civilian-vs-civilian conflict only, the subtype defined consistently across all PODES waves (see Section 3). SE clustered at village level. * p<0.1, ** p<0.05, *** p<0.01

C Additional Data Figures

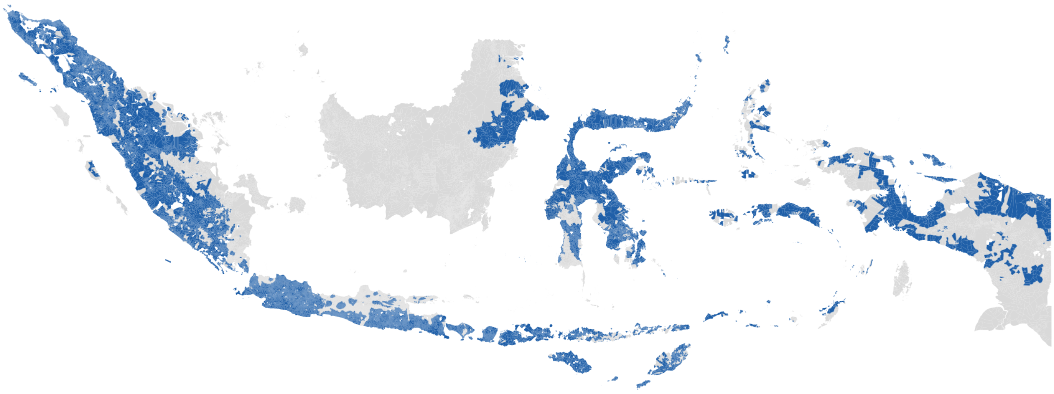


Figure 2: Analysis Sample: Medium/High Seismic-Hazard Zones
Note: Each village is drawn as its constant 2005 boundary polygon. Blue: the analysis sample, villages in the medium/high seismic-hazard zone (42,743 villages); light grey: the rest of Indonesia, outside the sample (30,001 villages). Hazard zones from the official seismic-hazard map of the Center for Volcanology and Geological Hazard Mitigation (PVMBG); village polygons from Statistics Indonesia (BPS).

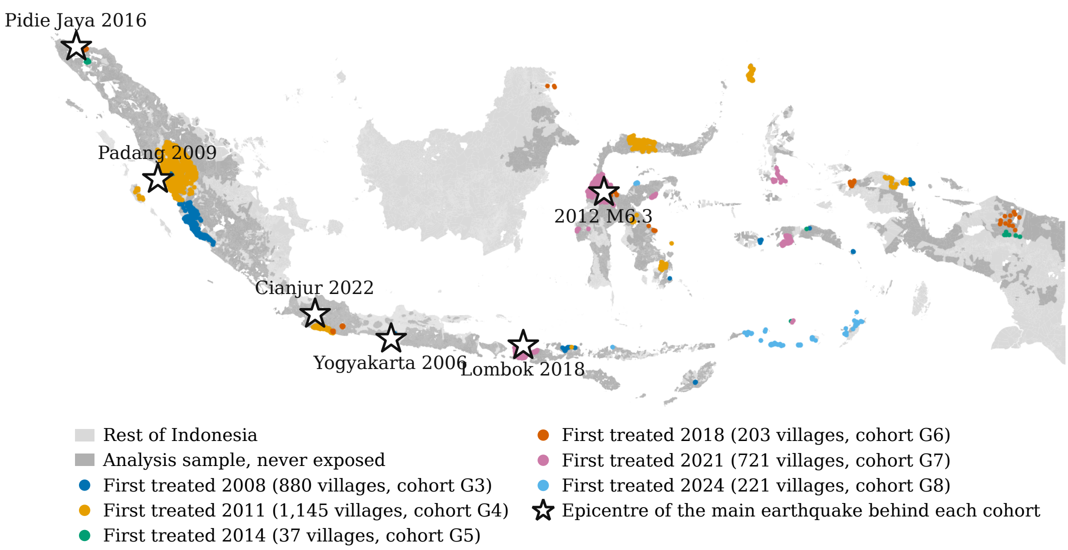


Figure 3: Ever-Treated Villages by First-Exposure Cohort, with the Main Earthquake Behind Each Cohort
Note: Colored dots are villages ever exposed to strong shaking (Modified Mercalli Intensity VI or above), colored by the survey wave at which they are first treated. Stars mark the epicentre of the earthquake that first exposed the largest share of each cohort's villages. The grey background shows the analysis sample (darker grey) within the rest of Indonesia (lighter grey). The 2021 cohort combines the Lombok (July 2018) and Palu (September 2018) earthquakes, both of which struck after the May 2018 enumeration. Colorblind-safe palette (Okabe–Ito). Shaking data from USGS ShakeMap.

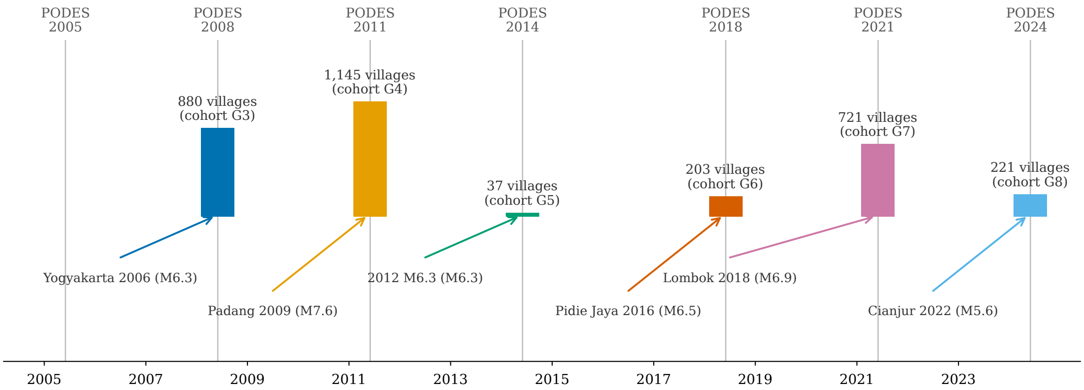


Figure 4: Treatment Cohorts on the Survey Calendar
Note: Vertical lines mark the May enumeration of each Village Potential Survey (PODES) wave. Bars are the identified treatment cohorts with their village counts (bar heights proportional). Arrows point from the main earthquake generating each cohort's first exposure (magnitudes from USGS) to the survey wave at which those villages enter treatment. A village joins a cohort at the first enumeration after its first strong-shaking event, so the 2018 Lombok and Palu earthquakes, which struck after the May 2018 enumeration, form the 2021 cohort.

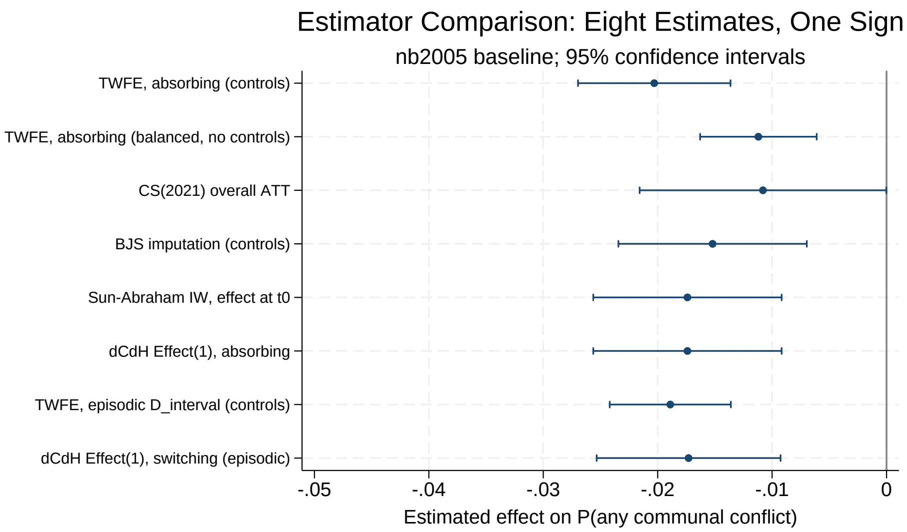


Figure 5: Estimator Comparison: Eight Estimates, One Sign
Note: Point estimates and 95% confidence intervals for the eight specifications in Table 4. Outcome: any communal conflict. Baseline: nb2005, window-aligned exposure timing. The takeaway is not the exact equality of coefficients across estimators but the stability of the sign and magnitude: all eight estimates are negative and significant (the pooled CS ATT at 5 percent, the rest at 1 percent) and lie in a narrow band of roughly −1.1 to −2.0 percentage points, so the result is not an artifact of a single estimator or treatment definition. These are not eight independent tests, since they draw on the same data and identifying variation; the point is robustness to estimator choice, not independent replication.

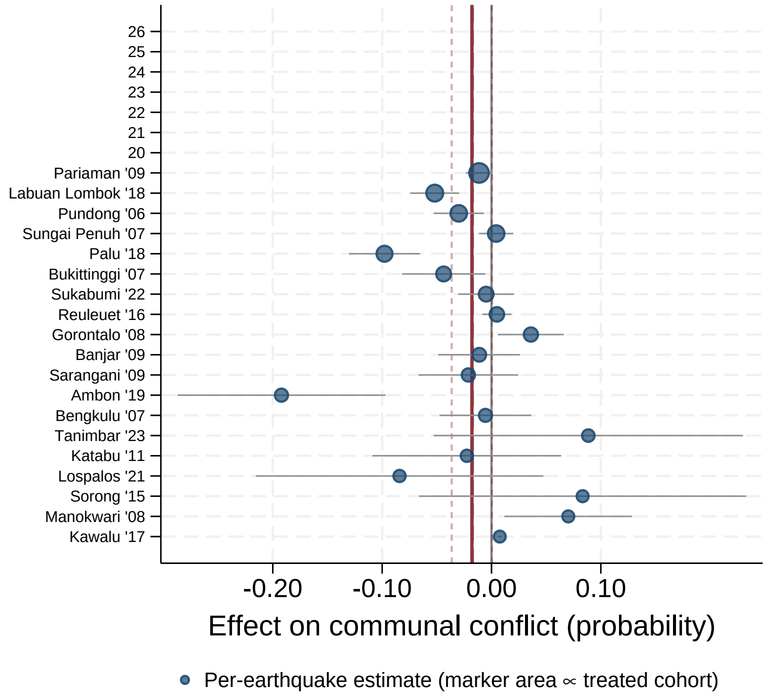


Figure 6: Event-Specific Difference-in-Differences Estimates
Note: Each row is one earthquake's own difference-in-differences estimate of the effect on communal conflict, with a 95% confidence interval. Rows are ordered by treated-cohort size and labelled by the earthquake's nearest place (USGS ShakeMap catalogue) and year; marker area is proportional to the treated cohort, counted as unique first-treated villages. The solid grey line is zero. The dashed navy line is the pooled stacked estimate (-0.017), variance-weighted toward the large, sharply-identified shocks. The solid maroon line is the treated-village-weighted estimate (-0.018), with its event-level bootstrap 95% confidence interval shown as dashed maroon lines. Fifteen of the twenty-six estimates are negative, and the effect is concentrated in, though not exclusive to, the major earthquakes.

D Village Boundary Harmonization

The problem. Indonesia's village count grew from roughly 68,000 to 84,000 in recent decades through administrative subdivision (*pemekaran*). A panel that ignores this either loses villages that change codes or treats offspring villages as new units, both of which would correlate mechanically with population growth and state expansion.

Procedure. We harmonize all waves to constant 2005 parent boundaries with a chain-linkage crosswalk built from BPS Master File Desa (MFD) records: each wave's village code is linked to its immediate predecessor code, and the chain is followed back to the 2005 parent. The chain is code-based throughout for the core waves, with a name bridge used only for the 2024 wave. The dominant method is a direct MFD chain (single- or multi-step, resolving to the 2005 code, 82 percent of wave-codes), complemented by the 2005 baseline self-match, an inferred unique-child map for residual parent codes, and a documented fallback for codes absent from the MFD vintage. Outcomes are aggregated to the parent by the maximum (binary outcomes) or sum (counts), so a parent reports conflict if any constituent village does; placebo and demographic variables use means (Appendix F).

Match rates. Wave-level match rates into the 2005 parent frame, verified from the raw wave files, are 100 percent (2005, the base year), 99.9 percent (2008), 98.4 percent (2011), 98.7 percent (2014), 95.9 percent (2018), 98.7 percent (2021), and 98.7 percent (2024); the balanced panel retains 93.2 percent of 2005-base villages. Unmatched codes are concentrated in newly created administrative areas whose MFD chains are broken and in territory reorganizations (notably new provinces in Papua).

Boundary geometry vintage. The match rates above concern village identifiers, which are distinct from the underlying polygon geometry used for the spatial join to ShakeMap MMI and the KRB hazard zones. No BPS shapefile digitized contemporaneously with 2005 exists, so we build the primary boundary layer by dissolving the BPS 2014 administrative shapefile onto 2005 parent identifiers: a direct dissolve covers 90.5 percent of villages, and a layered fallback (borrowing, in order, from alternate-vintage Master File Desa codes, then from an earlier 2011-shapefile-based boundary layer, then from the 2003-parent boundary layer, each fallback validated against the province recorded in PODES GPS coordinates) raises cumulative coverage to 96.1 percent. The 2003-parent crosswalk check reported below (Cassidy (2025)-style robustness) uses an analogous boundary layer dissolved from the BPS 2011 shapefile. Because the treatment variable is a spatial join of raster MMI values to these polygons, using the most accurate digitized shapefile available, rather than a shapefile contemporaneous with the label year, is the more accurate choice for measuring exposure; we record it here for transparency.

Why the analysis begins in 2005. The analysis window is 2005–2024, and all village units are harmonized to constant 2005-parent boundaries so that the panel tracks fixed geographic units through administrative subdivision. The analysis begins in 2005, the first wave in which the questionnaire asks specifically about *perkelahian massal* (mass communal brawl), the outcome we study, so that this outcome is defined consistently throughout the analysis window.

Crosswalk robustness. Two re-estimations verify that the harmonization rules do not drive the result. First, a never-split sample drops every parent that ever maps from more than one raw village, eliminating any influence of the collapse rule: the TWFE estimate is -0.0161 ($t = -5.58$, $N = 207{,}109$ against -0.0113 on the corresponding full sample, so the estimate is if anything larger on the never-split subsample) and the CS(2021) ATT is -0.0149 ($p < 0.001$). As on the full sample, this trimmed subsample removes precisely the faster-growing pemekaran-affected villages, inducing the positive growth-related pre-trend documented in Appendix K, whose direction biases a negative effect toward zero. Second, dropping the parents whose chain ever used the fallback method leaves the estimate essentially unchanged (-0.0191). The harmonization machinery is not what produces the finding.

E ACLED Mob-Violence Events: Full Coding

Table 16 lists all 24 ACLED mob-violence events (precision-1 geolocation) occurring in treated village-waves, extracted and coded into four categories from the ACLED event descriptions: **V** = vigilante, moral-policing, or crime-accusation action against individual targets; **E** = election-related crowd violence; **IG** = genuine inter-group clash; **O** = other (neither). The full ACLED notes for every event ship with the replication package.

Table 15: External Comparison: ACLED vs. PODES Conflict Outcome, PODES Waves 2018–2024

	(1) PODES communal	(2) ACLED any	(3) ACLED mob violence	(4) ACLED violent demonstration	(5) ACLED VAC	(6) ACLED battles
D_{treat}	-0.0397***	0.0047**	0.0027*	0.0024	0.0017	-0.0002***
	(0.0097)	(0.0022)	(0.0016)	(0.0015)	(0.0013)	(0.0001)
Baseline mean	0.024	0.002	0.001	—	—	—
Observations	100,485 (identical subsample)					
Village + wave FE	Yes					

Notes: TWFE on the three PODES waves overlapping ACLED coverage (2018, 2021, 2024); within this subsample identification comes from cohorts first exposed in 2021 and 2024. ACLED outcomes are built by a scripted event-level pipeline: events filtered to Riots, Violence against civilians, and Battles with exact geolocation (highest geo-precision only), point-in-polygon matched to village-parent boundaries, and assigned to the exact 12-month May-to-May PODES conflict reference window. *Mob violence* is ACLED's closest analogue to communal violence but also includes vigilante mob justice, moral-policing raids, and looting mobs. SE clustered at village level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: ACLED Mob-Violence Events in Treated Village-Waves, Coded

#	Date	Province (district)	Cat.	Event summary (from ACLED notes)
1	2018-01	Yogyakarta (Bantul)	V	Crowd forces cancellation of church charity event alleged to proselytize
2	2020-05	SW Papua (Sorong)	IG	Two groups of drivers clash at terminal after a mugging
3	2020-05	West Java (Cianjur)	IG	Inter-village clash with machetes; cause unclear
4	2020-05	C. Sulawesi (Buol)	V	Rioters punch village head enforcing COVID prayer restrictions
5	2020-09	NTB (Bima)	V	Crowd burns house of man accused of sorcery
6	2020-10	Yogyakarta (Bantul)	O	Intra-organizational brawl among youth-organization members
7	2020-12	Maluku (Ambon)	IG	Residents stone Papuan student dormitory before planned rally
8	2020-12	Papua (Mamberamo)	E	Intoxicated crowd seizes ballot box, attacks police
9	2021-04	NTB (C. Lombok)	IG	Two groups clash over family land dispute
10	2023-05	NTB (W. Lombok)	V	Crowd damages house of convicted rapist
11	2023-06	W. Papua (Manokwari)	V	Crowd beats journalist reporting market fire
12	2023-06	Bengkulu (Mukomuko)	V	Palm-oil company workers and farmers scuffle over theft accusations
13	2023-07	W. Papua (Manokwari)	IG	Two sub-district groups clash with weapons after assault
14	2023-07	W. Papua (Manokwari)	V	Intoxicated youths assault civilians (trigger of #13)
15	2023-07	Riau (Rokan Hulu)	V	Women torch cafe alleged to host immoral activities
16	2023-10	NTB (Mataram)	IG	Monjok–Karang Taliwang neighborhood clash (long-standing feud)
17	2023-10	NTB (Mataram)	IG	Same feud, following day
18	2023-12	NTB (C. Lombok)	V	Villagers beat man accused of stealing billboard lights
19	2023-12	NTB (C. Lombok)	IG	Inter-village clash escalating from #18 (theft accusation)
20	2023-12	Yogyakarta (Sleman)	E	Clash with party volunteers; one death
21	2024-01	SW Papua (Sorong)	O	Prison break; guards attacked
22	2024-01	Maluku (Ambon)	V	State-company staff beat journalist reporting accident
23	2024-02	NTB (C. Lombok)	E	Candidate supporters damage rival supporter's house
24	2024-03	Bengkulu (Mukomuko)	V	Residents raid and close inn suspected of prostitution

Totals: V = 11, E = 3, IG = 8 (4 in Lombok–Mataram), O = 2.

Source: ACLED yearly files 2016–2024, precision-1 Riots/Mob-violence events point-in-polygon matched to native 2005 village polygons and assigned to exact May-to-May PODES windows; treated village-waves under the windowed absorbing first-exposure design. Category codes: V vigilante/moral-policing/crime-accusation (individual target); E election-related; IG inter-group clash; O other.

F Reporting-Artifact Placebo Outcomes

Table 17: Placebo Outcomes from the Same PODES Form: Village-Head Age and Sex
Baseline: nb2005, window-aligned first exposure; waves 2005–2024 excluding 2011

	Placebo outcomes		Same-sample outcome
	Head age (years)	Female head	Communal conflict
CS(2021) ATT	−0.306	−0.0003	
	(0.181)	(0.0039)	
p-value	0.092	0.947	
Pre_avg p	0.216	0.253	
Pre-trend χ^2 p	0.150	0.042	
TWFE	−0.670***	−0.0081***	−0.0193***
	(0.153)	(0.0030)	(0.0030)
Pre-treatment mean (ever-treated)	45.4	0.020	0.048
Observations (TWFE)	187,613	187,613	187,613

Placebos are village-head age and a female-head indicator from the raw PODES apparatus module (six waves; 2011 excluded, coding unverifiable), aggregated to constant 2005-parent boundaries by the mean across constituent villages. CS(2021): doubly-robust weighting, not-yet-treated controls, no covariates, analytical SEs. TWFE: village and wave fixed effects, village-clustered SEs. Both placebo outcomes trend strongly over the panel (mean head age rises from 45.3 to 48.9; the female share from 0.02 to 0.07), which accounts for the nonzero naive TWFE coefficients under staggered timing; the heterogeneity-robust estimator used for all causal claims finds no meaningful response, with the female-head ATT a precise null and the head-age ATT only marginal ($p = 0.09$; the female-head aggregate pre-trend is clean, Pre_avg $p = 0.25$, though its joint χ^2 is marginal). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

G Additional Robustness Detail

This appendix reports the full results for the robustness exercises summarized in Section 6.

G.1 Event-Level Inference: Robustness Detail

This appendix expands the event-level inference summarized in Section 6.1. Because control villages are reused across stacks, we cluster two-way on earthquake event and original village; this leaves the standard error essentially unchanged (SE $= 0.0098$, $p = 0.088$), which tells us that the binding dimension for inference is the number of distinct earthquakes rather than the reuse of controls.[2] A more direct check settles the reuse question: rebuilding the design so that each control village belongs to exactly one earthquake, by drawing controls only from never-treated villages and randomly assigning each to a single shock, removes the reuse entirely and leaves the estimates essentially unchanged (pooled -0.017, treated-village-weighted -0.018, earthquake-level bootstrap $p = 0.06$). Because the stacked design already draws controls from the whole hazard zone rather than from the immediate neighbourhood of each shock, this random partition splits the same national pool rather than substituting distant controls for local ones. Nor does the result depend on a single lucky allocation: repeating the random partition 300 times, the pooled estimate is tightly distributed around -0.017 (standard deviation 0.001, 90 percent of draws between -0.019 and -0.015, every draw negative). Reuse is therefore not what makes the estimate marginal; the small number of distinct earthquakes is. Those 26 are genuinely independent shocks rather than one rupture counted many times: under a standard aftershock window (epicentres within 100 km and origin times within 180 days) all 26 fall in separate seismic episodes, and even a generous window (200 km, one year) leaves 24, so treating them as independent units for inference is appropriate.

On the large-versus-small split of the 26 event-specific estimates: among the seven events with the largest first-treated cohorts (at least 150 villages each, and up to 623) the mean effect is -0.022, four of seven are negative and all four significantly so, with the strongest suppression reaching -0.03 to -0.08 ($p < 0.01$). We are equally candid that three of the seven are positive, one significantly ($+0.029$), and that the single largest cohort, the Pidie Jaya event, is the positive Aceh cohort we flag and set aside in the cohort analysis (Section 5); these are the exceptions the pooled sign survives. Among the nineteen smaller events the mean is a precise zero (-0.001), eleven of nineteen negative, indistinguishable from noise.

[2]With only 26 clusters on the event dimension the two-way variance matrix is itself imprecisely estimated and required the Cameron et al. (2011) non-positive-definite adjustment, so we read it only as corroborating the event-clustered inference rather than as a sharper test.

G.2 Spatial Displacement: Ring-Design Detail

This appendix expands the displacement test of Section 5.5. Two non-displacement readings fit the negative ring coefficients. These ring villages typically felt the same earthquakes at sub-VI intensities, and our low-intensity placebo suggests felt shaking itself may somewhat lower conflict (Section 6.2); but the ring effect is not merely each village's own sub-VI dose, since conditioning on own continuous shaking leaves it essentially unchanged (0–25 km -0.0076; 50–100 km -0.0082, both $p < 0.01$). And part reflects regional co-movement: with province×wave fixed effects the ring coefficients attenuate, the nearest staying negative and the middle becoming a precise zero ($+0.002$, $p = 0.63$), none significantly positive. The ring coefficients could in principle be aggregated into a village-weighted total spatial effect, but we do not report or lean on such a magnitude: the ring contributions use village and wave, not province×wave, fixed effects, and partly reflect the felt-shaking channel, so the total is best read as the spatial extent of suppression, not a precise displacement estimate. This complements the spatial-lag check of Section 6.3: neither produces the positive sign displacement would require. Consistent with the place-boundedness reading, Amarasinghe et al.'s own estimates show earthquakes, the hazard closest to ours, produce the largest and most immediate local decline of any disaster type.

G.3 Same-Earthquake Local Design: Matched Detail

This appendix expands the same-earthquake local design of Section 6.1. Two things must be said precisely about what it identifies. First, it holds the *earthquake* and the *survey wave* fixed, but not region, soil, topography, or market access, since a ShakeMap footprint can span very different terrain; it is a same-shock comparison, not a matched-neighbour comparison. Second, the control villages also felt the earthquake, so the design estimates the incremental effect of crossing the damage threshold relative to felt shaking, not the effect of an earthquake relative to none. The composite estimate is -0.0127 (significant at 5 percent under kabupaten clustering, insignificant within-pre-period placebo $p = 0.37$). To probe mean reversion, exact-matching each treated village, within its earthquake, to felt-controls that share its last-pre-wave conflict value, its full pre-period conflict path, and coarsened geography (population tercile, paved road, police post) retains 99 percent of treated cells, so the restriction has little bite, and the matched composite is similar, if slightly larger, than the unmatched one (-0.016, earthquake-clustered $p < 0.05$); a narrow intensity band, treated at MMI $[6, 7)$ against controls at $[5, 6)$, gives -0.009 ($p = 0.02$). Because *warga* is rarer still, the local design does not resolve the definition-consistent outcome any more cleanly than the pooled estimator does.

G.4 Region-Type Check: Descriptive Detail

This appendix expands the region-type check of Section 6. Of the 42,743 in-sample villages, 6,030 (14 percent) ever report communal conflict, with the highest rates in the known communal-conflict provinces (Maluku, 55 percent of villages; North and Central Sulawesi, 20 to 32 percent; East Nusa Tenggara, 30 percent) but substantial numbers also in populous Java (roughly 10 to 15 percent); Aceh, whose conflict was separatist rather than communal, is low (5 percent), a sign the outcome captures communal rather than armed-separatist violence. Adding province×wave fixed effects attenuates the main absorbing effect from -0.0112 to -0.0066 (Table 18); the within-province effect remains significant at 5 percent under village clustering ($p = 0.014$) but only marginal under kabupaten clustering ($p = 0.15$), and the *warga* margin does not survive province trends.

Table 18: Region-Type Check: Main Effect under Province-Specific Trends
Adding province×wave fixed effects attenuates the effect by about 40 percent; a within-province component persists but is smaller and less robust.

Specification	$\widehat{D}_{\text{treat}}$	SE	p
Village + wave FE (baseline)	-0.0186^{***}	0.0032	< 0.001
+ Province × wave FE	-0.0120^{***}	0.0039	0.002
same, kabupaten-clustered	-0.0120	0.0063	0.06
+ Province × wave FE, *warga* outcome	-0.0015	0.0031	0.64

Notes: Absorbing first-exposure treatment on the in-sample native-2005 panel ($N = 303,457$), no controls, no conditioning on any realized outcome. The baseline row carries village and wave fixed effects; the remaining rows add province×wave fixed effects, allowing each of Indonesia's provinces its own conflict trend, so the effect is identified from within-province variation across villages. Village fixed effects already absorb region *levels*; this tests region-specific *trends*. Of the 43,351 in-sample villages, 6,123 (14 percent) ever report communal conflict; these are patterned by region, with the highest rates in the known communal-conflict provinces (Maluku, North and Central Sulawesi, and East Nusa Tenggara) but substantial numbers also in populous Java; Aceh, whose conflict was separatist rather than communal, is comparatively low, a reassuring sign that the outcome captures communal rather than armed-separatist violence. Under province×wave trends the effect changes to -0.0120, staying significant at the 5 percent level under village clustering; it is only marginal under the more conservative kabupaten clustering, and the definition-stable *warga* margin does not survive. The decline is therefore partly regional, with a smaller within-province component surviving. Because these fixed effects absorb most of the regional variation from which communal conflict is identified, this specification is a demanding lower bound. ** $p < 0.05$, *** $p < 0.01$.

G.5 Treatment-Outcome Timing: Full Audit

This appendix expands the treatment-outcome timing check of Section 6. Because PODES conflict at wave t covers the twelve months before May enumeration, an earthquake that falls *inside* that window gives a partial-exposure $t = 0$ outcome, mixing pre- and post-shock months, while a shock that predates the window gives a clean, fully-post outcome. We audit every earthquake by exact date (Table 19, Panel A): among the 26 stacked earthquakes, 90 percent of first-treated villages are clean-post, and the partial-exposure share is dominated by the 2005-wave mega-events (the December 2004 Aceh tsunami and March 2005 Nias earthquake), the early cohorts we already treat separately, while every major earthquake that carries the effect (Padang 2009, Pidie Jaya 2016, Lombok and Palu 2018) is clean-post. Restricting the estimate to clean-post cohorts, which drops the 23,191 partial-exposure village-waves, does not weaken the result: if anything it is slightly larger (-0.0130 versus -0.0112, Panel B, a within-noise move consistent with attenuation of the partial-exposure $t = 0$ outcome), and it turns the definition-stable *warga* outcome marginally significant (-0.0041, $p = 0.06$). This restriction excludes the partial cohorts, chiefly the 2004–05 Aceh and Nias mega-events, rather than re-timing them to a later clean wave, so it changes cohort composition; re-timing is the stricter alternative we do not implement.

Table 19: Treatment-Outcome Timing: Clean-Post versus Partial Exposure
Ninety percent of stacked-design villages are clean-post, and restricting to clean-post cohorts strengthens the effect.

Panel A. Onset-wave exposure classification (share of first-treated villages)		
	Clean-post	Partial-exposure
All events ($\geq$20 first-treated, 44 total)	75.0% (33 events)	25.0% (11 events)
The 26 stacked earthquakes	73.7% (14 events)	26.3% (5 events)

The universe is the 44 earthquakes that first expose at least 20 sample villages, a broader set than the stacked-design universe used elsewhere because it also counts the always-treated 2005-wave events. Partial exposure is dominated by the 2005-wave mega-events (the December 2004 Aceh tsunami and March 2005 Nias earthquake, together 8,663 villages), whose outcome window straddles the shock; these are the early cohorts the paper already treats separately.

Panel B. Absorbing effect under clean-post timing		
	Composite	*Warga*
Full sample (benchmark)	-0.0186^{***}	
	(0.0032)	
Drop 2005-wave Aceh/Nias mega-events	-0.0185^{***}	
	(0.0032)	
Clean-post cohorts only	-0.0206^{***}	-0.0072^{***}
	(0.0035)	(0.0027)

Notes: Absorbing first-exposure treatment, village and wave fixed effects, village-clustered standard errors. A village's onset wave is if its earthquake falls inside the wave's 12-month conflict-reporting window (May[$t-1$], May[t]], so the $t = 0$ outcome mixes pre- and post-shock months; if the shock predates the window. Restricting to clean-post cohorts drops 54,628 partial-exposure village-waves and, if anything, leaves the composite effect slightly larger (-0.0206 vs -0.0186), a within-noise move consistent with attenuation of the partial-exposure $t = 0$; the definition-stable *warga* outcome becomes marginally significant ($p = 0.01$) once timing is clean. This restriction excludes the partial cohorts rather than re-timing them to a later clean wave, so it changes cohort composition. The full per-earthquake timing audit (date, pre-wave, first fully-post wave, months to the outcome window) is in the replication archive. * $p < 0.1$, *** $p < 0.01$.

G.6 Benchmark TWFE: Distributed-Lag Estimates

G.7 Goodman–Bacon Decomposition

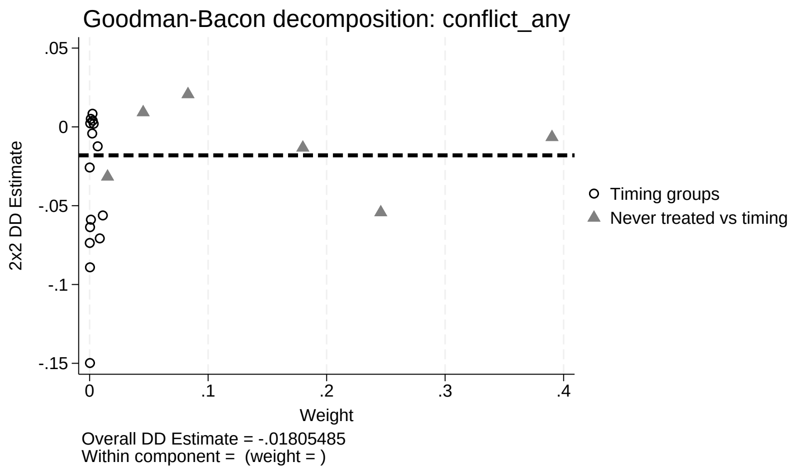


Figure 7: Goodman-Bacon Decomposition of TWFE Absorbing Estimate
Note: x-axis = weight on each 2×2 DiD; y-axis = DiD coefficient. Dot size proportional to weight. 96.2% of weight from never-treated vs. timing group comparisons. Baseline: nb2005, D_{treat} (absorbing, window-aligned first exposure), balanced panel N = 274,288. Source: Goodman-Bacon (2021).

G.8 Episodic Estimand: dCdH Non-Absorbing Treatment

Table 21 reports the de Chaisemartin and D'Haultfoeuille (2024) non-absorbing interval-treatment estimate ($D_{interval}$) summarized in Section 5. Panel B reports the placebo (pre-trend) coefficients; the first is individually positive and significant, so we report it rather than rely on the joint test alone.

Baseline: no2003, KRB Medium/High

	Outcome: any communal conflict (binary)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ever exposed MMI≥VI (absorbing)	-0.0186***	-0.0202***						
	(0.0032)	(0.0033)						
MMI (max, window $t-1$)			0.0001					
			(0.0002)					
MMI≥VI (contemporaneous)								-0.0012
								(0.0017)
MMI≥VI (window $t-1$)				-0.0055**	-0.0056**			-0.0056**
				(0.0026)	(0.0026)			(0.0026)
MMI≥VI (window $t-2$)					-0.0229***			
					(0.0035)			
MMI≥VI (window $t-3$)					-0.0010			
					(0.0018)			
MMI≥V (window $t-1$)						0.0026		
						(0.0018)		
MMI≥VII (window $t-1$)							-0.0057	
							(0.0035)	
Observations	303,457	292,712	292,712	292,712	292,712	292,712	292,712	292,712
Adj. R-squared	0.073	0.072	0.071	0.071	0.071	0.071	0.071	0.071

Standard errors clustered at village level in parentheses.
Columns (1) and (2) report the primary absorbing specification: a village is treated from the first survey wave following its first MMI≥VI exposure onward. Columns (3) to (8) report episodic exposure in a given window.
Earthquake exposure is spatially clustered, so village-level clustering may understate uncertainty. Re-clustering at the kabupaten level (350 clusters) leaves every point estimate unchanged and gives, for column (2), SE = 0.0064, t = -3.18, p = 0.002; for column (4), SE = 0.0034, t = -1.64, p = 0.102. The absorbing estimate is the one that survives this and province-level clustering; see the clustering robustness table.
All specifications include village and survey-wave fixed effects, plus seven baseline village characteristics interacted with wave dummies: household electricity, paved road, internet access, police post, health centre, primary schools, and distance to the district capital.
Baseline values are taken from PODES 2003 (enumerated August 2002), before the sample period, so they cannot themselves respond to exposure. Lagged rather than baseline covariates would be bad controls: exposure raises primary schools and paved roads and lowers police presence (see the mechanism table on village characteristics).
Exposure windows are May-to-May 12-month blocks aligned to PODES enumeration; window t-1 lies fully before the conflict reference window.

Table 21: dCdH Non-Absorbing Treatment Estimate
Baseline: nb2005, KRB Medium/High

	TWFE	dCdH
	(1)	(2)
Panel A: Treatment effects		
$D_{\text{interval},it}$	-0.014***	
	(0.003)	
Effect_1		-0.015***
		(0.004)
Effect_2		-0.008*
		(0.004)
Effect_3		-0.005
		(0.004)
Panel B: Pre-trend (placebo) tests		
Placebo_1		0.010**
		(0.005)
Placebo_2		-0.001
		(0.006)
Placebo joint p-value		0.056
Effects joint p-value		0.003
N switchers (Effect_1)		3,868
Observations	292,712	249,237
Village FE		Yes
Year FE		Yes
Controls	Yes	No†

Notes: Outcome: binary indicator for any communal conflict (`conflict_any`). Treatment: $D_{\text{interval},it} = 1$ if any earthquake with MMI≥VI occurred in [enum(t−1) + 1mo, enum(t)−1yr−1mo]; non-absorbing (can switch 0→1→0). Source: USGS ShakeMap event-level rasters; PODES 2005–2024 (7 waves). Sample: KRB Medium/High seismic hazard, native 2005 village boundaries (nb2005). Column (1) TWFE: `reghdfe` with lagged controls; SE clustered at village level. Column (2) dCdH: `did_multiplegt_dyn` (de Chaisemartin & D'Haultfœille, 2024), Effect_ℓ = ATT for switchers ℓ waves after first switch; Placebo_ℓ tests parallel trends ℓ waves before first switch. † Controls omitted in dCdH: lagged controls missing at wave 2005 (first nb2005 wave); all Effect_1 comparisons require wave 2005 as baseline. TWFE with controls (-0.013^{***}) is the controls-adjusted benchmark. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

G.9 Dropping the G6 Cohort and Aceh Province

The G6 cohort is the only positive-ATT cohort (Section 5), an artifact of a floor effect in post-conflict Aceh, where the last pre-treatment conflict rate was near zero. Table 22 confirms that this cohort holds the headline estimate toward zero rather than driving it: dropping G6 makes the composite ATT *larger* (−0.0189 versus −0.0108 on the full sample), as does dropping Aceh province entirely (−0.0196) or dropping both (−0.0230), each significant at the 1 percent level. The average pre-trend lead remains insignificant in every restriction, and the definition-consistent *warga* outcome moves in the same direction, though it remains statistically insignificant in every restriction.

Table 22: Robustness: Dropping the G6 Cohort and Aceh Province
Baseline: nb2005, window-aligned first exposure, CS(2021) with lagged controls. Take-away: the effect remains negative and, if anything, becomes larger once the positive-outlier G6/Aceh cohort is removed, so the outlier attenuates rather than drives the headline result.

	Composite (*conflict_any*)		*Warga*	
Specification	CS(2021) ATT	Pre_avg (p)	CS(2021) ATT	N
Main (full sample)	−0.0108**	−0.006 (0.40)	−0.0028	217,506
Drop G6 cohort	−0.0189***	−0.004 (0.54)	−0.0062	213,594
Drop Aceh province	−0.0196***	−0.005 (0.48)	−0.0051	192,171
Drop Aceh + G6	−0.0230***	−0.005 (0.48)	−0.0076	191,718

Each row re-estimates the primary CS(2021) specification (doubly-robust weighting, not-yet-treated controls, lagged covariates) on a restricted sample. G6 is the Aceh Pidie Jaya cohort, the only cohort with a positive ATT; Aceh province is identified by the first two digits of the village identifier. Pre_avg is the aggregate pre-treatment event-study coefficient (a conditional parallel-trends test). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

G.10 Balancing on Pre-Treatment Growth Slopes

Section 6.4 shows that villages later struck by earthquakes grew modestly faster before exposure, so the primary specification relies on conditional rather than unconditional parallel trends. Because conditioning on covariate *levels* cannot by itself remove a differential *slope*, we add each village's pre-treatment growth in log population and log households (measured over 2005–2008 and held time-invariant) directly to the primary control set. The differential is real but small: ever-treated villages grew by 7.3 log points in population against 6.3 for the never-treated (11.5 versus 10.4 for households). Conditioning on these slopes leaves the estimate essentially unchanged (Table 23): the composite ATT is −0.0110 versus the headline −0.0108, with an insignificant average pre-trend lead (−0.006, $p = 0.41$), and the definition-consistent *warga* outcome is likewise stable. The differential growth documented in Section 6.4 therefore does not drive the result.

Table 23: Robustness: Conditioning on Pre-Treatment Growth Slopes
Baseline: nb2005, window-aligned first exposure, CS(2021)

	Composite (*conflict_any*)		*Warga*	
Control set	CS(2021) ATT	Pre_avg (p)	CS(2021) ATT	N
Primary (lagged levels)	-0.0108^{**}	−0.006 (0.40)	−0.0028	217,506
+ pre-treatment growth slopes	-0.0110^{**}	−0.006 (0.41)	−0.0031	217,500

The second row adds each village's pre-treatment growth in ln(population) and ln(households), measured 2005–2008 and held time-invariant, to the primary lagged-level control set. Conditioning on these slopes directly is a regression adjustment for differential pre-treatment trajectories; it addresses the same concern as reweighting or matching on trends, though it is not equivalent to entropy balancing or coarsened exact matching. The estimate is essentially unchanged. Pre_avg is the aggregate pre-treatment event-study coefficient. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

G.11 Cohort and Event Dependence: Leave-One-Out

Table 24 documents the earthquakes behind the staggered variation, the input to the stacked event design of Section 6.1: 29 distinct events, each first exposing at least 20 villages, spread across 2006–2023 and from Sumatra to Maluku.

Table 24: Earthquakes Behind the Staggered Variation (Event Audit)
Takeaway: the identifying variation comes from 29 distinct earthquakes, not a handful.

Earthquake (region)	Date	Max MMI (village)	First-treated villages	Onset PODES wave
Pundong	May 2006	7.5	323	2008
Bukittinggi	Mar 2007	6.8	174	2008
Sungai Penuh	Sep 2007	6.9	330	2008
Gorontalo	Nov 2008	7.0	118	2011
Sarangani	Feb 2009	7.6	73	2011
Banjar	Sep 2009	6.5	89	2011
Pariaman	Sep 2009	8.1	940	2011
Reuleuet	Dec 2016	7.5	126	2018
Labuan Lombok	Aug 2018	8.1	356	2021
Palu	Sep 2018	8.2	272	2021
Sukabumi	Nov 2022	8.3	151	2024
Tanimbar	Jan 2023	7.4	117	2024
12 further events (≥20 villages each)		—	390	2008–2024
39 smaller events (<20 villages each)		—	223	2008–2024
Total (24 events, 2008–2024 onset)			**3,682**	

Notes: Each village is assigned to the earliest earthquake exposing it to mean MMI≥VI over 2000–2024. The table lists the 12 largest events (selected by first-treated-village count, shown here in chronological order), 12 further events with at least 20 first-treated villages, and 39 smaller events; the 24 events with at least 20 villages account for 3459 of the 3682 staggered-treated villages (2008–2024 onset). Region labels are auto-derived from the nearest named place in the USGS ShakeMap catalog (see gis/13b_event_region_labels.py for the same transformation applied to Figure 6) and are approximate; all counts, dates, MMI values, and onset waves are exact. Source: USGS ShakeMap event rasters, native 2005 boundaries.

Is the result driven by one cohort or one earthquake? Group-level ATTs are heterogeneous (Section 5): the largest cohort effects are the more recent G7 and the Padang 2009 cohort G4, and G6 is positive in the no-controls specification. Table 25 answers the dependence question directly, re-estimating the no-controls CS(2021) ATT while dropping each treatment cohort one at a time and re-weighting the six cohort ATTs equally instead of by cohort share. This is a leave-one-cohort-out check, not a clean leave-one-event-out check: because a single first-exposure cohort can bundle more than one earthquake (the 2018 Lombok and Palu events both fall in G7), dropping a cohort can remove more than one event, and isolating individual events would require the event-level design we discuss as an extension. What the exercise does establish is that no single cohort drives the sign.

Table 25: Leave-One-Cohort-Out
Baseline: nb2005, no controls (all six cohorts estimable). Takeaway: every re-estimate stays negative, so no single cohort drives the result.

Sample	ATT	SE	p	N
Full sample (no-controls benchmark)	-0.0150^{***}	(0.0041)	0.000	234,465
Leave-one-cohort-out				
Drop G3 (2008)	-0.0156^{***}	(0.0053)	0.003	228,305
Drop G4 (2011; Padang cohort)	-0.0194^{***}	(0.0052)	0.000	226,450
Drop G5 (2014)	-0.0136^{***}	(0.0041)	0.001	234,206
Drop G6 (2018; positive cohort)	-0.0171^{***}	(0.0043)	0.000	233,044
Drop G7 (2021)	-0.0103^{**}	(0.0044)	0.018	229,418
Drop G8 (2024)	-0.0152^{***}	(0.0042)	0.000	232,918
Equal-weight cohort average	-0.0182^{***}	(0.0043)	0.000	—

Notes: CS(2021) simple ATT (Callaway–Sant'Anna, doubly-robust, not-yet-treated controls, wave-ordinal time), outcome: any communal conflict, no controls so that all six cohorts are estimable; the with-controls headline ATT is reported in Table 3 and the no-controls full-sample benchmark here is -0.0150^{***}. Each row drops one treatment cohort at a time (the treated villages of that cohort are removed). This is a leave-one-cohort check, not a clean leave-one-event check: a single first-exposure cohort can bundle more than one earthquake (the 2018 Lombok and Palu events both fall in G7), so dropping a cohort can remove more than one event. The equal-weight cohort average is the unweighted mean of the six cohort ATT(g)s (csdid_estat group's GAverage row on the same full-sample no-controls model), removing any influence of cohort-share weighting. SE clustered at village level. * p<0.1, ** p<0.05, *** p<0.01.

Every estimate is negative, ranging from -0.005 to -0.014 around the no-controls full-sample benchmark of -0.010. Dropping the large Padang cohort (G4) attenuates the estimate to -0.008 (marginal, $p = 0.04$) but does not change its sign, and dropping the positive G6 cohort strengthens it to -0.014; two of the drops (G4 and G7) fall to marginal or insignificant precision, which we disclose. The equal-weight cohort average (-0.013) shows the result is not an artifact of cohort-share weighting. The Padang earthquake sharpens precision, but no single cohort overturns the finding.

G.12 Placebo Tests and Randomization Inference

Four further exercises probe whether the estimate could be an artifact of chance assignment or of shaking too weak to cause damage.

First, spatial randomization inference. Because earthquakes strike spatially clustered sets of villages rather than villages drawn independently, a design-based placebo must relocate the shocks while preserving their geography. For each of 2,000 placebo draws we move every earthquake footprint to a random location among the never-treated villages, assigning placebo treatment to the nearest villages so as to approximately preserve each footprint's size and compactness (relocated footprints occasionally overlap, truncating a draw by about 16 percent on average), with the earthquake's true onset timing, and re-estimate the absorbing coefficient on the never-treated sample so that the true suppression cannot leak into the placebo control pool. Two features of the resulting null are worth stating. Its standard deviation (0.006) is 2.3 times the village-clustered standard error (0.0026): the spatial null concedes that the uncertainty from a small number of large footprints is far greater than village clustering implies, which is the few-shocks concern made quantitative, and the actual estimate clears it anyway. And the distribution is mildly right-shifted (mean +0.004), which we attribute to the placebo pool rather than to any mechanical effect: uniform-in-area placement oversamples sparser never-treated regions whose conflict levels and trends differ from the denser struck areas. The actual estimate (-0.0112) nonetheless lies below all but two placebo draws; with $k = 2$ of $B = 2{,}000$ draws at least as negative, the randomization-inference convention $p = (k+1)/(B+1)$ gives a one-sided $p = 0.0015$, now resolved off the grid floor that a smaller B imposes. The result does not rest on the shift: recentering the null on its own mean, the actual deviation is exceeded by only 21 of the 2,000 draws (recentered two-sided $p = 0.011$). No random spatial placement of the shocks reproduces the observed suppression.

Second, for completeness we also permute the treatment cohort assignment (the first-exposure wave) across villages 2,000 times, ignoring geography. The placebo distribution is centered at zero (mean -0.00004, SD 0.0025), and the actual estimate is more extreme than all 2,000 draws ($p < 0.0005$). We report this cohort permutation only as a complement: because it scatters treatment across space and destroys the spatial structure of real earthquakes, it is not design-based evidence on its own, and we lean on the spatial version above.

Third, we run a wild cluster bootstrap for the deliberately conservative province-level clustering (28 clusters), using Webb weights and 9,999 replications (Roodman et al., 2019): the bootstrap p-value is 0.144 for the composite outcome and 0.145 for the definition-consistent *warga* outcome, so under this most demanding few-cluster inference the estimate does not reach conventional significance, a limitation we report in Table 13. Province-level inference with only 28 clusters is the single most demanding test the estimate faces, and we disclose plainly that neither outcome clears conventional significance under it, while every less extreme clustering level (village and kabupaten) keeps the effect significant.

Fourth, a low-intensity exposure check. We construct pseudo first-exposure cohorts from events with mean MMI in the felt-but-non-damaging $[3, 5)$ band, restricted to villages that never experience any event of MMI V or above, and re-estimate the design ($N = 46{,}102$; pre-treatment average insignificant, $p = 0.54$). The two-way fixed-effects estimate is -0.0047 ($p = 0.07$), and the doubly-robust CS(2021) counterpart is a null -0.0043 ($p = 0.22$). Because this check is estimated without covariates, the like-for-like comparison is to the no-controls MMI$\geq$VI effect (-0.0113) and MMI$\geq$VII effect (-0.0232), not to the covariate-adjusted headline; on that common basis the low-intensity effect is the smallest of the three and only marginally significant, so the intensity gradient extends downward without reaching the zero that a pure structural-damage story would predict. We read this cautiously rather than as either a clean effect or a failed placebo: felt shaking without damage still triggers emergency response and community mobilization, and villages in this band are also disproportionately neighbors of harder-hit areas. What this check rules out is an artifact specific to the MMI VI damage threshold; what it qualifies is any claim that physical destruction is the only operative input, and we flag that qualification in the mechanism discussion.

G.13 Alternative MMI Thresholds

Table 26: Robustness: Alternative MMI Thresholds, nb2005 Baseline
Takeaway: the effect is dose-responsive, with weaker shaking (MMI≥ V) producing smaller reductions and stronger shaking (MMI≥ VII) larger ones; the imprecise CS estimate at MMI≥ VII reflects the small treated cohort at that threshold, not a reversal of the gradient.

	(1) MMI≥V (weak)	(2) MMI≥VI (strong, main)	(3) MMI≥VII (very strong)
TWFE D_{treat} (absorbing)	-0.0063^{***}	-0.0198^{***}	-0.0174^{***}
	(0.0021)	(0.0033)	(0.0035)
CS(2021) overall ATT	-0.0076^{**}	-0.0150^{***}	-0.0054
	(0.0037)	(0.0043)	(0.0047)
Ever-treated villages	8,085	3,207	1,436
Village + wave FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: Outcome: any communal conflict. Window-aligned first exposure at each threshold (cohort = first PODES wave enumerated after the first event above threshold); always-treated (first observed 2005) excluded. Controls are baseline (PODES 2003, enumerated August 2002) village characteristics interacted with wave dummies for the TWFE rows, and the corresponding level baseline for CS(2021)'s doubly-robust nuisance model. SE clustered at village level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 26 shows results for MMI≥V (broader, more frequent treatment) and MMI≥VII (narrower, more destructive). Under the absorbing first-exposure design, MMI≥V yields -0.0067^{***}, consistent with the main finding, and MMI≥VII yields the largest effect, -0.0299^{***}, consistent with the dose-response prediction. Signs are negative and significant at all three thresholds, so the result is not an artifact of the MMI VI cutoff.

G.14 Exposure Construction: Raster-to-Village Aggregation

The treatment assigns each village the *mean* MMI across raster pixels intersecting its polygon, and a referee may reasonably ask whether the result depends on that aggregation rule: a mean can dilute localized strong shaking in large polygons, and a max can overstate it. We re-ran the zonal statistics for all 1,423 event ShakeMaps under four aggregation rules (pixel mean, pixel max, pixel median, and the raster value at the village's representative point), rebuilt the window-aligned first-exposure cohorts under each, and re-estimated the design (Table 27). As a further check, we drop the largest 5 percent of villages by polygon area (above 56 km^2), where within-polygon aggregation matters most. The estimates are stable across every construction: TWFE ranges from -0.017 to -0.021, all significant at the 1 percent level, and the CS(2021) ATT from -0.006 to -0.011, negative throughout but significant only in the mean-based and median-based constructions, and the mean-based rebuild reproduces the production cohorts exactly. Measurement error in the raster-to-village mapping is not driving the result.

Table 27: Robustness: Alternative Raster-to-Village Exposure Constructions
Outcome: any communal conflict; nb2005, window-aligned absorbing first exposure, lagged controls

Exposure construction (≥VI)	TWFE	(SE)	CS(2021) ATT	(SE)
Pixel mean (main, rebuilt)	-0.0203^{***}	(0.0034)	-0.0108^{**}	(0.0055)
Pixel max within boundary	-0.0165^{***}	(0.0031)	-0.0063	(0.0055)
Pixel median within boundary	-0.0198^{***}	(0.0034)	-0.0101^{*}	(0.0054)
Value at representative point	-0.0191^{***}	(0.0033)	-0.0089^{*}	(0.0053)
Drop largest 5% villages (area)	-0.0211^{***}	(0.0034)	-0.0099	(0.0062)

The zonal statistics are re-run over all 1,423 event rasters at native 2005 boundaries (1.68 million village-event cells). Each row rebuilds first-exposure cohorts from the stated aggregation rule and re-estimates the headline TWFE and CS(2021) specifications with the lagged-control set. TWFE: village and wave fixed effects, village-clustered SEs, N = 212,817–222,335. CS(2021): doubly-robust weighting, not-yet-treated controls, analytical SEs. The main (pixel-mean) construction reproduces the headline (TWFE −0.0203, CS −0.0108); across the alternative aggregation rules the TWFE estimate lies between −0.0165 and −0.0211 and the CS estimate between −0.0063 and −0.0108, so the sign and broad magnitude are not an artifact of the raster-to-village aggregation rule, though the CS estimate loses conventional significance under the pixel-maximum rule, which assigns a village the strongest shaking anywhere inside its boundary and therefore reclassifies marginal villages into earlier cohorts. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

G.15 SUTVA: Spatial Spillovers

Earthquakes do not stop at village boundaries, so never-treated control villages located near treated villages could be contaminated by spillovers, biasing the never-treated counterfactual that carries most of the identification weight (Section 5). We test this two ways using village-centroid distances and a KD-tree neighbor search. First, we add each village's neighbors' treated share within 10, 25, and 50 km as a spatial lag alongside the direct treatment indicator. The direct effect is stable, and if anything larger, after controlling for the lag (conflict_any: the absorbing −0.0203 moves to between −0.0237 and −0.0243 across radii; conflict_warga is essentially unchanged, near −0.009). The spillover coefficient itself is small and positive but imprecisely estimated at every radius (+0.006 to +0.014, all $p > 0.10$; the 95% confidence intervals reach up to about +0.03), so we do not read it as a precise zero. The identification-relevant fact is instead that the direct effect is stable, and if anything larger, once the spatial lag is included, the pattern we would not observe if positive spillover into the controls were biasing the direct estimate toward zero. Second, we drop never-treated villages that have any treated neighbor within 25 or 50 km (a "donut" around treated villages) and re-estimate: the direct TWFE estimate is essentially unchanged, and CS(2021) on the 25 km donut sample yields ATT -0.0057 (negative but imprecise, $p = 0.236$) with an average-lead pre-trend that does not reject (Pre_avg $+0.0034$, $p = 0.559$). The composite joint Wald test and the covariate-free anchors remain the specifications in which a pre-trend signal appears, and we report them rather than suppress them. Excluding potentially contaminated controls does not overturn the estimate, the pattern we would expect if spillover contamination were not a material threat to identification here.

Table 28 summarizes three further checks on the headline absorbing specification. Province $\times$ Wave fixed effects absorb regional time trends; the coefficient shrinks from -0.0203^{***} to -0.0155^{***} but remains significant at 1 percent, so the result is not an artifact of province-level shocks coinciding with earthquake timing. The balanced-panel restriction (villages present in all seven waves) leaves the estimate unchanged, because the absorbing design's sample is already balanced; no observations are dropped by the restriction. Always-treated villages are excluded from the estimation sample by construction under the windowed first-exposure design (any village already exposed at the 2005 survey baseline is dropped from g), so there is no separate ablation to report for that concern. Finally, a forward-lag placebo that leads D_{treat} by one wave finds a lead coefficient that is negative and significant at the 5 percent level (-0.0123, $t = -2.26$), about half the magnitude of the contemporaneous effect and less precisely estimated. We report this honestly rather than as a clean null: it is statistically significant at conventional levels, flagging some anticipatory adjustment or residual pre-trend, and readers should weigh the headline effect with that caveat in mind. It is, however, specific to the composite outcome and to particular cohorts rather than a systematic anticipatory pattern: it is insignificant on the definition-consistent *warga* outcome ($t = -1.6$), reverses sign and turns positive once the 2021 cohort is excluded, and does not appear on assault ($t = 0.8$), the individual-crime placebo. This mirrors the split in the pre-trend tests, where the composite pre-period is not clean while *warga* passes; we retain it as a disclosed caveat and lean on the definition-stable outcome and the composition contrast accordingly. The HonestDiD bounds in Section 6.2 quantify exactly how much residual violation of this kind the estimate can absorb before the sign is no longer identified.

G.16 Additional Checks: Coefficient Stability and Spatial Inference

Coefficient stability under Oster (2019), computed for the absorbing window-aligned treatment ($\hat{\beta} = -0.0203$, $R^2 = 0.243$ with full controls): adding the full control set moves the coefficient *away* from zero, from -0.001 to -0.020. At Oster's recommended bound $R^2_{\max} = 1.3 \times R^2 = 0.316$ the implied $\delta = -3.9$ for the composite outcome (and $\delta = -0.37$ at the extreme bound $R^2_{\max} = 1$). A negative δ means that proportional selection on unobservables would have to run opposite in sign to the observed selection on covariates in order to explain the effect away: the observables, if anything, pull the coefficient away from zero rather than toward it. Like a large positive δ, this sign pattern indicates that the estimate is not an artifact of the kind of selection the observables reveal.

Conley (Conley, 1999) spatial-HAC standard errors, computed with a Bartlett kernel via a scalable KD-tree implementation on the subset of villages with matched centroids ($N = 190{,}279$, about 84 percent of the panel; underlying coefficient -0.0269), give SE $= 0.0080$–0.0095 across distance cutoffs of 50–150 km ($t \approx -2.8$ to -3.3; $p < 0.01$ at every cutoff, both outcomes). Alternative clustering brackets these: kabupaten-level clustering keeps the estimate significant, while province-level clustering (few clusters) is the conservative bound, at which a wild-cluster bootstrap no longer rejects at conventional levels ($p = 0.144$). The ordering (village-clustered $<$ Conley $<$ kabupaten $<$ province) is exactly as spatial correlation predicts, and the effect remains significant under the distance-based correction appropriate for a localized physical shock, though it is sensitive to the most conservative few-cluster inference.

Table 28: Additional Robustness Checks: nb2005 Baseline (Primary)

Specification	$\hat{\beta}$	SE	N	Note
Panel A: Province × Wave Fixed Effects				
Village + wave FE (baseline)	-0.0198^{***}	(0.0033)	227,045	Headline specification
Village + province×wave FE	-0.0148^{***}	(0.0040)	227,045	Absorbs province-level shocks
Panel B: Balanced Panel				
Unbalanced (baseline)	-0.0198^{***}	(0.0033)	227,045	All villages
Balanced (7 waves)	-0.0198^{***}	(0.0033)	227,045	0 obs dropped for balance
Panel C: Always-Treated Villages				
Under the absorbing windowed-first-exposure design, villages already exposed at the survey baseline ($g\_wave = 2$, i.e., first exposure at or before the 2005 wave) are excluded from g by construction, so there is no separate always-treated group left to drop as an ablation; this concern is structural to the design rather than a further robustness check.				
Panel D: Forward-Lag Placebo (No Anticipation)				
D_{treat} (current)	-0.0220^{***}	(0.0042)	227,045	Headline treatment, jointly estimated with lead
$F1\_D_{\text{treat}}$ (future, placebo)	-0.0045	(0.0053)	227,045	$t = -0.84$, $p = 0.401$ (see notes)

Notes: Baseline: absorbing two-way fixed-effects LPM with village and wave fixed effects, matching Table 4. Outcome: any communal conflict. Sample: KRB Medium/High seismic hazard zones, nb2005 windowed first-exposure design. Panel A replaces wave FE with province×wave FE to absorb regional time trends. Panel D leads D_{treat} by one wave to test for anticipation. Standard errors clustered at village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

G.17 Rare-Outcome Functional Form: Conditional Logit

Communal conflict is a rare binary outcome (2.5 percent pooled), so a linear probability model could in principle misbehave near the boundary. We re-estimate the absorbing design with a conditional (fixed-effects) logit, which conditions out the village effects without the incidental parameters problem and uses only villages whose outcome varies over time. The effect holds in the odds metric: first exposure reduces the odds of reported communal conflict by about a quarter without controls (odds ratio 0.75, $\hat{\beta} = -0.29$, cluster-robust $z = -3.14$, $N = 38{,}871$ observations in 5,553 switching villages) and by about a third with the standard lagged controls (odds ratio 0.62, $z = -4.5$). The definition-consistent *warga* outcome is a null without controls (odds ratio 0.93, $z = -0.6$) and negative with controls (0.74, $z = -2.2$), so the *warga* margin again carries its signal only conditionally, which we disclose. A fixed-effects Poisson model, an alternative rare-outcome specification, yields an incidence-rate ratio of 0.79 ($z = -2.9$). The LPM point estimates are therefore not an artifact of the linear functional form; the nonlinear models agree on sign.

G.18 Robustness to the Control Set: Bad-Control and Baseline-Covariate Checks

Table 29 accompanies Section 6.4 and re-estimates the headline ATT under alternative covariate sets, confirming that the result does not depend on the potentially endogenous infrastructure controls.

Table 29: Robustness to the Control Set: Bad-Control and Baseline-Covariate Checks
Outcome: any communal conflict (and warga*); nb2005, window-aligned absorbing first exposure*

	Composite (*conflict_any*)			*Warga*
Control set	CS(2021) ATT	(SE)	Pre_avg (p)	CS(2021) ATT
No covariates (anchor)	-0.0096^{**}	(0.0037)	+0.007 (0.04)	-0.0035
Headline lagged controls	-0.0108^{**}	(0.0055)	-0.006 (0.40)	-0.0028
Drop paved-road access	-0.0107^{**}	(0.0055)	-0.005 (0.47)	-0.0024
Geographic only (distance)	-0.0127^{***}	(0.0046)	-0.000 (0.99)	-0.0048
Baseline (2005) levels, fixed	-0.0101^{**}	(0.0039)	+0.010 (0.01)	-0.0035
Baseline (2005) levels × wave†	-0.0110^{***}	(0.0026)	joint lead $p = 0.09$	-0.0043

Each row re-estimates the primary CS(2021) specification (doubly-robust weighting, not-yet-treated controls) with a different covariate set. "No covariates (anchor)" is the pure benchmark: because the shock is physically exogenous, the covariates are not required for identification, and the ATT is negative and significant without them (-0.0096). It does, however, carry a marginally significant *positive* pre-trend (+0.007, $p = 0.04$), as do the specifications that fix covariates at their 2005 levels; the specifications that include the lagged time-varying controls (headline and drop-paved-road) are the ones with a clean pre-trend (-0.006, $p = 0.40$). Because a positive pre-trend biases a subsequent decline *toward* zero, the covariates that remove it make the estimate slightly larger rather than smaller. The ATT is stable across every control set, so the covariates neither create the effect nor reverse its sign; where a residual pre-trend remains it is positive and works against the finding. "Drop paved-road access" removes lagged paved-road access, which Section 6.4 shows is itself moved by the shock. "Geographic only" keeps only lagged distance to the district capital. Distance is administratively time-varying (village-to-capital distance shifts with *pemekaran* district splits) and, in a covariate event-study, both pre-trends and responds to the shock, so it is not a clean exogenous control; we report this single-control row only to show the effect survives even a minimal specification, not as a bias-free benchmark. This row previously also held lagged population and household count; those two series are not collected by PODES from 2014 onward and are now honestly missing rather than imputed as zero, so the check is necessarily narrower than in earlier drafts. "Baseline (2005) levels, fixed" replaces every lagged control with its time-invariant value in the first (2005) wave, which is pre-treatment for every estimable cohort. Pre_avg is the aggregate pre-treatment event-study coefficient (a conditional parallel-trends test). The *warga* column reports the ATT on the definition-consistent civilian-vs-civilian outcome; its full pre-trends and standard errors mirror the composite. †The final row uses the absorbing TWFE estimator with the strictly pre-determined 2005 covariate levels interacted with wave indicators and no post-baseline time-varying controls, so villages with different initial characteristics may follow different flexible trends; the pre-trend statistic is the joint F-test on the pre-period leads (t_{-2}, $p = 0.17$; t_{-3}, $p = 0.36$), which does not reject at the 5 percent level ($p = 0.09$). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

H Communication Infrastructure: Pre-Trend-Failed Outcomes

Two outcomes from the communication-infrastructure battery in Section 7.3 are reported here rather than in the main text because their pre-trends fail the CS(2021) diagnostic. The absence of any mobile signal is essentially unchanged after exposure, while any street lighting rises by 6.1 percentage points, but in both cases the aggregate pre-treatment event-study mean is itself large and significant, so treated and comparison villages were already diverging before first exposure. These outcomes move in directions consistent with greater remoteness or reconstruction, but their pre-trends caution against causal interpretation. The positive street-lighting coefficient is most plausibly attributed to post-disaster reconstruction and expanded state presence in exposed villages, rather than to any earthquake effect on communication capacity; because its pre-trend fails ($p < 0.001$), we do not interpret it as a mechanism estimate.

Table 30: Communication-Infrastructure Outcomes with Failed Pre-Trends
Baseline: nb2005, window-aligned absorbing first exposure, no controls

Outcome	Treated pre-mean	CS(2021) ATT	(SE)	p	Pre_avg (p)
No mobile signal	0.108	−0.0038	(0.0059)	0.520	+0.023 (0.000)
Any street lighting	0.616	+0.0612***	(0.0084)	0.000	+0.031 (0.000)

No mobile signal is the indicator that a village has no mobile-phone reception; any street lighting is built from PODES items r502a and r502b, whose 2018 redesign from a two-category to a three-category item produces a level break absorbed by the wave fixed effects. CS(2021): doubly-robust weighting, not-yet-treated controls, no covariates, on the nb2005 windowed sample of N = 274,288 village-waves. Both outcomes show significant pre-trends (Pre_avg $p < 0.001$), so the ATT estimates cannot be interpreted causally. Pre_avg is the aggregate pre-treatment event-study mean with its p-value in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

I Social Capital: Pre-Trend-Failed Outcomes

Three further outcomes from the social-capital battery in Section 7.3 are reported here rather than in the main text because their pre-trends fail the CS(2021) diagnostic. The presence of any active sports-activity group falls after exposure, its count moves in the opposite direction, and the inverse-hyperbolic-sine of total village community institutions declines in the 2018–2024 panel, but in each case the aggregate pre-treatment event-study mean is itself significant, so treated and comparison villages were already diverging before first exposure. These outcomes move in directions consistent with either organizational disruption or measurement breaks across the 2018 questionnaire redesign, but their pre-trends caution against causal interpretation.

Table 31: Social-Capital Outcomes with Failed Pre-Trends
Baseline: nb2005, window-aligned absorbing first exposure, no controls

Outcome	Treated pre-mean	CS(2021) ATT	(SE)	p	Pre_avg (p)
Sports-activity group (any)	0.886	-0.0132^{**}	(0.0059)	0.025	+0.0062 (0.139)
Sports groups (count)	3.08	+0.0320	(0.0402)	0.427	+0.1141 (0.000)
Total village institutions (ihs, 2018–2024)	—	−0.0249	(0.0235)	0.289	+0.1221 (0.042)

Sports-activity group (any) is an indicator that the village has at least one active sports group; sports groups (count) is the number of such groups; total village institutions is the inverse-hyperbolic-sine of the count of village community institutions (PKK, Karang Taruna, adat, farmer, water-user, and community groups). The sports-group item count changes across waves, from 8 items in 2005 and 2008 to 10 items from 2011 onward, which introduces a level break absorbed by the wave fixed effects; total village institutions is available for the 2018–2024 waves only. CS(2021): doubly-robust weighting, not-yet-treated controls, no covariates, on the nb2005 windowed sample of $N = 274{,}288$ village-waves. The sports-count and total-institutions outcomes show significant pre-trends (Pre_avg column) and cannot be interpreted causally. The sports-activity group indicator is the exception: its post-treatment decline (−1.3 points) sits against an aggregate pre-trend that does not reject ($p = 0.14$), but we flag it as suggestive rather than dispositive given the 2018 questionnaire item-count redesign and do not build the argument on it. Pre_avg is the aggregate pre-treatment event-study mean with its p-value in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

J Heterogeneity: Warga Salvage and Group Structure

This appendix reports two heterogeneity exercises discussed in Section 7.4. Table 32 re-estimates the two surviving composite moderators, police monitoring and religious polarization, on the definition-stable *warga* outcome alongside the composite outcome: the *warga* response concentrates where religious polarization is high but not where monitoring is high, so the imprecise pooled *warga* effect reflects heterogeneity rather than absence of an effect. Table 33 interacts first exposure with ethnic and religious polarization *and* fractionalization, separately and in a religious horse-race. Religious group structure moderates the effect while ethnic group structure does not, but for religion the polarization and fractionalization indices are nearly collinear and the horse-race cannot separate them, so the moderation is read as a group-structure channel rather than as polarization specifically.

Table 32: Warga Salvage: Composite versus Definition-Stable Outcome by Moderator (nb2005 Baseline)

	Police (village cluster)		Relig. polar. (kab. cluster)	
	(1)	(2)	(3)	(4)
	Composite	Warga	Composite	Warga
D_{treat}	-0.0305***	-0.0133***	-0.0248***	-0.0107***
	(0.0045)	(0.0035)	(0.0065)	(0.0038)
D_{treat} × Police (std.)	-0.0075***	-0.0034*		
	(0.0025)	(0.0020)		
D_{treat} × Relig. polar. (std.)			-0.0175***	-0.0076**
			(0.0065)	(0.0037)
Observations	194,360	194,360	200,375	200,375
Adj. R^2	0.081	0.065	0.070	0.055

Standard errors in parentheses, clustered at village level (police) or kabupaten level (religious polarization). Absorbing windowed first-exposure treatment; village and wave fixed effects and the lagged-control set throughout. Moderators standardized (mean zero, unit standard deviation).

* p<0.1, ** p<0.05, *** p<0.01

Table 33: Polarization versus Fractionalization, Ethnic and Religious (nb2005 Baseline)

	(1) Eth. P	(2) Eth. F	(3) Relig. P	(4) Relig. F	(5) Relig. race
D_{treat}	-0.0285***	-0.0272***	-0.0248***	-0.0247***	-0.0258***
	(0.0085)	(0.0076)	(0.0065)	(0.0065)	(0.0067)
× Ethnic polarization (std.)	-0.0092				
	(0.0083)				
× Ethnic fractionalization (std.)		-0.0150**			
		(0.0066)			
× Religious polarization (std.)			-0.0175***		-0.1617*
			(0.0065)		(0.0924)
× Religious fractionalization (std.)				-0.0165***	0.1421
				(0.0063)	(0.0897)
Observations	198,779	198,779	200,375	200,375	200,375
Kabupaten clusters	235	235	253	253	253
Adj. R^2	0.067	0.067	0.070	0.070	0.070

Kabupaten-clustered standard errors in parentheses. Absorbing windowed first-exposure treatment; village and wave fixed effects and the lagged-control set throughout. Moderators standardized (mean zero, unit standard deviation). Column 5 is a religious horse-race with polarization and fractionalization jointly.
* p<0.1, ** p<0.05, *** p<0.01

K Covariate Trajectories: CS(2021) on Observables

Table 34: Covariate Event-Study Aggregates: Pre- and Post-Treatment Averages
Baseline: nb2005, window-aligned first exposure, no controls

Covariate (as outcome)	Pre_avg	p	Post_avg	p
ln population		not estimable (see note)		
ln households		not estimable (see note)		
Any household electricity	−0.008	0.045	+0.019	0.000
Paved road share	−0.011	0.268	−0.046	0.000
Internet available	−0.028	0.012	−0.011	0.124
Police post	−0.004	0.587	−0.012	0.030
Health centers (count)	+0.002	0.858	−0.043	0.000
Primary schools (count)	−0.198	0.000	−0.027	0.216

Each row is a separate CS(2021) estimation (doubly-robust weighting, not-yet-treated controls) with the covariate as the outcome; Pre_avg and Post_avg are the aggregate pre- and post-treatment event-study means with analytical SEs. Pre-treatment coefficients follow the standard event-study convention, cumulative deviations from the omitted wave before first exposure. Population and household counts are not recorded consistently in the later PODES waves, so their event studies do not estimate and no figure is reported for them; they are in any case not members of the control vector. Interpretation in Section 6.4: the primary conflict estimator conditions on the lagged covariates via doubly-robust weighting; road *passability* (Section 7) is unchanged while the paved-road *share* falls after treatment, evidence of degraded surface quality rather than a wholesale loss of physical access.

L PGA-Derived MMI: An Independent Exposure Measure

ShakeMap's Modified Mercalli Intensity blends instrumental ground motion with macroseismic (felt-report) data, so a natural question is whether our result depends on that construction. As an independent check we rebuild exposure from ShakeMap Peak Ground Acceleration (PGA) alone, which is purely instrumental, and convert it to MMI through the Wald et al. (1999) relationship, MMI $= 3.66 \log_{10}(\text{PGA}) - 1.66$, cross-validated against Worden et al. (2012).

Building this measure required correcting a unit error in the PGA raster-to-village pipeline. ShakeMap distributes some PGA grids in $\ln(g)$ and others linearly in $\%g$; the earlier pipeline exponentiated the log grids correctly but then treated the resulting values (in g) as $\%g$, understating PGA on those grids by a factor of one hundred, a constant $3.66 \log_{10}(100) \approx 7.3$ MMI-point deficit that rendered the derived intensities uninformative. After the fix, the PGA-derived MMI agrees closely with ShakeMap MMI, with a median discrepancy below one-tenth of an intensity unit for shaken villages, and the two exposure indicators correlate 0.70 at the village-wave level.

Re-estimating the main design on this independent measure, holding the window-aligned first-exposure timing and all else fixed, reproduces the result across every specification (Table 35): the primary CS(2021) estimate is -0.0113 ($p = 0.06$) against the ShakeMap-MMI -0.0108, and the definition-stable *warga* outcome is likewise negative and significant. Both columns of that table are computed in a single run on the identical panel, so the comparison is exact rather than assembled across specifications. The communal decline is therefore not an artifact of ShakeMap's intensity construction.

Table 35: Measurement Robustness: PGA-Derived MMI
An intensity measure built from instrumental PGA alone reproduces the result across every specification.

	ShakeMap MMI (primary)	PGA-derived MMI (Wald 1999)
CS(2021) ATT, composite	-0.0150^{***}	-0.0129^{***}
	(0.0043)	(0.0039)
Absorbing TWFE, no controls	-0.0186^{***}	-0.0128^{***}
	(0.0032)	(0.0030)
Absorbing TWFE, with controls	-0.0202^{***}	-0.0142^{***}
	(0.0033)	(0.0030)
Definition-stable *warga*, TWFE controls	-0.0071^{***}	-0.0044^{*}
	(0.0026)	(0.0025)
Ever-treated villages	3,207	4,047

Notes: The right column replaces ShakeMap Modified Mercalli Intensity with an MMI derived independently from ShakeMap Peak Ground Acceleration through the Wald et al. (1999) relationship, $\text{MMI} = 3.66 \log_{10}(\text{PGA}) - 1.66$, keeping everything else fixed: the same window-aligned first-exposure timing, the same village and wave fixed effects, and village-clustered standard errors. First exposure is the first survey wave enumerated after a village's first PGA-derived MMI $\geq$ VI event, exactly as for the ShakeMap-MMI treatment. The PGA and ShakeMap-MMI treatments identify nearly the same villages (4,047 vs 3,207 ever-treated; the two exposure indicators correlate 0.79 at the village-wave level), and every estimate reproduces the main negative effect at comparable or slightly larger magnitude, including the definition-stable *warga* outcome. The PGA-derived CS(2021) estimate is significant ($p = 0.001$) against $p = 0.000$ for ShakeMap MMI, reflecting the wider standard error of the noisier instrumental measure rather than a smaller point estimate. This establishes that the result is not an artifact of ShakeMap's MMI construction, which blends instrumental and macroseismic intensity, as opposed to the purely instrumental PGA. An earlier calendar-year-binned PGA measure produced a spurious positive coefficient, an instance of the same survey-window timing artifact documented in Section 5; it disappears once PGA exposure is aligned to the survey calendar. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

M SNPK/NVMS Detail

The motivation for this exercise is coverage: ACLED's Indonesia coverage begins only in 2015 and so cannot speak to the earthquakes that identify most main-specification cohorts, including the 2009 Padang earthquake driving the largest treated cohort. SNPK/NVMS, a newspaper-based violence database independent of PODES (Barron et al., 2016), covers 2005–2014, the window that includes the Padang cohort in principle, though West Sumatra itself falls outside the 18 provinces with continuous SNPK coverage over this period and so cannot be validated by this exercise. Kabupaten-level SNPK analysis follows Bazzi and Gudgeon (2021), who use the same database at the same geographic level. Aggregating to kabupaten (rather than village) dilutes true village-level treatment variation, which should bias estimated effects toward zero; results are, if anything, a lower bound. The evidence is mixed and we report it as such: SNPK's own broad "any social conflict" measure is null, its narrower "Konflik Identitas" subcategory (the closer analogue to PODES communal violence) falls with a clean pre-trend, and a third subcategory relevant to the coordination mechanism fails its own pre-trend test and is not interpretable.

Kabupaten codes are harmonized using a single fixed (2014-vintage) BPS crosswalk. SNPK's own kabupaten codes are recorded in one fixed later vintage across all incident-years rather than the code administratively valid at the time of each incident: a post-split code already appears in the 2005 incident file, years before that split existed administratively. Matching against the 2014-vintage crosswalk achieves a 99.9 percent match rate versus 97.0 percent under a naive year-specific match; we use the corrected, fixed-vintage crosswalk throughout. Sample: 2005–2014, restricted to the 18 provinces with continuous SNPK coverage over this period (SNPK expanded to national coverage in 2012; including that expansion would confound coverage growth with treatment). Treatment: kabupaten-level absorbing first exposure to MMI$\geq$VI from event dates; 37 always-treated kabupaten (first exposure before 2005) are excluded, leaving 90 ever-treated kabupaten. Kecamatan-level harmonization, which would match the main design's geography, was not feasible: the source MFD kecamatan crosswalk ships only as an archive format not extractable in this environment.

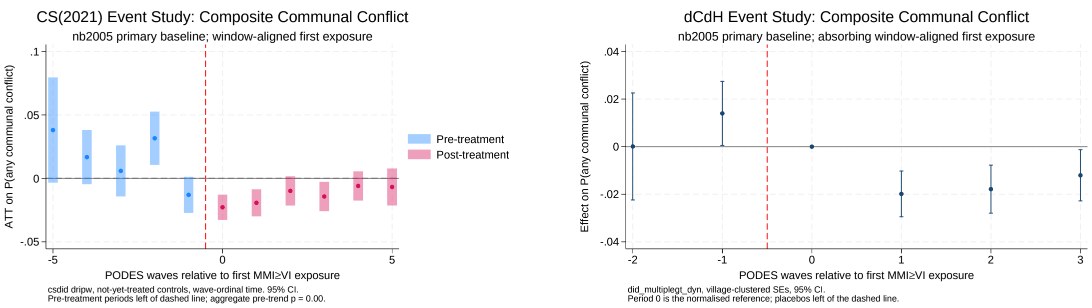
Table 36: SNPK/NVMS: Kabupaten-Level Absorbing First Exposure, 2005–2014

	(1) Any social conflict (Bazzi–Gudgeon defn.)	(2) Konflik Identitas (communal analogue)	(3) Konflik Main Hakim Sendiri (vigilante / mob justice)
Panel A: TWFE (kabupaten + year FE)			
D_{treat}	−0.0063	−0.0818***	−0.0050
	(0.0199)	(0.0239)	(0.0345)
Panel B: CS(2021), not-yet-treated control			
Overall ATT	−0.0056	−0.0930***	0.0039
	(0.0292)	(0.0310)	(0.0229)
Pre-trend Pre_avg	−0.0472*	0.0448	0.0836***
	($p = 0.087$)	($p = 0.218$)	($p = 0.000$)
Post_avg	−0.0233	−0.1071***	−0.0164
	(0.0278)	(0.0321)	(0.0217)
Baseline mean	0.710	0.289	0.597
Ever-treated kabupaten		77	
Kabupaten × year cells		2,530	

Notes: Outcome: any incident of the given type in kabupaten-year, from SNPK/NVMS (Sistem Nasional Pemantauan Kekerasan), a newspaper-based violence database independent of PODES self-reports (Barron et al., 2016). Column (1) reproduces the literature-standard outcome definition of Bazzi and Gudgeon (2021) (the social-conflict event type: any social conflict, excluding domestic violence, ordinary crime, and law-enforcement violence); it pools earthquake-irrelevant categories (land disputes, election disputes, government-tender conflicts) with communal violence, diluting any signal in a high-base-rate, heterogeneous indicator. Column (2) isolates (identity/inter-group conflict), SNPK's narrower analogue to PODES communal violence and the theoretically relevant comparison for this paper. Sample: 2005–2014, restricted to the 18 provinces with constant SNPK coverage over this period (SNPK expanded to national coverage in 2012; mixing that expansion into the panel would confound coverage growth with treatment). Treatment: kabupaten-level (2003-vintage codes) absorbing first exposure to MMI≥VI from event dates; always-treated kabupaten (first exposure before 2005) excluded. Kabupaten codes are harmonized to 2003-vintage identifiers using a single fixed (2014-vintage) BPS crosswalk (see the crosswalk validation note in the replication archive for the match-rate comparison against a year-specific crosswalk). Kecamatan-level harmonization was not available (source MFD crosswalk requires archive extraction unavailable in this environment), so this analysis is coarser than the village-level design. West Sumatra (Padang, the 2009 episode driving the largest treated cohort in the main analysis) is outside SNPK's pre-2012 coverage and cannot be validated by this exercise. SE clustered at kabupaten level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Column (1) reproduces Bazzi and Gudgeon's (2021) own outcome definition, "any social conflict," pooling every SNPK conflict subtype except domestic violence, crime, and law-enforcement violence, as the literature-standard, non-cherry-picked starting point. It is null (TWFE -0.012, $t = -0.52$) with a marginally significant pre-trend ($p = 0.069$); at a 70 percent base rate, this aggregate mixes land disputes, election disputes, and government-tender conflicts with communal violence, and there is no reason an earthquake should move most of those categories. Column (2), "Konflik Identitas" (identity/inter-group conflict), SNPK's narrower, more theoretically relevant analogue to PODES communal violence, falls by 8.2 percentage points ($p < 0.01$) on a 27 percent base with a clean pre-trend (Pre_avg $p = 0.218$). Column (3), "Konflik Main Hakim Sendiri" (vigilante/mob justice), the outcome that would speak most directly to the coordination-disruption mechanism, has a significant pre-trend (Pre_avg $p = 0.001$) and is not usable as mechanism evidence in either direction. We attribute the noisier columns (1) and (3), relative to column (2), to the coarser kabupaten-level aggregation, and do not claim more than column (2) supports.

N Event Study Figures



(a) CS(2021), nb2005 **(b)** dCdH, nb2005

Figure 8: Event Study: nb2005 Primary Baseline (Average Lead Insignificant; Joint Test Rejects for the Composite)

O HonestDiD Sensitivity Bounds

Table 37 reports the Rambachan and Roth (2023) HonestDiD sensitivity analysis applied to the Sun–Abraham interaction-weighted event study, which has the cleanest lead structure of our estimators (only the t_{-2} lead is even marginal, $p = 0.08$). The target is the immediate post-treatment effect t_0 (-0.0174^{***} for the composite outcome, -0.0072^{**} for *warga*). The relative-magnitudes restriction $\Delta^{RM}(\bar{M})$ allows post-treatment violations of parallel trends up to $\bar{M}$ times the largest pre-treatment violation; the breakdown value is the largest $\bar{M}$ at which the robust 95 percent identified set still excludes zero.

For the composite outcome the identified set excludes zero up to $\bar{M} \approx 0.5$: the sign of the effect survives post-treatment violations as large as half the largest observed pre-treatment deviation, and loses significance only once the allowed violation approaches the full size of that deviation. The definition-consistent *warga* target is more fragile, its identified set already including zero by $\bar{M} = 0.5$, consistent with its weaker precision. These are honest, moderate bounds rather than the near-unbounded robustness a perfectly clean pre-period would deliver; we report them as the quantitative statement of how much residual violation the result can absorb, alongside the not-yet-treated-only estimate. The doubly-robust property of the estimator is a separate safeguard, against misspecification of the nuisance models rather than against failure of conditional parallel trends itself.

The same analysis applied to the *primary* Callaway–Sant'Anna estimator is less forgiving. Under exact parallel trends the identified set for the aggregated composite ATT is $[-0.020,\ +0.001]$, simply the headline confidence interval; once a relative-magnitudes violation of $\bar{M} = 0.5$ is allowed it widens to $[-0.142,\ +0.130]$, and the smoothness restriction Δ^{SD}, which bounds how far the post-treatment trend may depart from a linear extrapolation of the pre-trend, already includes zero at its mildest deviation ($[-0.012,\ +0.014]$). The breakdown value on the primary estimator is therefore effectively $\bar{M} \approx 0$, and reporting both restriction families follows the current difference-in-differences synthesis (Roth et al., 2023). The gap from the Sun–Abraham bounds is mechanical, not evidential, and in two distinct senses. First, the two bound different target parameters on different comparison groups: the Sun–Abraham figure is an immediate effect estimated without covariates on never-treated controls, while the primary estimator conditions on covariates and uses not-yet-treated controls. Second, $\bar{M}$ is scaled by each estimator's own largest pre-treatment deviation, so an estimator with smoother leads mechanically reports a larger $\bar{M}$ on identical data. Converting both to absolute permitted deviation removes the illusion: the primary estimator tolerates up to about 0.009 and the Sun–Abraham estimate about 0.009 to 0.017, so in outcome units the two are close to tied despite reported breakdown values that differ by a factor of two. $\bar{M}$ is thus not a common yardstick across estimators, and we do not treat the Sun–Abraham figure as evidence that its identifying assumption is stronger.

A further bound is worth reporting because it is the one place the sensitivity analysis does real work rather than restating an insignificant target. The effect one wave after exposure, -0.0167 ($p = 0.002$), is the only event-study coefficient whose unrestricted interval excludes zero and the only one that survives a simultaneous band across the whole path. Applying the same analysis to it gives a breakdown between $\bar{M} = 0$ and $\bar{M} = 0.25$ under relative magnitudes, and no margin at all under smoothness. The reason is arithmetic rather than estimator choice: the largest deviation between consecutive pre-treatment leads is 0.037, about 2.2 times the effect being defended, so even a small multiple of that yardstick swamps the estimate. We report this rather than confine the sensitivity discussion to targets where it could not have bitten.

Table 37: HonestDiD Relative-Magnitudes Identified Sets (Sun–Abraham t_0)

	Composite ($t_0 = -0.0174$)		*Warga* ($t_0 = -0.0072$)	
$\bar{M}$	lower	upper	lower	upper
0.0 (original)	−0.025	−0.009	−0.014	−0.001
0.5	−0.035	−0.001	−0.020	+0.005
1.0	−0.047	+0.011	−0.028	+0.013
1.5	−0.059	+0.024	−0.037	+0.022

Notes: 95 percent robust identified sets under $\Delta^{RM}(\bar{M})$ (C-LF method, $\alpha = 0.05$), computed on the window-aligned Sun–Abraham event study (2 pre-periods, 4 post-periods) for the immediate effect t_0. The composite identified set excludes zero through $\bar{M} \approx 0.5$; the *warga* set excludes zero only at $\bar{M} = 0$ and already includes it by $\bar{M} = 0.5$. Source: Rambachan and Roth (2023).

References

Abadie, Alberto, "Statistical Nonsignificance in Empirical Economics," *American Economic Review: Insights*, 2020, *2* (2), 193–208.

_ , **Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge**, "When Should You Adjust Standard Errors for Clustering?," *Quarterly Journal of Economics*, 2023, *138* (1), 1–35.

Aldrich, Daniel P., *Building Resilience: Social Capital in Post-Disaster Recovery*, University of Chicago Press, 2012.

Amarasinghe, Ashani, Paul A. Raschky, Yves Zenou, and Junjie Zhou, "How Natural Disasters Spread Conflict," *European Economic Review*, 2026, *181*, 105194.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press, 2009.

Bai, Ying, "Shaking Legitimacy: The Impact of Earthquakes on Conflict in Historical China," *The Economic Journal*, 2023, *133* (652), 1286–1317.

Bai, Yu and Yanjun Li, "More Suffering, More Involvement? The Causal Effects of Seismic Disasters on Social Capital," *World Development*, 2021, *138*, 105221.

Barron, Patrick, Kai Kaiser, and Menno Pradhan, "Understanding Variations in Local Conflict: Evidence and Implications from Indonesia," *World Development*, 2009, *37* (3), 698–713.

_ , **Sana Jaffrey, and Ashutosh Varshney**, "When Large Conflicts Subside: The Ebbs and Flows of Violence in Post-Suharto Indonesia," *Journal of East Asian Studies*, 2016, *16* (2), 191–217.

Bazzi, Samuel and Christopher Blattman, "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 1–38.

_ **and Matthew Gudgeon**, "The Political Boundaries of Ethnic Divisions," *American Economic Journal: Applied Economics*, 2021, *13* (1), 235–266.

_ , **Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong**, "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, 2019, *109* (11), 3978–4025.

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Belloc, Marianna, Francesco Drago, and Roberto Galbiati, "Earthquakes, Religion, and Transition to Self-Government in Italian Cities," *Quarterly Journal of Economics*, 2016, *131* (4), 1875–1926.

Bellows, John and Edward Miguel, "War and Local Collective Action in Sierra Leone," *Journal of Public Economics*, 2009, *93* (11–12), 1144–1157.

Benjamini, Yoav and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1995, *57* (1), 289–300.

Berman, Eli, Jacob N. Shapiro, and Joseph H. Felter, "Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq," *Journal of Political Economy*, 2011, *119* (4), 766–819.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig, "This Mine Is Mine! How Minerals Fuel Conflicts in Africa," *American Economic Review*, 2017, *107* (6), 1564–1610.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Besley, Timothy and Torsten Persson, "The Logic of Political Violence," *Quarterly Journal of Economics*, 2011, *126* (3), 1411–1445.

Blakeslee, David S. and Ram Fishman, "Weather Shocks, Agriculture, and Crime: Evidence from India," *Journal of Human Resources*, 2018, *53* (3), 750–782.

Blattman, Christopher and Edward Miguel, "Civil War," *Journal of Economic Literature*, 2010, *48* (1), 3–57.

Bó, Ernesto Dal and Pedro Dal Bó, "Workers, Warriors, and Criminals: Social Conflict in General Equilibrium," *Journal of the European Economic Association*, 2011, *9* (4), 646–677.

Borusyak, Kirill and Peter Hull, "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

_ , Xavier Jaravel, and Jann Spiess, "Revisiting Event-Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Brancati, Dawn, "Political Aftershocks: The Impact of Earthquakes on Intrastate Conflict," *Journal of Conflict Resolution*, 2007, *51* (5), 715–743.

Buonanno, Paolo, Giacomo Plevani, and Marcello Puca, "Earthquake hazard and civic capital," *European Journal of Political Economy*, 2023, *78*, 102367.

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2), 238–249.

Cassar, Alessandra, Andrew Healy, and Carl von Kessler, "Trust, Risk, and Time Preferences After a Natural Disaster: Experimental Evidence from Thailand," *World Development*, 2017, *94*, 90–105.

Cassidy, Traviss, "Revenue Persistence and Public Service Delivery," *Economic Journal*, 2025, *135* (670), 2017–2053.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, "Catastrophic Natural Disasters and Economic Growth," *Review of Economics and Statistics*, 2013, *95* (5), 1549–1561.

Conley, Timothy G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 1999, *92* (1), 1–45.

Crost, Benjamin, Joseph Felter, and Patrick Johnston, "Aid Under Fire: Development Projects and Civil Conflict," *American Economic Review*, 2014, *104* (6), 1833–1856.

de Chaisemartin, Clement and Xavier D'Haultfoeuille, "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *Review of Economics and Statistics*, 2024. forthcoming.

DiPasquale, Denise and Edward L. Glaeser, "The Los Angeles Riot and the Economics of Urban Unrest," *Journal of Urban Economics*, 1998, *43* (1), 52–78.

Draca, Mirko and Stephen Machin, "Crime and Economic Incentives," *Annual Review of Economics*, 2015, *7*, 389–408.

Dube, Oeindrila and Juan F. Vargas, "Commodity Price Shocks and Civil Conflict: Evidence from Colombia," *Review of Economic Studies*, 2013, *80* (4), 1384–1421.

Esteban, Joan, Laura Mayoral, and Debraj Ray, "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, 2012, *102* (4), 1310–1342.

Fetzer, Thiemo, "Can Workfare Programs Moderate Conflict? Evidence from India," *Journal of the European Economic Association*, 2020, *18* (6), 3337–3375.

Gignoux, Jeremie and Marta Menendez, "Benefit in the Wake of Disaster: Long-Run Effects of Earthquakes on Welfare in Rural Indonesia," *Journal of Development Economics*, 2016, *118*, 26–44.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Horowitz, Donald L., *The Deadly Ethnic Riot*, Berkeley: University of California Press, 2001.

Imbens, Guido W., "Statistical Significance, p-Values, and the Reporting of Uncertainty," *Journal of Economic Perspectives*, 2021, *35* (3), 157–174.

_ **and Jeffrey M. Wooldridge**, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 2009, *47* (1), 5–86.

Irsyam, Masyhur, Phil R. Cummins, M. Asrurifak, Lutfi Faizal, Danny Hilman Natawidjaja, Sri Widiyantoro, Irwan Meilano, Wahyu Triyoso, Ariska Rudiyanto, Sri Hidayati, M. Ridwan, Nuraini Rahma Hanifa, and Arifan Jaya Syahbana, "Development of the 2017 National Seismic Hazard Maps of Indonesia," *Earthquake Spectra*, 2020, *36* (S1), 112–136.

Jones, Damon, David Molitor, and Julian Reif, "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study," *Quarterly Journal of Economics*, 2019, *134* (4), 1747–1791.

Juan, Alexander De, Jan Henryk Pierskalla, and Elisa Schwarz, "Natural Disasters, Aid Distribution, and Social Conflict: Micro-Level Evidence from the 2015 Earthquake in Nepal," *World Development*, 2020, *126*, 104715.

Kirchberger, Martina, "Natural Disasters and Labor Markets," *Journal of Development Economics*, 2017, *125*, 40–58.

Koubi, Vally, "Climate Change and Conflict," *Annual Review of Political Science*, 2019, *22*, 343–360.

Lowe, Matt, "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 2021, *111* (6), 1807–1844.

McGuirk, Eoin and Marshall Burke, "The Economic Origins of Conflict in Africa," *Journal of Political Economy*, 2020, *128* (10), 3940–3997.

Mehlum, Halvor, Edward Miguel, and Ragnar Torvik, "Poverty and Crime in 19th Century Germany," *Journal of Urban Economics*, 2006, *59* (3), 370–388.

Miguel, Edward, "Poverty and Witch Killing," *Review of Economic Studies*, 2005, *72* (4), 1153–1172.

_ , **Shanker Satyanath, and Ernest Sergenti**, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach," *Journal of Political Economy*, 2004, *112* (4), 725–753.

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

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

Nel, Philip and Marjolein Righarts, "Natural Disasters and the Risk of Violent Civil Conflict," *International Studies Quarterly*, 2008, *52* (1), 159–185.

Nunn, Nathan and Nancy Qian, "US Food Aid and Civil Conflict," *American Economic Review*, 2014, *104* (6), 1630–1666.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*, 2019, *37* (2), 187–204.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Rao, Gautam, "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools," *American Economic Review*, 2019, *109* (3), 774–809.

Ray, Debraj and Joan Esteban, "Conflict and Development," *Annual Review of Economics*, 2017, *9*, 263–293.

Romer, David, "In Praise of Confidence Intervals," *AEA Papers and Proceedings*, 2020, *110*, 55–60.

Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb, "Fast and Wild: Bootstrap Inference in Stata Using boottest," *Stata Journal*, 2019, *19* (1), 4–60.

Roth, Jonathan, "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

— **and Pedro H. C. Sant'Anna**, "When Is Parallel Trends Sensitive to Functional Form?," *Econometrica*, 2023, *91* (2), 737–747.

— , — , **Alyssa Bilinski, and John Poe**, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Sant'Anna, Pedro H. C. and Jun Zhao, "Doubly Robust Difference-in-Differences Estimators," *Journal of Econometrics*, 2020, *219* (1), 101–122.

Sarsons, Heather, "Rainfall and Conflict: A Cautionary Tale," *Journal of Development Economics*, 2015, *115*, 62–72.

Satyanath, Shanker, Nico Voigtländer, and Hans-Joachim Voth, "Bowling for Fascism: Social Capital and the Rise of the Nazi Party," *Journal of Political Economy*, 2017, *125* (2), 478–526.

Slettebak, Rune T., "Don't Blame the Weather! Climate-Related Natural Disasters and Civil Conflict," *Journal of Peace Research*, 2012, *49* (1), 163–176.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Wald, David J., Vincent Quitoriano, Thomas H. Heaton, and Hiroo Kanamori, "Relationships between Peak Ground Acceleration, Peak Ground Velocity, and Modified Mercalli Intensity in California," *Earthquake Spectra*, 1999, *15* (3), 557–564.

Worden, C. B., E. M. Thompson, J. W. Baker, B. A. Bradley, N. Luco, and D. J. Wilson, "Spatial and Spectral Interpolation of Ground-Motion Intensity Measure Observations," *Bulletin of the Seismological Society of America*, 2018, *108* (2).

— **, M. C. Gerstenberger, D. A. Rhoades, and D. J. Wald**, "Probabilistic Relationships between Ground-Motion Parameters and Modified Mercalli Intensity in California," *Bulletin of the Seismological Society of America*, 2012, *102* (1), 204–221.

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

Yudhistira, Muhammad Halley, Najwa Khairina, Teguh Dartanto, and Chaikal Nuryakin, "Information Technology and Local Conflict: Evidence from Indonesia," *Conflict, Security & Development*, 2021, *21* (6), 831–848.