Earthquakes and Communal Conflict: Evidence from Indonesia's Seismic Zones

Billy Thandy Tulungen[a,c,*] Muhammad Halley Yudhistira[a,b]

Jahen Fachrul Rezki[a,b]

[a]Department of Economics, Faculty of Economics and Business, Universitas Indonesia, Depok 16424, Indonesia

[b]Institute for Economic and Social Research (LPEM), Faculty of Economics and Business, Universitas Indonesia, Jakarta 10430, Indonesia

[c]Government of North Sulawesi Province, Manado, Indonesia

[*]Corresponding author: Billy Thandy Tulungen, Department of Economics, Faculty of Economics and Business, Universitas Indonesia, Depok 16424, Indonesia. Email: billy.thandy21@ui.ac.id

This version: 23 September 2026.

Abstract

Communal violence destroys assets, deters investment, and erodes the local cooperation on which public goods depend, and it is common in the places large earthquakes strike. Does a strong earthquake raise communal conflict or suppress it? The empirical record disagrees on the sign. We combine seven waves of Indonesia's village census (2005–2024) with USGS ShakeMap ground-shaking intensities for the 43,351 villages in the country's official seismic-hazard zones. A village's first exposure to damaging shaking lowers the probability that it reports a mass brawl by 1.7 percentage points, against a pre-exposure mean of 6.7 percent. Such a decline is equally consistent with less fighting, less reporting or better local order, and no single outcome separates these readings. We therefore use several. The same village head, on the same form and recall window, also reports theft, robbery and assault, and these acts differ in what they demand of offenders. Communal violence falls while theft does not follow it down, and robbery and assault do not follow theft. We defend this divergence between the four acts rather than the size of theft's own rise. The evidence is consistent with a loss of the organisational capacity that collective violence requires and stealth theft does not.

Keywords: earthquakes, communal conflict, property crime, composition of disorder, Indonesia, difference-in-differences

JEL Codes: D74, K42, Q54, O12, O18, R12

Highlights

- First damaging shaking cuts reported mass brawls by 1.7 points on a 6.7% base
- Theft does not fall with it, and robbery and assault do not follow theft
- One reporter and one form: common reporting shifts drop out of the contrast
- The pattern fits a loss of organisational capacity, not a change in motive
- A first-time-conflict margin fails its own pre-exposure falsification test

1 Introduction

When a strong earthquake strikes a village, those who respond must allocate a fixed budget almost immediately. The village head, the district government and the arriving relief agencies have to divide limited resources between rebuilding and maintaining local order. The allocation matters, because communal violence destroys assets, deters investment, and erodes the local cooperation on which public services depend (Barron, Kaiser, and Pradhan, 2009). Indonesia faces this decision more often than most. It is the most seismically active of the world's large economies, and it has a long record of localised communal violence (Barron, Kaiser, and Pradhan, 2009).

Thirty years of research has not settled how that budget should be divided. Does a strong earthquake raise communal conflict, or suppress it? The sign itself is still disputed. Brancati (2007) and Nel and Righarts (2008) find that disasters raise conflict risk. Slettebak (2012) finds the opposite. Newer evidence finds that disasters can lower conflict where they strike while raising it nearby, through spatial displacement (Amarasinghe et al., 2026). The reviews divide the same way: Koubi (2019) finds no robust general effect, while the meta-analysis of Burke, Hsiang, and Miguel (2015), pooling 55 studies, finds that climatic deviations systematically raise conflict risk.

Our own estimate takes a side. In Indonesia's seismic hazard zones, the first time a village is shaken hard enough to crack its masonry, the probability it reports a mass brawl falls by 1.7 percentage points against a 6.7 percent pre-exposure mean. A decline of about a quarter. We defend that number. It is measured on the one census item whose wording is identical in all seven waves, it survives province-by-wave fixed effects almost intact, and its pre-exposure path is flat.

What the number cannot do is say *why*, and that is the difficulty this paper addresses. A fall in reported mass brawls after an earthquake admits three readings, and all three are equally plausible. Fighting may genuinely have declined. The village head, occupied with displaced families, may have recorded the form less carefully. Or local order may have improved generally, under the supervision of arriving responders. One conflict measure cannot separate them, and one measure at a time is exactly how the disaster-conflict literature has been asking the question.

We use several. That is this paper's contribution.

The key is the census form itself. The village head who answers the mass-brawl question also answers, on the same page, about theft, robbery and assault. One reporter. One form. The same twelve-month recall window. And those four acts demand very different things of the people who commit them. Communal violence needs coordination and confrontation together. Robbery needs an encounter but no coordination. Theft needs neither. Assault's encounters are largely among co-resident acquaintances.

That difference becomes an identification device, because the rival explanations predict different things across the four. Whatever moves all four acts alike within a village drops out of a contrast taken inside that village: its fixed characteristics, and any reporting shift common to all four items. Reporting responses specific to a single item do not drop out, and we examine that possibility below. That contrast is the paper's central result. What is left is an assumption, not a guarantee: that untreated *relative* incidence would have moved in parallel. We test it directly.

We build the data from seven waves of Indonesia's village census (PODES, 2005–2024) for the 43,351 villages in the official medium and high seismic-hazard zones, combined with USGS ShakeMap ground-shaking intensities. We date each village's first exposure to Modified Mercalli Intensity (MMI) VI or above, the level at which ordinary masonry buildings begin to crack.

Here is what we find. On a single common sample of 300,018 village-waves, first exposure lowers reported communal conflict by 2.2 percentage points against a 2.6 percent base. Reported theft does not fall with it. The five outcomes are stacked with outcome-interacted fixed effects, so the standard errors respect the correlation of a village's own outcomes. Under the specification our selection rule prefers, communal conflict falls by 1.7 points, and the same model estimated as a Poisson puts that at 28 percent of the base. Equality of the theft and communal responses is rejected on both scales ($p = 0.003$ linear, $p < 0.001$ proportional). Robbery and assault do not follow theft. Taken to the earthquake, the unit at which treatment is actually assigned, the gap between theft and the definition-stable communal series reaches 7.6 points under treated-village weighting and 6.3 points under an equal-event mean ($p = 0.023$). It is positive in 12 of the 19 usable shocks.

It is worth being explicit about which half of that contrast carries the argument. It is the communal decline, not theft's level. Theft does rise sharply under wave effects, by 4.3 percentage points on a 42 percent base. But once province-by-wave effects are imposed, its own coefficient is no longer separately identified: the rise falls to 1.2 points and loses its precision. What the data establish is that the two series part company, and that they part because communal violence falls while acquisitive crime does not. Nothing below requires theft to have risen.

That pattern is what discriminates. Consider what each explanation would imply. If reporting habits alone changed, all four acts would move together. If local order genuinely collapsed, all four would rise. If the motive for predation strengthened, theft and robbery would rise together, since both are acquisitive. What we observe is none of these. What we observe is coordination-intensive violence down, theft not following it down, robbery down by less, and assault unmoved. Whether theft also *rises* depends on the specification, and nothing below turns on it. The pattern is most consistent with a loss of the organisational capacity that collective violence requires and stealth theft does not.

One thing we cannot separate: which part of that capacity gave way, the ability to coordinate or the confrontational contact coordination operates on. We use *capacity* for the pair. A second piece of evidence points the same way. The decline concentrates in villages that already had mobile-phone signal at the 2005 baseline, an interaction worth 2.1 percentage points. Because the signal was there before any earthquake, that moderator requires no parallel trend in itself, and it survives the province-by-wave specification. We do not claim to have measured any channel's share, nor to have identified capacity as a mediator. Our claim is narrower: of the alternatives these data can distinguish, none fits the observed pattern.

The aggregate decline holds, and we defend it, with its one limit stated ourselves. It clears village-clustered and earthquake-clustered inference, so long as shocks are weighted by the villages they treat ($p = 0.046$). It does not clear an equal-event mean across the nineteen shocks. So the estimand is the average exposed village, not the average earthquake.

A further result appears initially compelling but fails a pre-exposure falsification test. Among villages with no conflict recorded in any earlier observed wave, exposure is associated with a 0.76 percentage point rise in first recorded conflict against a 2.2 percent counterfactual. The number survives a long battery of checks. Then we ran the same design entirely inside the period *before* the earthquake, where it should return zero. It returned 1.5 percentage points, and the gap did not widen when the earthquake actually arrived. These data suggest the margin; this design cannot identify it. Section 8 gives the estimate, the battery and the falsification together, and the falsification decides.

So what the composition establishes is not that earthquakes pacify or inflame. It is that one shock pushes different kinds of disorder in different directions, and that the direction it pushes them discriminates among candidate mechanisms.

This paper contributes to the economics of conflict in three senses. *First and foremost*, the composition of disorder is informative where the level alone is not. Because one official reports four acts on one form, a shock's effect on their *relative* movement is measured within the village. The differences between shaken and unshaken places that defeat a single-outcome comparison fall away with it. The disaster-conflict literature has studied one outcome at a time. Here several outcomes answer jointly, and their pattern fits a loss of organisational capacity rather than a reporting shock, a collapse of order, or a rise in the motive for predation.

Bazzi and Blattman (2014) make the separation of onset from continuation central to their study of economic shocks. We extend it to physical shocks. A damaging earthquake changes two things at once: the incentives for disorder, and the technology of collective action. That lets the two margins move in opposite directions. It also explains why the literature has not converged, since an aggregate estimate then depends on the mixture of previously peaceful and conflict-affected places a study happens to observe. Whether this composition accounts for the existing cross-national record remains unknown.

Second, we provide evidence on mechanisms of a kind rarely available here, and one that both Koubi (2019) and Burke, Hsiang, and Miguel (2015) call for. We observe not only *whether* the shock moves conflict, but *which forms of disorder* move, in which directions, and for *whom*. Theft does not fall on the same census forms on which communal conflict falls, robbery falls by less, assault does not move, and high-involvement collective action weakens. The pattern fits degraded capacity for organised collective action. It is hard to reconcile with a uniform reporting shock, or with an opportunity-cost account in which income destruction raises all forms of violence (Becker 1968; Dube and Vargas 2013; within the framework of Blattman and Miguel 2010). We also confront the displacement concern head on: exploiting the geography of each shock, we find no detectable rise in conflict among nearby never-treated villages. That bounds, without eliminating, the worry that the suppression we estimate is merely the spatial relocation Amarasinghe et al. (2026) document (Supplementary Material Section S1.5).

Third, we contribute the measurement design, and it travels. Using the composition of offences to separate deterrence from motive is long-standing in the economics of crime (Draca and Machin, 2015); what is new is applying it to a collective outcome, where the offence that requires coordination and the offence that requires none respond to one shock in opposite directions. Nothing in the design is specific to Indonesia. All it requires is a source that records several offence types from one reporter, on one form, over one recall window. The inputs each offence demands differ, so the predictions differ, and reporting shocks operating at the level of the reporter drop out by construction. No single-outcome estimate can make these distinctions. In our data theft and robbery differ at $p < 0.001$.

One methodological point follows, smaller than the rest but worth recording. In panels with unequally spaced waves, calendar-year exposure lags can misalign with the enumeration month. On an identical sample, dating exposure by calendar year attenuates the two-wave effect from 2.2 to 0.3 percentage points, and produces a significant coefficient one lag later that the aligned specification shows to be zero (Appendix A, Table A.3). We also provide the first village-level quasi-experimental estimate of the earthquake–conflict relationship for Indonesia.

What this design can and cannot establish. Three limits govern the interpretation. Theft's own coefficient is not separately identified under the preferred specification, so the contrast is carried by the communal decline; we claim the divergence, not theft's rise. The dynamic path is imprecise: the joint test over the interior pre-exposure leads does not reject ($p = 0.31$), but the binned far-lead term remains positive and significant, and the jump at exposure reaches only $p = 0.09$. And a first-time-conflict margin appears in these data and then fails its own placebo; Section 8 reports it and nothing rests on it. Section 9 sets out what remains.

The rest of the paper runs as follows. Section 3 sets out the conceptual framework and the predictions it generates. Section 4 describes the village census, the crime module and the ShakeMap exposure measure. Section 5 presents the empirical framework and the identification assumptions. Section 6 reports the level effect, the composition of disorder, a first-time margin the design cannot identify, and the limits of the evidence. Section 10 concludes.

2 Institutional Background

2.1 Communal conflict in Indonesia

Communal conflict in Indonesia is local, recurrent, and mostly small. It is fought between residents of the same or neighbouring villages rather than between an insurgency and a state, and its typical form is the mass brawl: a fight involving groups rather than individuals, lasting hours or days, and resolved or suppressed locally. Episodes of this kind are common enough that the national village census asks every village about them every three years, which is what makes the question in this paper answerable at all. They also carry real economic cost, destroying assets, deterring investment, and eroding the local cooperation on which public services depend (Barron, Kaiser, and Pradhan, 2009).

The large episodes that reached international attention, in Maluku and Poso, ran along religious lines, and that is the dimension along which the polarisation theory of conflict predicts intensity should vary: polarisation rather than fractionalisation drives conflict where the contested prize is public, such as local political control or cultural dominance, rather than private and divisible (Esteban, Mayoral, and Ray, 2012; Ray and Esteban, 2017). Those episodes are not our variation. Ours is the ordinary, recurrent, sub-national kind that the census records.

Tajima (2014) gives the account of where such violence appears. Communal violence erupted, on his reading, where the withdrawal of authoritarian security after 1998 exposed communities that had come to rely on the state rather than build arrangements of their own. What matters for this paper is the implication: the capacity to keep order locally, and the capacity to mobilise a crowd, are not separate things held by separate people.

2.2 The village as an administrative and security unit

The Indonesian village, *desa*, is the lowest administrative tier with its own head, budget, and records. The head is an elected official, the *kepala desa*, who both answers to the district administration and leads the village's response when something happens, including a disaster. This dual role matters twice over in what follows. It is why the census can ask one official about several kinds of disorder on one form, and it is why a decline in what that official reports is not automatically a decline in what occurred.

Order at village level rests substantially on informal arrangements rather than on formal policing. The village night watch, *siskamling*, and its posts, *poskamling*, are the institution through which residents take rotating responsibility for local security, and they have carried that function since the New Order (Barker, 1998). The same standing arrangements that let a village staff a night watch are the ones that let it assemble a crowd. That observation is the hinge of this paper's argument, and it is an institutional fact about Indonesia before it is a hypothesis: the two capacities are not merely correlated, they are the same people, meeting in the same way, under the same schedule.

2.3 Seismic exposure and hazard zoning

Indonesia is the most seismically active of the world's large economies. It sits where several plates converge, so damaging shaking is not a rare national event but a recurrent local one: in any given year some villages somewhere are shaken hard enough to crack masonry, and which villages those are is settled by fault geometry and wave propagation rather than by anything about the villages.

The government maps this exposure. The national geological agency publishes earthquake hazard zones, *Kawasan Rawan Bencana* (KRB), which classify areas by tectonic setting, namely fault and subduction-zone geometry, in the same spirit as Indonesia's national probabilistic seismic-hazard analysis (Irsyam et al., 2020). The classification describes where damaging shaking is *possible*, not where it has occurred, so a village's band reflects its tectonic setting rather than whether a particular earthquake happened to strike it. We use those maps to fix the comparison set: the paper is estimated on the 43,351 villages in the official medium and high hazard zones, so that treated and untreated villages share a common expected exposure band and the comparison is between places that could have been shaken, not between places that could and places that could not.

Within that set, the treatment is a threshold on realised shaking. Modified Mercalli Intensity VI is the level at which ordinary masonry buildings begin to crack (Wald et al., 1999), which is also the level at which the physical disruption the next section builds on begins: buildings unusable, routines interrupted, and the places where people ordinarily gather closed.

3 Conceptual Framework

A strong earthquake could raise conflict. It could also suppress it. What settles the question is which mechanism dominates, and this section sets two candidates against each other.

The inputs collective violence requires. To arrive at a mass brawl, a village needs four things. A grievance or economic stress. An opportunity for the groups to meet. Someone able to gather and direct a crowd. And working against all three, a state and a community that watch (Besley and Persson, 2011). The four work as a package. Organised violence needs grievance, contact and mobilisation capacity present together while monitoring is weak. Remove any one of the capacity inputs and it goes out, however strong the grievance (Blattman and Miguel, 2010; Esteban, Mayoral, and Ray, 2012; Ray and Esteban, 2017).

Consider what the earthquake does to these inputs. The village hall cracks, roads break, the places where people normally gather can no longer be used. Survivors spend their time clearing rubble and looking after their families. Assembling a crowd becomes much harder. The shock may also deepen grievance, through destroyed assets and lost income, though our data cannot show whether it does, and the argument below does not need it to.

The design rests on the following idea. The capacity that gets damaged is not dedicated to fighting. The same standing arrangements that let a village assemble a crowd also let it staff a night watch. Degrade them and two things go at once: the ability to mobilise collective violence, and the ability to deter a thief. In Indonesia these are literally the same institution, the village night watch, *siskamling* and its *poskamling* posts (Barker, 1998). That observable guardianship deters property crime, and does so locally, is established by Di Tella and Schargrodsky (2004).

Tajima (2014) offers an instructive comparison. His shock, the collapse of authoritarian rule, withdrew monitoring from above while leaving village mobilisation intact. Disorder rose. An earthquake withdraws the mobilisation as well, and the same account predicts the opposite.

A second movement runs alongside. A strong earthquake disperses households and damages dwellings, so the chance that an offender meets a resident falls. Both are classical inputs of the routine activity approach (Cohen and Felson, 1979): activity dispersing away from households raises property crime with motive held fixed. Responders and security forces do arrive, and their presence can suppress organised violence (Berman, Shapiro, and Felter, 2011). But they watch crowds and fights, not the empty houses people left behind. So their presence reinforces the first movement without offsetting the second.

The four acts then sort themselves by what each requires, and that sorting is the identification idea behind the mechanism analysis in Section 7. Collective violence needs both coordination and confrontation, so it should fall. Theft needs neither. It is a stealth act, deterred by guardianship rather than prevented by any coordination constraint, and an emptied dwelling is a better target rather than a lost one. Both movements push it up.

Robbery is the complicated one, and that is precisely its value. Lapsed guardianship pushes it up; scarcer encounters push it down. Theory signs only the comparison, that robbery must rise by less than theft, and leaves its level to the data. Assault is unsigned for the same reason with one more added: its encounters are largely among co-resident acquaintances, which dispersal thins far less than it thins offender-stranger encounters. We count a null there as uninformative rather than as a success.

Note what this establishes. The four acts separate without our assuming anything about whether the motive for predation has changed. That is what makes their composition a test of capacity rather than of grievance.

One boundary should be fixed here rather than conceded later. The account has two inputs, the ability to coordinate and the confrontational contact coordination operates on, and a strong shock degrades both at once. We use *capacity* for the pair, because the four outcomes cannot tell them apart. A village whose residents can no longer organise, and a village whose residents simply meet less often, generate exactly the same ranking across the four acts. Nothing below separates them, and the reader should take every later use of the word in that composite sense.

Two competing channels, and why a brief shock leaves a long mark. The first channel is the canonical opportunity-cost one (Becker, 1968; Miguel, Satyanath, and Sergenti, 2004). Destroyed assets imply a positive reduced-form effect, and the net sign of predation depends on wages relative to contestable rents (Dube and Vargas, 2013). But our shock is physical rather than a growth shock, and we find the opposite sign on the collective margin. The composition evidence below tries to explain why.

The competing channel runs the other way. Disrupted contact suppresses violence regardless of grievance. Temporary external monitoring raises the cost of organising it (Berman, Shapiro, and Felter, 2011). And recovery pulls residents' time toward reconstruction. Disasters can even redirect participation toward recovery-oriented coordination (Bai and Li, 2021; Cassar, Healy, and von Kessler, 2017). In our setting, though, the observed PODES proxies do not support a durable social-capital surge (Section 7.2).

A direct objection belongs here rather than in a footnote, because it runs against our account at its centre. Bellows and Miguel (2009) find that Sierra Leonean communities more exposed to wartime violence went on to *higher* local collective action, and work since reports the same direction after violent and natural shocks. If exposure raises the propensity to act collectively, a fall in collective violence is the last thing one should expect.

The two claims are about different objects, and that distinction is what our design turns on. That literature measures the *willingness* to participate, elicited or reported at the individual level, often years after the shock. What an earthquake destroys in the months we observe is the *infrastructure* of assembly: the hall, the road, the routine gathering, the night-watch rota, and the hours that clearing rubble takes from all of them. Willingness can rise while the standing arrangements that turn willingness into a crowd are unavailable, and mass violence needs the arrangements. Note also what the rival predicts for the other three acts. A village more willing to act together guards itself better, so theft should fall. It does not.

The shaking lasts seconds; structural damage at MMI VI lasts years. Exposure changes a village's *state* rather than delivering a pulse, which is where the absorbing first-exposure model comes from, with an episodic design alongside it and a direct test of whether the effect decays (Section 5). Supplementary Material Section S3 develops each channel and its prior evidence.

Five things to test. The framework yields five predictions, and we test all of them below.

First, the effect should be *state-dependent*, with onset and continuation moving in opposite directions. Sustaining violence already under way draws on standing arrangements the shock degrades, while a *first* episode requires less organisation and more of a trigger. The identifying assumption is stated plainly: continuation responds to capacity, not to triggers, and onset is the reverse. Prior evidence separates the two the same way, with price shocks leaving onset unmoved (Bazzi and Blattman, 2014) and Indonesian case evidence documenting both halves (Barron and Sharpe, 2008). We test the margins separately (Section 8) rather than reporting the average they combine into.

Second, and most sharply, the four acts should separate by input rather than by motive (Section 7.1). A rise in predatory motive alone predicts theft and robbery moving together; the robbery margin fits our account and conflicts with that prediction. It cannot yet separate our account from one combining greater motive with physical displacement, which we test below. The comparative static has been observed with the opposite sign on the input (García-Hombrados, 2020; Cromwell et al., 1995). Our contribution is not that earthquakes raise theft, which is not general, but that the sign of the theft response tracks the sign of the community-capacity response (Supplementary Material Section S3).

Third, the capacity inputs should be observably degraded. This is the prediction our data test least well, and we say so plainly. Collective participation is measured directly only from 2018. Formal policing does not move. The community security post trends throughout the panel. Each direct proxy is therefore short, flat or confounded (Section 7.2). What carries this channel is the composition of crime, not the input measures.

Fourth, because communal violence is fought *between* groups, the effect should be larger where pre-treatment polarisation is higher (Supplementary Material Section S4.4). *Fifth*, stronger shaking should have a larger effect, which we test with the same-earthquake band comparison rather than the threshold ladder (Supplementary Material Section S2). At the earthquake level, an unweighted event mean should be attenuated relative to one weighted by exposed villages (Supplementary Material Section S4.6).

4 Data and Measurement

4.1 Conflict Outcome: PODES

Every three years, enumerators from BPS visit every village in Indonesia. Among the hundreds of questions they put to the village head is the one at the centre of this paper: in the past twelve months, has any intergroup conflict or mass brawl occurred here? The answer is yes or no. That is our main outcome, and it comes from the Village Potential Survey, PODES. The panel opens in 2005 rather than earlier because 2005 is the first round whose questionnaire asks specifically about *perkelahian massal*, a mass brawl. PODES itself runs back further, and the 2003 round does carry a conflict item, but it asks about *konflik* in general, a broader and differently bounded concept. Splicing it on would introduce exactly the kind of definitional break we document below for the subtype composite, so we use 2003 only for baseline covariates. What makes it valuable is that it is a census, not a sample. Every village is asked in every wave. Coverage therefore does not depend on whether a journalist happened to visit, or on whether the road was passable. That dependence is the central weakness of event-based conflict data in rural areas.

Its shortcomings are equally clear. The answer is only yes or no, so there is no intensity margin; it is retrospective over twelve months, so there are no exact dates; and it is self-reported, so a village head may under-report events that reflect poorly on the village. Idiosyncratic under-reporting unrelated to the earthquake simply pulls our estimate toward zero. The serious threat is under-reporting correlated with treatment, which we test directly with a four-part reporting-artifact analysis in Section 7 and an external event-data comparison in Supplementary Material Section S5.1.

We use seven waves enumerated each May between 2005 and 2024, beginning in 2005 because that is the first wave whose questionnaire asks specifically about mass communal brawls. The window covers six major seismic events of national significance, from the 2004–05 Sumatra-Andaman sequence to West Java in 2022. The conflict question's wording changes across waves, so we construct a crosswalk to harmonise the definitions, with exact wording and coding for each wave in the replication package.

Two questions, two measures, and why the difference matters. The questionnaire asks twice. First a gate: did any mass brawl occur, yes or no. If yes, the village head is asked to sort the episodes into named types, such as between groups of residents, between villages, between residents and the security forces, between ethnic groups. Two questions, two measures, and the difference is not innocuous.

We take the **gate** as our primary outcome, for a simple reason: its wording is identical in all seven waves, so it is comparable across the panel by construction. The alternative, which we report alongside it, is the union of the named communal subtypes. In 2005 that union coincides with the gate, because the parent item is the bare gate. Then subtypes are added one by one, and the union narrows. The wedge is 0 percent in 2005, then 11.7 in 2008, 17.1 in 2014, 23.7 in 2018 and 33.7 in 2024. So the composite carries a purely definitional trend, from no gap to a gap of one third. Wave effects absorb its level but cannot absorb a component that moves with treatment.

That choice has consequences for identification (Section 6.1, Table 2). Under province-by-wave effects, the gate retains 94 percent of its estimate, the composite 68 percent, and the *warga* subtype only 33. The gate is the measure least dependent on comparisons made across provinces, and the only one whose wording does not move. We lead with it, report the composite and *warga* beside it, and note where the three disagree.

Building a panel from villages that keep multiplying. Over recent decades Indonesia has gone through extensive administrative subdivision. The number of villages grew from roughly 68,000 to 84,000, which makes panel construction difficult for anyone. Our fix is a chain-linkage crosswalk that maps every village in each wave back to the boundary it occupied in 2005, its "2005 parent". We call the result the *2005-boundary panel.* Villages created by later pemekaran are returned to their parent polygon, so administrative splits cannot slip into or out of the sample. Appendix B summarises the construction, and Supplementary Material Section S6 documents the procedure, match rates by wave, and two crosswalk-robustness re-estimations.

Controls are baseline village characteristics from PODES 2003, enumerated August 2002. We enter them in levels for the Callaway–Sant'Anna nuisance model and interact them with wave dummies in two-way fixed-effects specifications. After dropping missing controls or outcomes, the sample has $N = 292{,}712$ village-waves. The benchmark Callaway–Sant'Anna estimator uses $N = 225{,}372$ and identifies villages first shaken in waves 3–8 (2008–2024).

4.2 Crime Outcomes: The Rest of the Security Module

The composition argument needs more than communal conflict, and PODES supplies it on the same page. Immediately after the mass-brawl item, the security module asks whether each of ten categories of criminal act occurred in the village, over exactly the same recall window. We use three: theft, robbery, which the questionnaire defines as theft with violence, and assault. These are item codes 01, 02 and 04, and the codes are stable across all seven waves. The remaining seven categories, which include fraud, arson, sexual offences, drugs, gambling, homicide and trafficking, we do not use. None is signed by the framework of Section 3.

Three features make these items work. *First*, they share a reporter, instrument and recall window with communal conflict, so a common shift in village reporting moves all four and differences out of the within-village contrast. *Second*, the column conventions in the released microdata differ across waves even though the questionnaire item numbers do not; Appendix B describes the mapping and Supplementary Material Section S6 verifies it item by item. *Third*, the module is fielded in every wave, so crime and conflict are observed on the same seven-wave grid and event time is contiguous.

The base rates differ sharply. Theft is common, at roughly two villages in five. Robbery and assault are rare, at three and five percent. Communal conflict is rarer still, at between two and three percent (Appendix Table B.1). That gap is why Section 7.1 reports the contrast on a proportional scale as well as a linear one; Appendix B reports incidence by wave.

The composition analysis runs on the subsample of village-waves for which all five outcomes are non-missing: $N = 300{,}018$ village-waves covering 43,351 villages.

It is smaller than the panel used for the level analysis because the security module has item-level non-response. It is the sample on which every number in Section 7.1 is computed.

4.3 Seismic Exposure: USGS ShakeMap MMI

To measure how hard a village was shaken, we use the Modified Mercalli Intensity scale, MMI. It runs from I to XII. It describes how strongly an earthquake is felt at a given point, and how much damage it does there. We take the values from USGS ShakeMap, which stitches instrument readings and felt reports into a fine grid for each earthquake (Worden et al., 2018). For every event we average MMI across the grid cells that fall inside a village's boundary. When several earthquakes strike in one year, we keep the largest value.

The USGS ComCat catalogue spans 1990 to 2024 and holds 72,745 events. It begins fifteen years before our first wave, and that is deliberate. Older shaking cannot then push a village into a spuriously late cohort, or into the never-treated group.

Treatment is defined two ways. The **continuous** version is simply the village-year MMI value. The **binary** version is $D_{mmi6} = \mathbf{1}\{\text{MMI} \geq \text{VI}\}$. That threshold corresponds to "Strong" shaking on the EMS-98 scale, the level at which ordinary masonry begins to crack (Wald et al., 1999). We map exposure to the first PODES enumeration after the earthquake. Because the conflict item looks back twelve months, this can leave the first treated wave only partly exposed. Section 6.2 audits the event dates and reports a clean-post restriction that drops those cases.

For the staggered estimators we need a date of first exposure. We take each village's first calendar year reaching MMI VI or above, then map it to the corresponding wave. Note what this captures. It is first *observed* strong-shaking exposure during the catalogue period, not a village's first strong earthquake in history. Villages that never reach the threshold between 2005 and 2024 form the never-treated comparison group. Cutoff robustness at MMI V and VII appears in

Section 6.2, and robustness to the intensity *measure* in Supplementary Material Section S7.

Why a threshold and not a dose. MMI is an ordinal scale. One step corresponds to a factor of about 1.9 in peak ground acceleration, so entering it linearly would be wrong. Treatment is therefore a threshold, and intensity contrasts compare bands within the same earthquake. Averaging raster cells within a village does still treat the scale as cardinal, and Appendix B explains that choice and bounds what it costs. The estimate keeps its sign under the pixel maximum, the median, and a representative point. Exposure rebuilt from cardinal Peak Ground Acceleration yields a decline of 1.3 percentage points, significant at 1 percent.

Where the shaking data come from. ShakeMap coverage for Indonesian events breaks at 2013, and the break matters. Products for earlier events were not produced at the time. They are reconstructions, generated later in a single batch from the ShakeMap Atlas, built from modelled ground-motion relations rather than recorded intensities, with no station data behind them. From 2013 onward the products are instrumental. So the 2008 and 2011 cohorts, and most of 2014, rest on modelled shaking, while the 2018, 2021 and 2024 cohorts rest on measured shaking.

The census offers an independent check. From 2008 onward, PODES asks whether an earthquake occurred in the village during its recall window. We can therefore ask whether ShakeMap predicts the village's own report. It does, in both eras. The correspondence is weaker in the reconstructed era by roughly a quarter, which is the size and the direction that greater measurement error would produce (Appendix B, Appendix Table B.2).

That validation supports a narrow conclusion, and we will not stretch it. Our treatment is a threshold, so error in the underlying intensity does not enter as classical error in a continuous regressor. It misclassifies the treatment itself, and it does so most often for villages whose true intensity sits near the cutoff.

Under classical misclassification this would tend to attenuate the contrast. But non-classical error can produce more complicated bias, and error correlated with magnitude or with distance from the epicentre is exactly the kind this catalogue might carry. What the validation establishes is that the reconstructed products carry substantial information about which villages were shaken. It does not establish the direction of any bias that remains.

4.4 Descriptive Statistics

Table 1 presents the summary. The pooled conflict probability is 2.6 percent, and roughly 1.2 percent of village-waves experience MMI $\geq$ VI in the prior wave. That low base matters for reading magnitudes later: communal conflict is a rare event, so an effect of a few percentage points is large in relative terms. But the pooled mean understates the benchmark that actually matters. Ever-treated villages reported conflict at 6.2 percent across their pre-treatment waves, roughly twice the pooled rate, and Section 6 interprets every estimate against that baseline.

Table 1: Descriptive Statistics: Medium and High Seismic-Hazard Zones

Variable	Mean	Std. Dev.	Min	Max	*N*
Panel A: Outcomes					
Any communal conflict (composite)	0.026	0.158	0	1.000	303,457
Communal conflict, definition-stable (*warga*)	0.016	0.126	0	1.000	303,457
Panel B: Treatment					
Ever exposed MMI$\geq$VI (absorbing)	0.273	0.445	0	1.000	303,457
Strong shaking, MMI$\geq$VI (window t-1)	0.012	0.109	0	1.000	303,457
MMI intensity (mean-then-max, window t-1)	2.298	1.903	0	8.330	303,457
Panel C: Baseline Control Variables (PODES 2003, enumerated August 2002)					
Any household electricity	0.945	0.225	0	1.000	294,357
Paved road access	0.562	0.494	0	1.000	294,357
Internet available	0.032	0.175	0	1.000	294,357
Police post present	0.102	0.301	0	1.000	294,357
Health centre present	0.414	0.490	0	1.000	294,357
Number of primary schools	2.480	2.365	0	32.0	294,357
Distance to district capital (km)	44.1	54.6	0	985.0	292,712

Sample: KRB Medium/High seismic-hazard zones, PODES waves 2005–2024, native 2005-boundary baseline (43,351 villages $\times$ 7 waves). Panel C reports the baseline controls used throughout: PODES 2003 village characteristics, entered in levels for the Callaway–Sant'Anna nuisance model and interacted with wave dummies in the two-way fixed-effects specifications. Control N falls below the outcome N because baseline characteristics are missing for a small number of villages.

5 Empirical Framework

5.1 Choosing Among Specifications

The comparison this design needs is not just any village. It is a village in the same province, observed in the same survey wave. That is not statistical convenience; it is what the economics demands. Indonesia's seismic zones lie disproportionately outside Java, in post-conflict provinces already on their own declining conflict trajectories. Village fixed effects absorb the level of such a trajectory but not its slope. Pool across provinces and the earthquake gets credited with a decline the province was already experiencing without it.

So the preferred specification absorbs village and province × wave fixed effects and excludes the always-treated cohort.

Concretely, for village v observed in PODES wave t we estimate

$$y_{vt} = \beta\, D_{vt} + \boldsymbol{\gamma}'\mathbf{X}_{vt} + \mu_v + \lambda_{p(v)t} + \varepsilon_{vt}, \qquad g_v \neq 2005, \tag{1}$$

where y_{vt} is the binary communal-conflict indicator and $D_{vt} = \mathbf{1}\{t \geq g_v\}$ is an absorbing indicator equal to one from the wave of first exposure at MMI $\geq$ VI, g_v, onward. The term μ_v denotes village fixed effects and $\lambda_{p(v)t}$ province × wave fixed effects, where $p(v)$ is the province containing village v. It is this second term, rather than a common wave effect, that lets each province follow its own conflict trajectory. $\mathbf{X}_{vt}$ collects the seven PODES 2003 baseline characteristics, each interacted with the six post-baseline wave dummies, so villages differing in initial infrastructure are permitted to diverge over time. Standard errors are clustered at the village (Abadie et al., 2023). The restriction $g_v \neq 2005$ drops the always-treated cohort, which contributes no pre-exposure wave, so every identified cohort is first exposed between 2008 and 2024. Replacing $\lambda_{p(v)t}$ with a common wave effect λ_t recovers the conventional two-way benchmark of Supplementary Material Equation (S1); the difference between the two is precisely the specification choice at issue here.

To test the pre-period and trace the path of the effect, we replace the single absorbing indicator with a set of lead and lag dummies in event time,

$$y_{vt} = \sum_{k\in\mathcal{K}} \beta_k \, \mathbf{1}\{t - g_v = k\} + \boldsymbol{\gamma}'\mathbf{X}_{vt} + \mu_v + \lambda_{p(v)t} + \varepsilon_{vt}, \tag{2}$$

with $\mathcal{K} = \{\leq -5, -4, -3, -2, 0, 1, 2, 3, 4, \geq 5\}$. The wave immediately before first exposure, $k = -1$, is the omitted category, and the endpoint cells are binned so that this reference contains only pre-exposure observations rather than a mixture that includes post-treatment waves. The joint test of the pre-exposure coefficients $\beta_{-4}, \beta_{-3}, \beta_{-2}$ in Equation (2) is the pre-trend test reported below. Raw cohort paths are consistent with both features, and we read them as motivation for the fixed-effect design rather than as evidence on parallel trends (Supplementary Material Figure S1).

There is a temptation here we have to avoid. Staggered designs admit many defensible specifications, and choosing among them after seeing the estimates invites exactly the criticism such designs attract. We have no pre-analysis plan to appeal to, so we do not ask the reader to take the ordering on trust. The argument for province-by-wave effects is the economic one made just above, and it can be weighed without reference to any estimate: seismic zones sit disproportionately in provinces already on their own conflict trajectories, and a common wave effect credits the earthquake with a decline the province was experiencing anyway. Two things make the choice auditable. It is conservative on magnitude, since of the two it yields the *smaller* estimate. And we report the alternative throughout rather than setting it aside, so a reader who rejects our reasoning can read the wave-effects column in every table and form their own view.

The diagnostics then follow rather than lead, and they do not agree. Under the preferred specification the joint test of the pre-treatment leads does not reject for either outcome ($p = 0.68$ composite, $p = 0.66$ *warga*). Under wave fixed effects the Callaway–Sant'Anna joint Wald test rejects for both ($p < 0.001$ and $p = 0.014$). These are different statistics, for a reason given in Appendix A. That the specification we prefer on economic grounds is also the one that passes is convenient for us, and a reader is entitled to discount it accordingly. The sensitivity bounds of Section 6.2, which do not turn on which pre-test one credits, are the more informative constraint.

We still report the estimators below on the wave-fixed-effects panel, both because they are the conventional benchmarks and because the contrast between the two is itself informative. The Callaway–Sant'Anna estimator, its weighting, and the implementation choices this panel forces are described in Supplementary Material Section S4. Villages enter unweighted, so the target is the effect on the average exposed *village*, not the average exposed *person*. A population-weighted version would answer a different and equally reasonable question, but cannot be reported: BPS stopped disseminating village population counts after the 2011 round (Supplementary Material Section S4).

5.2 Identification Assumptions

The earthquake is physically exogenous. Where earthquakes occur and how hard they shake is determined by fault geometry, seismic-wave propagation and local geophysical conditions. All of it is plausibly orthogonal to short-run village-level social and political conditions, conditional on the hazard-zone restriction, village fixed effects and pre-treatment trajectories (Kirchberger, 2017; Gignoux and Menendez, 2016). One qualification matters: even within an exogenous earthquake, realised shaking varies nonrandomly. Villages near faults face higher expected exposure (Borusyak and Hull, 2023). Restricting comparisons to the same broad hazard band addresses that, while village fixed effects absorb fixed differences in fault proximity.

Parallel trends, and a mixed verdict. 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). We assess it two ways, and they do not agree.

The first looks at the average. Pre-treatment coefficients are cumulative deviations from the omitted wave immediately before first exposure, and their average is small and insignificant for either outcome. We give the signs, because they matter for what the drift would do: the average is -0.8 percentage points ($p = 0.33$) on the composite and -0.4 points ($p = 0.56$) on *warga*. Conditional on the covariates, that is, the pre-period drifts in the *same* direction as the estimated effect rather than against it, which is the less comfortable of the two possibilities and not the one the unconditional series suggests. The second is less reassuring. A joint Wald test of all pre-treatment group-time coefficients, which does not net offsetting signs against each other, *rejects* for the composite ($\chi^2(15) = 37.5$, $p = 0.001$) and, less strongly, for *warga* ($\chi^2(15) = 28.8$, $p = 0.017$). That verdict is not invariant to the covariates, and we report the whole surface rather than one cell of it (Appendix A, Table A.2): without regional controls the rejection is decisive on both outcomes, and once province and region effects enter it is marginal, with the composite at $p = 0.053$. The direction of that movement is what the specification argument of Section 5.1 predicts, and we draw no comfort from the marginal cell: on any of these covariate sets the wave-effects panel fails a test the preferred specification passes.

Two features bound that rejection. It is specific to wave fixed effects. And the residual drift runs against the finding rather than toward it, so it biases a subsequent decline toward zero rather than manufacturing one. Appendix A sets out both, with the caveat that a transitory elevation that would have reverted on its own is not excluded. So we treat conditional parallel trends as an assumption to probe rather than assert, and do not read either pre-period as clean. Supplementary Material Section S8 documents the covariate trajectories being conditioned on (Supplementary Material Table S10) and shows the residual growth imbalance is conservative for our finding. We also report HonestDiD sensitivity bounds (Rambachan and Roth, 2023): the sign survives only moderate relative-magnitude bounds, and on both the Sun–Abraham event study and the CS(2021) benchmark the identified set admits zero by $\bar{M} = 0.5$ (Section 6.2).

The always-treated 2005 cohort is excluded; identified cohorts enter between 2008 and 2024. In 2005 the composite uses the broader gate, so we check it against both the definition-stable *warga* series and the identical-wording gate. Both move in the same direction (Table S12).

Figure 1 shows the event study on that preferred specification. It is the exhibit on which a reader should form their own view of the pre-period. The five leads are individually indistinguishable from zero, the largest reaching only $t = 1.4$, and they do not reject jointly ($\chi^2(5) = 4.67$, $p = 0.46$). The decline appears in the wave of exposure at 1.8 percentage points, holds at about two points for three waves, then fades to zero by the fifth. We nonetheless do not rest the argument on this figure. The reason was just given: the same series does not pass under wave fixed effects, and a pre-test that one specification passes and another fails is weak evidence either way.

One technical note on the figure. It and the other event studies plot pointwise intervals. Simultaneous bands are about 35 percent wider (sup-t multiplier 2.65, versus pointwise 1.96). Under those bands all pre-exposure coefficients remain indistinguishable from zero, the effect three waves later becomes marginal, and the exposure-wave and first post-exposure effects remain significant (Supplementary Material Section S4.5).

Because the pre-period is not clean, we rest the composition result on the composition contrast, whose own identifying assumption we state in Section 7.1 rather than treat as absent.

No anticipation. Seismic events at the frequency of PODES surveys, every three years, are unforecastable. Villages cannot systematically adjust conflict behaviour in anticipation of an earthquake that has not happened.

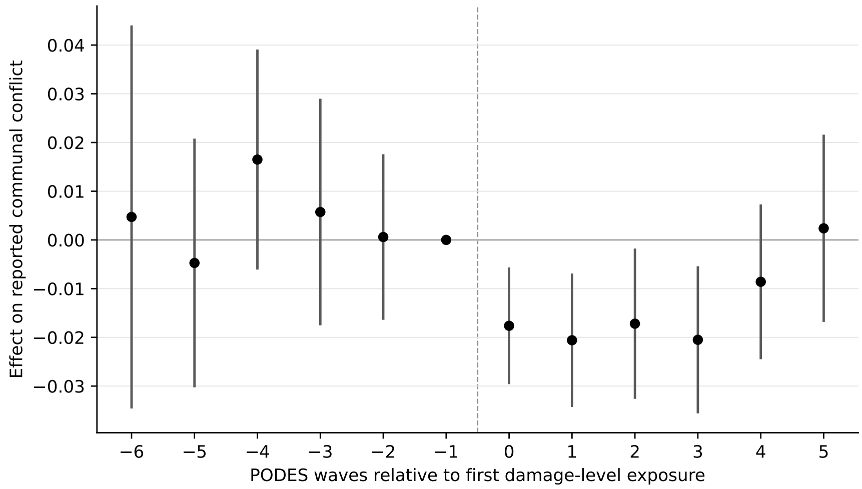

Figure 1: Event Study on the Preferred Specification
Note: Sun–Abraham (2021) interaction-weighted event study of first exposure at MMI≥VI on the composite communal-conflict indicator, with village and province-by-wave fixed effects and 95 percent *pointwise* confidence intervals clustered at the village. Bands valid for the entire path simultaneously are wider by the sup-t factor reported in Supplementary Material Section S4.5. Event time is in PODES waves relative to first exposure; the endpoint cells are binned so the omitted category is the wave before exposure alone rather than a mixture that includes post-treatment cells. The always-treated 2005 cohort is excluded, so every identified cohort is first exposed between 2008 and 2024. Appendix C reports the Callaway–Sant'Anna counterpart on the wave-fixed-effects panel, where the joint pre-trend test behaves differently, and Section 6.2 reports the sensitivity bounds that we, rather than this figure, treat as the binding constraint.

Reporting validity. The outcome is reported by the village head, who also leads the local disaster response, so it is reasonable to suspect that what fell is not conflict but the capacity to report it. We treat this as a threat to identification, not merely to interpretation. The central defence is behavioural: theft is recorded by the same respondent, on the same form, over the same recall window, and it does not fall when communal conflict does. No uniform reporting shock produces that composition. The defence needs only the divergence, which holds under every specification in the main text and in all but one of the appendix. The exception is the switcher-only sample under province-by-wave effects, where village and wave effects and then province-by-wave effects together leave 8.1 percent of the treatment variance and the contrast is no longer estimable rather than rejected (Supplementary Material Section S10.9). It does not need theft's own rise, which is not separately identified once province-by-wave effects are imposed (Section 7.1). Section 7 develops the argument, adds three further observational checks, and reports placebo items from the same form, where the sex placebo is a precise null and head age moves opposite to the direction a reporting story predicts.

One pre-period signal we traced. An earlier calendar-year coding of first exposure produced a marginally significant pre-period coefficient, traced to wave-boundary timing that misplaced onset by one wave for about 21 percent of ever-treated villages; re-timing exposure to the first wave enumerated after the event removes it (Appendix A). The joint Wald rejection under wave fixed effects remains, and we carry it as an open caveat, but the binding constraint is the HonestDiD bound (Section 6.2), not the pre-trend tests.

Pre-treatment balance. In the 2005 KRB Medium/High sample, eventually-treated and never-treated villages are close on nine of ten normalised differences, each below the 0.25 threshold (Imbens and Wooldridge, 2009); the exception is the health-centre count (0.423), and interacting it with wave moves every estimate by less than one percent (Appendix Table D.1). Remaining level differences are absorbed by village fixed effects, and the pre-trend diagnostics of Appendix A address differential trends.

6 The Level Effect on Communal Conflict

The results come in two designs, and we keep them apart because they answer different questions and stand on different assumptions. This section reports the first: the effect on the outcome the design intends, whether any mass brawl was reported (Section 6.1), followed by the threats that estimate has to survive (Section 6.2). Section 7 then turns to the second design, which asks what that number means, since the level alone cannot say. Section 8 sets out a margin these data will not identify, and reports it in full precisely because its failure is instructive.

One thing is better said now than at the end. The level effect is a magnitude we report and defend within stated bounds, not the claim the paper rests on.

Its sensitivity bound is the constraint that binds, and it is not generous: the identified set admits zero once pre-trend violations half the size of the largest observed pre-period deviation are allowed (Section 6.2). What the paper claims is the *composition* of disorder in Section 7, which is identified by a contrast taken within the village and does not inherit that bound. A reader who discounts this section entirely still has the paper's argument.

6.1 The Effect on Reported Mass Brawls

Start with the number. Under the village and province-by-wave effects of Equation (1), estimated without its covariate block so that the gate is measured on the widest sample the design allows, first exposure at MMI $\geq$ VI lowers reported mass brawls by 1.73 percentage points ($p < 0.001$). Adding the baseline-by-wave controls is what Table C.2 reports for the subtype composite, and it moves that estimate from 1.5 to 1.6 points; we note the difference rather than let Equation (1) stand for a specification it did not produce. The pre-exposure mean among eventually exposed villages is 6.7 percent. So the decline is 26 percent of what these villages reported before they were shaken.

Three features of Table 2 carry the argument, and each is worth taking in turn.

The first is the comparison across measures. Under village and wave fixed effects the subtype composite gives the larger decline, 1.9 percentage points against the gate's 1.6. That is what a definition narrowing over the panel mechanically delivers. Now impose province-by-wave effects, and the ranking reverses. The gate does not attenuate at all; it strengthens slightly, to 1.7 points, 106 percent of its wave-effects estimate. The composite falls to 1.5 points and retains 80 percent. The *warga* subtype falls to 0.3 points, retains 49 percent, and loses significance entirely. The concern that motivated the province-by-wave specification, that seismic zones sit disproportionately in provinces already on declining trajectories, turns out to bite hard on the narrower measures and barely at all on the gate.

The second is the pre-exposure record. Panel B traces the outcome in event

Table 2: The Level Effect on Three Measures of Communal Conflict

The definition-stable gate is the measure least dependent on comparisons made across provinces, and the only one whose near pre-exposure path is flat.

	Mass-brawl gate (definition-stable)	Subtype composite (`conflict_any`)	*Warga* subtype
Panel A. Effect of first exposure at MMI $\geq$ VI			
Village + wave FE	-0.0163	-0.0193	-0.0070
	(0.0035)	(0.0033)	(0.0026)
Village + province $\times$ wave FE	-0.0173	-0.0154	-0.0035
	(0.0042)	(0.0040)	(0.0032)
p	0.0000	0.0001	0.2836
Pre-exposure mean	0.0670	0.0624	0.0372
Effect as % of that mean	-25.8	-24.7	-9.3
% of wave-FE estimate retained	105.8	79.7	49.5
Panel B. Gate in event time, endpoints binned, reference t_{-1}			
	Village + wave FE	*Village + prov. $\times$ wave FE*	
$t_{\leq -5}$ (binned)	-0.0238	-0.0070	
	(0.0087)	(0.0098)	
t_{-4}	-0.0118	-0.0033	
	(0.0088)	(0.0095)	
t_{-3}	-0.0037	-0.0086	
	(0.0088)	(0.0099)	
t_{-2}	0.0120	0.0048	
	(0.0066)	(0.0079)	
t_0	-0.0215	-0.0183	
	(0.0054)	(0.0063)	
t_{+1}	-0.0195	-0.0242	
	(0.0057)	(0.0069)	
joint p, interior leads	0.067	0.594	
joint p, incl. binned	0.005	0.682	
Observations	232,778	232,778	232,778

The mass-brawl gate is the survey's own yes/no item on whether any mass brawl occurred in the village in the past twelve months, asked in identical terms in every wave from 2005 to 2024. The subtype composite is the union of the named communal subtypes, which coincides with the gate in 2005 and is progressively narrower thereafter, so it carries a definitional trend that wave effects absorb in level but not in interaction with treatment. Panel A reports the absorbing first-exposure effect under the two fixed-effect structures the paper compares; the row labelled % of wave-FE estimate retained is the ratio of the province-by-wave coefficient to the wave-FE coefficient, and is the quantity of interest for how much of each estimate is between-province rather than within. Panel B reports the gate in event time on the same sample with the always-treated cohort dropped, endpoints binned rather than discarded so that the reference period is t_{-1} alone. Standard errors clustered at the village.

time with the endpoints binned rather than discarded, so the reference is the wave before exposure alone. Under the preferred specification all four pre-exposure coefficients are individually insignificant, the joint test over the interior leads does not reject ($p = 0.59$), and the joint test including the binned far endpoint does not reject either ($p = 0.68$). The outcome then falls by 1.8 points in the wave of exposure and 2.4 points in the wave after. Under wave effects alone the same leads are jointly marginal ($p = 0.07$) and reject once the binned endpoint is included ($p = 0.005$). That is a further reason to prefer the province-by-wave counterfactual, not a reason to doubt it.

But we will not call that record clean, for two reasons. First, non-rejection buys only what the test's power is worth. A linear differential trend of half a percentage point per wave escapes this pre-test half the time, and on the gate a trend that size would have to run the length of the observed pre-period to accumulate to the estimated effect (Appendix A, Table A.1). That makes passing informative here, and close to uninformative on *warga* (Roth, 2022). Second, the constraint that actually binds is not the pre-trend test but the relative-magnitudes bound. And there the gate is no more robust than the subtype composite. The identified set for its first post-exposure effect excludes zero at $\bar{M} = 0.25$ ($[-0.034, -0.003]$) but admits zero at $\bar{M} = 0.50$ ($[-0.039, +0.001]$). The composite breaks down there too, and its aggregated target admits zero already at $\bar{M} = 0.25$. So the gate buys definitional stability and near-immunity to province trends. It does not weaken the parallel-trends requirement.

The third is the unit of inference, and here the result is only partial. Treatment is assigned by earthquakes, not by villages, so village clustering overstates precision. We estimate effects across the 19 usable shocks and resample whole earthquakes. The treated-village-weighted mean is a decline of 1.6 percentage points, with a block-bootstrap 95 percent interval of $[-0.036, -0.000]$ and $p = 0.046$. It is negative in 12 of 19 events, with a leave-one-out range of $[-0.023, -0.011]$. But give every earthquake equal weight, so that a shock treating twenty villages counts the same as one treating two thousand, and the decline is only 0.4 points and does not approach significance ($p = 0.78$). The effect concerns the average exposed *village*. It is not detectable as an average across *earthquakes* (Table 3).

Leave-one-out understates how concentrated the weighted mean is, so we give the decomposition rather than let a reader infer it from the forest plot. Three shocks carry it: Lombok 2018 (−5.2 points on 353 treated villages), Palu 2018 (−6.3 on 252) and Ambon 2019 (−12.2 on 56). Together they supply 88.7 percent of the weighted sum. Dropping all three leaves −0.24 points on the remaining 2,136 treated villages, and the equal-event mean of the other sixteen shocks is +1.05 points. Two further facts about those three belong with the number. They share a first post-exposure wave, 2021, whose twelve-month recall window runs from May 2020 to April 2021 and therefore sits inside Indonesia's PSBB and PPKM mobility restrictions; the 2018-onset cohort, measured before the pandemic, is positive. And the 2018 round was enumerated in May 2018, which is before the August Lombok and September Palu earthquakes, so for those two the wave we treat as the last pre-exposure observation is the last one genuinely prior to the shock. We return to what this concentration does and does not license in Section 9.

Appendix C reports the estimator comparison, the full event study and the specification-selection arithmetic.

So far we know how much, but not why. A decline in reported mass brawls remains consistent with three things at once: less fighting, less reporting, or better local order. No single outcome can separate them. That is what the next section is for.

6.2 Threats to Identification and Robustness

This section asks whether the level result survives the main threats to the design. It is organised around those threats rather than around specifications. The estimate stays negative and close to the benchmark across most margins, though not under every weighting and unit of assignment. Supplementary Material Section S15

Table 3: Earthquake-Level Inference on the Mass-Brawl Gate
The effect holds when earthquakes are weighted by the villages they treat, and is not detectable as an unweighted average across shocks.

	Mass-brawl gate
Treated-village-weighted mean	-0.0165
earthquake block-bootstrap SE	(0.0091)
95% CI	[-0.0359, -0.0002]
two-sided p	0.046
Equal-event mean	-0.0036
SE	(0.0125)
two-sided p	0.777
Leave-one-earthquake-out range	[-0.0227, -0.0114]
Earthquakes with a negative effect	12 of 19
Usable earthquakes	19

For each usable earthquake, defined as one whose first-treated cohort contains at least twenty in-sample villages and whose onset wave falls in 2008–2024, we estimate the absorbing treated-cohort effect on the definition-stable mass-brawl gate against never-treated controls, with village and wave fixed effects. The treated-village-weighted mean weights each earthquake by the number of villages it first treats; the equal-event mean weights every earthquake alike. The block bootstrap resamples whole earthquakes with replacement (9,999 draws, seed 42) and recomputes the weights in each draw, which is the correct unit of assignment and gives a materially larger standard error than village clustering.

reports every check in full. Here we state the question, the answer, and what it bears on.

Timing, measurement, and functional form. PODES conflict at wave t covers the twelve months before May enumeration. An earthquake inside that window therefore gives a partially exposed $t = 0$ outcome. An exact-date audit shows that 74 percent of the 19 stacked earthquakes are clean-post, and every major earthquake carrying the effect is clean-post. Restricting to clean-post cohorts leaves the estimate essentially unchanged, 2.1 against 2.0 percentage points. Measurement is the next question. Restricting to village-waves with a measured rather than imputed first lag leaves the composite at 1.5 percentage points ($p = 0.01$). Functional form is the third. The estimate keeps its sign across four raster-to-village aggregation rules and a five-percent area trim, and keeps its significance except under the pixel-maximum rule. Instrumental Peak Ground Acceleration reproduces it, as does a conditional logit. That last check matters, because parallel trends in levels and after a nonlinear transformation cannot generally both hold (Roth and Sant'Anna, 2023).

Parallel trends and region effects. Section 5 states the assumption and the evidence in full. The residual drift under wave fixed effects is positive, so it biases a subsequent decline toward zero rather than manufacturing one. Region effects need a sharper test. Village fixed effects already absorb region levels, so what matters is region-specific trends. Province×wave effects attenuate the no-controls absorbing effect on the subtype composite by about a fifth, from 1.9 to 1.5 percentage points. It stays significant at 1 percent under village clustering and at 5 percent under kabupaten clustering ($p = 0.026$). The *warga* margin does not survive ($p = 0.28$). On the definition-stable gate the same comparison does not attenuate the estimate at all (Table 2), which is why Section 6.1 rests on it.

Cohort influence and spatial coincidence. No single cohort drives the result. Dropping each in turn leaves every estimate negative and significant, ranging from 1.3 to 2.5 percentage points. A forward-lag placebo is a null (0.3 points, $p = 0.60$). We then relocate each earthquake footprint to a random never-treated village, five hundred times. Only one draw is at least as negative as the actual estimate of 2.0 points (one-sided $p = 0.004$). One caution follows from the same exercise. Placebo dispersion is 2.1 times the clustered standard error, which indicates that village clustering understates uncertainty when earthquakes assign treatment.

HonestDiD sensitivity. This is the constraint that actually binds, not the pre-trend tests. On the preferred specification, the composite identified set excludes zero at $\bar{M} = 0.25$ ($[-0.032, -0.004]$) and admits it at $\bar{M} = 0.50$ ($[-0.037, 0.000]$). The definition-stable *warga* target breaks down an order of magnitude earlier, at $\bar{M}$ between 0.05 and 0.10. That bound comes from Sun–Abraham, however, and $\bar{M}$ is not comparable across estimators (Appendix E).

We draw the consequence rather than leave it implicit. A design whose sign requires $\bar{M}$ below one half is not one on which to rest a causal claim about the level, and on the definition-stable outcome the requirement is far tighter still. We therefore report the level as a bounded magnitude and rest the paper on the composition contrast, which rests on a *different* assumption, that the theft-minus-communal difference would have moved in parallel, and which we test directly (Section 7.1). Different is not the same as weaker, and we ran the same sensitivity analysis on the contrast rather than assert an advantage for it. It has a bound of its own. In $\bar{M}$ units that bound looks smaller, 0.15 at the wave after exposure against 0.25 to 0.50 for the gate, but $\bar{M}$ is scaled by each outcome's own pre-period and the contrast's is the noisier one, so the two are not comparable as stated. Converted to the absolute violation each tolerates, as Appendix E does for the estimator comparison, they are close: 0.0034 to 0.0068 for the gate against 0.0053 at the wave after exposure and 0.0064 for the aggregate. Two concessions follow. At the wave of exposure the contrast does not separate from zero even under exact parallel trends, so its evidence lives one wave later and in the aggregate rather than on impact. And the definition-stable version, theft against *warga*, does not separate at that later wave either. We report those rather than claim the contrast escapes the constraint. One further defence is available in principle, and we ran it rather than appeal to it. Rambachan and Roth (2023) allow the relative-magnitudes class to be restricted by a sign or a monotonicity condition on the differential trend, and a reader might expect that to rescue the level. It does not. Imposing a non-negative differential trend pushes the breakdown value past $\bar{M} = 64$ on the composite, but that number is an artefact rather than a finding: with a negative estimate and a non-negative bias imposed, the restriction bounds the effect above by the estimate itself, so the identified set stops widening upward as $\bar{M}$ grows: its upper end holds about 0.4 of a percentage point below zero and never reaches it. The sign then follows from the restriction rather than from the data, and the breakdown statistic stops measuring anything. The opposite restriction lowers the bound slightly, to 0.44. On the definition-stable series every restricted class breaks down at $\bar{M} = 0$, because that coefficient is not separated from zero even under exact parallel trends.

We therefore report the unrestricted bound as the honest one (Supplementary Material Section S15.3).

7 The Composition of Disorder: One Shock Across Four Acts

Why would one shock raise reported theft while communal fighting falls? This is the paper's central result. PODES records not only communal violence but also village-level crime, and competing theories of the earthquake–conflict link make sharply different predictions across those outcomes.

One caution before we go in. The evidence establishes which *margins* of disorder move together, and that is enough to discriminate among theories of the sign. It does not *identify* the underlying economic channel, because PODES measures neither welfare nor labour-market outcomes. So we read these results as reduced-form patterns consistent with a capacity account rather than as proof of it, and we flag direct welfare measures, for example from Susenas, as what would convert that consistency into identification.

7.1 The Composition Result: Communal Conflict Falls Where Crime Does Not

We estimate five outcomes at once on a single *common sample*: 43,351 villages over seven waves, with the same absorbing treatment and no controls. All five are stacked with outcome-interacted fixed effects, so the standard errors account for the correlation of a village's own outcomes (Table 4). One difference from the level specifications: there we exclude the cohort first exposed in 2005, here we keep it, because it is the sample the crime module defines. Dropping it moves the estimates in the direction that favours our reading rather than against it (Supplementary Material Section S10).

Like the conflict outcome, the crime outcomes are constructed from wave-specific PODES crime-module items. Section 4.2 gives the item for each wave and Supplementary Material Section S6 the verification. A change in an item's wording is common to every village in that wave, so it is absorbed by the outcome-specific wave effects rather than left in the contrast. What the design cannot absorb is a treatment-correlated shift specific to one item. We take that up next. The contrast

Table 4: 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}	-0.0218***	0.0429***	-0.0141***	-0.0008	-0.0085***
	(0.0034)	(0.0080)	(0.0035)	(0.0045)	(0.0027)
FDR q (Benjamini-Hochberg)	0.000	0.000	0.000	0.864	0.002
Baseline mean	0.026	0.418	0.031	0.055	0.016
Observations	300,018	300,018	300,018	300,018	300,018

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 all seven crime-module waves (2005–2024), restricted to village-waves where all five outcomes are non-missing. Communal is the collective violent composite; 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. The Westfall–Young family-wise correction is not reported: it requires a separate bootstrap run and is available on request. 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= 57.70 p= 0.000; theft vs. warga F= 38.78 p= 0.000; theft vs. assault F= 26.47 p= 0.000; theft vs. robbery F= 48.83 p= 0.000, the comparison the framework signs. Kabupaten-clustered robustness: theft vs. communal F= 7.07 p= 0.008; theft vs. warga F= 5.04 p= 0.025; theft vs. robbery F= 10.89 p= 0.001. Standard errors clustered at the village level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

requires two things: parallel trends in the theft-minus-communal difference, and a reporting response common to the two items. We test the first directly rather than assert it. Under the preferred specification the joint test over the interior pre-exposure leads does not reject ($p = 0.31$; $p = 0.18$ against the definition-stable series), and the two leads nearest exposure are precise nulls. Under wave effects the same test rejects ($p = 0.018$). One term is unstable in both: the binned coefficient on periods five waves and more before exposure is 4.5 percentage points and individually significant ($p = 0.025$), so a joint test including it sits at $p = 0.06$.

That cell is narrow and late. Only the 2021 and 2024 cohorts are observed five or more waves before their own exposure at all, and between them the 2021 cohort supplies 694 of its 1,096 village-waves. No earlier cohort contributes a single observation. So the term compares those two cohorts with the rest of the panel fifteen years or more before they were shaken, which is a level difference between late-treated and early-treated villages rather than a trend running into exposure. Supplementary Material Table S13 reports both tests. Supplementary Material Section S10 states the assumption, the three threats that survive the differencing, and what bounds each.

The result follows. On identical data, communal conflict falls (-0.0218^{***}) while theft does not follow it down and instead rises, by 4.3 percentage points. The two move in opposite directions (Figure 2, Panel A), and a test of their equality is decisively rejected ($p < 0.001$), on a gap of 6.5 percentage points. This is what we lean on, and within it the communal decline, the half that survives every specification we report.

There is an objection to answer first. The composite changes definition in 2008, so could the result be an artefact of that? We anchor the contrast on *warga*, the civilian-versus-civilian subtype with near-stable wording. It falls by 0.8 percentage points ($p = 0.002$). Equality with theft is rejected ($p < 0.001$), including under kabupaten clustering (Supplementary Material Section S10).

The same holds when the communal outcome is the mass-brawl gate that Section 6.1 makes primary. On the identical common sample the gate falls by 1.8 percentage points under wave effects and 1.9 under province-by-wave effects, both significant at 1 percent. Theft is 4.3 and 1.2 points. Equality is rejected at $F = 50.8$ and $F = 10.0$ ($p = 0.002$), a sharper rejection than the composite delivers. So the contrast is not an artefact of which communal measure is used. It holds on all three.

The next objection is about scale. The four outcomes have very different base rates, so a linear equality test cannot rule out a common proportional disturbance.

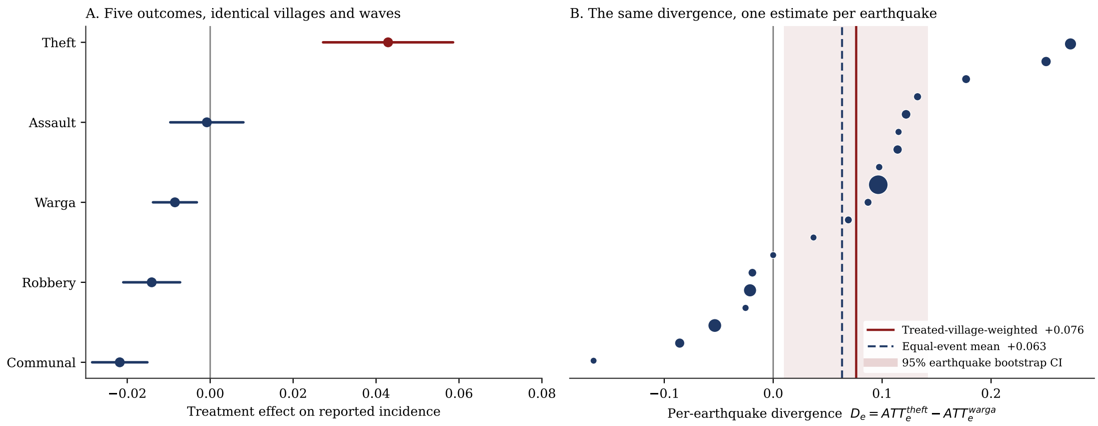

Figure 2: The Composition of Disorder
Communal violence falls where acquisitive crime does not, and the divergence holds when the earthquake, rather than the village, is the unit of inference.
Note: Panel A plots the absorbing treatment effect on each of five outcomes estimated on the single common sample described in this section, with village and wave fixed effects, no covariates, and 95 percent village-clustered confidence intervals, so a coefficient of -0.022 is a fall of 2.2 percentage points. Theft is drawn in maroon. Panel B plots the per-earthquake divergence $D_e = ATT_e^{theft} - ATT_e^{warga}$ for each of the 19 usable earthquakes, sorted by magnitude, with marker area proportional to the earthquake's number of unique first-treated villages. The solid maroon line is the treated-village-weighted mean, the dashed navy line the equal-event mean, and the shaded band the 95 percent earthquake block-bootstrap interval for the treated-village-weighted estimate (9,999 draws, seed 42). Individual earthquakes are not labelled here because the common-sample restriction changes the treated counts; the named per-event estimates for communal conflict appear in Figure S2 and the full event list in Supplementary Material Section S15.

If every act moved by the same percentage, theft on a 42 percent base would move by far more percentage points than communal conflict on a 2.6 percent base. We re-estimate the same stacked specification as a fixed-effects Poisson model, in which coefficients are proportional responses. The divergence survives. Under wave effects, communal conflict falls by 31 percent while theft rises by 14 percent; under the province-by-wave effects our selection rule prefers, the same contrast is -28 percent against $+6$ percent. Equality is rejected on both ($p < 0.001$), as is equality against the definition-stable series ($p = 0.003$). Two cautions belong here rather than in an appendix. First, the proportional gap is about half as wide under the preferred effects as under wave effects, so the headline proportional figures should be read from the former. Second, robbery, assault and *warga* are each individually indistinguishable from zero under the preferred effects ($p = 0.31$, 0.85 and 0.23), and assault in particular *rises* by 15 percent under wave effects while sitting at -1 percent under province-by-wave ones. What the proportional scale establishes is therefore the theft-communal divergence, not a ranking of the other three. The Poisson figure is the proportional effect to read, not the ratio of the linear coefficient to the pooled base. Villages that are eventually exposed report communal conflict at 6.2 percent before exposure rather than at the pooled 2.6 percent, and 2.2 points against that base is a decline of about a third, which is what the Poisson returns.

All three equality contrasts survive kabupaten clustering. And correcting the five per-outcome p-values for multiple testing leaves the communal, *warga* and robbery declines and theft's rise significant at 5 percent, with assault the single null ($q = 0.864$; Table 4). Supplementary Material Section S10 reports the full proportional estimates, the robbery and assault coefficients, and the clustered and corrected variants.

Estimated outcome by outcome under the CS(2021) benchmark, the same gradient appears (Supplementary Material Table S16): theft rises by 6.4 percentage points, robbery does not rise, assault is null and communal conflict falls. But theft's own pre-trend fails there (3.0 percentage points, $p < 0.01$). Theft can also rise from weaker guardianship, abandoned dwellings, aid inflows, or changed reporting. So the benchmark supports the *contrast* with communal conflict rather than a claim about theft's level (Supplementary Material Section S10).

Three inferences follow. *First*, the decline is not part of a general improvement in local order, since property crime moves sharply the other way. *Second*, the asymmetry is exactly what a single degraded input predicts. Theft is a stealth act that a lapsed night watch stops deterring; mass fighting is a coordinated act that the same lapsed capacity stops enabling; and police-post presence does not rise. The two part company without any change in the motive for predation. Studies of income shocks find the same asymmetry (Mehlum, Miguel, and Torvik, 2006; Blakeslee and Fishman, 2018; Bignon, Caroli, and Galbiati, 2017; Hardiman, 2025), though we cannot separate that reading from a strengthening economic motive, because household welfare is unobserved here. *Third*, with individual violence flat while theft rises, the decline is concentrated in *organised, collective* violence, which requires assembly and arenas of inter-group contact (Ray and Esteban, 2017; Esteban, Mayoral, and Ray, 2012; Yanagizawa-Drott, 2014). Assault, the closest individual analogue of fighting, does not move, but the framework leaves it unsigned, so that null is uninformative.

Robbery is the margin that really discriminates. A raised-motive story predicts the two acquisitive acts move together. They do not: under wave effects theft rises while robbery falls, and under the preferred specification theft does not move while robbery still falls. The framework signs the theft–robbery comparison but leaves robbery's level to the data. Their effects differ at $F = 48.8$ ($p < 0.001$) under village clustering and $F = 10.9$ ($p = 0.001$) under kabupaten clustering, the strongest pairwise contrast in Table 4. They also differ on the proportional scale ($p = 0.017$), which matters given base rates of 0.031 and 0.418. So the pattern does not require a decline in the motive for conflict. What it selects is disruption of the social technology of mass fighting. We do not read robbery's level as a causal estimate, any more than theft's, because its own pre-trend rejects (Supplementary Material Section S10).

The contrast under the preferred specification. Consistency requires applying the province-by-wave standard of Section 5.1 to the contrast and not only to the level, so we replace outcome×wave effects with outcome×province×wave effects in the stacked model. The communal decline attenuates from 2.2 to 1.7 percentage points and stays significant at 1 percent, while theft's rise falls from 4.3 to 1.2 points and is no longer distinguishable from zero. The equality tests survive: theft and communal conflict differ at $F = 8.61$ ($p = 0.003$) and proportionally ($p < 0.001$), communal conflict falling by 28 percent of its base while theft moves 6 percent the other way. Against the definition-stable *warga* series the linear test sits just outside 5 percent ($p = 0.061$), the weakest of the three comparisons, and we do not present it otherwise. For theft against robbery, the comparison the framework actually signs, equality is rejected on the linear scale ($p = 0.036$) but not on the proportional one ($p = 0.143$) (Supplementary Material Section S10, Table 5).

Table 5: The Composition Contrast Under the Preferred Specification
One common sample, all five outcomes stacked; the theft-versus-robbery comparison the framework signs is the one that survives.

	Linear		Proportional
	(1) Wave	(2) Province × wave	(3) Province × wave
Panel A. Effect of first exposure, outcome by outcome			
Communal conflict	-0.0218***	-0.0169***	-0.3345***
	(0.0034)	(0.0041)	(0.1047)
Theft	0.0429***	0.0123	0.0568**
	(0.0080)	(0.0093)	(0.0231)
Robbery	-0.0141***	-0.0075**	-0.1363
	(0.0035)	(0.0037)	(0.1330)
Assault	-0.0008	-0.0044	-0.0148
	(0.0045)	(0.0053)	(0.0764)
Warga conflict	-0.0085***	-0.0057*	-0.1629
	(0.0027)	(0.0033)	(0.1348)
Panel B. Equality of treatment effects across outcomes			
Theft = communal, p	0.000	0.003	0.000
Theft = *warga*, p	0.000	0.061	0.104
Theft = robbery, p	0.000	0.036	0.143
Observations	1,500,090	1,500,090	457,943

All five outcomes are stacked and interacted with first exposure so that each column is one regression, with outcome-by-village effects throughout. Columns (1) and (2) run on one common sample of 43,351 villages across 28 provinces and all seven crime-module waves (2005–2024). Column (3) does NOT: the Poisson estimator drops any outcome-village group whose outcome is zero in every wave, and because the base rates differ by more than an order of magnitude, theft retains almost every village while the rarer communal series retain only those with at least one recorded episode. The proportional equality tests in column (3) therefore compare outcomes measured on different village sets, which is why we read them alongside the linear tests rather than in place of them. Columns (1) and (2) differ only in whether the time effects are outcome-by-wave or outcome-by-province-by-wave; column (3) repeats column (2) as a Poisson model, so its coefficients are proportional rather than in percentage points. Panel B tests equality of the treatment effect between theft and each of the other outcomes within the same regression. The theft-versus-robbery comparison is the one the framework signs, since both are property offences recorded on the same form. Standard errors are clustered at the village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Note what the preferred specification actually removes: theft's own rise, not the communal decline. The decline attenuates by about a fifth and stays significant; theft's coefficient loses three quarters of its size and all of its precision. A substantial part of theft's movement is between-province rather than within, which is what one would expect of a shock whose footprint is regional, and here we cannot separate that from province-specific trends in reporting or in acquisitive crime. Nothing in the mechanism argument requires theft's level. It requires only that theft fails to fall, and that holds under every specification we report except the switcher-only province-by-wave cell, where too little treatment variance survives to estimate it either way (Supplementary Material Section S10.9).

The divergence at the earthquake level. Because treatment is assigned by earthquakes, we also take the divergence to the shock level. That is the honest unit, given how concentrated the communal effect is in a minority of earthquakes (Supplementary Material Section S4.6). For each of the 19 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 (Table S15).

The treated-village-weighted mean divergence is 7.6 percentage points. And unlike the communal decline, it is significant under earthquake-level inference (Figure 2, Panel B). A bootstrap resampling whole shocks returns a 95 percent interval of $[+0.010, +0.142]$ ($p = 0.025$), with a leave-one-out range of $[+0.057, +0.095]$. The divergence is positive in 12 of the 19 events, and the equal-event mean is positive and significant in its own right (0.063, $p = 0.023$). So the result does not depend on weighting toward the larger earthquakes. The compositional shift is the part of our result that holds up at the level of the shock itself under either weighting, even where the communal decline alone is only marginal and, on the equal-event estimand, absent. Randomisation inference points the same way with less margin: relocating earthquake footprints at random leaves the divergence significant at a one-sided $p = 0.044$, against $p = 0.004$ for the level result, so the composition contrast clears the design-based test by the narrower of the two margins (Supplementary Material Section S15.18).

The *warga* response is negative and separated from zero at the shock level. On the full *warga* sample its treated-village-weighted effect is a decline of 1.1 points, with an earthquake block-bootstrap 95 percent interval of $[-0.0260, -0.0007]$. Its failure to rise is therefore informative rather than a mere failure to reject

(Supplementary Material Section S10).

The same divergence under the preferred counterfactual. Correct inference and a credible counterfactual are different requirements, and we impose them jointly: we re-estimate divergence event by event using province-by-wave rather than wave effects, and compare above-threshold villages with those that felt the same shock below the damage threshold. The result must be reported as it stands. Divergence remains positive but insignificant in both designs, the weighted mean falling from 7.6 to 1.6 percentage points and the within-earthquake estimate reaching 2.2 points, neither distinguishable from zero. The pre-period placebo for that design, by contrast, is clean. Run entirely inside the pre-exposure waves it returns 0.4 points with a standard error of 1.5, a null under village clustering ($p = 0.77$) and under earthquake clustering ($p = 0.87$). So the within-earthquake contrast is not disqualified by a contaminated pre-period. What limits it is precision at the unit of assignment: the 2.2-point estimate reaches $p = 0.03$ when villages are the clusters and only $p = 0.47$ when earthquakes are, and earthquakes are what assign treatment. We therefore read it as uninformative in either direction rather than as a failed test. What survives at the shock level is the between-earthquake divergence under wave effects; what does not is the within-footprint intensity contrast (Supplementary Material Section S10).

One clarification about what we are doing. This is outcome discrimination, not mediation. The crime outcomes are alternative responses to the same shock, and their signs select among theories instead of decomposing 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. So 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.

Reporting artifacts. All four outcomes are reported by the same village head on the same form, so it is fair to ask whether post-earthquake administrative disruption suppresses the *reporting* of communal conflict rather than conflict itself. Four features of the data argue against it. The items sit in the same security module with the same respondent and format, yet move in opposite directions, where a disruption to reporting capacity would depress them together. Under-reporting depresses items, it does not inflate them, and theft rises on the same form in the same wave. Salience runs against the artifact too, since a mass brawl is more memorable than diffuse petty theft, so degraded recall would understate the theft increase. And the conflict decline is negative at all six post-treatment horizons and significant as late as t_{+3}, well beyond any plausible horizon of administrative disruption. The ACLED cross-check adds an independent media-based lens whose incident-level reading does not show the broad masked rise the artifact story requires (Supplementary Material Section S5.1). A reporting story consistent with all four facts would have to be as specific as the behavioural one it replaces.

There are also two placebo items on the same form from the same respondent: the village head's age and a female-head indicator. Under the causal estimator the latter is a precise null. Head age does move significantly, but opposite to the direction a reporting explanation predicts (Supplementary Material Section S11). The channels through which a shock can move each outcome, and why the framework does not sign them without the composition contrast, are set out in Supplementary Material Section S4.

7.2 Which Capacity Inputs Move

The next question: of the four inputs the framework names, which does the shock actually move? We state the evidentiary standard once here. Where an input can be used as a *pre-determined moderator*, the treatment is interacted with its pre-exposure value. Because the moderator is fixed before any identified cohort is treated, it cannot itself respond to the shock, so no parallel trend in the input is required. What is required instead is parallel trends *within* each stratum of the moderator, which is stronger than the pooled assumption rather than the same one. Where an input is instead treated as an outcome in its own right, that is a second causal claim requiring its own parallel trends, and none of the input series PODES offers passes that test. So we lead with the moderator evidence and report the outcome regressions as descriptive.

Mobilisation capacity: the moderator test. Strong mobile-phone signal proxies contact and mobilisation capacity (Yudhistira et al., 2021). Mobile coverage is associated with protest in Africa (Manacorda and Tesei, 2020), and social media with protest in Russia (Enikolopov, Makarin, and Petrova, 2020). In those settings it lowers coordination costs rather than simply spreading information. If that is the operative channel here, the decline should concentrate in villages that had signal to lose. It does.

Table 6 interacts first exposure with signal presence at the 2005 baseline, a value fixed before any identified cohort is treated and held by 76.4 percent of villages. The interaction is negative and significant on both outcomes and under both fixed-effects regimes. The split is stark: among villages with no baseline signal the effect is 0.1 percentage points with a standard error of 0.7, indistinguishable from zero, against a decline of 2.2 points among villages that had it. Two features make this the strongest mechanism evidence in the paper. The moderator is pre-determined, so it cannot have responded to the shock, which is the standard route in this literature (Bai 2023; Amarasinghe et al. 2026). And it survives the province-by-wave specification, which most of the paper's other margins do not. It is not free of the parallel-trends requirement, only of the input's own version of it, and the horse race below shows how far that buys us.

But it is not a decomposition, and we will not pretend otherwise. Panel B races signal against the other 2005 infrastructure inputs the design can measure. The signal interaction weakens once the others enter. It reaches only 10 percent significance on the composite (1.7 percentage points, $p = 0.088$) and does not clear

Table 6: Mobilisation Capacity: The Pre-Determined Signal Moderator
The decline is concentrated in villages that had mobile signal to lose, and the moderator is fixed before any identified cohort is treated.

	Any conflict		*Warga* conflict	
	(1)	(2)	(3)	(4)
Panel A. First exposure interacted with 2005 mobile signal				
First exposure	0.0030	-0.0014	0.0078	0.0058
(villages without 2005 signal)	(0.0062)	(0.0065)	(0.0049)	(0.0052)
First exposure × 2005 signal	-0.0307***	-0.0205***	-0.0203***	-0.0135**
	(0.0072)	(0.0074)	(0.0057)	(0.0058)
Implied effect, villages with signal	-0.0277***	-0.0219***	-0.0126***	-0.0077**
	(0.0038)	(0.0046)	(0.0030)	(0.0037)
Village fixed effects	Yes	Yes	Yes	Yes
Wave fixed effects	Yes	—	Yes	—
Province × wave fixed effects	—	Yes	—	Yes
Observations	232,778	232,778	232,778	232,778
Panel B. Horse race against other 2005 infrastructure, province × wave fixed effects				
First exposure × 2005 signal	-0.0205***	-0.0165*	-0.0135**	-0.0126
	(0.0074)	(0.0096)	(0.0058)	(0.0079)
× paved road	—	-0.0112	—	-0.0157***
		(0.0071)		(0.0052)
× police post	—	-0.0044	—	-0.0028
		(0.0103)		(0.0085)
× distance to district capital	—	0.0027	—	-0.0023
		(0.0038)		(0.0032)
Observations	232,778	210,294	232,778	210,294

The moderator is an indicator for mobile-phone signal recorded in the 2005 PODES wave, fixed before any identified cohort is treated and held by 76.4 percent of villages. Because it is pre-determined, the test does not require a parallel trend in the moderator itself. All columns absorb village fixed effects and cluster standard errors at the village. Panel A reports the effect of first exposure among villages without 2005 signal, its interaction with signal, and the implied effect among villages with signal. Panel B re-estimates the province-by-wave columns adding interactions of first exposure with three other 2005 inputs: an indicator for a paved road, an indicator for a village police post, and distance to the district capital standardised to mean zero and unit standard deviation. The signal indicator itself is absorbed by the village fixed effects. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

conventional levels on *warga* (1.3 points, $p = 0.11$). Meanwhile the paved-road interaction independently predicts *warga* (1.6 points, $p = 0.003$). More than one input is doing work, and signal is not cleanly separable from road access. Neither a police post nor distance to the district capital moderates anything. Supplementary Material Section S17 reports the moderators the design cannot cleanly separate.

The same inputs as outcomes, and why they do not corroborate. Treating the same inputs as outcomes in their own right is a second causal claim, and we report the result rather than only the verdict. Estimating each input on all seven waves under the preferred specification, one input passes its own pre-exposure test, *gotong royong* participation, and it is a ceiling variable held by 97 percent of villages returning a precise null. Every input with a detectable effect fails that test, and the inputs that move most move *against* the account we favour: places of worship and the *linmas* count rise, migrant workers fall, and theft rises by more where no security post existed to lose. The *linmas* rise is the one we can partly place. That count is a registered civil-protection roster, mobilised administratively after a disaster, so it tracks the state's response rather than the nightly rotation our account is about. The reading is consistent with the coefficient but not established by it, and a roster that grows while theft rises is not evidence for us. The security-post result we cannot place at all: if lapsed guardianship drove theft, the rise should concentrate where there was a post to lose, and it does not. One null is precise rather than uninformative. First exposure does not produce a disaster early-warning system: 0.9 percentage points with a standard error of 1.1, against a base of 11.6 percent, and a pre-trend that passes ($p = 0.45$). The competing state-capacity reading is therefore not driving the result. The conclusion has to be stated plainly. The direct input measures do not corroborate the capacity reading; what carries our evidence is the pattern across the four acts and the pre-determined moderator. A reader who requires direct measurement of the channel should treat this paper as not providing it (Supplementary Material Section S4, Table S3). Supplementary Material Section S12 reports the moderator horse-race and the guardianship falsification in full. A PODES module on *gotong royong* corroborates the capacity reading on the three waves that carry it, with the power limits that implies, and heterogeneity by pre-treatment polarisation points the same way (Supplementary Material Section S4).

The contrast in event time. The falsification the first-time margin fails has to be put to the contrast this paper rests on. It does better, but not cleanly. Under the preferred effects, the three interior pre-exposure coefficients of the theft-minus-communal difference are 2.7 percentage points, then 0.9 and 0.1 points in the opposite direction, and their joint test does not reject ($p = 0.31$). The difference then rises to 2.5 percentage points in the exposure wave and 4.0 points afterward. The binned distant endpoint remains 4.5 points (Section 7.1), and that term suggests a pre-exposure level difference more than a trend. Supplementary Material Section S10 reports the full event-time path under both specifications and records how this table came to differ from earlier drafts.

8 A First-Time Margin This Design Cannot Identify

There is a cleaner-looking way to compare: hold the earthquake itself fixed. For each earthquake, villages shaken at MMI≥VI are compared with villages in the same footprint that felt sub-damaging shaking (mean MMI in $[4, 6)$). The design then identifies the *incremental* effect of crossing the damage threshold, and we take same-earthquake-and-same-province as its primary counterfactual for the reason set out in Supplementary Material Section S13.

Splitting that sample by pre-period conflict cuts against the average. Roughly nine in ten treated villages have no conflict recorded in *any* earlier observed wave. Among them, exposure is associated with a 0.76 percentage point rise in first recorded conflict ($p = 0.001$). The counterfactual onset rate among matched controls is 2.2 percent, so this is an increase of about a third. Meanwhile the aggregate local decline loads on the minority with a prior conflict history, where it reaches 6.2 points.

Taking it to the earthquake. Village clustering is not the right unit here either, so we re-estimate the margin earthquake by earthquake and resample whole earthquakes. The treated-village-weighted mean is +0.0080, with an earthquake bootstrap standard error of 0.0029 and a 95 percent interval of [+0.0018, +0.0133] that excludes zero ($p = 0.013$). No single earthquake carries it: leaving each out in turn moves the mean only within [+0.0061, +0.0089]. The equal-event mean, which gives a shock that treated twenty villages the same weight as one that treated two thousand, is larger at +0.0101 but not distinguishable from zero ($p = 0.24$, positive in 13 of 23 events). On this margin the weighted estimate therefore survives the unit of assignment where the level estimate did not, which is part of what made it tempting.

A falsification test this design does not pass. None of the placebos reported so far tests whether the onset comparison itself is confounded, and on the onset arm the pre-exposure outcome is zero by construction. So we run the same design inside the pre-exposure period, moving exposure back one wave: a village counts as previously peaceful if it reports no conflict through two waves before its earthquake, and the outcome is first recorded conflict in the last wave *before* the earthquake. If the design were sound, this coefficient would be zero. It is not. Villages about to be shaken are 1.5 percentage points more likely to report a first episode in the pre-exposure wave ($p = 0.004$), and the gap widens under the local restrictions (Supplementary Material Section S13, Table S19). These are two to three times the effect we estimate at exposure, and they are the sharpest limit on the onset result.

So we report the margin and do not carry it. Three things point the same way: the gap predates the earthquake, it does not widen when the earthquake arrives, and it is confined to villages on firm ground. Supplementary Material Section S13 gives the matched design, the classification rule behind "no prior conflict" and the differential-reversion caveat; the full battery is in Supplementary Material Section S14.

9 Discussion

Two questions remain: what the evidence supports, and where it stops. We take them in that order, beginning with the results that are genuinely sensitive, which are better set out together than scattered. HonestDiD binds hardest. Inference then depends on the outcome and the estimator. A province-level wild-cluster bootstrap gives the composite $p = 0.030$, but *warga* reaches only $p = 0.215$. On the composite, the level holds when earthquakes are weighted by the villages they treat ($p = 0.01$) though not under an equal-event mean ($p = 0.32$); on the definition-stable gate the same pair is $p = 0.046$ and $p = 0.78$ (Section 6.1). The onset margin fails the test that matters most: run inside the pre-exposure period, the design returns 1.5 percentage points where it should return zero. Supplementary Material Section S13 documents three further limits, and none of them rescues it.

Three weaknesses recur besides (Table S2). Conditional pre-trends pass only with province-by-wave effects. Treatment comes from 19 to 24 earthquakes, so village clustering overstates precision. And the definition-stable *warga* outcome moves in the same direction but reaches only 10 percent on the specification that passes the pre-trend test. Population displacement is real, but far too small to carry the result (Supplementary Material Section S4).

A fourth belongs with them, and it is the one we would put first if forced to rank. The treated-village-weighted level result is concentrated in three shocks, which supply 88.7 percent of it; without them the weighted mean is -0.24 percentage points and the equal-event mean of the remaining sixteen is positive (Section 6.1). We state plainly what that does and does not license. It does not license reading the number as the effect of a typical earthquake: it is not that, and the equal-event mean already said so. Neither, in our view, does it license the opposite conclusion that the result is an artifact of three places. Palu and Lombok were the most destructive events in the panel, and destruction, not shaking intensity, is what the capacity mechanism actually predicts, so these are the shocks where it should bite hardest. The honest position is that with three events we cannot separate a genuine effect concentrated in catastrophic shocks from reversion in three high-conflict footprints that happen to be the catastrophic ones, and that a design with more large shocks, or an outcome observed continuously rather than in three-year waves, is what would separate them.

So we do not rest the paper on the average level effect. The paper's claim is the composition of disorder: the relative movement of four acts, recorded by one reporter on one form, identified by a contrast taken within the village. That contrast is better behaved in the pre-exposure period than the first-time margin is, with a flat near pre-period, and under wave effects it survives clustering at the kabupaten rather than the village ($p = 0.008$). Under the preferred specification we have only village-clustered inference for it, and we state that gap rather than fill it. It is not clean either. A distant binned endpoint remains positive and significant, and we report it rather than drop it.

Finally, on scope. The estimand is local and intent-to-treat. The sample is restricted to villages inside Indonesia's officially mapped medium and high seismic-hazard zones. Within that sample, treated villages are more conflict-prone and faster-growing, reporting communal conflict at 6.2 percent before exposure against a pooled 2.6 percent. The estimate concerns first observed damage-level exposure among villages actually struck, in a seismically active setting. It does not imply the same effect for an arbitrary Indonesian village, or for settings with different conflict institutions.

The aggregate decline is a result and we defend it, with its limits stated by us. On the definition-stable measure it survives province-by-wave effects almost intact, and shows no statistically detectable differential pre-exposure trend. It also clears earthquake-level inference when shocks are weighted by the villages they treat. It does not clear an unweighted average across shocks. So what we estimate is the effect on the average exposed village. The composition result carries two boundaries of its own. Theft's own rise is not separately identified once province-by-wave effects are imposed, so the divergence is carried by the communal decline rather than by theft climbing, and nothing in the mechanism argument requires theft's level. At the shock level the divergence survives between earthquakes but not within a single footprint, where the design is too coarse. A first-time-conflict margin that these data suggest fails its own falsification, returning before exposure the gap it should return only after (Section 8).

Every sign we estimate is consistent with two consequences of one shock. On the reading we favour, a strong earthquake degrades the village's capacity for organised collective action, and that capacity was doing two jobs at once. It supplied the mobilisation communal violence requires. It also supplied the routine guardianship, the night-watch rotations, that deters property crime. Remove it and coordination-intensive violence falls while stealth theft does not follow it down, whereas acts that require an encounter fall or stay put: robbery declines under both, by 1.4 percentage points under wave effects and by about half that under the preferred specification, and assault does not move. That asymmetry is the part a rising-motive story alone does not deliver, since with encounters and guardianship unchanged a stronger motive would raise theft and robbery together.

A second, narrower piece of direct evidence points the same way, on the footing the data allow: high-involvement public-purpose *gotong royong* falls by 3.1 to 3.2 points while emergency mutual aid does not. We do not observe household income, so the theft rise remains consistent with lapsed guardianship, with economic stress, or with both, and the sharpest test our data permit of the guardianship reading fails (Section 7.2). What we can say is only this: no change in motive is needed to generate the pattern we find.

Two limits of scope remain. The identified cohorts enter treatment between

2008 and 2024, so villages already exposed before the panel opens are not estimated. And the crime comparison discriminates among mechanisms without decomposing the effect into income, guardianship, aid and mobilisation shares. Direct household evidence on welfare and labour-market margins would turn the motive side of the argument from consistency evidence into a measured channel.

The estimates come from rural and peri-urban villages in Indonesia's seismic hazard zones, a setting with dense village-level social organisation and a state that mobilises visibly after disasters. The production-function logic travels as a *prediction*; the net sign does not necessarily travel with it. Displacement into camps can concentrate rival groups. Weak or predatory state presence can favour violence. Under either condition the framework predicts more violence, as in the legitimacy-and-coordination channel through which earthquakes *raised* conflict in imperial China (Bai, 2023).

10 Conclusion

A strong earthquake does not simply raise or lower communal conflict in an Indonesian village. It moves several kinds of disorder at once, and the pattern of that movement is what these data identify.

Consider the estimates. On one common sample of village-waves, first exposure at MMI $\geq$ VI lowers reported communal conflict by 2.2 percentage points against a 2.6 percent base, and by 1.7 points once province-by-wave effects are imposed. The same specification estimated as a Poisson, which reports proportional rather than absolute responses, puts that at rather more than a quarter of the base. Reported theft does not fall with it. Robbery falls by less than communal violence does, and assault does not move. Under the preferred specification, theft and communal responses differ on the linear ($p = 0.003$) and proportional ($p < 0.001$) scales. Against definition-stable *warga*, equality rejects under wave effects but sits just outside 5 percent under preferred effects.

What makes those numbers mean something is how they were measured. All of it comes from inside the village, on one form, from one reporter. Whatever moves all four acts alike, the village's fixed characteristics and any common reporting disturbance, drops out. What does not drop out is a differential trend in theft *relative* to communal conflict. The design assumes its absence and cannot guarantee it. Under the preferred specification the near pre-exposure period is flat, the two closest leads are precise nulls, and the interior leads do not jointly reject. At distant horizons a level difference remains, but only the two latest cohorts are observed that far back, and one of them supplies most of that cell. The first-time margin in the same data fails a sharper version of that test, returning before exposure the gap it should return only after.

The contrast also discriminates. Consider each explanation in turn. A uniform reporting shock would move all four acts together. A general collapse of order would raise all four. A rise in the motive for predation would raise theft and robbery together. A loss of the organisational capacity that collective violence needs, and stealth theft does not, would produce exactly the pattern we observe. A pre-determined moderator agrees: the decline concentrates in villages that had mobile-phone signal at the 2005 baseline, an interaction worth 2.1 percentage points that needs no parallel trend in itself and survives the specification most other margins do not.

Two implications follow, and the design's limits keep both modest.

The first concerns the post-disaster window, when governments, security forces and aid agencies divide scarce resources between reconstruction and conflict prevention. The message is this: a quiet communal-conflict count in that window is not reassurance. The decline concentrates where conflict was already established, and it fits a loss of the capacity to organise better than an easing of grievances. Latent tensions could therefore resurface as external monitoring and reconstruction wind down, a dynamic our design does not test. The first-time margin yields no implication at all. Its design does return a positive coefficient, but the gap it reports is already present before the earthquake arrives, so it cannot tell a planner where a first episode becomes more likely.

The second is for research. An aggregate disaster-conflict estimate is an average over places with and without a conflict history, and its sign can move with that mixture. Separating onset from continuation may well be the right question for this literature. But our evidence shows how demanding that question is: the design that appears to answer it fails a falsification that the composition contrast survives, run on the same villages and the same waves.

CRediT author statement

Billy Thandy Tulungen: Writing – original draft, Writing – review & editing, Software, Formal analysis, Conceptualization, Validation, Project administration, Methodology, Data curation, Funding acquisition. **Muhammad Halley Yudhistira:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization. **Jahen Fachrul Rezki:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization.

Funding

This document is an output from the research initiative "Reducing Conflict and Improving Performance in the Economy" (ReCIPE), a programme funded by the Foreign, Commonwealth & Development Office (FCDO), contract reference ReCIPE_LOA_303020_Tulungen. The views expressed are not necessarily those of FCDO. FCDO had no role in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

Declarations

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Billy Thandy Tulungen is an employee of the Government of North Sulawesi Province, one of the administrative units covered by the village census data used here. The research question, the specification rule and the results were not reviewed or approved by that employer, and the employer had no role in study design, data analysis, or the decision to submit. This study uses secondary and publicly available data and de-identified restricted survey aggregates; it did not require ethics-board approval.

Data availability

This paper combines confidential BPS microdata (PODES) with public geophysical data (USGS ShakeMap and PVMBG seismic-hazard maps) and licensed conflict data (ACLED). The confidential microdata cannot be redistributed, and nothing containing the BPS village identifier is deposited. The replication package is therefore a partial, restricted reproduction: it provides in full the code that builds the analysis panel from source, so that a researcher with BPS access can rebuild the panel and reproduce every number in the paper. The package contains a complete data-availability and provenance statement, following the AEA Data Editor template, that classifies all thirteen sources by provider, licence, access route, and the code that consumes it.

Acknowledgements

We thank Rizki Nauli Siregar, Milda Irhamni, Yohanna M. L. Gultom, Chaikal Nuryakin, I Dewa Gede Karma W., and Saut Aritua H. A. Sagala for comments that improved the paper, and BPS-Statistics Indonesia for access to the Village Potential Census microdata and to the village-boundary products on which the spatial matching depends. All remaining errors are our own.

References

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

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

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

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

Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2026. "Difference-in-Differences Designs: A Practitioner's Guide." *Journal of Economic Literature* 64 (2):498–557.

Barker, Joshua. 1998. "State of Fear: Controlling the Criminal Contagion in Suharto's New Order." *Indonesia* 66:7–42.

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

Barron, Patrick and Joanne Sharpe. 2008. "Local Conflict in Post-Suharto Indonesia: Understanding Variations in Violence Levels and Forms Through Local Newspapers." *Journal of East Asian Studies* 8 (3):395–423.

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

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

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

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

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

Bignon, Vincent, Eve Caroli, and Roberto Galbiati. 2017. "Stealing to Survive? Crime and Income Shocks in Nineteenth Century France." *Economic Journal* 127 (599):19–49.

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

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

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

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

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

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. "Climate and Conflict." *Annual Review of Economics* 7:577–617.

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

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

Cohen, Lawrence E. and Marcus Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44 (4):588–608.

Cromwell, Paul, Roger Dunham, Ronald Akers, and Lonn Lanza-Kaduce. 1995. "Routine Activities and Social Control in the Aftermath of a Natural Catastrophe." *European Journal on Criminal Policy and Research* 3 (3):56–69.

de Chaisemartin, Clement and Xavier D'Haultfoeuille. 2026. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *Review of Economics and Statistics* 108 (4):863–880.

Di Tella, Rafael and Ernesto Schargrodsky. 2004. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack." *American Economic Review* 94 (1):115–133.

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

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

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2020. "Social Media and Protest Participation: Evidence from Russia." *Econometrica* 88 (4):1479–1514.

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

García-Hombrados, Jorge. 2020. "The Lasting Effects of Natural Disasters on

Property Crime: Evidence from the 2010 Chilean Earthquake." *Journal of Economic Behavior & Organization* 175:114–154.

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

Hardiman, Allen. 2025. "The Impact of Rainfall-Induced Income Shocks on Crime: Evidence from Indonesia." *Journal of International Development* 37 (5):1104–1115.

Imbens, Guido W. and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 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. 2020. "Development of the 2017 National Seismic Hazard Maps of Indonesia." *Earthquake Spectra* 36 (S1):112–136.

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

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

Manacorda, Marco and Andrea Tesei. 2020. "Liberation Technology: Mobile Phones and Political Mobilization in Africa." *Econometrica* 88 (2):533–567.

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

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

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

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

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

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

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

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235 (2):2218–2244.

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

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

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

Tajima, Yuhki. 2014. *The Institutional Origins of Communal Violence: Indonesia's Transition from Authoritarian Rule*. Cambridge: Cambridge University Press.

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

Worden, C. B., E. M. Thompson, J. W. Baker, B. A. Bradley, N. Luco, and D. J. Wald. 2018. “Spatial and Spectral Interpolation of Ground-Motion Intensity Measure Observations.” *Bulletin of the Seismological Society of America* 108 (2):866–875.

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

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

Appendix

Earthquakes and Communal Conflict: Evidence from Indonesia's Seismic Zones

Supplementary data

Supplementary material related to this article, comprising the online Supplementary Material document cited throughout the text, can be found in the online version of this article and in the replication package. The publisher inserts the DOI link at typesetting.

A Pre-Trend Diagnostics Across Estimators

This appendix collects the pre-trend evidence that Section 5 summarises: what bounds the rejection under wave fixed effects, the power of the pre-test itself, and how the verdict moves across estimator families.

What bounds the wave-effects rejection

Two things bound what that rejection can imply. It is specific to wave fixed effects, though the comparison cannot use the identical statistic: `csdid` cannot absorb interacted fixed effects, so the province-by-wave evidence comes from the joint test of the event-study leads, which does not reject ($p = 0.68$ composite, $p = 0.66$ *warga*), and the saturated Sun–Abraham event study, which can absorb them, agrees ($\chi^2(5) = 4.67$, $p = 0.46$). And the drift the data do exhibit runs *against* the finding rather than toward it: the unconditional group-time average pre-trend is positive ($+0.0107$, $p = 0.01$), as is that of the specification least exposed to bad-control concerns ($+0.0151$, $p < 0.01$), and a positive pre-existing drift biases a subsequent decline toward zero instead of manufacturing it. That argument holds against a drift that *continues* into the post-treatment period, not against a transitory elevation that *reverts* by the time of exposure; conditioning on a passed pre-test is itself a selection step, and it displaces the wave-of-exposure estimate by about 0.004, again toward zero.

Pre-test power

A test that passes is worth only as much as its power, and a weak test always passes (Roth, 2022). We therefore ask directly how large a differential trend this pre-test would miss, on the specification we actually use rather than on the wave-effects panel it rejects. The calculation refers to the pre-test the `pretrends` package implements, which flags a violation when any individual pre-treatment coefficient is significant, and not to the joint Wald test reported above. Table A.1 reports it for the three communal outcomes, and Figure A.1 shows what that means for the two that matter.

The answer differs sharply by outcome, and it lines up with the choice Section 6.1 makes. On the gate, a linear trend of half a percentage point per wave escapes the test half the time, and a trend of that size needs about three and a half waves, which is the whole observed pre-period, to accumulate to the estimated effect. The composite behaves the same way. Passing there is therefore informative, though not decisive. On *warga* it is close to uninformative: the trend the test misses half the time accumulates to the whole estimated effect in under a wave and a half. 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 composition contrast of Section 7.1, which rests on a different assumption rather than on none, and which we state there.

The likelihood ratios in the same table sharpen that comparison and then qualify it. Under the linear trend the coefficients we observe are three times more likely under parallel trends than under the violation on the composite, and twice as likely on the gate; on *warga* they are two and a half times more likely under the *violation*. That is a stronger statement against reading *warga*'s passing pre-test as evidence than the ratio of magnitudes alone delivers. The qualification is that this vulnerability belongs to a straight line. The competing account this appendix names is mean reversion, which is not linear, so we repeat the exercise for a convex violation calibrated to the same power. That reverses the ranking: the observed coefficients are then far more likely under parallel trends on all three outcomes, and on *warga* most of all. What the *warga* pre-test cannot rule out is a trend that runs straight through the pre-period, not one that accelerates.

Figure A.1 puts the same calculation in event time, and it shows one thing the ratios cannot. The path the coefficients would be expected to trace under that hypothesised trend, *conditional on the pre-test passing*, lies almost on top of the trend itself, never separating from it by more than a quarter of a percentage point on either outcome. Conditioning on a passed pre-test therefore removes very little of the violation the test was meant to catch, and that is the precise sense in which a passing test is weak evidence rather than none. What separates the two outcomes is how far the coefficients we actually observe fall below that expected path once exposure arrives: on the gate the gap is more than twice the one on *warga*, in the wave of exposure and in the wave after alike. That is the same contrast the ratios state, read off the event-time path rather than off a single summary.

One direction is worth recording alongside this. A pre-trend large enough to survive the test would displace the estimate *toward* zero rather than away from it, so it would understate rather than manufacture the communal decline, the same direction as the growth imbalance in Supplementary Material Section S8.

The pre-trend verdict across estimator families

The pre-trend verdict turns on the fixed-effect structure rather than on the estimator, and we state the pattern in one place because it is the evidence behind the specification choice of Section 5.1. Under wave fixed effects the joint pre-treatment test rejects in every estimator family we run: Callaway–Sant'Anna ($\chi^2(15) = 37.5$, $p = 0.001$ composite; $\chi^2(15) = 28.8$, $p = 0.017$ *warga*; Table A.2 reports every covariate set, and the rejection weakens to marginal only once province and region effects enter), the imputation estimator of Borusyak, Jaravel, and Spiess (2024) (a lead two waves before exposure of 2.3 percentage points, $p = 0.04$), a Sun–Abraham lead two waves before exposure significant at 5 percent, and a nearest placebo

Table A.1: What the Pre-Trend Test Would Miss

Takeaway: on the gate and the composite a trend big enough to explain the effect would have to run the length of the pre-period to escape the test; on warga *it would not, which is why passing there carries little weight.*

Outcome	Trend detected		Effect	Waves to	Likelihood ratio	
	half	80%			linear	quadratic
	(points per wave)		(points)	match it		
Subtype composite	0.51	0.78	1.77	3.5	0.33	0.11
Mass-brawl gate	0.54	0.82	1.83	3.4	0.50	0.07
Warga subtype	0.41	0.62	0.56	1.4	2.41	0.01

Notes: Power of the pre-trend test under the preferred specification, computed with the `pretrends` package of Roth (2022) on the three interior pre-exposure leads. Columns two and three give the linear differential trend, in percentage points per wave, that the test would reject with probability 0.5 and 0.8. The fifth column divides the exposure-wave effect by the trend detected half the time: it is the number of waves such a trend would need in order to accumulate to the estimated effect, so a small number means passing the pre-test carries little information. The likelihood ratio is the likelihood of the observed coefficients under the hypothesised trend relative to under parallel trends, so a value below one favours parallel trends and a value above one favours the trend. The quadratic column repeats the exercise for a convex violation calibrated to the same power, because mean reversion, the competing account this appendix names, is not linear. The Bayes factor the package also returns is omitted: at a fixed target power it is constant across outcomes by construction and carries no information here. The binned endpoint is excluded because it occupies no single position in event time.

of 1.0 point under the episodic estimator of de Chaisemartin and D'Haultfoeuille (2026).

Under province-by-wave effects it passes in every family that can absorb them: the dynamic specification ($p = 0.68$ composite, $p = 0.66$ *warga*), the saturated Sun–Abraham event study of Figure 1 ($\chi^2(5) = 4.67$, $p = 0.46$), and the imputation estimator, whose three leads are individually insignificant there. That the verdict flips with the fixed effects and not with the estimator is what the economic argument of Section 5.1 predicts, and it is a stronger statement than the diagnostic alone could support: the specification is not merely the one that passes, it is the one that passes across estimators that make different assumptions about treatment-effect heterogeneity. What this does not buy is precision. The imputation estimator that clears the pre-test under province-by-wave effects cannot estimate the effect there (+0.013, standard error 0.031), because village and province-by-wave effects together absorb nearly all the variation its imputation step uses. The point

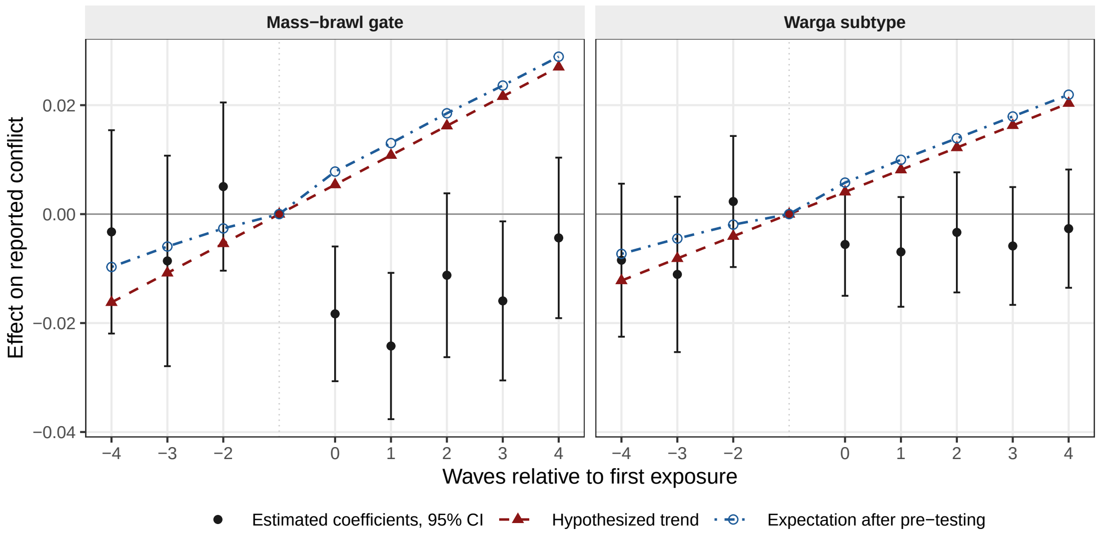

Figure A.1: What a Passing Pre-Test Still Permits
Both outcomes fall below the trend the test would miss, but the gate falls more than twice as far, which is the contrast Table A.1 states in numbers.
Note: Three series, following Roth (2022). Filled circles are the event-study coefficients on the preferred specification, with 95 percent village-clustered intervals; the wave before first exposure is the omitted category. The dashed line is the linear differential trend that the pre-test would reject only half the time, computed separately for each outcome. The dot-dashed line is what the coefficients would be expected to look like under that trend *conditional on the pre-test passing*; it sits almost on top of the trend itself, which is the point of the exercise, because conditioning on a passed pre-test removes very little of the trend it was supposed to detect. The reading is the vertical distance between the observed points after exposure and those two lines: 2.6 percentage points on the gate in the wave of exposure and 3.7 in the wave after, against 1.1 and 1.7 on *warga*. The binned endpoint is excluded because it occupies no single position in event time.

Table A.2: The Pre-Trend Verdict Under Callaway–Sant'Anna, by Covariate Set
Takeaway: on the wave-effects panel the pre-trend test rejects for both outcomes under every covariate set, decisively without regional controls and marginally once province and region effects enter.

		Subtype composite		*Warga* subtype	
Covariates in the CS(2021) estimator	df	χ^2	p	χ^2	p
No covariates	15	51.4	< 0.001	35.0	0.003
Baseline characteristics	15	42.9	< 0.001	—	—
+ province dummies	10	50.0	< 0.001	—	—
+ region effects	15	37.5	0.001	28.8	0.017
+ province and region	15	24.7	0.053	25.5	0.044

Notes: Joint Wald test that all pre-treatment group-time coefficients are zero, under the Callaway and Sant'Anna (2021) estimator on the wave-fixed-effects panel, for each set of covariates the project estimates. Baseline characteristics are the seven PODES 2003 items. The χ^2 statistics are *not* comparable across rows because the province-dummy row rests on ten degrees of freedom rather than fifteen; the comparable column is p. The composite is estimated without a *warga* companion in the two province-dummy specifications, and those cells are left empty rather than filled from another specification. The verdict is specification-dependent: the rejection is decisive without regional controls and marginal once province and region effects enter, which is the pattern the specification argument of Section 5.1 predicts.

estimates converge and the diagnostics agree on the specification; the cost is that the design has less power the more of the regional counterfactual we absorb, and we report the bounds of Section 6.2 rather than read any single pre-test as settling the matter.

Exposure timing and the wave-boundary artifact

Two timing points that Sections 1 and 5 state in one sentence each are set out here.

The calendar-year artifact. We document a pitfall for work that merges the timing of physical shocks onto periodic surveys. In panels with unequally spaced waves, calendar-year exposure lags misaligned with the enumeration month both conceal the true response and manufacture a spurious one: on an identical sample, dating exposure by calendar year attenuates the two-wave effect from -0.022 to -0.003 and renders it insignificant, while producing a significant coefficient one lag later that the aligned specification shows to be zero (Table A.3). We also provide the first village-level quasi-experimental estimate of the earthquake–conflict relationship for Indonesia.

Resolving a pre-period signal. An earlier calendar-year coding of first exposure produced a marginally significant pre-period coefficient; diagnostics traced it to wave-boundary timing that misplaced onset by one wave for about 21 percent of ever-treated villages, and re-timing exposure to the first wave enumerated after the event removes it, as the paragraph below sets out. The joint Wald rejection under wave fixed effects remains and we carry it as an open caveat; the HonestDiD bounds (Section 6.2), not the pre-trend tests, are the binding constraint.

Table A.3: Exposure Timing and the Distributed Lag: Calendar Year versus Survey Window
Dating exposure by calendar year rather than by the enumeration window both conceals the true response and manufactures a spurious one.

	Composite		*Warga*	
Exposure lag	Calendar-year	Survey-aligned	Calendar-year	Survey-aligned
Exposure, one wave before	−0.0017	−0.0032	−0.0019	−0.0023
	(0.0018)	(0.0027)	(0.0015)	(0.0022)
Exposure, two waves before	−0.0028	-0.0218^{***}	−0.0019	-0.0074^{***}
	(0.0025)	(0.0034)	(0.0021)	(0.0028)
Exposure, three waves before	-0.0063^{***}	−0.0000	−0.0016	+0.0011
	(0.0020)	(0.0017)	(0.0015)	(0.0013)
Observations	285,263	285,263	285,263	285,263

Each column is a distributed-lag two-way fixed-effects regression with village and wave fixed effects and standard errors clustered at the village level. The calendar-year columns date exposure by the calendar year of the earthquake; the survey-aligned columns date it by the May-to-May window that matches the PODES enumeration month, so that each lag falls entirely before the twelve-month conflict reference window. Both timings are estimated on the identical sample, so the columns differ only in how exposure is dated. Misalignment does not reverse any sign; it attenuates the response and displaces it by one lag. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

B Data and Exposure Detail

This appendix carries the measurement detail that Section 4 states in brief: the incidence of the four acts by wave, the column conventions of the crime module, why exposure is a threshold rather than a dose, and the validation of the pre-2013 ShakeMap reconstructions against the village's own earthquake report.

Incidence of the four acts by wave

Table B.1 reports incidence by wave. Theft is common, at roughly two villages in five, while robbery and assault are rare, at three and five percent, and communal conflict rarer still, at between two and three percent. The gap in base rates is the reason Section 7.1 reports the composition contrast on a proportional scale as well as a linear one: a shock that moved every act by the same proportion would move theft by far more percentage points than communal conflict, so a linear test of equal effects is not by itself evidence against a common proportional disturbance. All four series decline after 2018, which the wave effects absorb.

Table B.1: Incidence of the Four Acts of Disorder, by Wave
All four are recorded by the same official, on the same form, over the same recall window; theft is an order of magnitude more common than the rest.

Wave	Communal conflict	*Warga* subtype	Theft	Robbery	Assault
2005	0.024	0.018	0.425	0.035	0.054
2008	0.027	0.016	0.474	0.037	0.074
2011	0.029	0.017	0.402	0.032	0.062
2014	0.032	0.018	0.465	0.041	0.059
2018	0.032	0.021	0.505	0.038	0.073
2021	0.017	0.011	0.306	0.022	0.033
2024	0.019	0.012	0.353	0.012	0.034
All waves	0.026	0.016	0.418	0.031	0.055

Share of village-waves reporting each act in the preceding twelve months, on the composition common sample: village-waves inside the medium and high seismic-hazard zones for which all five outcomes are non-missing. Communal conflict is the composite mass-brawl measure; *warga* is the civilian-versus-civilian subtype whose wording is near-stable across waves; theft, robbery and assault are items 01, 02 and 04 of the same security module. N = 300,018 village-waves, 43351 villages.

Village boundary harmonisation

The 2005-boundary panel maps every village in each wave back to its 2005 parent village. It is built from BPS Master File Desa records and supplemented by GPS spatial joins and manual verification where name matching fails. Because villages created by later pemekaran are mapped back to their 2005 parent polygon, administrative splits cannot mechanically enter or exit the sample. Supplementary Material Section S6 documents the procedure, match rates by wave, and two crosswalk-robustness re-estimations.

Column conventions in the crime module

The questionnaire item numbers of the crime module are stable across the seven waves, but the column conventions in the released microdata are not. Two waves interleave the occurrence column with a follow-up asking whether the act became more or less frequent than the year before, so the position of an item in the data file is not its position on the form. Supplementary Material Section S6 gives the exact variable name used for each outcome in each wave, together with the cross-tabulation against the follow-up item that verifies each mapping.

Why a threshold and not a dose

Modified Mercalli Intensity is an ordinal scale, and that governs what can be done with it. Its categories are defined by what people feel and what buildings do, not by a measured quantity, and the instrumental relations that map ground motion onto it are logarithmic: Wald et al. (1999) estimate $\mathrm{MMI} = 3.66 \log_{10}(\mathrm{PGA}) - 1.66$ over $\mathrm{V} \leq \mathrm{MMI} \leq \mathrm{VIII}$, so one step on the scale corresponds to a factor of about 1.9 in peak ground acceleration rather than to a fixed increment. The scale does not even rest on one physical quantity throughout its range: below about MMI VII it tracks acceleration and felt reports, above it velocity and damage, which is why Wald et al. (1999) recommend different relations on either side.

A coefficient on MMI entered linearly would therefore assert that a move from IV to V means the same thing as a move from VIII to IX, which is false in ground motion and unsupported in damage. We do not make that assertion. Treatment is a threshold, $\mathbf{1}\{\mathrm{MMI} \geq \mathrm{VI}\}$, which is an operation an ordinal scale admits, and the intensity contrasts we report compare bands within the same earthquake rather than fitting a slope in MMI.

One step in our construction is nonetheless cardinal, and we flag it rather than let it pass. Averaging intensity across the raster cells inside a village treats MMI as though differences were comparable, which by the logarithmic relation above they are not: the average of MMI V and MMI VII understates the intensity implied by the average acceleration. Three things bound the consequence. The bias grows with within-village dispersion, which is small for the village polygons here; the estimate keeps its sign under the pixel maximum, the median and a representative point as well as the mean; and an exposure rebuilt from Peak Ground Acceleration, which is cardinal, reproduces the result at -0.0129 (standard error 0.0039).

Validating the reconstructed ShakeMap products

Section 4 notes that USGS ShakeMap products for Indonesian events before 2013 are later reconstructions built from modelled ground-motion relations rather than recorded intensities. The usual response is to concede the point and restrict the sample. We can do better, because the village census contains an independent check that we had not previously used. From 2008 onward PODES asks each village, in its disaster block, whether an earthquake occurred there over the recall window. That is the village head's own report, collected by a different agency for a different purpose, and it is available for the same village-waves as the outcome.

Table B.2 asks whether the ShakeMap treatment predicts it. It does, strongly and in both eras. Villages whose ShakeMap intensity reaches MMI $\geq$ VI within the window are 34 percentage points more likely to report an earthquake themselves (standard error 0.7), against a base of 7 to 13 percent. Splitting at the provenance break, the relationship is +0.302 (standard error 0.011) in the reconstructed era and +0.398 (0.011) in the instrumental era. The difference is statistically clear, and its size and direction are what greater measurement error would produce: the correspondence is weaker in the reconstructed era by roughly a quarter.

Table B.2: Does the ShakeMap Treatment Predict the Village's Own Report?
An independent check on the pre-2013 reconstructed products, from the village census disaster block.

	Reconstructed 2008–2014	Instrumental 2018–2024	All waves
Panel A. Raw agreement			
P(village reports a quake | ShakeMap MMI$\geq$VI)	0.422	0.517	
P(village reports a quake | no ShakeMap MMI$\geq$VI)	0.069	0.126	
difference	0.352	0.391	
Panel B. Village and wave fixed effects			
ShakeMap MMI$\geq$VI in the window	0.3019	0.3979	0.3411
	(0.0107)	(0.0114)	(0.0072)
Observations	128,658	130,046	258,904
p, equality of the two eras		0.000	

The outcome is the village's own report, in the PODES disaster block, that an earthquake occurred over the recall window; it is collected independently of USGS. The regressor is episodic ShakeMap exposure at MMI$\geq$VI within the same window rather than the absorbing indicator, because the survey item asks about the window and not about history. USGS ShakeMap products for events before 2013 in this catalogue are model reconstructions built years afterwards from no station data; from 2013 they are instrumental. Columns one and two split the panel at that break. A relationship of similar size on both sides is evidence that the reconstructed products carry real information about which villages were shaken; a much weaker one on the left would quantify how much the early treatment is attenuated. Standard errors clustered at the village.

C The Level Estimate in Full

This appendix carries the detail of the average effect that Section 6.1 states in brief.

Table C.2 reports the specification selected by the rule of Section 5.1: village and province × wave fixed effects, always-treated cohort excluded. First exposure to damage-level shaking lowers the probability that a village reports communal conflict by 1.6 percentage points against a pre-exposure mean of 6.2 percent, a reduction of roughly a quarter. The estimate is 1.5 points without covariates and 1.6 with them, so it does not rest on the control set.

It does not rest on *how* the covariates enter either, which is a separate question and one the design can answer directly. Covariates can be brought into a DiD estimate in three ways, and each is biased under a different failure: an outcome-regression adjustment fails if the model for untreated outcome trends is misspecified, inverse probability weighting fails if the propensity-score model is, and the doubly robust combination is consistent when either one is right (Baker et al., 2026). Table C.1 reports all three on the same specification. They agree to the fourth decimal, at −0.0176, −0.0177 and −0.0177 for the composite and −0.0074, −0.0075 and −0.0075 for *warga*, so no working model is carrying the result.

The event-study rows show where the effect sits in time. It is present in the wave of exposure (−0.0194) and the wave after (−0.0227), and the three subsequent lags are jointly indistinguishable from zero ($p = 0.38$). The wave of exposure is the first PODES enumeration after the shock, which the survey calendar can place anywhere from a few months to three years later, so t_0 is a first observed post-exposure effect and not a short-run one; we use that language throughout rather than calling it immediate. Because the absorbing indicator averages over later waves in which villages have already recovered, the pooled estimate is best read as smaller than the wave-of-exposure response.

Both outcomes pass the joint pre-trend test in this specification ($p = 0.68$ composite, $p = 0.66$ *warga*), as the economic reasoning of Section 5.1 anticipates.

Table C.1: Covariates Entered Three Ways

Takeaway: regression adjustment, inverse probability weighting and the doubly robust combination give the same estimate, so it does not depend on which working model is assumed correct.

	Regression adjustment	Inverse probability weighting	Doubly robust
Any communal conflict	-0.0176^{***}	-0.0177^{***}	-0.0177^{***}
	(0.0046)	(0.0047)	(0.0047)
Warga conflict	-0.0074^{**}	-0.0075^{**}	-0.0075^{**}
	(0.0037)	(0.0037)	(0.0037)
Observations		225,372	

Notes: Callaway–Sant'Anna aggregate ATT on the preferred panel, not-yet-treated controls, `long2` base period, baseline covariates fixed at the 2005 wave. The three columns differ only in how those covariates enter. Regression adjustment models the untreated outcome trend given covariates and is biased if that working model is wrong; inverse probability weighting models the propensity score instead and is biased if *that* model is wrong; the doubly robust estimator combines the two and is consistent if either is correct. Reporting all three follows Baker et al. (2026): agreement across the columns means the estimate does not rest on any one working model. Standard errors are analytical, based on the Callaway–Sant'Anna influence function and clustered at the village. * p<0.1, ** p<0.05, *** p<0.01.

The definition-stable *warga* outcome moves in the same direction throughout but does not attain conventional significance here. Its pre-exposure mean is 3.6 percent against 6.1 for the composite, so it carries roughly half the power; we read it as corroborating in sign rather than as independent confirmation.

Because treated villages enter the window from elevated conflict levels, the obvious competing account is mean reversion, and it is worth stating the floor it leaves. Letting villages with different fixed 2005-baseline conflict follow their own wave trends, which is precisely the design that allows conflict-prone villages to revert along their own trajectory, still leaves the effect at -0.0131 (standard error 0.0028), and at -0.0140 (0.0035) when province-by-wave effects are added as well. Reversion can absorb part of the decline but not the whole of it. What it does bound is where the decline lives: the effect loads on the fewer than one in ten villages with a prior conflict episode, so the average decline is concentrated where conflict had already occurred; it does not describe prevention of first-time conflict.

Three further comparisons bear on this estimate and are reported in the Supplementary Material rather than here. The Callaway–Sant'Anna benchmark reproduces the sign and broad magnitude on both outcomes. Absorbing province trends as fixed effects and conditioning on them inside the estimator's nuisance models give different answers, and the two routes are set out side by side. And the event study places the response in the wave of exposure and the wave after, with later horizons imprecise; its pre-treatment coefficients are the object the pre-trend tests interrogate.

Table C.2: Preferred Specification: Province × Wave Fixed Effects, Always-Treated Cohort Excluded

	Any communal conflict	*Warga* (definition-stable)
Absorbing exposure, no covariates	-0.0154^{***}	-0.0035
	(0.0040)	(0.0032)
Absorbing exposure, baseline covariates	-0.0165^{***}	-0.0042
	(0.0041)	(0.0033)
Event study: wave of exposure	-0.0194^{***}	-0.0070
	(0.0063)	(0.0050)
Event study: one wave after	-0.0227^{***}	-0.0064
	(0.0069)	(0.0054)
Pre-trend joint p (leads $-4, -3, -2$)	0.678	0.660
Pre-exposure mean	0.0624	0.0372
Effect as % of pre-exposure mean	-26.4	-11.3
Observations, covariate rows	225,372	225,372
Observations, no-covariate row	232,778	232,778

All columns include village and province × wave fixed effects, with the always-treated cohort excluded, so no already-treated village serves as a control. Standard errors clustered at the village level. Covariates are PODES 2003 baseline characteristics interacted with wave dummies. The two event-study rows come from a single regression that also includes leads at -4, -3 and -2 and lags at $+2$, $+3$ and $+4$; the latter are jointly indistinguishable from zero (composite $p = 0.38$), so the effect is confined to the two waves following exposure. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

D Pre-Treatment Balance

How comparable were eventually-treated and never-treated villages before any treatment occurred? Table D.1 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 D.1: Pre-Treatment Balance at the 2005 Wave: Eventually-Treated vs. Never-Treated Villages, KRB Medium/High

	Eventually treated	Never treated	Norm. diff.
Communal conflict (binary)	0.040	0.026	0.081
Population	4308.494	3655.616	0.148
Number of households	1071.929	920.837	0.139
Paved road access	0.669	0.604	0.136
Household electricity	0.947	0.960	-0.061
Distance to district capital (km)	26.350	29.344	-0.126
Number of primary schools	3.374	2.672	0.246
Police post present	0.175	0.114	0.176
Number of health centers	0.757	0.456	0.423
Internet available	0.040	0.035	0.025

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 = 2995); never treated: villages in the same KRB Medium/High hazard zones with no MMI≥VI exposure through 2024 (n = 30259). 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 whose 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.

Nine of the ten normalised differences (Imbens and Wooldridge, 2009) fall below the conventional 0.25 threshold. The exception is the number of health centres, at 0.423, which we flag rather than average away; the next largest are the primary-school count (0.246) and the police post (0.176). Eventually treated villages are larger than never-treated ones (4,308 against 3,656 residents, normalised difference 0.148) and better provisioned on several margins, which is what one would expect if seismically active districts differ systematically from the rest of the country. Baseline conflict differs by 1.4 percentage points (0.040 against 0.026, normalised difference 0.081).

Because the health-centre difference sits above the conventional threshold, we ask directly whether it is load-bearing. Adding the baseline health-centre count interacted with wave moves the estimate by less than one percent on every outcome, and trimming villages in the top decile of baseline health-centre provision leaves every estimate significant, with a largest movement of 10 percent on inter-village conflict. The imbalance is visible but it is not doing the work: remaining level differences are absorbed by village fixed effects, and the pre-trend tests of Section 6.2 address differential trends.

E HonestDiD Sensitivity Bounds

The paper states in Section 6.2 that the relative-magnitudes bound, not the pre-trend test, is the constraint that actually binds. This appendix reports that bound for the specification the paper prefers, and then for the Sun–Abraham event study alongside it.

Table E.1: HonestDiD Under the Preferred Specification
The composite sign survives post-treatment violations of a quarter of the largest pre-treatment deviation, and not of a half.

$\bar{M}$	Lower	Upper	Excludes zero
Original	−0.029	−0.006	Yes
0.00	−0.029	−0.006	Yes
0.25	−0.032	−0.004	Yes
0.50	−0.037	0.000	No
0.75	−0.042	0.006	No
1.00	−0.048	0.012	No

Relative-magnitudes identified sets (Rambachan and Roth, 2023) for the preferred specification of Section 6.1. $\bar{M}$ is the permitted post-treatment violation as a multiple of the largest deviation between consecutive pre-treatment leads. In this specification that yardstick is 0.0118, which is 0.67 times the effect being defended; under the wave-fixed-effects specification it was 0.037, or 2.2 times the effect, which is why the breakdown point moves. The identified set excludes zero through $\bar{M} = 0.25$ and admits zero at $\bar{M} = 0.50$.

Table E.1 gives the identified set under $\Delta^{RM}(\bar{M})$ for the composite outcome on the preferred village and province-by-wave specification. It excludes zero at $\bar{M} = 0.25$ ($[-0.032, -0.004]$) and admits it at $\bar{M} = 0.50$ ($[-0.037, 0.000]$). That is the number quoted in Section 6.2, and it is the binding constraint on the level result: a post-treatment violation of parallel trends half as large as the largest pre-treatment deviation is enough to admit a zero effect.

Table E.2 reports the same exercise for the mass-brawl gate, the definition-stable outcome the design intends and the one the level result is stated on. Its first post-exposure effect excludes zero at $\bar{M} = 0.25$ ($[-0.034, -0.003]$) and admits it at

$\bar{M} = 0.50$ ($[-0.039, +0.001]$), so the gate is no more robust to relative-magnitude violations than the subtype composite: definitional stability and near-immunity to province trends do not buy a weaker parallel-trends requirement. Aggregating over the five post-exposure periods with uniform weights is more demanding still, admitting zero already at $\bar{M} = 0.25$.

Table E.2: HonestDiD for the Mass-Brawl Gate
The headline outcome is no more robust to relative-magnitude violations than the subtype composite.

	First post-exposure effect			Aggregated ATT		
$\bar{M}$	Lower	Upper	Excl. 0	Lower	Upper	Excl. 0
Original	-0.031	-0.006	Yes	-0.026	-0.004	Yes
0.00	-0.031	-0.006	Yes	-0.026	-0.004	Yes
0.25	-0.034	-0.003	Yes	-0.041	0.011	No
0.50	-0.039	0.001	No	-0.062	0.032	No
0.75	-0.045	0.007	No	-0.084	0.054	No
1.00	-0.052	0.014	No	-0.106	0.076	No

Relative-magnitudes identified sets (Rambachan and Roth, 2023) for the mass-brawl gate, the definition-stable outcome the design intends, on the preferred village and province-by-wave specification. $\bar{M}$ is the permitted post-treatment violation as a multiple of the largest deviation between consecutive pre-treatment leads. The left block targets the first post-exposure effect; the right block targets the ATT aggregated over the five post-exposure periods with uniform weights. Read with Table E.1, which reports the same exercise for the subtype composite.

The remainder of this appendix reports the same exercise on the Sun–Abraham interaction-weighted event study, which is a different estimator with a different target parameter, so its $\bar{M}$ is not comparable with the one above.

Table E.3 reports the Rambachan and Roth (2023) HonestDiD sensitivity analysis applied to the Sun–Abraham interaction-weighted event study. We report it because its leads are tighter than the pooled estimators', not because they are clean: the composite t_{-2} lead is $+0.0138$ and significant at 5 percent, which is exactly the violation the sensitivity analysis is meant to price. The target is the first post-exposure effect t_0 (-0.0223^{***} for the composite outcome, -0.0096^{*} 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.

On a grid fine enough to locate it, the composite breakdown value lies between $\bar{M} = 0.40$, where the set is $[-0.040,\ -0.000]$, and $\bar{M} = 0.45$, where it is $[-0.042,\ +0.001]$: the sign survives post-treatment violations up to about two fifths of the largest observed pre-treatment deviation, not up to half. The definition-stable *warga* target is far more fragile, breaking down between $\bar{M} = 0.05$ and $\bar{M} = 0.10$, consistent with its weaker precision; reporting only that it admits zero by $\bar{M} = 0.5$ would understate how early it fails. These are moderate bounds, not 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. 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.

A second safeguard addresses a different worry. Because the never-treated pool is large and disproportionately Javanese, one might suspect the level decline reflects that pool's own trajectory rather than the effect of exposure. Dropping the never-treated villages entirely, so that identification comes only from differences in exposure timing among the 2,995 villages that are eventually treated, strengthens rather than weakens the estimate: the composite ATT is -0.0390 (standard error 0.0096) with an average pre-treatment coefficient of $+0.0025$ ($p = 0.61$), and the definition-stable *warga* outcome, which does not attain conventional significance on the level anywhere else in the paper, reaches -0.0196 (standard error 0.0082, $p = 0.017$) with an average pre-treatment coefficient of -0.0033 ($p = 0.45$). We do not promote this to the lead result, because the not-yet-treated estimand answers a narrower question than the one the paper asks, and because the composite's individual lead one wave before exposure is -0.0236 ($p = 0.05$). But the direction of the check matters: the comparison group most exposed to the objection is the one whose removal makes the result larger.

The same analysis applied to the Callaway–Sant'Anna estimator on the wave-fixed-effects panel is less forgiving. Under exact parallel trends the identified set for the aggregated composite ATT is $[-0.025,\ -0.008]$, essentially the level confidence interval, and it excludes zero; once a relative-magnitudes violation of $\bar{M} = 0.5$ is allowed it widens to $[-0.057,\ +0.024]$ and admits zero, so the breakdown value for that target lies between $\bar{M} = 0$ and $\bar{M} = 0.5$ rather than at zero. The target that isolates the wave-of-exposure effect is more durable on the same panel, still excluding zero at $\bar{M} = 0.5$ ($[-0.042,\ -0.004]$). What binds on this panel is the smoothness restriction Δ^{SD}, which bounds how far the post-treatment trend may depart from a linear extrapolation of the pre-trend: it reaches zero at its mildest deviation ($[-0.041,\ 0.000]$). Reporting both restriction families follows the current difference-in-differences synthesis (Roth et al., 2023). Under the preferred specification the same target instead breaks down between $\bar{M} = 0.25$ and $\bar{M} = 0.50$, for the reason set out immediately below.

That advantage, however, belongs to the relative-magnitudes class and does not carry over to the other one. Applying the smoothness restriction to the preferred specification, rather than to the wave-effects panel, the identified set still excludes zero when the counterfactual trend is required to be exactly linear, at $[-0.036, -0.007]$ on the gate, but admits zero as soon as any curvature is permitted: at $M = 0.005$, half a percentage point of departure from a linear extrapolation per period, it is $[-0.043, 0.003]$. The subtype composite behaves the same way. So on smoothness the preferred specification is no more durable than the Callaway–Sant'Anna benchmark whose fragility we reported above, and we report that rather than let the relative-magnitudes comparison stand for both families. Figure E.1 shows the two classes side by side for both outcomes. 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 a wave-of-exposure effect estimated without covariates on never-treated controls, while the CS(2021) benchmark 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 CS(2021) benchmark 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.

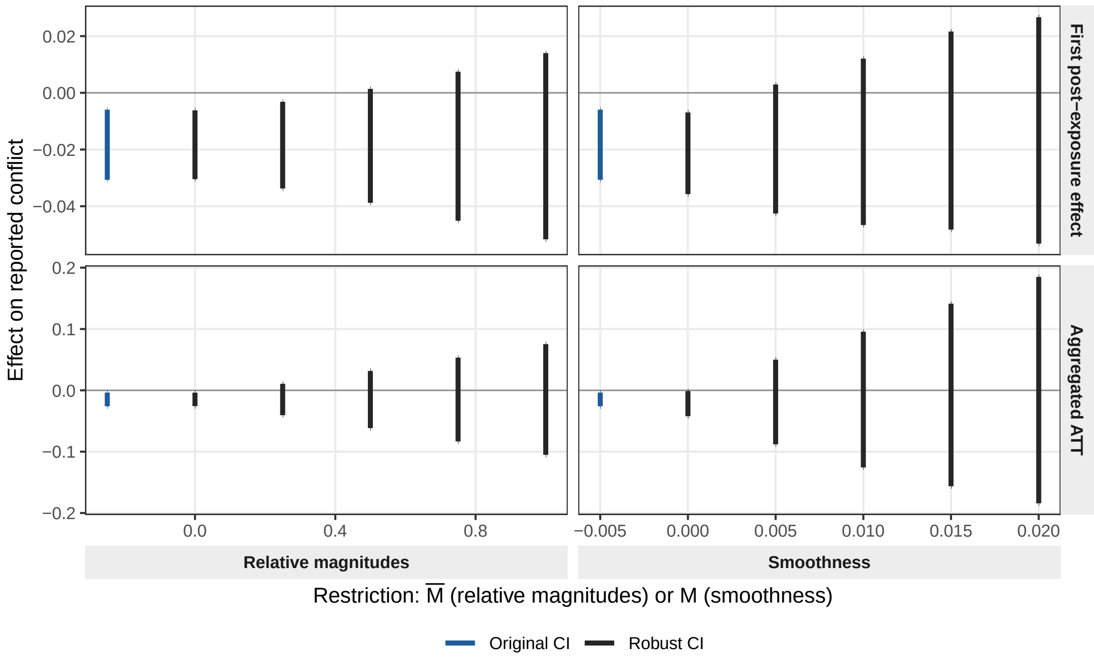

Figure E.1: Sensitivity of the Level Effect, by Target and Restriction Class
The durability the specification shows under relative magnitudes belongs to the wave-of-exposure effect and to that class alone: it does not survive the move to smoothness, and it does not survive aggregation across post-exposure waves.
Note: Robust confidence intervals of Rambachan and Roth (2023) for the first post-exposure effect on the preferred specification, alongside the original interval computed from the same event study. The left column varies $\bar{M}$, which caps the post-treatment violation of parallel trends as a multiple of the largest pre-treatment violation between consecutive periods. The right column varies M, which caps how far the counterfactual trend may depart from a linear extrapolation of the pre-trend between consecutive periods, so $M = 0$ imposes exact linearity rather than exact parallel trends. An interval that crosses the horizontal line admits a zero effect. Rows are the mass-brawl gate; the subtype composite gives the same breakdown values in all four cells and is omitted only to keep the figure legible. Two cautions on reading it. The two classes are not on a common horizontal scale, so $\bar{M}$ and M should not be compared across columns. And the two rows are different estimands on different vertical scales, so the visual width of an interval is comparable within a row but not between rows.

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, -0.0173 ($p < 0.001$), is the event-study coefficient 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.

Table E.3: HonestDiD Relative-Magnitudes Identified Sets (Sun–Abraham t_0)

	Composite ($t_0 = -0.0223$)		*Warga* ($t_0 = -0.0096$)	
$\bar{M}$	lower	upper	lower	upper
0.0 (original)	−0.032	−0.012	−0.018	−0.002
0.5	−0.047	0.001	−0.024	0.005
1.0	−0.064	0.018	−0.034	0.014
1.5	−0.083	0.037	−0.045	0.025
2.0	−0.103	0.056	−0.056	0.036

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 wave-of-exposure effect t_0, under wave fixed effects. Both outcomes admit zero at $\bar{M} = 0.5$. The corresponding analysis on the preferred province-by-wave specification, where the pre-treatment leads are smaller and the breakdown point correspondingly higher, is reported in Supplementary Material Section S7.