Supplementary Material

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

Submitted with the article as online Supplementary Material, and included in the replication package. The article cites this document in 66 places; where it does so, it cites it for detail behind a result already stated and evidenced in the article itself.

This document has two parts, and the Contents page below separates them. The first fifteen sections carry material the manuscript relies on and cites for detail: the benchmark estimators, the mechanism narrative, the identifying assumption behind the composition contrast, and the design of the first-time margin. The remaining sections either explore heterogeneity the paper does not claim to identify, document a mechanism proxy the paper reports as inconclusive, or record why an alternative data source cannot serve as a check; no claim in the paper rests on those. Where the manuscript cites a section here, it cites it for detail behind a result already stated and evidenced in the submitted text.

Contents

S1 Benchmark Estimators, Province Trends, and Displacement

This appendix holds the estimator comparisons and two checks that the main text summarises in one sentence each: the two-way benchmark and its decomposition, the episodic estimand, the Callaway–Sant'Anna benchmark in full, the two routes for province trends, and the displacement check.

S1.1 Benchmark Estimator: TWFE

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

$$y_{vt} = \alpha\, D_{treat,vt} + \boldsymbol{\gamma}'\mathbf{X}_{vt} + \mu_v + \lambda_t + \varepsilon_{vt}, \tag{S1}$$

where $D_{treat,vt} = \mathbf{1}\{t \geq g_v\}$ is a binary absorbing treatment indicator equal to one from the first treatment wave onward, $\mathbf{X}_{vt}$ is a vector of seven village characteristics measured at the PODES 2003 baseline (household electrification, paved road access, internet availability, police post, health-centre presence, number of primary schools, and distance to the district capital), entered as levels in the doubly-robust estimators and interacted with wave dummies in the two-way specifications, μ_v are village fixed effects, and λ_t are wave fixed effects. Standard errors are clustered at the village level following Abadie et al. (2023).

We also estimate a distributed-lag TWFE with exposure lags aligned to May-to-May windows, so each lag falls fully before the conflict reference window (full specification in Appendix S2). Alongside CS(2021) we report the interaction-weighted event-study estimator of Sun and Abraham (2021).

The distributed-lag results, and the timing artifact that unequally spaced survey panels create when exposure windows are misaligned, are documented in Appendix S2.

Two terms recur below and we fix them here. The preferred *counterfactual specification* is the province-by-wave two-way fixed-effects design of Section 5.1. Our primary *heterogeneity-robust staggered estimator* is the interaction-weighted event study of Sun and Abraham (2021). The reason is the design, not the

diagnostics: it is the heterogeneity-robust estimator that can absorb interacted fixed effects, so it implements the preferred counterfactual directly rather than approximating it. CS(2021) can hold province trends fixed only by carrying province indicators inside the nuisance models it fits cell by cell, which is a different operation and, as Section S1.4 shows, one this panel does not support well. We report CS(2021) throughout as the conventional benchmark and for its explicit treatment of staggered timing, and we report what the two routes give side by side rather than selecting between them on the basis of which pre-trend test passes. Every estimate below is labelled by the design it uses.

S1.2 Complementary Estimand: Episodic Exposure (dCdH)

Because earthquakes are episodic physical events, we complement the absorbing design with an estimand in which treatment can switch off. We estimate the de Chaisemartin and D'Haultfoeuille (2026) estimator with a non-absorbing treatment indicator: $D_{interval,vt} = \mathbf{1}\{\exists \text{ month in interval}[t-1, t) : \text{MMI} \geq 6\}$, which switches between zero and one across waves. The dCdH Effect_1 captures the ATT one wave after switching to $D_{interval} = 1$, that is, the effect of fresh exposure in the current inter-wave window regardless of the village's exposure history. Together, the two designs answer two distinct questions: what is the effect of having entered the post-earthquake state (absorbing), and what is the effect of being freshly shaken (episodic)?

S1.3 The Conventional Benchmark: CS(2021)

Table S1 reports the benchmark for two outcomes. Panel A gives the CS(2021) aggregate ATT: -0.0177^{***} for the composite "any communal conflict" indicator (SE = 0.0047, N = 225,372) and a smaller -0.0075^{**} (SE = 0.0037, $p = 0.042$) for the definition-stable *warga* indicator. Panel B gives the Sun–Abraham (2021) interaction-weighted estimator, with significant effects in the wave of exposure ($t_0 = -0.0223^{***}$ composite; -0.0096^{**} *warga*). Neither outcome has a clean joint pre-trend under wave fixed effects; Appendix A sets out the diagnostics across estimator families and Section 6.2 bounds what the residual violation can do to the estimate.

We report the two columns together rather than as a lead estimate plus a robustness check, because each trades one property for another. The composite is the more powerful but its coding rule changes at 2008, in a direction that makes the 2005 wave broader rather than narrower. *Warga* is definition-stable across all seven waves but rarer, and on the level effect that imprecision binds: it agrees in sign with the composite and is not significant under the preferred specification (-0.0031, standard error 0.0031). We therefore do not present it as a second level estimate. What it delivers is an outcome whose survey definition is fixed, which is what the composition contrast of Section 7.1 requires and where it is decisive.

Economic magnitude. First strong exposure lowers the communal-conflict probability by 1.8 percentage points, with a 95 percent interval of $[-2.69, -0.85]$ points. The benchmark is not the pooled sample mean of 2.6 percent but the treated villages' own counterfactual: ever-treated villages reported conflict at 6.2 percent across their pre-treatment waves, roughly twice the pooled rate, because earthquakes strike the more conflict-prone, faster-growing parts of the hazard zone. The estimand is accordingly local to hazard-zone villages that are actually struck. The identifying variation comes from cohorts G3–G8 (first treated 2008–2024), principally the Padang 2009 earthquake and later events. Post-treatment dynamic ATTs are negative at all six estimable horizons and individually significant at t_{+0}, t_{+1} and t_{+3}; the rest are negative but imprecise, so the longer-run path shows no reversal without establishing persistence.

Group-level heterogeneity. No single cohort drives the effect. The one positive cohort is post-Helsinki Aceh, and dropping it, or Aceh entirely, leaves the estimate unchanged or larger (Appendix S27).

The conventional benchmarks. The two-way fixed-effects benchmark gives -0.0213 with controls and -0.0193 without; the latter is not a negative-weighting artifact, since the already-treated comparisons carry only 3.8 percent of the Goodman–Bacon weight and return -0.0376 against -0.0186 for the clean ones, pulling in the same direction rather than against it, and the episodic estimator of de Chaisemartin and D'Haultfœuille (2026) gives -0.015^{***} with matching dynamics.

Table S1: Conventional Staggered-DiD Benchmarks: Effect of First MMI≥VI Exposure on Communal Conflict
2005-boundary panel, 2005–2024, KRB Medium/High

	Warga conflict (consistent def.)	Any communal conflict (composite)
Panel A: Callaway–Sant'Anna (2021), not-yet-treated control		
Overall ATT (simple)	-0.0075^{**}	-0.0177^{***}
	(0.0037)	(0.0047)
Pre-trend, average lead (p)	-0.0038 (0.56)	-0.0077 (0.33)
Pre-trend, joint Wald (p)	$\chi^2(15) = 29.5$ (0.014)	$\chi^2(15) = 42.9$ (0.000)
Panel B: Sun–Abraham (2021) IW, never-treated control		
Effect at t_0	-0.0096^{**}	-0.0223^{***}
	(0.0041)	(0.0051)
Effect at t_{+1}	-0.0067	-0.0194^{***}
	(0.0043)	(0.0054)
Pre-lead t_{-2} (placebo)	$+0.0077$	$+0.0138^{**}$
Panel C: Earthquake-level inference (stacked-event design, 19 events)		
Treated-village-weighted ATT	-0.0114	-0.0268
event block-bootstrap SE [p]	(0.0065) [0.034]	(0.0105) [0.006]
Equal-event mean, secondary [p]	-0.0080 [0.49]	-0.0155 [0.32]
Earthquake events (clusters)	19	19
Baseline mean (pooled)	0.016	0.026
Pre-treatment mean, ever-treated	0.037	0.062
Observations	225,372	225,372

Notes: Panel A reports analytical influence-function SEs and Panel B village-clustered SEs in parentheses. The average pre-trend does not reject, but a joint Wald test of all pre-treatment group-time ATT(g,t) coefficients rejects for both outcomes under wave fixed effects (the aggregated event-study leads oscillate in sign rather than trending toward the post-treatment decline); the test is passed once province-by-wave effects are absorbed (the robustness discussion in the main text). Panel C treats the earthquake, not the village, as the unit of assignment: the treated-village-weighted ATT is stacked across the 19 earthquakes and its SE comes from an event-level block bootstrap that resamples whole earthquakes (9,999 draws); the naive village-clustered SE for this estimand understates uncertainty because treatment is driven by 19 shocks. The equal-event mean weights each earthquake equally (13 of 19 negative). Conley spatial-HAC SEs at 50–150 km cutoffs and kabupaten-level clustering leave the composite reduced-form estimate significant at the 1 percent level (the robustness discussion in the main text); under a deliberately conservative province-level wild-cluster bootstrap (few clusters), the pooled staggered ATT is marginal under province clustering ($p = 0.064$), which we report rather than suppress. Pre-treatment mean = mean outcome among ever-treated villages before first exposure. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Across heterogeneity-robust and conventional estimators, and across absorbing and episodic treatment definitions, the estimate ranges from −0.014 to −0.022 and every confidence interval lies below zero (Appendix Table S57); the decompositions, distributed lags, the timing artifact for unevenly spaced panels, and the placebo panel are in Appendix S2. CS(2021) is the estimator we carry through the robustness battery because it has the largest set of published diagnostics.

Table S2 makes the point on a common sample and control set: six designs, including the same-earthquake stack that holds the shock itself fixed, range from −0.0165 to −0.0227, between 26 and 36 percent of the pre-exposure mean. The stacked row targets the *incremental* effect of crossing the damage threshold, smaller than the total effect by construction, and we do not call it the strongest evidence: its own pre-lead is marginal (+0.0136, $p = 0.06$) and its earthquake-clustered standard error is far larger. The intensity ladder is not monotone under CS(2021) and changing the threshold also changes which cohorts and earthquakes enter, so we rest the dose argument on the same-earthquake band comparison instead (Appendix S2).

Table S2: Six Designs, One Sign: Estimates Across Identification Strategies

Design	Estimate (SE)	% of pre-exposure mean	Pre-trend joint p	N
Absorbing TWFE, province × wave FE	−0.0165*** (0.0041)	-26.4	0.678	225,372
Finite duration (two waves)	−0.0173*** (0.0037)	-27.8	0.678	225,372
Event study: wave of exposure	−0.0194*** (0.0063)	-31.0	0.678	225,372
Event study: one wave after	−0.0227*** (0.0069)	-36.4	0.678	225,372
Callaway–Sant'Anna (2021)	−0.0177*** (0.0047)	-28.4	—	225,372
Stacked within-earthquake	−0.0174*** (0.0040)	-27.9	0.063	147,905

Rows 1–5 are estimated on the same sample (KRB Medium/High, always-treated cohort excluded) with baseline PODES 2003 controls; standard errors clustered at the village level. Pre-trend joint p tests leads at −4, −3 and −2 against zero. The pre-exposure mean is 0.0624. Callaway–Sant'Anna uses wave fixed effects, since the estimator does not accommodate province × wave absorption; its pre-trend is reported separately via `estat pretrend`. The stacked design compares villages exposed at MMI ≥ VI to villages that felt the *same* earthquake at MMI ∈ [4, 6), with event × village and event × wave fixed effects over 24 earthquakes; clustering by earthquake widens its standard error to 0.0098. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S1.4 Province Trends: Absorbed or Conditioned On

Section 5.1 argues that the counterfactual should hold province-specific trajectories fixed; that argument does not say how. Province trends can be absorbed as saturated fixed effects or conditioned on inside the estimator's nuisance models, and the two routes disagree: absorbing them leaves the result intact, conditioning on them does not. Either the level decline is partly carried by provincial trajectories, or conditioning on thirty-two province and island indicators per group-time cell breaks the estimator's overlap; our data do not separate the two, and Section S15 reports both routes in full. What follows limits the level estimate, not the paper: the composition contrast differences out disturbances common to the outcomes recorded on the same form by the same respondent, and the earthquake-level divergence is estimated event by event rather than from this panel at all. This is the second reason, alongside the pre-trend record, why the level stands as a motivating average and why the composition evidence rests on the divergence rather than on either level.

S1.5 Does the Decline Displace? Conflict in Neighbouring Villages

Disasters can *lower* conflict where they strike while *raising* it nearby (Amarasinghe et al., 2026). Assigning each never-treated village to a mutually exclusive proximity ring by distance to the nearest treated village, displacement predicts positive ring coefficients; we find the opposite, with all three rings negative under village and wave fixed effects and no ring ever significantly positive under either fixed-effect structure (Section S15, Section S15). The local decline shows no detectable offset onto neighbours; we read this as bounding, not eliminating, the displacement concern, since PODES cannot track individuals who relocate across village boundaries.

S2 Supporting Detail on Design and Robustness

This appendix holds detail supporting the design and robustness discussion of the main text: raw cohort paths, the case for the absorbing treatment definition, the

hazard-zone restriction, the intensity ladder, and the full robustness battery.

Benchmark estimators in full

The two-way fixed-effects benchmark (Section S15) tells the same story and is not a negative-weighting artifact: a Goodman–Bacon decomposition assigns 3.8 percent of the weight to forbidden comparisons, but they return -0.0376 against -0.0186 for the clean ones, pulling in the same direction. The episodic estimator of de Chaisemartin and D'Haultfœuille (2026), which identifies the effect only when a village switches into fresh exposure, gives -0.015^{***} (SE $= 0.004$) with dynamics decaying from -0.015 to -0.005, matching the absorbing event study; its joint placebo does not reject ($p = 0.056$), though the nearest individual placebo, 0.010, is significant with the opposite sign. Full decompositions, distributed lags, the timing artifact we document for unequally spaced panels, and the placebo panel are set out below.

Intensity gradient. The framework predicts a dose-response, and the CS ladder does not deliver one: it is not monotone across MMI≥V, ≥VI and ≥VII. We do not claim to have identified a dose-response, and the ladder is not the reason we could not: changing the threshold also changes which cohorts and earthquakes enter the treated group, so the ladder is not a pure intensity contrast holding composition fixed, and no regression discontinuity is available at MMI VI because the cutoff is a classification convention rather than an assignment rule. We rest the intensity argument on the same-earthquake stacked design instead.

Estimator implementation

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

Unbalanced panel and software. PODES does not observe every village in every wave, so the estimation panel is unbalanced. We use the default pairwise-

balancing rule of the `csdid` implementation (Callaway and Sant'Anna, 2021), under which each $ATT(g,t)$ is computed on the villages observed in both of the two periods it compares, rather than restricting the whole analysis to villages observed in all seven waves. This retains cohorts that a fully balanced restriction would drop and is the panel analogue of the estimator in Callaway and Sant'Anna (2021). All CS(2021) estimates in the paper use `csdid` version 1.72; the replication package records the version of every user-written command.

S2.1 Raw Outcome Paths by Cohort

Before any specification is imposed, Figure S1 plots the raw share of villages reporting communal conflict, wave by wave, separately for each treatment cohort and for the never- and always-treated groups. Cohorts differ in level: the 2021 and 2024 cohorts sit well above the never-treated line throughout, which is the composition fact that motivates village fixed effects. The never-treated path is close to flat and drifts down slightly after 2018, which is the trend that province-by-wave effects are meant to absorb. And the cohorts are of very unequal size, from 668 villages first exposed in 2008 to 37 in 2014, so the raw paths for the smaller cohorts are correspondingly noisy; the 2014 cohort is omitted from the figure for that reason.

S2.2 The Intensity Ladder and the Regression-Discontinuity Alternative

The conceptual framework predicts a dose-response relationship. We report the threshold estimates but do not claim to have identified one. Under the absorbing first-exposure design the effect strengthens from -0.0069^{***} (SE $= 0.0021$) at MMI$\geq$V to -0.0213^{***} (SE $= 0.0034$) at MMI$\geq$VI, and then stops strengthening: at MMI$\geq$VII ("Very Strong") it is -0.0174^{***} (SE $= 0.0035$), while the CS(2021) counterpart in the same table falls from -0.0177^{***} to -0.0054 and loses significance. The pattern is therefore not monotone across the full range.[1] We withdraw the dose-

[1] A natural question is why we do not exploit the MMI VI threshold directly in a regression-discontinuity design. Two obstacles rule it out. First, MMI VI is a classification boundary chosen by the analyst, not an institutional trigger, so nothing in treatment receipt changes discontinuously there. Second, the running variable is not smooth in the neighbourhood of the

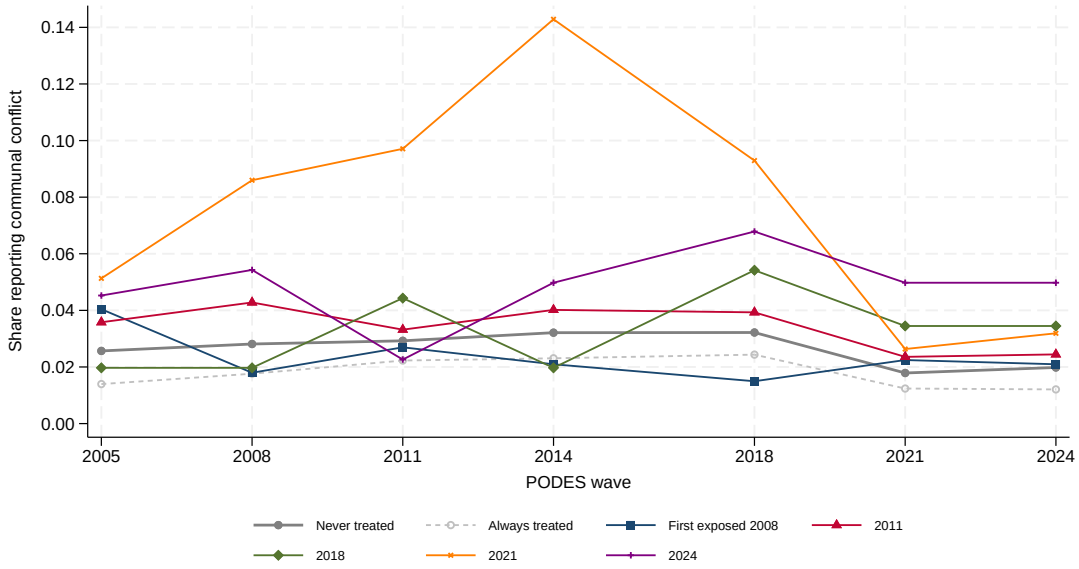


Figure S1: Raw Conflict Rates by Treatment Cohort
Note: Unadjusted means, no covariates and no fixed effects. Cohort sizes: never treated 30,259 villages; always treated 10,097; first exposed 2008, 668; 2011, 1,145; 2014, 37 (omitted); 2018, 203; 2021, 721; 2024, 221. The always-treated cohort is excluded from the preferred specification.

response reading for two reasons. First, the CS(2021) benchmark does *not* separate the two higher bins: at MMI≥VII its estimate is −0.0054 and insignificant, against −0.0177*** at MMI≥VI, reflecting the smaller and less precisely estimated cohort at that threshold (1,436 ever-treated villages against 2,995). Second, changing the threshold also changes which cohorts and which earthquakes enter the treated group, so the comparison is not a pure intensity contrast holding composition fixed. A cleaner test would compare villages struck by the *same* earthquake at different intensities; the same-earthquake design of Section 8 does exactly that, and we rest the intensity argument on it rather than on the threshold ladder. We report the threshold estimates for completeness and draw no dose-response conclusion from them.

cutoff. ShakeMap intensities reach villages through contoured raster bands, and the resulting village-event distribution is visibly heaped: 5.4 percent of the 1,961,850 in-sample village-event observations fall within 0.01 of an integer against roughly 2 percent under continuity, and MMI 6.0 is itself a mass point. The density is not so degenerate as to make local-polynomial estimation mechanically impossible, but combined with the first objection, that the cutoff is a classification convention rather than a treatment assignment rule, there is no discontinuity for such an estimator to recover. The band comparison reported in Section S15, which contrasts villages at MMI $[6, 7)$ with those at $[5, 6)$ struck by the same earthquake, is the defensible version of the same intuition and is unaffected by heaping within bands.

S2.3 dCdH episodic estimator detail

The de Chaisemartin and D'Haultfoeuille (2026) episodic estimator identifies the effect only when a village switches into fresh exposure in the current interval, not the cumulative ever-treatment effect. It gives -0.015^{***} (SE = 0.004); the full estimates and placebo panel are in Section S15. Its joint placebo test does not reject at conventional levels ($p = 0.056$), but we report the individual placebos rather than the joint test alone: the nearest one is +0.010 (SE = 0.005), which is significant at 5 percent and carries the opposite sign to the effect. The episodic effect, −0.015, is close to the absorbing estimates (TWFE −0.019, pooled CS ATT −0.0177). That closeness does not discriminate between a persistent effect and one concentrated early: an absorbing average over horizons that are strongly negative at first and near zero later can match a fresh-exposure estimate without either implying persistence. The direct evidence on duration is the event study, which places the effect in the wave of exposure and the wave after and finds the later horizons individually imprecise (Section 6). The benchmark TWFE is larger in magnitude (−0.019), but it weights cohorts differently and is the estimator the heterogeneity-robust methods are designed to correct, so we do not read that contrast as evidence on duration. The episodic dynamics speak to duration directly: the effect falls from −0.015 at the first horizon to −0.008 and then −0.005, matching the decay in the absorbing event study.

S2.4 TWFE benchmark, distributed lag, timing artifact, Goodman-Bacon

The two-way fixed-effects benchmark, reported in full in Section S15, tells the same story as CS(2021) and is included only for comparability with the older literature; no causal claim rests on it alone. Its distributed lag traces the CS(2021) dynamic path: the effect is already significant one window after exposure (-0.0056^{**}, SE = 0.0026), deepens to -0.0229^{***} (SE = 0.0035) two windows later, and returns to −0.0010 (SE = 0.0018) by the third, so mass fighting responds quickly, builds, and fades. This distributed lag captures fresh exposure in specific May-to-May windows, whereas the absorbing CS(2021) design estimates the effect of entering

the post-earthquake state; the fading of the episodic lag should therefore not be read as contradicting the absence of reversal in the absorbing post-exposure effect. The response is nonlinear in shaking intensity, not proportional to it, and it is not a knife-edge at MMI VI: weaker shaking produces smaller but nonzero reductions. The threshold ladder does not establish a monotone dose-response, for reasons we set out in Appendix S2. Dating exposure by calendar year rather than by the May-to-May survey window displaces the distributed-lag response by one wave. The two-wave effect of -0.0218 falls to -0.0028 and loses significance, while a coefficient of -0.0063 appears at three waves where the aligned specification gives zero. Both timings use the identical sample, so the difference is entirely one of dating (Table A.3). We document this as a cautionary point for unequally spaced panels, not a feature of the data.

Finally, the absorbing benchmark is not a negative-weighting artifact, although the reason is not that the forbidden comparisons are absent. The Goodman–Bacon decomposition we report is computed on the sample the preferred specification actually uses, which drops the always-treated cohort because those villages have no pre-period at all. On that sample 96.2 percent of the weight falls on treated-versus-never-treated comparisons and 3.8 percent on the timing comparisons that use an earlier-treated village as a control for a later-treated one. That residual weight is small, and it is also uninformative about the sign: the forbidden comparisons return -0.0376 against -0.0186 for the clean ones, so they pull the estimate in the same direction, not against it. Dropping the always-treated cohort removes the units without a pre-period rather than every forbidden comparison, which is why we rest the argument on the heterogeneity-robust estimators instead (Table S58; Appendix S15, Section S15).

S2.5 Why the Absorbing Treatment Definition

An earthquake is an episodic physical event, so treating a village as exposed forever after its first strong shock requires justification. Three arguments support it. First, the conceptual framework (Section 3) predicts hysteresis: the shock disrupts physical and institutional inputs to collective violence that remain disrupted until reconstruction, so the treated state may outlast the shaking. Second, the definition

does not require the effect to persist, and the estimates indicate that it does not. The dynamic ATTs are concentrated in the wave of exposure and the wave after, with the three subsequent lags jointly indistinguishable from zero (Section 6), and the episodic estimator decays along the same path (-0.015, then -0.008, then -0.005; Section S15). That profile governs how the pooled estimate should be read, not whether the absorbing definition is admissible. Because the absorbing indicator averages over later waves in which villages have largely recovered, it attenuates the effect toward zero, so the pooled estimate is smaller than the response at its peak. Third, the absorbing and episodic definitions answer different questions and we estimate both: the absorbing design identifies the average post-entry effect of being in the post-earthquake state, the episodic design (Section S1.2) the effect of fresh exposure regardless of history, and the two agree in sign and are close in magnitude (-0.019 versus -0.015).

S2.6 Is the Hazard-Zone Restriction Endogenous?

A natural concern is whether the hazard classification is itself endogenous to the earthquakes we study. The KRB zones we use are dated to 2010 in the source files; on that vintage they postdate the 2004–2009 events (Aceh, Yogyakarta, Padang) and predate the later ones (Aceh Singkil 2011, Pidie Jaya 2016, Lombok and Palu 2018, Cianjur 2022). The KRB classification is a seismic-hazard assessment based on tectonic setting, namely fault and subduction-zone geometry, in the same spirit as Indonesia's national probabilistic seismic-hazard analysis (Irsyam et al., 2020), rather than the realized damage of any particular event, so a village's hazard band reflects its tectonic setting rather than whether a specific earthquake happened to strike it. Second, the sample restriction defines a fixed set of villages and enters the design only through village fixed effects, which absorb any time-invariant hazard level; identification comes from the timing of exposure within this fixed set, not from the map's boundaries. We nonetheless treat the earliest cohorts, whose events predate the map, with the same caution we apply throughout, and the result is robust to dropping them one cohort at a time (Section 6.2). Two further facts show the map is not selecting the result. First, the restriction barely touches the treated group. Of the 85,306 treated village-waves in the national panel, 82,417 lie

inside the hazard zones, and of the 13,549 villages ever exposed nationally, 13,063 do; on both counts the zones hold about 96 percent of the treated group. The map therefore changes the comparison pool rather than which villages are treated. Second, dropping the hazard-zone restriction entirely and re-estimating on the full national sample leaves the effect intact. On all 441,805 village-waves, of which 138,348 lie outside the hazard zone, the absorbing TWFE estimate with baseline-by-wave controls is -0.0201 for the composite ($p < 0.001$) and -0.0069 for *warga* ($p = 0.010$), against -0.0217 and -0.0078 in-sample. Dropping the restriction attenuates the estimate mildly as the control pool becomes more heterogeneous, and does not remove it.

S2.7 The Robustness Battery

The estimate survives the exposure, measurement, clustering, and spatial choices, all detailed below. The MMI VI cutoff is not special: the effect stays negative and significant at broader MMI≥V (-0.0069^{***}) and at narrower MMI≥VII (-0.0174^{***}), though neither exceeds the MMI≥VI estimate and the ladder traces no dose-response (Appendix S2). Measurement error in the raster-to-village mapping is not driving it: four aggregation rules across all 1,423 event ShakeMaps, and dropping the largest 5 percent of villages by area, leave the estimate stable in sign and broad magnitude (TWFE -0.0073 to -0.0209; CS -0.0059 to -0.0169, falling to 10 percent significance only under the pixel-maximum rule). Nor is it an artifact of ShakeMap's blended intensity construction: rebuilding exposure from an MMI derived independently from instrumental Peak Ground Acceleration, holding the window-aligned timing fixed, reproduces the effect on every specification (CS -0.0129, standard error 0.0039, against the ShakeMap-MMI -0.0177; Appendix S7). Re-clustering at the kabupaten level (320 clusters) keeps the composite significant at 1 percent and *warga* at 5 percent, and the conservative 28-cluster province inference still leaves the composite significant at 5 percent ($t = -2.79$), consistent with the wild bootstrap above, while *warga* does not clear conventional levels. The single-item *warga* outcome gives -0.0073^{***} (Appendices S15 and S9).

The absorbing treatment definition also appears not to be costly on the dimension it most plausibly ignores, namely *repeated* exposure. Because it treats

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

On spatial spillovers, which could bias the never-treated counterfactual that carries most of the identification weight, we report what the tests actually return rather than the reassuring version. Adding neighbours' treated share within 10–50 km as a spatial lag does *not* leave the direct effect standing beside a small spillover: the own-exposure coefficient collapses from −0.0213 to −0.0024 at 25 km, −0.0066 at 50 km and −0.0001 at 10 km, while the neighbour share takes a large negative coefficient at every radius (−0.0242, $p = 0.006$; −0.0227, $p = 0.002$; −0.0241, $p = 0.026$). We do not read this as evidence that the effect is a spillover. A treated village's neighbours inside the same footprint are treated too, so own and neighbourhood shaking are close to collinear and the split between them is not identified; what the regression identifies is a footprint-level effect, and the two coefficients should be read together rather than against each other. The binary neighbour-exposure version, where the collinearity is weaker, keeps both terms negative and significant and is the form we rely on. A 25 km "donut" dropping never-treated villages with treated neighbours leaves the estimate essentially unchanged (−0.0218 under two-way fixed effects, −0.0183 under CS(2021), $p < 0.001$), but its average pre-trend now rejects (+0.0157, $p < 0.001$), so we do not present the donut as a clean pass either. One further limit belongs here: the donut is

built on a neighbour universe that excludes the always-treated 2005 cohort, so some contaminated controls survive it. Province×wave fixed effects, the preferred specification, move the coefficient from -0.0213^{***} to -0.0165^{***} and leave it significant; and a forward-lag placebo is a clean null (-0.0045, $t = -0.84$, $p = 0.40$), so the design shows no sign of anticipatory adjustment (Appendix S15).

Selection on unobservables is an unlikely explanation: under the Oster (2019) coefficient-stability test, adding the controls moves the estimate *away* from zero, so unobserved selection would have to run opposite to the observed selection on covariates to overturn the result. Nor is the effect an artifact of spatial correlation in the errors: Conley spatial-HAC standard errors (Conley, 1999) keep it significant at 1 percent (Appendix S15).

S3 Framework Detail: Channels, Predictions, and Prior Evidence

This appendix gives the full version of the framework that Section 3 states in compressed form: each channel with its prior evidence, and each of the five predictions with the literature it engages. Nothing here is a separate result. The main text states each prediction and the identifying assumption; what this appendix adds is the prior evidence each prediction engages and the timing detail behind the fifth.

Income-destruction channel. The canonical opportunity-cost model (Becker, 1968) predicts that any shock reducing expected income lowers the cost of violence, and the canonical development evidence is Miguel, Satyanath, and Sergenti (2004), who find that a negative growth shock of five percentage points raises civil-conflict risk by about one half. Dube and Vargas (2013) show that income shocks shift the calculus of predation through two opposing forces, a fall in wages lowering the opportunity cost of appropriation and a rise in contestable rents raising its return, so the net sign depends on which dominates. Applied to earthquakes the channel implies a *positive* reduced-form effect from asset destruction and labour-market disruption. Our shock is physical rather than a growth shock, and on the collective

margin we recover the opposite sign, which is what the composition evidence below is meant to explain.

Contact disruption and emergency coordination. The competing channel runs the other way: contact disruption suppresses potential violence regardless of grievance, and disaster response can bring a temporary surge of external monitoring that raises the cost of organising violence (Berman, Shapiro, and Felter, 2011); whether this external presence, rather than durable village policing, is what changes is an empirical question we return to below. Prior work shows that disasters can redirect social participation toward recovery-oriented coordination: Bai and Li (2021) find that seismic exposure raised community social participation for the same 2004–2007 Indonesian earthquakes, and Cassar, Healy, and von Kessler (2017) find that the 2004 tsunami raised prosocial behaviour in rural Thailand. In our setting the observed PODES proxies do not support a durable generalized social-capital surge (Section 7.2). And households in recovery mode concentrate resources on reconstruction rather than conflict, which itself changes the opportunity-cost calculus in the opposite direction from the income-destruction prediction (Dube and Vargas, 2013).

Why a temporary shock can have persistent effects. The shaking lasts seconds, but a strong earthquake is a temporary shock to the ground and a persistent shock to the village's physical and institutional state. Structural damage at MMI VI and above is not undone when the shaking stops: houses, markets, roads, and places of worship stay damaged until reconstruction, which in rural Indonesia takes years. Exposure is therefore a state change rather than a pulse. This motivates the treatment definition in Section 5, where we model first exposure as an absorbing state, estimate the episodic design alongside it, and test directly whether the effect decays.

Testable implications. The five predictions the framework yields, and where each is tested, are stated in Section 3 and not repeated here.

S4 Mechanism Narrative, Event Study, and Design Detail

Table S3: Capacity Inputs Treated as Outcomes
Only one input passes its own pre-exposure test, and the inputs that move most move against the account we favour.

	Waves obs.	Pre-exp. mean	Wave FE	Prov.×wave (preferred)	% of mean	Pre-trend p
Civil-defence members	7	16.702	0.1347	0.4133	2.5	0.000
			(0.1950)	(0.2200)		
Security post present	6	0.672	-0.0122	-0.0099	-1.5	0.000
			(0.0085)	(0.0099)		
Places of worship	7	7.146	0.3046	0.2639	3.7	0.000
			(0.0499)	(0.0618)		
Mosques	7	6.168	0.2297	0.1919	3.1	0.001
			(0.0418)	(0.0493)		
Migrant workers	7	74.164	-8.6067	-7.6899	-10.4	0.006
			(1.5645)	(1.3776)		
Agriculture main income	6	0.847	-0.0058	-0.0157	-1.9	0.000
			(0.0043)	(0.0048)		
Any *gotong royong*	6	0.969	-0.0021	-0.0022	-0.2	0.399
			(0.0032)	(0.0036)		
Disaster early-warning system	4	0.116	-0.0008	0.0086	7.4	0.448
			(0.0102)	(0.0112)		

Each row is a separate absorbing regression of the input on first exposure at MMI $\geq$ VI, with village fixed effects and either wave or province-by-wave effects. % of mean expresses the preferred-specification coefficient against the pre-exposure mean among villages that are eventually exposed. The final column is the joint test of the pre-exposure leads from an endpoint-binned event study on the preferred specification, with the always-treated cohort dropped; treating an input as an outcome is a second causal claim, and a small p there means the series was already diverging before exposure and the coefficient should not be read causally. Standard errors clustered at the village.

This appendix holds the material the main text summarises in one sentence each: the channel narrative, the *gotong royong* corroboration, the polarisation heterogeneity, the event study, the count of identifying shocks, regional trends, and the clustering and spatial-structure choices.

S4.1 Population displacement

If shaking drives residents out, the pool for collective violence shrinks and the decline is mechanical rather than behavioural. We can test this only in part. PODES records population and household counts through 2011 and not from 2014 onward, so the 2008 and 2011 cohorts can be checked while the later cohorts, which

supply most of the identifying variation, cannot. On the 130,052 village-waves with counts, the point estimates do point that way, but neither is distinguishable from zero: log population falls by -0.0072 (standard error 0.0060, $p = 0.23$) and log households by -0.0089 (0.0063, $p = 0.16$), about 0.7 and 0.9 percent. That residential location is the margin a disaster moves durably, while income recovers faster than expected, is what Deryugina, Kawano, and Levitt (2018) document for Hurricane Katrina, and it is why we read headcounts rather than income as the first stage available to us.

Displacement is therefore not established on the one sample where we can look for it, and even read at its point estimate it is far too small to carry the result. On the same restricted sample the communal-conflict effect is -0.0147 against a 5.1 percent pre-exposure rate, a fall of roughly 29 percent, so for a 0.7 percent loss of population to produce it mechanically the elasticity of reported conflict with respect to village population would have to be about forty. The PODES counts are a village-head report rather than an enumeration, but matching the 2010 Population Census to the 2011 wave for 35,628 in-sample villages gives a correlation of 0.99 in levels and logs and a log-log slope of 0.970 (standard error 0.001), so correcting for the implied compression would move the first stage from -0.72 to about -0.74 percent. The composition bounds the mechanical version directly: depopulation predicts that *every* category of reported disorder falls, yet theft rises sharply over the same villages and waves while communal conflict falls. That rules out depopulation as the sole account. It does not rule out selective displacement of particular groups, which neither PODES nor our design observes, and we flag that as an open threat rather than one we have closed (Section S15).

S4.2 Channels Consistent with the Evidence

What follows is mechanism-consistent evidence, not a mediation decomposition: we do not claim to identify any channel's share. The organising logic is that a strong earthquake *may* raise the economic *motive* for disorder, through asset and income losses PODES cannot observe, while degrading the *capacity* to organise it collectively. The two outcomes resolve that tension differently. Individual theft, which needs only motive, rises. Communal violence, which also needs contact,

coordination and the ability to evade monitoring, falls, because on this account the shock erodes those inputs faster than it may raise the motive.

The capacity-degrading channels. Three forces plausibly degrade the capacity for organised violence. *Contact disruption*: communal violence emerges from contested spaces and the inter-group contact central to relations in Indonesia (Bazzi et al., 2019; Lowe, 2021; Rao, 2019), which strong shaking can render unusable. We flag in advance that the one venue type we can count, places of worship, does not decline, so within this channel the evidence rests on communication rather than venue destruction. *Post-disaster monitoring*: a temporary surge of external actors raises the cost of organising violence, which the counterinsurgency evidence suggests can suppress it (Berman, Shapiro, and Felter, 2011), though the net effect of aid and presence is context-dependent and can be conflict-increasing where aid becomes a contestable prize (Crost, Felter, and Johnston, 2014; Nunn and Qian, 2014). This is the most tentative channel: PODES cannot measure the surge, and the one durable state-presence proxy it records, police-post presence, does not rise after exposure. *Recovery-period opportunity costs*: reconstruction can reallocate labour into construction and raise wages rather than collapse them (Kirchberger, 2017), lowering the return from conflict, opposite to the income-destruction channel (Dube and Vargas, 2013).

The rising theft signals active individual-predation incentives. Whether that reflects a lower return to honest work or a higher return to appropriation, the two income-side forces of Dube and Vargas (2013), neither requires inter-group coordination and both push individual predation upward. What the composition establishes is that these incentives, however generated, do *not* convert into collective violence. The -1.5 point communal estimate is therefore best read as consistent with a setting in which degraded collective capacity dominates any countervailing rise in the motive for disorder: if the shock also raises the motive it is a net effect in which capacity dominates, and if it does not it is the capacity effect alone. Our data do not separate these, and the composition is the same under both. Susenas microdata would turn the motive from consistency evidence into a measured channel; because the opportunity-cost force predicts *more* conflict, a

finding that welfare fell would sharpen this reading rather than weaken it.

Onset and continuation in prior evidence. The first prediction of Section 3 separates onset from continuation, and the prior evidence separates them the same way. Bazzi and Blattman (2014) make the separation of these margins an explicit methodological lesson, and their price shocks leave onset unmoved, which is what a shock that transfers income without destroying the physical settings of daily life should do; a damaging earthquake differs precisely in supplying the incidents an onset requires. Barron and Sharpe (2008) document both halves in Indonesia, describing established conflict as cyclical and retaliatory and tracing the path into first-time violence through isolated incidents that escalate into group clashes.

S4.3 Corroborating Evidence on Collective Capacity

The channels above are inferred from what the shock does to inputs. PODES also measures collective capacity directly. From 2018 onward the survey asks whether villagers customarily engage in *gotong royong*, mutual-help collective action, and how widely they participate, on an identical three-point scale in 2018, 2021 and 2024. Two items are asked separately: collective action for public purposes, which the questionnaire illustrates with communal labour, night-watch rotations and village festivals, and collective action to help residents struck by misfortune such as death, illness or accident. The first is a direct measure of the capacity to organise the village; the second is a measure of disaster response.

Table S4 estimates the effect of first strong exposure on both. High-involvement public-purpose *gotong royong* falls by 3.1 to 3.2 percentage points against a mean of 86 percent, significant under both sets of fixed effects. The misfortune-response item does not fall; its extensive margin rises slightly. The two move in opposite directions, which is what the capacity reading predicts and what a general collapse in village functioning would not: routine collective organisation weakens while emergency mutual aid, which needs no standing coordination, does not.

We put weight on this cautiously. The item exists in only three waves, so only the 2021 and 2024 cohorts are identified within the window, 942 villages in all. A Callaway–Sant'Anna estimate on the same item corroborates the two-way

specifications, at -0.0326 (standard error 0.0142, $p = 0.021$), and the single pre-treatment lead that three waves permit does not reject ($+0.022$, standard error 0.040, $p = 0.58$), although a test built on one lead has little power. The calibration row of the table makes the power problem explicit: on this restricted sample the conflict effect itself is no longer significant under the preferred fixed effects. Two further limits bound the reading. The estimand is not the paper's main one: it averages the 2021 and 2024 cohorts only, so it covers different earthquakes over a different period and corroborates the direction of the effect without being a component of its magnitude. And the 2021 wave was enumerated in May 2021, during the pandemic, which depresses collective activity in level; that wave is also the first post-exposure wave for the 2021 cohort. Wave fixed effects absorb that shock where it is common across villages, and it would bias the estimate only if restriction intensity covaried with earthquake exposure. We therefore read the result as corroborating the capacity interpretation, not as establishing it.

S4.4 Suggestive Heterogeneity

Interacting exposure with pre-treatment religious polarisation deepens the decline by 1.6 percentage points per standard deviation, as a contact-and-mobilisation account predicts, but the gradient does not survive the proportional-scale test that governs the rest of the paper, so we do not read it as evidence for the channel (Section S21).

S4.5 Event Study

Figure 1 presents the event study on the preferred specification. The diagnostics that bind are those of the CS(2021) event study on the conventional wave-fixed-effects panel (Appendix Figure S8): there the average lead is small and insignificant for both outcomes ($p = 0.33$ composite, $p = 0.56$ *warga*) but a joint Wald test over all pre-treatment group-time cells rejects for both ($\chi^2(15) = 37.5$ and $\chi^2(15) = 28.8$), so we do not read either pre-period as a clean pass. A violation the pre-test would miss half the time exceeds the pooled *warga* ATT (Roth, 2022), which is one reason we do not read a passed pre-test as evidence, and the HonestDiD bounds of Section 6.2, not the pre-trend tests, are the binding constraint

Table S4: Collective Capacity After Strong Shaking

Outcome	Wave FE	Province × wave FE	Pre-exposure mean	N
Public-purpose *gotong royong*, high involvement	−0.0311**	−0.0324**	0.863	78,225
	(0.0138)	(0.0164)		
any involvement	−0.0027	0.0030	0.994	78,225
	(0.0018)	(0.0026)		
Misfortune-response *gotong royong*, high involvement	−0.0111	−0.0033	0.911	78,225
	(0.0099)	(0.0127)		
any involvement	0.0023	0.0059**	0.996	78,225
	(0.0020)	(0.0025)		
Power calibration on the same restricted sample				
Any communal conflict	−0.0395***	−0.0148	0.024	93,603
	(0.0097)	(0.0112)		

Gotong royong is measured on an identical three-point scale in PODES 2018, 2021 and 2024 (1 = present, most residents involved; 2 = present, few involved; 3 = no such custom); the item is absent or differently coded in earlier waves, so the window is three waves. Only cohorts first exposed in 2021 and 2024 are identified within it (942 villages, 1.8 percent of village-waves treated); villages exposed at or before 2018 are always-treated here and dropped. The calibration row shows that the conflict effect itself loses significance on this restricted sample under the preferred fixed effects, so these estimates are low-powered by construction. A Callaway–Sant'Anna estimate on the same item gives −0.0326 (standard error 0.0142, $p = 0.021$), and the single pre-treatment lead the three waves permit does not reject (+0.0222, standard error 0.0404, $p = 0.58$), though a test built on one lead has little power. The 2021 wave was enumerated during the COVID-19 period, which depresses collective activity in level; wave fixed effects absorb that shock where it is common across villages. Standard errors are clustered at the village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

(Appendix A). We carry the definition-stable *warga* measure throughout because its wording is far closer to comparable across the seven waves and its pre-trend is the milder of the two; Section S9 quantifies its residual 2005 incomparability at 3.8 percent of *warga* villages against 22.9 percent for the composite. We do not read it as a second estimate of the level effect, for which it is underpowered. The longer-run estimates show no reversal, but the later lags are imprecise nulls rather than evidence of persistence, so we read the absorbing average as attenuated by them rather than as showing that the post-earthquake state carries the effect on its own.

Simultaneous inference on the event-study path. One inference point belongs with any event study and we report it rather than leave it implicit (Baker et al., 2026). The intervals plotted in Figure 1, and those in every event study we show, are pointwise. A statement about the whole path of coefficients, which is what a pre-trend reading is, compares several estimates at once and therefore needs a critical value that accounts for the multiplicity. On the Callaway–Sant'Anna counterpart we compute that value by simulating from the estimated covariance matrix of the event-study vector: the 95 percent sup-t multiplier is 2.65 against the pointwise 1.96, so bands valid for the entire path are about 35 percent wider. The pre-period conclusion survives the stricter standard, since all three pre-exposure coefficients remain indistinguishable from zero under it, which is the direction that matters for reading pre-trends. One post-exposure coefficient does not: the estimate three waves after exposure is -0.0167 with a pointwise interval of $[-0.029,\ -0.004]$ but a simultaneous interval of $[-0.034,\ +0.000]$, so it is significant read alone and marginal read as part of the path. The effect in the wave of exposure and one wave later remain significant under both.

Under wave fixed effects the joint pre-test rejects in every estimator family we run; under province-by-wave effects it passes in every family that can absorb them. The verdict turns on the fixed effects, not on the estimator, which is what Section 5.1 predicts (Appendix A).

Conflict in neighbouring villages does not rise as exposed villages' conflict falls, so the decline is not displacement (Appendix S1).

S4.6 How Many Shocks? Stacked Event Design and Event-Level Inference

The design has 2,995 ever-treated villages, but they are not 2,995 independent assignments: villages in the same footprint receive the same shock. Assigning each village to the earliest earthquake exposing it to MMI≥VI, **19 distinct earthquakes** with at least twenty in-sample first-treated villages each account for 93 percent of the staggered-treated villages (Appendix S23.1).

Treating the earthquake rather than the village as the unit of assignment separates two estimands, and that distinction is the substantive point of this subsection. The *treated-village-weighted* average weights each earthquake's effect by the villages it first exposed and answers what happened to the average exposed village; it is −0.0268 (standard error 0.0105, earthquake bootstrap interval [−0.047, −0.007]). The *equal-event* mean counts each earthquake once and answers what the average earthquake does to its footprint; it is −0.0155 (standard error 0.0150), not distinguishable from zero, and its coincidence with the village-level ATT is numerical rather than substantive.

We claim the first, because the question posed in the introduction is what strong shaking does to the villages it strikes; on the second the paper establishes no decline, and we regard that as the honest ceiling on what the staggered result can claim. The pooled evidence is carried by the six large-footprint earthquakes, but that should not be read as a footprint gradient: estimated directly the slope of the per-event effect on log footprint is −0.0004 (standard error 0.0016, $p = 0.81$), estimated across the 46 events that clear the lower inclusion threshold rather than the 19 the figure displays, and weighted by the precision of each event's own estimate. Unweighted the slope is −0.0097 (0.0074, $p = 0.20$), still not distinguishable from zero. Larger events carry the pooled estimate through their weight; the six largest-footprint events do average −0.038 against −0.004 for the rest (Section S15), a step the linear slope does not detect and which we read as post hoc. The composition result does not depend on the choice: the theft-versus-*warga* divergence is positive and significant under *both* estimands (Section 7). Section S15 collects every estimand, weighting and inference procedure, Figure S2 shows all 19 event-specific estimates, and Section S15 gives the stacked construction and the

clustering, disjoint-control and aftershock-independence checks.

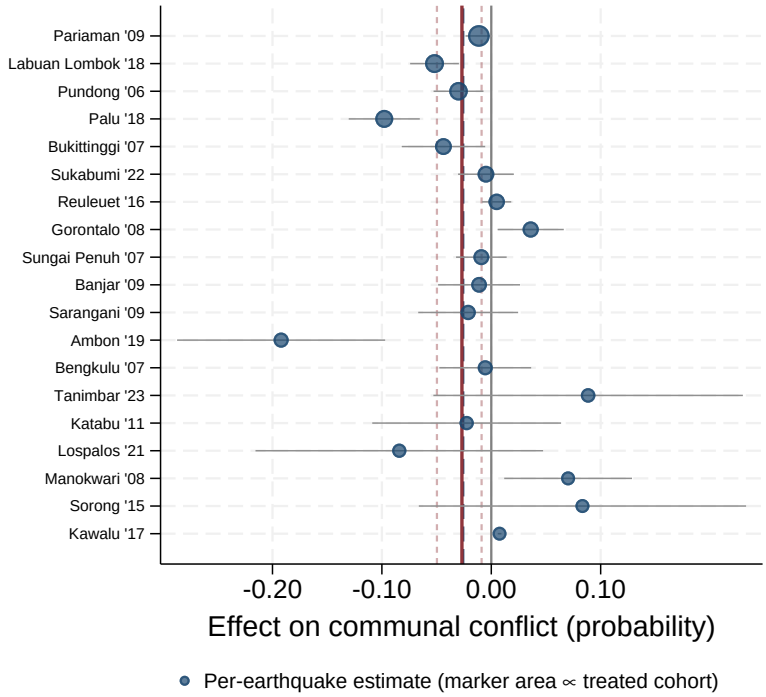


Figure S2: Event-Specific Difference-in-Differences Estimates
Note: Each row is one earthquake's own difference-in-differences estimate of the effect on communal conflict, with a 95% confidence interval. Rows are ordered by treated-cohort size and labelled by the earthquake's nearest place (USGS ShakeMap catalogue) and year; marker area is proportional to the treated cohort, counted as unique first-treated villages. The solid grey line is zero. The dashed navy line is the pooled stacked estimate (−0.0257), variance-weighted toward the large, sharply-identified shocks. The solid maroon line is the treated-village-weighted estimate (−0.0268), with its earthquake-level bootstrap 95% confidence interval shown as dashed maroon lines. Thirteen of the 19 estimates are negative, and the effect is concentrated in, though not exclusive to, the major earthquakes.

S4.7 Regional Trends and Forbidden Comparisons

Two objections apply to the absorbing design at once: the regional-trajectory problem of Section 5.1, and the already-treated comparison problem of de Chaisemartin and D'Haultfoeuille (2020), on which a Goodman–Bacon decomposition puts 3.8 percent of the weight, though those comparisons return −0.0376 against −0.0186 for the clean ones and so pull in the same direction.

The same critique can be put in terms of the weights the estimator places on the individual group-period effects rather than on comparison types, and that is the form in which de Chaisemartin and D'Haultfoeuille (2020) state it, so we report it too (Table S5). Of the 82,832 underlying effects, 32,347 receive positive weight

summing to 1.1775 and 50,485 receive negative weight summing to −0.1775, so negative weights are 13.1 percent of the gross weight mass. More effects carry a negative weight than a positive one, but the positive weights carry almost all of the magnitude. This bounds the concern rather than dismissing it: an estimate whose negative weights are this small cannot be produced by sign reversal alone, though the decomposition does not by itself establish convexity, which is why the heterogeneity-robust estimators and not this benchmark carry the paper.

Table S5: Weights the Two-Way Estimator Places on the Underlying Effects
Negative weights are a small share of the gross mass, so the estimate is not an artifact of sign reversal.

	Number of ATTs	Sum of weights
Positive weights	32,347	1.1775
Negative weights	50,485	-0.1775
Total	82,832	1.0000
Negative share of gross weight	13.10 percent	

Decomposition of the two-way fixed-effects coefficient into the weights it places on the underlying group-period average treatment effects, following de Chaisemartin and D'Haultfoeuille (2020). The weights are a property of the treatment design and the panel structure alone, not of the outcome, so a single decomposition characterises every outcome the paper reports; we verified that the composite returns identical weights. Negative weights arise because already-treated units serve as controls for later-treated ones. A decomposition in which the negative weights are a small share of the total, as here, means the estimate cannot be driven by sign reversal alone, though it does not by itself establish that the weighted sum recovers a convex average of the underlying effects.

Province×wave effects absorb the first; dropping the always-treated cohort removes the units with no pre-period and barely moves the estimate (−0.0199 to −0.0193), and the estimators of Section 5, none of which uses an already-treated unit as a control, answer the residual concern. The joint pre-trend clears in all four columns (Appendix Table S58), which it does not under wave fixed effects, and the kabupaten×wave column is reported for completeness but is uninformative, absorbing 62 percent of the remaining treatment variance (Appendix Table S59).

The effect is concentrated rather than general: inter-village conflict falls by 1.7 percentage points, 45 percent of its pre-exposure mean, while inter-group conflict within the village does not move detectably (−0.0035, standard error 0.0032, against a pre-exposure mean of 3.7 percent). That interval still admits a decline of about a quarter of the base, so we read it as an absence of evidence rather than

as a precise zero. That margin also decays, strongest within the first year and indistinguishable from zero beyond roughly six years, with earthquake fixed effects identifying each lag within the same event and the decay test rejecting equality between recent and distant bins (Appendix Table S60); neither the composite nor the mass-brawl gate shows a comparable profile, and identification here is exposed villages against themselves over time rather than exposed against unexposed.

S4.8 Design, Clustering and Spatial Structure

The estimate survives the exposure, clustering, and spatial choices: alternative MMI thresholds (≥V and ≥VII, though the ladder traces no dose-response), kabupaten clustering, Conley spatial-HAC errors, Oster (2019) coefficient stability under which adding controls moves the estimate *away* from zero, and spillover checks leave it intact, though a 25 km donut keeps the sign and loses precision (-0.0057, $p = 0.24$), while the conservative 28-cluster province bootstrap leaves the composite at 5 percent and *warga* short of it. Repeated exposure is rare and where it occurs we detect no additional effect (0.004, $p = 0.75$); statistical insignificance is not equivalence, so we read this as consistent with, not establishing, a first-shock state change (Sections S15 and S9).

S4.9 Conventional Staggered-DiD Benchmark: Callaway and Sant'Anna (2021)

Our conventional staggered-DiD benchmark is the group-time average treatment effect estimator of Callaway and Sant'Anna (2021), hereafter CS(2021):

$$ATT(g,t) = \mathbb{E}\Big[Y_t(g) - Y_t(0) \,\Big|\, G = g\Big], \tag{S2}$$

where $G = g$ denotes the cohort of villages whose first treatment wave is g, and $Y_t(0)$ is the potential outcome the same villages would have had at wave t absent any earthquake. Because it never uses already-treated villages as a comparison group, it avoids the contamination that biases conventional two-way fixed effects when treatment timing is staggered (Callaway and Sant'Anna, 2021; Roth et al., 2023). We aggregate $ATT(g,t)$ to a single average effect across post-treatment

periods using doubly-robust inverse-probability weighting (Sant'Anna and Zhao, 2020), taking villages not yet treated at each point in time as controls. Standard errors are analytical, based on the Callaway–Sant'Anna influence function. A wild cluster bootstrap is not available for this estimator, so its few-cluster check is province-clustered inference (Section 6.2); the fixed-effects estimates are checked with a wild cluster bootstrap (9,999 Webb-weight replications, seed 42).

Why absorbing treatment? An earthquake is an episodic physical event, so treating a village as exposed forever after its first strong shock requires justification: the framework of Section 3 predicts hysteresis, the precisely estimated response is concentrated in the first two waves while later estimates are imprecise, a profile that governs how the pooled estimate should be read rather than whether the absorbing definition is admissible, and the episodic design of Section S1.2 lets treatment switch off; Appendix S2 gives the full justification.

Implementation. Waves are indexed ordinally because the survey is unevenly spaced, and estimation uses the default pairwise-balancing rule of `csdid` on the unbalanced panel, which retains cohorts a fully balanced restriction would drop. All CS(2021) estimates use `csdid` version 1.72; Appendix S2 records the implementation details and the replication package the version of every user-written command.

Cohort composition. Villages first exposed before or during 2005 are always-treated in the 2005-boundary panel and are excluded from CS(2021) by the not-yet-treated control requirement. Cohorts G3–G8 (first treated 2008–2024) identify the estimator, with group-level ATTs for all six cohorts reported in Appendix Table S61. The exclusion is not a technicality at the margin: of the 13,092 in-sample villages ever exposed to damage-level shaking, 10,097, or three quarters, are always-treated, so the identified sample is the quarter first shaken after the panel opens and it contains neither the 2004 Sumatra–Andaman nor the 2005 Nias rupture. Because state capacity, reconstruction regimes, and conflict history differ across the period, treatment effects need not be homogeneous across cohorts; the estimand applies to cohorts entering treatment between 2008 and 2024, and we do not extrapolate it

to the excluded early cohorts.

Weighting and the estimand. One weighting choice shapes the estimand and we state it plainly, because it decides whose average the paper reports (Baker et al., 2026). Villages enter unweighted, so every group-time cell counts each exposed village equally and the target is the effect on the average exposed *village*, not on the average exposed *person*. Indonesian villages differ greatly in population, so the two are not the same quantity, and a population-weighted version would answer a different and equally reasonable question. We cannot report it. BPS stopped disseminating village population counts after the 2011 round, so population is structurally missing from 2014 onward and therefore unavailable for the 2018, 2021 and 2024 cohorts, which supply most of the identifying variation. Where population does enter, in the matched within-earthquake designs, we use the 2005 value fixed before any identified cohort is treated rather than a contemporaneous one, for the same reason the covariates are fixed at baseline.

The two-way fixed-effects benchmark and its Goodman–Bacon decomposition are set out in Appendix S1; no claim here rests on it.

The episodic estimand of de Chaisemartin and D'Haultfoeuille (2026), which lets treatment switch off, is described in Appendix S1.

S4.10 Sample Restriction: Seismic Hazard Zones

The analysis is restricted to villages in Indonesia's Kawasan Rawan Bencana (KRB) Medium or High seismic-hazard zones (Appendix Figure S4), mapped by the Center for Volcanology and Geological Hazard Mitigation (PVMBG/ESDM) and consistent with Indonesia's official national seismic-hazard assessment (Irsyam et al., 2020). Limiting to KRB zones ensures that all included villages have non-trivial earthquake exposure, so the not-yet-treated comparison villages are drawn from the same seismic risk set rather than from areas that face no earthquake hazard at all. The restricted sample covers Sumatra, Java, Bali, Nusa Tenggara, Sulawesi, Maluku, and Papua. Of the 63,115 villages in the 2005-boundary panel, 43,351 fall in a Medium or High zone and 19,362 do not; a further 402, six tenths of one percent, have no KRB record at all, because the polygon does not intersect the

published map, and we drop them rather than assign them to either side.

A natural concern is that the hazard classification is endogenous to the earthquakes we study. It is not: the KRB map is a hazard assessment based on fault and subduction-zone geometry rather than realized damage (Irsyam et al., 2020), the restriction enters the design only through village fixed effects, which absorb any time-invariant hazard level, and dropping it entirely and re-estimating on the full national sample leaves the effect intact (Appendix S2). Appendix Figures S5 and S6 map the treated villages by cohort and place the cohorts on the survey calendar.

S4.11 Alternative Conflict Measure: ACLED

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

S5 ACLED Mob-Violence Events: Full Coding

S5.1 What the ACLED Comparison Can and Cannot Show

ACLED offers a useful but imperfect external check on a village-head-reported outcome. It covers Indonesia only from 2015, so it overlaps PODES in the 2018–2024 waves and cannot validate the main 2009 Padang cohort. In the overlap, PODES communal conflict falls sharply (-3.97^{***} percentage points) while ACLED events rise slightly (Appendix Table S6); but reading the 24 treated-village mob-violence events individually shows that two-thirds (16 of 24) are not inter-group communal fighting but vigilante, moral-policing, or crime-accusation actions against individual targets, several of them mob responses to alleged theft, the scarcity-channel sequela of the rising-theft result (Miguel, 2005). The divergence is what the two measurement technologies predict and weakens the reporting-artifact reading:

ACLED's media-driven counts rise where post-disaster coverage concentrates, whereas PODES is a census independent of media attention. ACLED therefore neither replicates nor overturns the village-level result; it captures a different, media-visible tail of disorder. The full event coding appears in Table S7 below. Section S20 reports why the SNPK/NVMS incident database (2005–2014, Barron, Jaffrey, and Varshney 2016), the one violence series whose window reaches the 2009 Padang cohort, cannot serve as an external check here: its reporting expands sharply over the sample period, and the estimates it yields reverse sign when the same specification is estimated in proportional rather than level form.

Table S7 lists all 24 ACLED mob-violence events (precision-1 geolocation) occurring in treated village-waves, extracted and coded into four categories from the ACLED event descriptions: **V** = vigilante, moral-policing, or crime-accusation action against individual targets; **E** = election-related crowd violence; **IG** = genuine inter-group clash; **O** = other (neither). The full ACLED notes for every event ship with the replication package.

Table S6: External Comparison: ACLED vs. PODES Conflict Outcome, PODES Waves 2018–2024

	(1) PODES communal	(2) ACLED any	(3) ACLED mob violence	(4) ACLED violent demonstration	(5) ACLED VAC	(6) ACLED battles
D_{treat}	-0.0397***	0.0047**	0.0027*	0.0024	0.0017	-0.0002***
	(0.0097)	(0.0022)	(0.0016)	(0.0015)	(0.0013)	(0.0001)
Baseline mean	0.024	0.002	0.001	—	—	—
Observations	99,762 (identical subsample)					
Village + wave FE	Yes					

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

Table S7: ACLED Mob-Violence Events in Treated Village-Waves, Coded

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

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

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

S6 Village Boundary Harmonization

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

Table S8: From the 2005 Village Universe to the Analysis Sample

	Villages	%
Villages in the 2005 Master File Desa (universe)	69,957	
enumerated in PODES 2005	68,637	98.1
in KRB Medium/High hazard zones (risk set)	45,450	65.0
observed in all seven waves	42,333	93.1
less those with an incomplete conflict series	-231	
with a complete conflict series	42,102	92.6
Plus post-2005 units not linked to a 2005 parent	1,249	
Analysis sample	43,351	

The universe is the 2005 Master File Desa, not the set of parent identifiers in the constructed panel: the latter contains 1,249 units in the analysis sample (and 13,138 overall) that correspond to villages created after 2005 whose chain linkage did not resolve to a 2005 parent; these enter the panel under their own identifier and are retained, since the design needs only a consistent unit across waves, not a 2005 ancestor. The 231 villages observed in all seven waves but excluded have at least one wave with a missing conflict outcome, and the estimation sample requires a complete series. Percentages are shares of the universe except the last two, which are shares of the risk set.

Procedure. We harmonize all waves to constant 2005 parent boundaries with a chain-linkage crosswalk built from BPS Master File Desa (MFD) records: each wave's village code is linked to its immediate predecessor code, and the chain is followed back to the 2005 parent. The chain is code-based throughout for the core waves, with a name bridge used only for the 2024 wave. The dominant method is a direct MFD chain (single- or multi-step, resolving to the 2005 code, 82 percent of wave-codes), complemented by the 2005 baseline self-match, an inferred unique-

child map for residual parent codes, and a documented fallback for codes absent from the MFD vintage. Outcomes are aggregated to the parent by the maximum (binary outcomes) or sum (counts), so a parent reports conflict if any constituent village does; placebo and demographic variables use means (Appendix S11).

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

Unresolved parents and the coverage ceiling. Broken chains have a second consequence worth stating explicitly. When the chain cannot resolve a post-2005 village to a 2005 parent, that village enters the constructed panel as its own parent identifier. The panel therefore contains 13,138 parent units that do not correspond to any village in the 2005 Master File Desa. Most lack a complete wave series and are dropped by the balanced-panel requirement, but 1,249 of them are observed in all seven waves, including 2005, and are retained: the design needs a unit that is consistent across waves, not one with a 2005 Master File ancestor. This is why Table S8 takes the 2005 Master File Desa as its denominator rather than the set of parent identifiers in the panel: the latter exceeds the 2005 village universe by roughly ten thousand units and would understate coverage. Of the villages missing only the 2005 wave, 637 are on Nias, which PODES 2005 did not enumerate after the 2004 tsunami and which Cassidy (2025) likewise excludes; a further 1,180 appear in the 2005 Master File Desa but were not enumerated, so their outcomes are genuinely absent.

We also asked whether the remaining unresolved parents can be folded back by name. A layered procedure matching on exact village name within subdistrict and then within district, followed by name-root matching, folds 2,220 of the 13,138 units and would raise coverage of the risk set from 93.1 to 94.7 percent. We do not adopt it: 6.6 percent of the links it produces are ambiguous, because several 2005 villages

share a name within the same district, and a mislinked village injects measurement error into both treatment and outcome, which is worse than dropping it. The reported coverage is therefore a deliberate floor rather than a ceiling imposed by the data.

Boundary geometry vintage. The match rates above concern village identifiers, which are distinct from the underlying polygon geometry used for the spatial join to ShakeMap MMI and the KRB hazard zones. No BPS shapefile digitized contemporaneously with 2005 exists, so we build the primary boundary layer by dissolving the BPS 2014 administrative shapefile onto 2005 parent identifiers: a direct dissolve covers 90.5 percent of villages, and a layered fallback (borrowing, in order, from alternate-vintage Master File Desa codes, then from an earlier 2011-shapefile-based boundary layer, then from the 2003-parent boundary layer, each fallback validated against the province recorded in PODES GPS coordinates) raises cumulative coverage to 95.9 percent (65,794 of 68,637 villages). The alternate-vintage layer recovers 3,123 villages and the 2003-parent layer 528; the 2011-shapefile layer recovers none on this vintage, leaving 2,843 villages, 4.1 percent, without a polygon. The 2003-parent crosswalk check reported below (Cassidy (2025)-style robustness) uses an analogous boundary layer dissolved from the BPS 2011 shapefile. Because the treatment variable is a spatial join of raster MMI values to these polygons, using the most accurate digitized shapefile available, rather than a shapefile contemporaneous with the label year, is the more accurate choice for measuring exposure.

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

Crosswalk robustness. Two re-estimations verify that the harmonisation rules do not drive the result. First, a never-split sample drops every parent that ever

maps from more than one raw village, eliminating any influence of the collapse rule: the TWFE estimate is -0.0170 ($t = -4.74$, $N = 207{,}067$) against -0.0193 ($t = -5.86$, $N = 232{,}778$) on the corresponding full sample, and the CS(2021) ATT is -0.0157 ($p = 0.001$). The subsample estimate is marginally smaller in magnitude and less precisely estimated, as one would expect after discarding an eighth of the observations, but it is not attenuated toward zero in any way that would indicate the collapse rule is generating the result. Second, dropping the parents whose chain ever used the fallback method leaves the estimate essentially unchanged (-0.0182, $t = -5.69$). The harmonisation machinery is not what produces the finding.

S7 PGA-Derived MMI: An Independent Exposure Measure

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

Building this measure required correcting a unit error in the PGA raster-to-village pipeline. ShakeMap distributes some PGA grids in $\ln(g)$ and others linearly in $\%g$; the earlier pipeline exponentiated the log grids correctly but then treated the resulting values (in g) as $\%g$, understating PGA on those grids by a factor of one hundred, a constant $3.66 \log_{10}(100) \approx 7.3$ MMI-point deficit that rendered the derived intensities uninformative. After the fix, the PGA-derived MMI tracks ShakeMap MMI closely among shaken villages, and the two exposure indicators correlate 0.79 at the village-wave level. Across all matched village-earthquake cells, including the sub-damaging comparison villages, the median discrepancy is 0.23 of an intensity unit; Section 8 uses that discrepancy as a conditioning variable.

Re-estimating the main design on this independent measure, holding the window-aligned first-exposure timing and all else fixed, reproduces the result across every specification (Table S9): the CS(2021) benchmark estimate is -0.0129 ($p = 0.001$)

against the ShakeMap-MMI -0.0177, and the definition-stable *warga* outcome is likewise negative and significant. Both columns of that table are computed in a single run on the identical panel, so the comparison is exact rather than assembled across specifications. The communal decline is therefore not an artifact of ShakeMap's intensity construction.

Table S9: Measurement Robustness: PGA-Derived MMI
An intensity measure built from instrumental PGA alone reproduces the result across every specification.

	ShakeMap MMI (primary)	PGA-derived MMI (Wald 1999)
CS(2021) ATT, composite	-0.0177^{***}	-0.0129^{***}
	(0.0047)	(0.0039)
Absorbing TWFE, no controls	-0.0199^{***}	-0.0128^{***}
	(0.0033)	(0.0030)
Absorbing TWFE, with controls	-0.0217^{***}	-0.0142^{***}
	(0.0034)	(0.0030)
Definition-stable *warga*, TWFE controls	-0.0078^{***}	-0.0044^{*}
	(0.0027)	(0.0025)
Ever-treated villages	2,995	4,047
Observations	292,712	292,712

Notes: The right column replaces ShakeMap Modified Mercalli Intensity with an MMI derived independently from ShakeMap Peak Ground Acceleration through the Wald et al. (1999) relationship, $\text{MMI} = 3.66 \log_{10}(\text{PGA}) - 1.66$, keeping everything else fixed: the same window-aligned first-exposure timing, the same village and wave fixed effects, and village-clustered standard errors. First exposure is the first survey wave enumerated after a village's first PGA-derived MMI $\geq$ VI event, exactly as for the ShakeMap-MMI treatment. The PGA and ShakeMap-MMI treatments identify nearly the same villages (4,047 vs 2,995 ever-treated; the two exposure indicators correlate 0.78 at the village-wave level), and every estimate reproduces the main negative effect at comparable or slightly larger magnitude, including the definition-stable *warga* outcome. The PGA-derived CS(2021) estimate is significant ($p = 0.001$) against $p = 0.000$ for ShakeMap MMI, reflecting the wider standard error of the noisier instrumental measure rather than a smaller point estimate. This establishes that the result is not an artifact of ShakeMap's MMI construction, which blends instrumental and macroseismic intensity, as opposed to the purely instrumental PGA. An earlier calendar-year-binned PGA measure produced a spurious positive coefficient, an instance of the same survey-window timing artifact documented in Section 6; it disappears once PGA exposure is aligned to the survey calendar. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

S8 Covariate Trajectories: CS(2021) on Observables

Table S10: Covariate Event-Study Aggregates: Pre- and Post-Treatment Means
Baseline: 2005-boundary panel, CS(2021), always-treated cohort dropped

Covariate	Pre-treatment aggregate	p	Post-treatment aggregate	p
Any household electricity	-0.0155^{***}	0.001	0.0044	0.246
Paved road access	0.0284^{***}	0.007	-0.0599^{***}	0.000
Internet available	-0.0665^{***}	0.000	0.0346^{***}	0.000
Police post present	-0.0043	0.599	-0.0075	0.234
Number of health centres	-0.0057	0.696	-0.0422^{**}	0.020
Health centre present	0.0069	0.531	-0.0079	0.238
Number of primary schools	-0.2243^{***}	0.000	-0.0940^{***}	0.003
Distance to district capital	0.3841	0.460	1.4717^{***}	0.000

Each row runs the benchmark Callaway–Sant'Anna specification with the covariate as the outcome and no controls, on the analysis sample with the always-treated cohort dropped. A significant pre-treatment aggregate indicates that the covariate itself trends differentially before exposure, which is why the design conditions on covariates rather than assuming unconditional parallel trends. Population and household counts cannot be tested here: BPS stopped disseminating them after the 2011 wave, so they are missing for four of the seven waves and the event-study aggregate is not estimable. Their pre-treatment growth is examined separately over 2005–2008 in Supplementary Material Section S7. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S9 Alternative Clustering and Outcome Comparability

S9.1 The composite definition break at 2008

The composite outcome is an indicator equal to one if the village reports any intergroup conflict or mass brawl. The two definitions do not nest, and the direction of the resulting incomparability is not the obvious one. In 2005 the composite is the bare mass-brawl gate, so a village that reports a brawl enters regardless of the type it names; of the 1,588 gate-positive villages that wave, 363, or 22.9 percent, name a non-communal type or none at all and would be coded zero under the post-2008 rule. From 2008 the composite instead requires a positive count in one of three communal subtypes, which sets 258 of that wave's 2,278 gate-positive villages, or 11.3 percent, to zero. The 2005 measure is therefore narrower

in question wording but broader in what it counts. Wave fixed effects absorb the level shift this produces, but not any part of it that moves with treatment, so we address the problem directly in two ways: with a narrower subtype whose wording is near-stable across all seven waves, namely a civilian-versus-civilian brawl, which we call the *warga* outcome throughout, and with a robustness check reported below (Table S12). The revised manuscript goes one step further and leads with the gate item itself, whose wording is identical in all seven waves; the composite and the *warga* subtype are carried beside it, and this section keeps the comparability record that motivated the change. Closer, not identical: the 2005 item records only the most frequent conflict type, so a village reporting a *warga* brawl alongside a more frequent other type is coded zero, where from 2008 any positive *warga* count suffices. Applying the 2005 rule to the 2008 microdata, where subtype counts are observed, would drop 46 of 1,225 *warga* villages, 3.8 percent, against the 22.9 percent the same exercise costs the composite.

Table S11: Robustness: Kabupaten-Level Clustering
Takeaway: under kabupaten clustering the composite remains significant at the 1 percent level and warga *at the 10 percent level; under the more conservative province-level bootstrap the composite still clears 5 percent while* warga *does not.*

	Village cluster (32,196 clusters)	Kabupaten cluster (320 clusters)	Province cluster (28 clusters)
Outcome: composite			
D_{treat}	-0.0213^{***}	-0.0213^{***}	-0.0213^{***}
	(0.0034)	(0.0066)	(0.0076)
t-statistic	-6.18	-3.24	-2.79
Outcome: warga			
D_{treat}	-0.0073^{***}	-0.0073^{**}	-0.0073
	(0.0027)	(0.0037)	(0.0051)
t-statistic	-2.71	-1.99	-1.45
Observations		225,372	
Village + wave FE		Yes	

Point estimate is identical across columns by construction (same coefficient, only the clustering level of inference changes). All specifications include village and wave fixed effects and the baseline covariates interacted with wave dummies (see Supplementary Material Section S7). Because the province level has few clusters, we also report wild cluster bootstrap p-values (Webb weights, 9,999 replications): $p = 0.030$ (conflict_any) and $p = 0.215$ (conflict_warga), consistent with the analytic inference. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table S11 re-clusters standard errors at the kabupaten (district, 320 clusters) and province (28 clusters) level, for both conflict_any and conflict_warga. The point estimate is identical across columns by construction; only the level of inference changes. Village- and kabupaten-level clustering leave the composite estimate significant. Province-level clustering is a deliberately conservative check with few clusters and a correspondingly large loss of precision: the two-way fixed-effects composite still clears 5 percent analytically ($t = -2.79$) and under a wild-cluster bootstrap ($p = 0.030$), while *warga* does not ($p = 0.215$), and the pooled CS(2021) ATT still clears 5 percent under province clustering ($p = 0.043$). This pattern, in which standard errors rise monotonically from village to kabupaten to province, is

exactly what within-region spatial correlation in earthquake exposure and conflict predicts, and it does not overturn the level estimate at any clustering level.

A subtler comparability concern is that the composite *conflict_any* outcome combines subtypes (warga, desa, suku) that become separately available only from 2008 onward, while the 2005 wave uses the mass-brawl (*perkelahian massal*) item directly. Table S12 addresses this on the absorbing windowed-first-exposure design. Column (1) reproduces the *conflict_any* level estimate (-0.0213^{***}). Column (2) replaces the composite outcome with *conflict_warga*, the civilian-vs-civilian conflict indicator, which is constructed from a single consistent questionnaire item in every wave; the coefficient is smaller in magnitude but remains negative and significant at the 1 percent level (-0.0073^{***}), indicating the effect is not an artifact of the composite definition and is not driven solely by the inter-village or inter-ethnic subtypes that become available only from 2008 onward. Column (3) is the gate item itself (-0.0182^{***}), the measure the manuscript now leads with; measured against its own pre-exposure mean it falls by less than the composite, which is what a definition that narrows over the panel mechanically delivers for the composite and is the reason the manuscript treats the gate, not the composite, as primary. Within the analysis window (2005–2024) the sign of the result is therefore not an artifact of how conflict is defined.

Table S12: Robustness: Outcome Variable Comparability

Takeaway: all three outcomes decline, including the enumerator's gate question in column 3, whose wording is identical in every wave from 2005 to 2024 and which the manuscript now treats as the primary outcome. The sign of the result is therefore not an artifact of how conflict is defined. Part of the composite's magnitude is: measured against its own pre-exposure mean, the gate falls by about two thirds as much as the composite, because the composite narrows from the bare gate in 2005 to communal subtypes only from 2008.

	(1)	(2)	(3)
	Composite	*Warga*	Any mass brawl
D_{treat}	-0.0213***	-0.0073***	-0.0182***
	(0.0034)	(0.0027)	(0.0036)
t-statistic	-6.18	-2.71	-5.00
Observations	225,372	225,372	225,372
Village + wave FE	Yes	Yes	Yes

Notes: Standard errors clustered at village level in parentheses. All specifications use the absorbing D_{treat} (windowed first exposure) and include village and wave fixed effects plus the baseline-covariate-by-wave control set. Column 1, conflict_any, is the composite outcome (warga + antar-desa + suku) used in the headline result. Column 2 restricts to civilian-vs-civilian conflict, the subtype defined consistently across all PODES waves (see Section 4). Column 3 is the enumerator's gate question itself, recording whether any mass brawl occurred regardless of type, and so also counts brawls against security or government personnel, student brawls, and other types that columns 1 and 2 exclude by construction. It is the only one of the three whose definition is identical in every wave from 2005 to 2024: in 2005 the composite is the bare gate, whereas from 2008 it is restricted to communal subtypes, so the share of gate-positive villages that the composite records as zero rises from 11.3 percent in 2008 to 33.7 percent in 2024. On the same estimation sample its mean is 0.0333 against 0.0273 for the composite, so the estimate in column 3 is 54.8 percent of its own base while column 1 is 77.8 percent of its own. We report it because a decline measured on a narrowing definition would overstate the effect, not because it resolves the pre-trend evidence discussed in Section 5, which it does not. * p<0.1, ** p<0.05, *** p<0.01

S10 Composition of Disorder: Identifying Assumption and Benchmarks

This appendix states the identifying assumption behind the composition contrast of Section 7.1, names the threats that survive the differencing, and reports the benchmark that estimates each outcome separately.

The contrast carries its own identifying assumption, and we state it rather than present the design as assumption-free. Estimating the outcomes on a common sample with outcome-interacted fixed effects is equivalent to estimating the same difference-in-differences on a single differenced outcome, theft minus communal conflict. What it requires is parallel trends in that difference: absent the earthquake, the gap between reported theft and reported communal conflict in treated villages would have evolved as it did in villages not yet exposed, and exposure changes the reporting of the two items, if at all, only by a common amount. This is weaker than the level effect's assumption in one specific respect. Any village-wave shock that moves both items equally differences out, which covers a uniform disruption of the village head's reporting, post-disaster enumeration conditions, and mechanical depopulation. Three threats survive the differencing and we name them. A shock correlated with exposure that moves the two items differentially does not cancel: an incentive to document property losses for aid claims would inflate reported theft without inflating reported communal conflict, and what weighs against that account is that robbery, also a property loss on the same form, does not rise with theft. That comparison assumes the documentation incentive loads similarly on the two items, which is not automatic: an unwitnessed loss is easier to claim than one requiring an account of a violent encounter, so we treat this as weighing against the account rather than closing it. Because theft's pre-exposure rate is roughly ten times the *warga* rate, a reporting shock proportional to each item's level does not cancel in an absolute difference; the earthquake-level equivalence bound on *warga* excludes a proportional surge large enough to produce theft's rise. And the differenced outcome has a pre-trend of its own. We test it directly and report the result specification by specification, because it does not point the same way in both. Under the preferred province-by-wave effects the joint test over the interior

pre-exposure leads does not reject ($p = 0.31$ against the composite, $p = 0.18$ against the definition-stable series), and the two leads nearest exposure are precise nulls of -0.001 and -0.009. Under wave effects the same test rejects ($p = 0.018$ and $p = 0.040$). In both, the binned term collecting periods five waves and more before exposure is positive and significant, at $+0.045$ (standard error 0.020) under the preferred effects, so a joint test that includes it sits at $p = 0.06$. We do not claim the contrast dispenses with pre-trend concerns. We claim something narrower: under the specification our selection rule prefers, the near pre-exposure period is flat, and what remains is a level difference at horizons so distant that a single cohort supplies most of the cell.

An earlier draft of this paper reported a rejection at $p < 0.001$ for this object and rested the composition result on the earthquake-level divergence instead. Two things changed. The endpoints are now binned rather than dropped, which is the standard treatment and which we should have applied from the start. And the crime module turns out to be fielded in all seven waves rather than five: the two missing waves were absent from our extraction, not from the survey (Section 4.2). Recovering them removes the ten-year hole between 2008 and 2018 that made event time on this sample non-comparable in calendar time, and raises the sample from 211,082 to 300,018 village-waves. The earlier rejection was in part an artifact of that hole. We report the corrected tests here and in Table S13, and we continue to report the earthquake-level divergence alongside them rather than in place of them.

S10.1 The definition-stable anchor

Because the composite communal measure changes definition in 2008, we anchor the contrast on the fifth outcome, *warga*, the civilian-versus-civilian communal subtype whose survey wording is near-stable across waves. The *warga* coefficient on its own is -0.0085 (standard error 0.0027, $p = 0.002$), a decline of just over half a percentage point on a 1.6 percent base, and it survives the Benjamini–Hochberg correction across the five outcomes ($q = 0.002$). A test of the equality of the *warga* and theft effects is rejected ($p < 0.001$), and the rejection holds under kabupaten clustering ($F = 5.04$, $p = 0.025$). The composition result therefore does not depend

on the composite outcome or on its 2008 definition change: the definition-stable communal series falls while theft does not, and the equality of their effects is formally rejected.

S10.2 The contrast in event time

The first-time margin fails a falsification run inside the pre-exposure period, and the same test has to be put to the contrast this paper rests on. It does better, but not cleanly, and we report both halves. Table S13 traces the theft-minus-communal difference in event time with the endpoint cells binned rather than dropped, so the reference category is the wave before exposure alone and the coefficient at t_0 is itself the jump at exposure. Under the province-by-wave effects our selection rule prefers, the three interior pre-exposure coefficients are $+0.027$, -0.009 and -0.001 and their joint test does not reject ($p = 0.31$); the difference then rises to $+0.025$ in the wave of exposure and $+0.040$ in the wave after. The definition-stable version behaves the same way in the pre-period ($p = 0.18$). Under wave effects the same interior test rejects ($p = 0.018$) while the jump at exposure is larger and sharper ($+0.034$, $p = 0.007$).

What does not settle down in either specification is the binned term collecting periods five waves and more before exposure, which is $+0.045$ (standard error 0.020) under the preferred effects. A joint test that includes it sits at $p = 0.06$. That term is a single cohort observed in a single calendar wave nineteen years before its own exposure, which is the configuration endpoint binning exists to handle, and we regard it as weak evidence of a pre-exposure level difference rather than of a trend. But it is evidence, and we report the joint test both ways rather than only the version that passes.

We also correct the record on how this table came to read as it does. Earlier drafts of this paper reported a rejection at $p < 0.001$ for this object and explained it as contamination of the reference category. That explanation was wrong: the earlier specification gave every event time its own dummy and pooled nothing. Two things actually changed. The endpoints are now binned rather than dropped, which is the standard treatment; and the crime module turns out to be fielded in all seven waves rather than five, so the ten-year hole between 2008 and 2018 that

made event time non-comparable in calendar time is gone (Section 4.2). Before that hole was closed, the cells being differenced drew on largely non-overlapping cohorts. They no longer do.

Table S13: The Composition Contrast Before and After Exposure
The near pre-exposure period is flat under the preferred specification; the distant binned endpoint is not.

	Theft − communal		Theft − *warga*	
	Wave FE	Prov.×wave	Wave FE	Prov.×wave
$t_{\leq -5}$ (binned)	0.0545	0.0454	0.0416	0.0419
	(0.0172)	(0.0202)	(0.0169)	(0.0199)
t_{-4}	0.0053	0.0266	0.0041	0.0333
	(0.0171)	(0.0192)	(0.0166)	(0.0187)
t_{-3}	-0.0347	-0.0091	-0.0335	-0.0053
	(0.0167)	(0.0187)	(0.0159)	(0.0179)
t_{-2}	-0.0301	-0.0008	-0.0232	0.0009
	(0.0132)	(0.0154)	(0.0126)	(0.0148)
joint pre-trend p, interior leads	0.018	0.308	0.040	0.182
joint pre-trend p, incl. binned bin	0.000	0.061	0.000	0.066
t_0	0.0335	0.0246	0.0194	0.0126
	(0.0125)	(0.0146)	(0.0121)	(0.0141)
t_{+1}	0.0685	0.0402	0.0543	0.0259
	(0.0128)	(0.0155)	(0.0125)	(0.0150)
p on t_0 (the jump from the t_{-1} reference)	0.007	0.091	0.109	0.371
Memo: change t_{-2} to t_0	0.0636	0.0254	0.0426	0.0117
p	0.000	0.108	0.001	0.447

Event-study coefficients on the differenced outcome, the object the composition result rests on. The reference period is t_{-1}, and the endpoints are binned rather than dropped, so the coefficient on t_0 is itself the jump at exposure. We report the joint pre-trend test twice: over the interior leads, and again including the binned far-lead term, because excluding that term is what made an earlier draft of this table read as a clean pass. The memo row gives the change from t_{-2}, which is the most negative lead, and is therefore an upper bound on the movement at exposure rather than the movement itself. Standard errors, clustered at the village, in parentheses. N = 229,871 village-waves on the crime-module common sample, all seven waves.

S10.3 The divergence at the earthquake level, worked through

The panel evidence above and the earthquake-level divergence of Section 7.1 answer different questions, and a reader is entitled to ask them together. The earthquake-level estimate fixes the unit of inference, since it treats the earthquake rather than the village as the assignment, but it does not by itself impose the province-specific

counterfactual this paper prefers elsewhere. Correct inference and a credible counterfactual are not the same requirement, and we test them jointly rather than leave the gap open. Table S14 reports two ways of doing so. The first re-estimates D_e event by event with province-by-wave rather than wave fixed effects. The second compares, within the same earthquake, villages shaken at MMI≥VI against villages that felt that same shock at MMI in $[4, 6)$, taking theft minus *warga* as the dependent variable directly, so that province-specific trajectories difference out by construction rather than by assumption.

The divergence stays positive under both. It does not stay significant under either. Before reading that as a failed test rather than an imprecise one, note what the same within-earthquake design says about its own pre-period: running the differenced outcome on pre-exposure waves only, the placebo is +0.004 (standard error 0.015), which does not reject under village clustering ($p = 0.77$) and does not reject under earthquake clustering either ($p = 0.87$). The pre-period is therefore clean, and the reason the divergence loses significance under these two restrictions is precision rather than contamination: with earthquakes as the clusters the within-earthquake estimate itself reaches only $p = 0.47$. We read the contrast as uninformative in either direction rather than as a failed test, and we do not claim the architecture validated on this evidence. The treated-village-weighted mean falls from the +0.076 of Section 7.1 to +0.0155 with province-by-wave effects, and its bootstrap interval contains zero. The within-earthquake design in Panel B gives +0.0221 (bootstrap standard error 0.0332), positive in 13 of 18 earthquakes, with an equal-event mean of +0.0476 ($p = 0.05$). Pooling over all stacks gives +0.0217 (standard error 0.0100, $p = 0.03$) under village clustering and the same coefficient with $p = 0.47$ once the earthquake, which is the unit of assignment, is the cluster. We read the last of those as the governing number.

How much power this design has is a question we can answer rather than assert, and we answer it with the calculation we apply to the onset margin in Supplementary Material Section S14. At 80 percent power against a two-sided test at 5 percent, the smallest effect a specification can reject is 2.802 times its standard error. Taking the earthquake as the cluster, that minimum detectable effect is 0.0832 for the pooled within-earthquake estimate, 0.0930 for its treated-

village-weighted counterpart and 0.0723 under province-by-wave effects, against point estimates of 0.0217, 0.0221 and 0.0155. Every one of these specifications is therefore between 3.3 and 4.7 times too coarse to detect an effect the size of the one it produces, and even the tightest of them would need an effect nearly twice its own estimate before it would reject. Their failure to reject consequently carries little information about whether the divergence is present within earthquakes. Part of this is mechanical, since the median earthquake reaches only one or two provinces and the fixed effects absorb most of the within-event variation, and the within-earthquake design retains fifteen of 24 earthquakes once both sides need twenty villages. But we do not read the loss of significance as an artifact of power alone, and we do not claim the exercise rescues the result. What it establishes is narrower and worth stating plainly: the composition contrast is identified from comparisons across regions, it is positive in sign wherever we look, and it is not demonstrated within earthquake at conventional levels. The evidence for recomposition rests on the consistency of that sign across weightings, designs and inference units, not on a within-earthquake estimate that clears a threshold.

The *warga* arm of the divergence deserves its own statement. The *warga* response is not merely indistinguishable from zero: it is negative and separated from zero at the shock level. Estimated on the full *warga* sample rather than on the narrower crime-module common sample that the divergence uses, its treated-village-weighted effect is -0.0114 with an earthquake block-bootstrap 95 percent interval of $[-0.0260,\ -0.0007]$, which excludes zero; the corresponding arm inside Table S15, computed on the common sample, is -0.008. Civilian communal violence therefore falls by about a third of its 3.7 percent pre-exposure rate under earthquake-level inference. The finding that it does not rise is therefore an informative equivalence bound, not a mere failure to reject.

We do not attribute this to over-absorption; the over-absorption arithmetic below sets out why.

What survives the paper's own specification rule is therefore the narrower claim: theft and communal conflict move in opposite directions and differ significantly under both fixed-effect structures, while the sharper contrast with the definition-stable *warga* series holds under wave effects and not under province-by-wave. The

Table S14: Composition Divergence at the Earthquake Level, under the Preferred Counterfactual

	Panel A Province×wave	Panel B Within-earthquake
Earthquakes with support	17	18
Earthquakes dropped for lack of support	2	6
Treated-village-weighted mean D_e	+0.0155	+0.0221
bootstrap standard error	(0.0258)	(0.0332)
bootstrap 95% CI	[−0.0267, +0.0755]	[−0.0290, +0.1014]
leave-one-earthquake-out range	[−0.0009, +0.0322]	[+0.0000, +0.0445]
Equal-event mean D_e	+0.0202	+0.0476
two-sided p	0.38	0.05
events with $D_e > 0$	8 of 17	13 of 18
Pooled over all stacks, village-clustered	—	+0.0217 (0.0100), $p = 0.03$
earthquake-clustered	—	+0.0217 (0.0297), $p = 0.47$

The paper's headline composition result estimates the divergence $D_e = ATT_e^{\text{theft}} - ATT_e^{\text{warga}}$ earthquake by earthquake under village and wave fixed effects, and obtains +0.076 with a bootstrap standard error of 0.033. This table imposes the counterfactual the paper prefers elsewhere. Panel A repeats the event-by-event estimation with province-by-wave rather than wave fixed effects; because the median earthquake reaches only one or two provinces, the fixed effects absorb most of the within-event variation and two earthquakes lose support entirely. Panel B avoids that problem by comparing, within the same earthquake, villages shaken at MMI≥VI against villages that felt the same shock at MMI in $[4, 6)$, with the differenced outcome (theft minus *warga*) as the dependent variable, so province-specific trajectories difference out by construction; earthquakes with fewer than twenty treated or twenty felt-control villages are dropped. The divergence stays positive in sign in every column and under both weightings. The treated-village-weighted intervals contain zero in both panels. Within earthquake the equal-event mean and the village-clustered pooled estimate do reach conventional significance, but the earthquake is the unit of assignment, and clustering there returns $p = 0.47$. We therefore report this as a limit on the composition result, not as support for it.

main text states the claim at that width.

Table S15: Composition at the Shock Level: Theft versus *Warga* per Earthquake
*The theft-versus-*warga *divergence, aggregated to the earthquake, clears earthquake-level inference under both weightings, where the communal decline alone does not.*

	Theft	Warga	Divergence
Treated-village-weighted mean	0.068	-0.008	0.076
Equal-event mean	.	.	0.063
Equal-event two-sided p-value	.	.	0.023
Bootstrap standard error	.	.	0.033
95% CI lower bound	.	.	0.010
95% CI upper bound	.	.	0.142
Two-sided bootstrap p-value	.	.	0.025
Leave-one-out min	.	.	0.057
Leave-one-out max	.	.	0.095
Events with D_e greater than 0	.	.	12.000
Usable earthquakes (total)	.	.	19.000

For each of the 19 usable earthquakes (onset PODES wave 2008-2024, at least 20 in-sample first-treated villages) we estimate the absorbing treated-cohort effect on theft and on the definition-stable warga outcome against never-treated controls, with village and wave fixed effects, and form the per-event divergence: the theft effect for that earthquake minus its warga effect. The treated- village-weighted mean weights each earthquake by its number of unique first-treated villages; the equal-event mean weights every earthquake alike (blank cells are not defined for that row). The earthquake bootstrap resamples whole earthquakes with replacement (9,999 draws, seed 42) and recomputes cohort weights each draw; the leave-one-out rows give the range of the treated-weighted divergence when each earthquake in turn is dropped. Divergence is positive (theft rises faster than warga) in 12 of 19 events. Missing cells (Theft/Warga columns below row 1) are statistics defined only for the divergence.

S10.4 The proportional scale, in full

Section 7.1 summarises the proportional evidence; this appendix reports it in full, including the robbery and assault coefficients and the clustered and multiple-testing variants.

One objection to a contrast stated in percentage points is that the outcomes have very different base rates, so a difference in levels need not be a difference in proportional response. We therefore re-estimate the same stacked, outcome-interacted specification as a fixed-effects Poisson model, in which the coefficients are proportional rather than absolute effects. The divergence survives the change of scale, and on this scale it is the communal decline that is the larger movement: the composite communal measure falls by 31.4 percent (-0.378, standard error 0.087) and *warga* by 17.1 percent (-0.188, standard error 0.108), while theft rises by 14.0 percent (0.131, standard error 0.017). Equality of the theft and communal

proportional effects is rejected at $p < 0.001$, and equality of theft and *warga* at $p = 0.003$. The comparison the framework signs survives the same change of scale: theft and robbery differ proportionally at $p = 0.017$. This is the demanding version of the test, because robbery's base rate is an order of magnitude below theft's, so an account in which the shock raises predation proportionally across property crime would show up here and does not. Robbery's own proportional coefficient is −0.0590 (standard error 0.0800), a 5.7 percent decline that is not distinguishable from zero, which is why we rest this margin on the contrast rather than on robbery's level. Assault's proportional coefficient is +0.1397 (0.0616), a 15.0 percent rise, so the absolute null for assault should not be read as a tight zero on every scale. Read proportionally the communal decline is in fact the larger of the two movements, which is the opposite of the impression the absolute coefficients give, and the composition result does not rest on the scale in which it is expressed. The *warga* decline survives kabupaten clustering, at −0.0085 (standard error 0.0036), so unlike in earlier drafts we do not have to set that coefficient aside. Both contrasts survive spatial dependence: re-clustering at the kabupaten level, the equality of theft and each communal measure is still rejected (Table 4). The composite communal coefficient remains significant under kabupaten clustering (−0.0218, SE 0.0061) while the *warga* coefficient does not (SE 0.0037). Clustering is not earthquake-level inference, so we still do not read this table as fresh shock-robust evidence that communal conflict falls; that claim rests on the staggered estimators of Section 6. What the table does establish is the divergence: theft rises where communal violence, however measured, does not. Robbery, the property-*violent* act, does not rise (-0.0141^{***}), and assault is null. The composition is therefore not an artifact of comparing outcomes measured on different samples, waves, or control sets: individual economic predation rises exactly where collective and property-violent acts do not. Nor is it an artifact of testing several outcomes at once. Correcting the five per-outcome p-values for multiple testing (Benjamini and Hochberg, 1995) leaves four of them significant at the 5 percent level: theft's rise and the communal, robbery and *warga* declines, with q-values of 0.000, 0.000, 0.000 and 0.002. Only assault, at $q = 0.864$, adjusts to a clear null (Table 4). We none the less continue to rest the argument on the composition contrast rather

than on the *warga* coefficient, for the reasons given in Section 7.1: surviving a multiple-testing correction is not the constraint that binds it.

S10.5 The theft-versus-robbery margin

The contrast that carries the argument against a raised-motive account is theft against robbery, and it deserves the same treatment we gave theft against communal conflict rather than an appeal to the design. Because both items are recorded by the same official on the same form for the same village and wave, the comparison is a difference-in-differences in the relative outcome, and what it requires is parallel trends in that relative outcome, not in each level separately; the biases of the two component comparisons need to be equal, not zero (Olden and Møen, 2022). That is why we test the difference.

The result is mixed and we report both halves. Over the full event window the joint test of the pre-exposure leads of theft minus robbery rejects under the preferred effects ($F = 4.09$, $p = 0.001$) and under wave effects ($F = 5.22$, $p < 0.001$). The rejection is not spread across the pre-period: it is carried entirely by the most distant lead, whose coefficient is +0.129 (standard error 0.037) under the preferred effects, while the remaining four are individually insignificant and small (-0.035, $+0.013$, -0.013, $+0.007$, reading from the most distant). The reason is structural rather than substantive. That most distant lead is a single cohort observed in a single calendar wave, nineteen years before its own exposure, which is the configuration endpoint binning exists to handle; binned rather than dropped, it remains positive and significant, and we report the joint test both ways in Appendix S10 rather than only the version that passes.

An earlier draft defended this margin by restricting attention to a contiguous block of waves, on the grounds that the crime module skipped 2011 and 2014 and left event time irregularly spaced. That defence is no longer needed and we have withdrawn it: the module is fielded in every wave, so spacing across the panel is regular by construction and the full window is the right one to test. We nonetheless treat the theft-minus-robbery margin as not cleanly pre-tested, because the distant lead rejects under both specifications and we decline to read a rejection away.

Two further results belong here. Robbery's own level pre-trend rejects under

every window we examined, which is why we do not read its level causally, exactly as with theft. Assault's does not reject ($p = 0.27$ under the preferred fixed effects, and in every narrower window), so the assault comparison rests on a tested assumption rather than an assumed one. Finally, our specification rule selects the simplest specification that passes a conditional parallel-trends test, which is itself a pre-test selection rule, and we read the resulting intervals in that light rather than as exact (Roth, 2022).

S10.6 The benchmark estimated outcome by outcome

The same design estimated outcome by outcome under the CS(2021) benchmark, on four acts that differ in whether they are collective or individual and economic or violent, is reported in Appendix Table S16. The benchmark confirms the same gradient. Theft *rises* sharply, by 6.4 percentage points under CS(2021) ($p < 0.001$, about 14 percent of its 46 percent base) and by 4.3 points on the common sample, though its own pre-trend does not pass (-0.030, $p < 0.01$), which is one of the two reasons we do not rest the argument on theft's level. Because theft can also rise from weaker guardianship, abandoned dwellings, aid inflows or changed reporting, we do not read its magnitude as a standalone causal estimate; what we read from the comparison is the *contrast* with communal conflict, which does not require it. Robbery does not rise, at an insignificant -0.0008 under CS ($p = 0.86$) and negative on the common sample (-0.0141^{***}), so what climbs is specifically the non-violent, economic act. Assault is null in both. Communal conflict *falls*.

Table S16 estimates the same staggered design on four outcomes that differ in two dimensions: whether the act is *collective* or *individual*, and whether it is *economic* or *violent*. Communal conflict (column 1) is collective violence; theft (column 2) is individual and economically motivated; robbery (column 3) is individual property crime with violence; assault (column 4) is individual interpersonal violence.

S10.7 Retaining the always-treated cohort

One sample choice differs from the level analysis and we state it rather than leave it to be discovered. The level specifications exclude the cohort first exposed in

Table S16: Mechanism Discrimination: CS(2021) by Outcome
Window-aligned first exposure, 2005-boundary panel

	Communal conflict (collective, violent) (1)	Theft (individual, economic) (2)	Robbery (property, violent) (3)	Assault (individual, violent) (4)
CS(2021) overall ATT	−0.0177***	0.0642***	−0.0008	−0.0038
	(0.0047)	(0.0104)	(0.0044)	(0.0064)
Pre-trend Pre_avg (p)	0.016 (0.00)	−0.030 (0.00)	−0.013 (0.02)	0.009 (0.14)
Baseline mean	0.028	0.460	0.036	0.062
Waves	7	7	7	7
Observations	225,372	228,894	228,894	228,894
Controls	Yes	No	No	No

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

2005, which is always treated inside the panel; the composition sample retains it. Retaining it cannot bias the contrast directly, since a village whose exposure never changes contributes no within-village variation to the treatment coefficient, but it does help identify the wave and province-by-wave effects. Dropping the cohort moves the estimates in the direction that favours our reading rather than against it: under wave effects theft rises from +0.043 to +0.062, the communal decline is essentially unchanged at −0.021, and the theft-versus-communal F rises from 57.7 to 92.9, while under province-by-wave effects the theft-versus-robbery test is unchanged ($p = 0.031$ against $p = 0.036$). We report the fuller sample because it is the one the crime module defines.

S10.8 Over-absorption arithmetic

Section 7.1 reports that the contrast narrows under province-by-wave effects. We do not attribute that narrowing to over-absorption, because the arithmetic does not support that reading. Province-by-wave effects absorb 33.5 percent of the treatment variance that survives village and wave effects, which predicts a standard-error inflation of 1.23; the observed inflation for theft is 1.17. The loss of precision is therefore fully accounted for, and the movement in the point estimate is not.

Comparing each coefficient against its own wave-effects confidence interval, theft and communal conflict both move outside theirs, while robbery, assault and *warga* remain inside. A substantial part of both effects is accordingly between-province rather than within, which is what one would expect of a shock whose footprint is regional, but which we cannot separate here from province-specific trends in reporting or in disorder.

S10.9 Dropping the never-treated: the contrast on switchers alone

A referee will ask what the contrast looks like when the comparison group contains no never-treated villages at all, so that identification comes only from the staggering of exposure. We report it, and it is the least favourable cell in the paper. Restricting to the villages that switch inside the panel, dropping both the never-treated and the always-treated, the contrast is sharp under wave effects: theft against communal conflict rejects at $F = 26.4$ ($p < 0.001$) and against the definition-stable series at $F = 12.0$ ($p < 0.001$). Under province-by-wave effects on that same subsample neither contrast rejects: $F = 0.78$ ($p = 0.38$) and $F = 0.03$ ($p = 0.87$). We do not call that a precise null, because the arithmetic below shows it is not a precise anything.

Here, unlike in the full sample, the over-absorption reading is the right one and we show the arithmetic rather than assert it. On the switcher-only subsample, village and wave effects already remove 70.5 percent of the raw variance in exposure, and province-by-wave effects remove a further 72.5 percent of what survives, leaving 8.1 percent of the original treatment variance to identify five outcome-specific coefficients and three cross-outcome contrasts. That is a different situation from the full sample, where the same specification leaves enough variation that the standard-error inflation is fully accounted for by the absorption and the point estimates still move. We therefore read the switcher-only province-by-wave cell as uninformative rather than as a rejection, and we say so here instead of omitting it.

S11 Reporting-Artifact Placebo Outcomes

An empirical falsification test. The four arguments of Section 7.1 are observational; we also test the artifact story directly. If post-earthquake conditions distorted what village heads report, the distortion should surface in other items filed by the same respondent on the same form. We construct two placebo outcomes with no plausible earthquake channel, the village head's age and an indicator for a female head, and estimate them under both fixed-effect schemes alongside communal conflict on the identical subsample. The two specifications disagree, and the disagreement is itself informative.

Under wave fixed effects the placebo fails. Head age moves under both estimators, by -0.54 years under CS(2021) and -0.72 under two-way fixed effects, each significant at 1 percent. The female-head indicator moves under two-way fixed effects (-0.0094, significant at 5 percent) though not under CS(2021) (-0.0060, $p = 0.26$). Both items reject their joint pre-trend test ($p = 0.004$ for age, $p < 0.001$ for the female indicator), so under wave effects these series are not behaving like placebos at all. Under the province-by-wave effects that Section 5.1 selects, both pass their pre-trend test ($p = 0.76$ and $p = 0.59$) and the female-head placebo becomes a precise null (-0.0001, $p = 0.98$). Head age does not become a null: it reverses sign to $+0.45$ years, about five months, and is significant at 5 percent. We report that rather than describe the placebo as passing outright. A reporting-capacity story predicts that exposure degrades what the head records, and the sign reversal runs the other way; what the exercise establishes is that no common distortion moves the apparatus items and the conflict item together, not that the apparatus items are motionless. Communal conflict on the identical subsample falls by 2.7 percentage points under wave effects and 2.3 under province-by-wave, significant at 1 percent in both.

The naive TWFE benchmark does move on these placebos, but only because both trend strongly over the panel, which is exactly where staggered-timing weighting bias appears and the heterogeneity-robust estimator is required. Table S17 tabulates the full results.

This is the same pathology the conflict outcome shows, and it has the same source. Both apparatus variables trend strongly across the panel, head age from

Table S17: Placebo Outcomes from the Same PODES Form: Village-Head Age and Sex
2005-boundary panel, window-aligned first exposure; waves 2005–2024 excluding 2011

	Placebo outcomes		Same-sample
	Head age (years)	Female head	Communal conflict
Panel A: wave fixed effects			
CS(2021) ATT	-0.5392***	-0.0060	
	(0.2023)	(0.0054)	
pre-trend χ^2 p	0.004	<0.001	
TWFE	-0.7171***	-0.0094**	-0.0270***
	(0.1726)	(0.0042)	(0.0040)
Panel B: province × wave fixed effects			
TWFE	0.4452**	-0.0001	-0.0225***
	(0.2178)	(0.0049)	(0.0050)
pre-trend joint p	0.755	0.594	0.121
Observations	159,449	159,449	159,449

Placebos are village-head age and a female-head indicator from the raw PODES apparatus module (six waves; 2011 excluded, coding unverifiable), aggregated to constant 2005-parent boundaries. The third column re-estimates communal conflict on the identical subsample, so effect and placebo are read off the same villages and waves under the same specification. Pre-trend tests bin the endpoints (rel ≤ -5 and rel $\geq +5$) so that the reference category is rel $= -1$ alone rather than a mixture that includes post-treatment cells. Standard errors are clustered at the village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

45.0 to 48.3 years and the female share from 0.027 to 0.072, and those trends differ by region. A staggered design that absorbs only wave effects attributes part of such a trend to treatment whenever cohorts enter at different points on it; absorbing province-by-wave effects removes it. That the placebos fail exactly where the conflict pre-trend fails, and pass exactly where it passes, is independent support for the specification rule rather than a separate problem.

S12 Capacity Inputs Treated as Outcomes

Table S18 lists every capacity proxy the village census makes available, all fourteen, on the same specification as the level result. We report the whole battery rather than the rows that move, because the share surviving both filters is what decides how much weight the mechanism discussion can carry. Two filters matter and a row has to clear both: the effect must be distinguishable from zero, and the series must pass its own aggregate pre-treatment test. Two of the fourteen do. Public-purpose *gotong royong* with high involvement falls by 3.3 percentage points on a clean

pre-period, which is the one result that points where the capacity reading expects. Worship venues rise on a clean pre-period, which does not. Everything else is either a null or fails its own pre-test, including both security proxies: the count of village security personnel rises sharply but its pre-trend fails decisively, and the community security post is a precise null whose pre-trend also fails.

That count is the honest summary of the mechanism evidence, and it is why we do not claim to have identified a channel. A reader who wants to know whether we tested only the proxies that worked can see from this table that we did not.

Table S18: Every Capacity Proxy, and Which Ones Can Be Read
Fourteen proxies; two clear both the effect and the pre-trend filter.

Capacity proxy	ATT	(SE)	Own pre-trend p	Reads as evidence?
Collective-action capacity				
Gotong royong, public purpose, high involvement	-0.0332^{***}	(0.0125)	0.382	yes
Gotong royong, any involvement	$+0.0081$	(0.0049)	0.283	no
Village security personnel (asinh count)	$+0.1639^{***}$	(0.0208)	0.000	no
Community security post (*poskamling*)	-0.0010	(0.0111)	0.000	no
Worship venues (log count)	$+0.0167^{***}$	(0.0049)	0.749	yes
Social capital				
Sports-activity group, any	-0.0036	(0.0053)	0.949	no
Sports groups (count)	-0.0149	(0.0477)	0.000	no
Karang taruna, any	$+0.0006$	(0.0114)	0.114	no
Village institutions (asinh count)	-0.0076	(0.0244)	0.001	no
Communication and access				
Strong mobile signal	-0.0459^{***}	(0.0092)	0.005	no
No mobile reception at all	$+0.0013$	(0.0062)	0.000	no
Any street lighting	$+0.0003$	(0.0082)	0.483	no
Road passable, four-wheel	-0.0022	(0.0040)	0.669	no
Road passable year-round	$+0.0055$	(0.0084)	0.753	no

Notes: Every capacity proxy the village census makes available, estimated as an outcome on the same Callaway–Sant'Anna specification as the level result, not-yet-treated controls, no covariates. The fourth column is the p-value of the outcome's *own* aggregate pre-treatment test. The last column marks a row as evidence only if the effect clears 5 percent *and* the series passes its own pre-trend test: 2 of 14 rows do. We report the whole battery rather than the rows that move, because the share that survives both filters is the fact that decides how much weight the mechanism discussion can bear. * p<0.1, ** p<0.05, *** p<0.01.

Estimated as outcomes rather than moderators, the same series point the same way but none supports a causal reading. Strong mobile signal falls by 4.6 percentage points after first exposure (CS -0.0459, $p < 0.001$, against a 72 percent pre-exposure mean and despite secular coverage expansion), while the complementary indicator for no reception at all does not move (Section S17); the count of village security personnel (*linmas*) rises (0.16 in asinh terms), and road

passability does not move (four-wheel −0.2, 95% CI [−1.0, 0.6]; year-round 0.6, [−1.1, 2.2]) even as the paved-road *share* falls by 6.0 points in the covariate event study reported in the paper's appendix (−0.0599, $p < 0.001$). Every one of these series that moves fails its own pre-trend test, the *linmas* series decisively (−0.12, $p < 0.001$) and the paved-road share as well (+0.028, $p = 0.007$), while the two passability series, which do not move, pass theirs ($p = 0.67$ and $p = 0.75$), so we report them as descriptive corroboration of direction and nothing more. Two further communication outcomes sit in Section S17 on the same footing: street lighting does not move at all (+0.0003, $p = 0.97$) and passes its own pre-trend test, while the absence of mobile signal is likewise flat but fails its pre-test decisively, so neither adds anything either way.

The night watch is the guardianship input our reading leans on, and PODES records it: whether the village maintains a community security post (*poskamling*), asked in six of seven waves. The two estimators disagree. The two-way benchmark shows a fall of 6.3 percentage points after exposure (−0.0628, $t = -7.37$) against a 67 percent pre-exposure mean, the sign the capacity reading predicts, but under Callaway–Sant'Anna the estimate is a precise null (−0.0010, standard error 0.0111). We read neither causally: the pre-trend fails decisively (−0.143, $p < 0.001$), and the pre-treatment movement is larger than either estimated effect. Security infrastructure in these villages is trending for reasons the design cannot separate from exposure. We therefore report guardianship as the reading most consistent with the composition of crime, not as a channel we have identified, and note that every direct proxy PODES offers is either trending or measured too coarsely to carry the claim.

One further test discriminates between our reading and its closest rival, and it does not go our way. If the theft rise works through a lapsed night watch it should concentrate in villages that had one to lose; if it works through raised motive and emptied dwellings it should be roughly uniform in prior guardianship. Interacting the absorbing treatment with an indicator for a security post in the village's last pre-exposure wave, a predetermined moderator, the interaction is negative and significant: theft rises by 8.8 percentage points where no post was present and by 3.3 where one was (−0.055 on the interaction, standard error 0.018,

$p = 0.002$; $N = 13{,}320$ across 2,695 exposed villages), while the communal decline does not vary with prior guardianship at all ($p = 0.15$). The theft response is therefore *larger* where there was no watch to lose, the opposite of what the dual-use reading predicts. Either villages with a standing post are simply more resilient on every margin, or the theft rise is driven by displacement and motive rather than withdrawn guardianship, and we cannot separate them. Two caveats attach: the test is identified off exposed villages only, since the moderator is undefined otherwise, and *poskamling* was a state-coordinated programme as much as a village institution (Barker, 1998), so its presence proxies state penetration as well as collective capacity. We report it because it is the sharpest test our data permit of the mechanism we favour, and it fails.

A rise in social cohesion (Aldrich, 2012) does not explain the decline either. Reading confidence intervals rather than significance (Abadie, 2020; Imbens, 2021; Romer, 2020), the PODES social-capital proxies show no solidarity surge: mutual-help participation, sports groups, Karang Taruna presence and the count of village institutions all have intervals that exclude movement of the magnitude a cohesion mechanism would require, and worship-venue counts are higher (+0.0167 in logs, standard error 0.0049, on a pre-trend that passes at $p = 0.75$), which counts against venue destruction and leaves the contact channel resting on mobile signal. Conceptually, denser associational life would if anything *raise* the capacity to organise violence (Satyanath, Voigtländer, and Voth, 2017), so evidence that these same earthquakes raised participation elsewhere (Bai and Li 2021; Cassar, Healy, and von Kessler 2017) is not in tension with a communal decline: it measures recovery-oriented participation, not the contact and communication capacity inter-group violence requires (Section S18).

S13 The First-Time Margin: Design Detail

This appendix carries the design detail behind the first-time margin that Section 8 states in brief and sets aside: the matched same-earthquake design and its composite estimate, the choice of counterfactual within that design, the split by prior conflict history, and the three narrower boundaries that Section 9 summarises. The

Table S19: Falsification: the Onset Design Run Before the Earthquake
Villages about to be shaken were already more likely to report a first episode, by more than the effect estimated at exposure.

Comparison set	Estimate	(SE)	p	p, quake
Same earthquake	0.0149***	(0.0051)	0.004	0.010
and same province	0.0164***	(0.0061)	0.007	0.030
and within 30 km	0.0186**	(0.0078)	0.017	0.084
Panel B. Treated-minus-comparison gap, wave by wave				
One wave before the earthquake (t_{-1})	0.0145	(0.0051)	0.005	
Wave of exposure (t_0)	0.0038	(0.0041)	0.347	
t_{+1}	0.0073	(0.0046)	0.116	
t_{+2}	0.0149	(0.0061)	0.015	
Change from t_{-1} to t_0	-0.0107	(0.0065)	0.101	

The same-earthquake matched design run entirely inside the pre-exposure period: exposure is moved 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 arrives. A positive coefficient would mean villages about to be shaken were already starting conflict sooner than their comparison villages, which is the confounding the onset design cannot rule out from its own pre-period. Panel B places that gap wave by wave on one sample and one regression, so the pre-exposure and post-exposure gaps are directly comparable: the last row asks whether the earthquake widens the gap it inherits. Stack-by-village and stack-by-wave effects throughout; the last column clusters on the earthquake. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

full battery, the longer look-back, the local-geography calipers and their power calculation, and the independent soil measurements are in Section S14.

S13.1 The matched design and the split by prior conflict

A matched counterfactual holds the shock itself fixed: for each earthquake, treated villages (mean MMI≥VI) are compared with villages in the same footprint that felt sub-damaging shaking (mean MMI in $[4, 6)$). Because the controls also felt the earthquake, this identifies the *incremental* effect of crossing the damage threshold, smaller than the total effect by construction. The composite is -0.0174, significant at 1 percent under village and kabupaten clustering but not under 24-cluster earthquake-level inference ($p = 0.09$); the equal-event mean over the same 24 events is -0.0143 ($p = 0.11$), with 14 of 24 event estimates negative. Its pre-period is not clean either: the lead two waves before exposure is $+0.0136$ ($p = 0.06$), so part of a same-window comparison can reflect mean reversion, and exact-matching on the full pre-period conflict path and coarsened geography sharpens rather than removes that caveat.

One design choice has to be stated before any result is read, because it decides which comparison the estimates describe. The identifying variation here is *within* an earthquake, so the residual threat is that villages above and below the damage threshold sit on different regional trajectories. Two devices address it, and they are not interchangeable. Restricting the matched cells so that treated and comparison villages share a province removes the between-province component while leaving the estimator doing what the design was built to do, comparing villages that felt the same shock in the same place. Absorbing province-by-wave effects *on top of* the stack-by-village and stack-by-wave effects the design already carries is different: an earthquake's footprint falls almost entirely inside one or two provinces, so province-by-wave effects are close to collinear with the earthquake-by-wave effects already absorbed, and what they remove is largely the identifying variation rather than a confound. We therefore take same-earthquake-and-same-province as the primary counterfactual for the matched design on that reasoning, not on the results, and report the province-by-wave version alongside it wherever it is estimated, including where it weakens the estimate.

Splitting the matched sample by pre-period conflict is the informative step, and it cuts against the average. Since the claim is about *onset*, the classification has to earn that word: a village counts as having no prior conflict when it reports none in *any* PODES wave observed before its earthquake, which for the later cohorts is a look-back of up to nineteen years, though the median unit has two observed pre-exposure waves and the mean 2.8, and the matching cells are built under that same flag so the split and the comparison set use one variable (Table S21, column 2). Among the roughly nine in ten treated villages with *no* such history, exposure *raises* the onset of communal conflict: the matched effect is +0.0076 (standard error 0.0023, $p = 0.001$) against a counterfactual onset rate of 2.2 percent among matched controls, a relative increase of about a third. Estimated earthquake by earthquake and aggregated with a bootstrap over earthquakes, the unit at which treatment is assigned, it is +0.0082 ($p = 0.017$), positive in 16 of 23 events. Requiring treated and comparison villages to share a province as well as an earthquake, the counterfactual set out above, leaves +0.0088 ($p = 0.008$ with the earthquake as the cluster). The aggregate local decline instead loads on the minority with a prior

conflict history (−0.0620, standard error 0.0180). Cells match on the full observed pre-period conflict path, so reversion conditional on that path is differenced out; what cannot be separated there is *differential* reversion, treated villages' pre-period elevation being more transitory than their matched controls'.

S13.2 Boundaries of the onset margin

The first and most serious boundary is the pre-exposure falsification reported in Section 8. The remaining three are narrower. The first is the wild cluster bootstrap, $p = 0.18$ on the risk-set sample under the full-history classification, against $p = 0.017$ under the earthquake-level aggregation that matches the assignment mechanism and a within-earthquake randomisation test that places the estimate outside all 500 draws. The second is spatial: when the comparison set is made local, by distance caliper or by absorbing epicentral-distance bins, the estimate stays positive but is no longer significant, and the specifications that absorb the most geography retain as little as 4.4 percent of the treatment variance (Table S22). The confidence interval of every one of those specifications contains the baseline, and the minimum effect each could detect at conventional power exceeds the effect we estimate (Table S23), so what they establish is imprecision rather than absence. Topography is still unobserved, but shear-wave velocity is not, and both it and the site-response proxy leave the margin intact in sign (Tables S24 and S25), so the specific reading in which site conditions rather than damaging shaking drive the estimate is not supported. The third is the specification: absorbing province-by-wave effects on top of the matched design leaves +0.0042 ($p = 0.10$ under village clustering, $p = 0.27$ with the earthquake as the cluster), so the onset level is not established under the rule Section 5.1 applies to the level outcome. We take the shared-province matching restriction, which leaves the estimate at +0.0088 ($p = 0.008$ with the earthquake as the cluster), as the more informative test of the same concern inside a design that is already within earthquake, and we report both. Across all three boundaries the contrast between margins is the firmer object than either level: under the full observed history it clears the wild cluster bootstrap at $p = 0.026$.

S14 The First-Time Margin: Full Battery

This appendix collects the estimates for the first-time margin that Section 8 sets aside. We report them in full because a margin abandoned on the strength of one falsification deserves to be shown at its strongest first.

Classifying instead on the one or two pre-exposure waves inside the matching window, which earlier drafts used and which is the weaker test, gives +0.0080 (column 1); the two definitions differ for 869 of the 33,045 village-earthquake units observed at the last pre-exposure wave, 2.6 percent. We report the short-window version alongside because it is the more common convention in panel work of this kind, not because the paper rests on it. One limit stays open: the *perkelahian massal* item begins in 2005, so this is the full *observed* history and not a complete one, and episodes before 2005 are invisible to every source available for a village panel of this size. For the modal treated village, then, the shock does not pacify. The aggregate estimate from this design corroborates the *sign* of the average at most, and we do not read it as independent identification. The split by prior conflict describes what the average conceals, but the timing placebo below shows that this design cannot identify the level of its onset arm, so we report it as a pattern in the data and not as a result the paper rests on.

One thing we cannot show belongs here rather than in a referee report. The framework of Section 3 holds that onset is trigger-elastic and that the shock supplies the triggers, among them the rise in individual predation, which implies that theft should rise in the very villages where first conflict starts. We test that inside the same matched design and cannot establish it. Among the villages with no pre-exposure conflict, exposure raises reported theft by 1.5 percentage points (standard error 1.3) against a comparison rate of 47 percent, positive but far from conventional levels; among the minority with prior conflict the theft response is larger at 7.2 points (standard error 4.1), and the two arms are not distinguishable ($p = 0.18$), with the point estimates ordered against the trigger reading rather than for it. Estimating both responses earthquake by earthquake, the correlation between an earthquake's onset effect and its theft response is +0.31 across 22 events, and the weighted slope is +0.032 (standard error 0.041). The signs are those the framework predicts and the magnitudes are not informative, so the

composition evidence explains the average decline without our having connected it to the onset margin. We report the mechanism as the class of explanation the sign pattern selects, not as a channel we have traced to the result the paper leads with.

Because treatment is assigned by earthquakes rather than by villages, we take the onset margin to that unit as well. The inference battery that follows was built on the short-window classification and we report it on that basis; Table S21 gives the full-history analogues, which differ little, and we flag where they differ. Estimating the onset effect separately for each of the 23 earthquakes with matched support and aggregating across them, 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 ($p = 0.013$), and no single earthquake carries it: leaving each out in turn moves the mean only within [+0.0061, +0.0089]. The equal-event mean is larger at +0.0101 but not distinguishable from zero ($p = 0.24$, 13 of 23 events positive). The direction is worth noting: for the level and the composition contrast the equal-event mean is *smaller* than the village-weighted one, which is what a footprint-size reading predicts, whereas for onset it is larger, so that reading does not carry over. We report the discrepancy rather than set it aside. Table S20 reports the split, the pooled matched estimates under both clustering choices, the earthquake-level aggregation, and the raw onset rates behind them; Section S15 reports the fixed-effect structure and the matching.

Two further checks cut against us, and we report them here. First, with 24 earthquakes the natural request is a wild cluster bootstrap, and the pooled matched estimate does not clear 5 percent under one ($p = 0.054$, Webb weights, 9,999 replications), against $p = 0.023$ under earthquake-clustered standard errors. We prefer the aggregation of Panel B of Section S15 because it matches the assignment mechanism directly, estimating one effect per earthquake and resampling earthquakes, rather than relying on the pooled regression's error structure when cluster sizes are as unequal as ours; but the two procedures disagree at conventional thresholds and we report both.

Section S15 therefore collects all four inference choices in one place rather than scattering them, and we add to them a test that does not resample at all. Holding the number of villages each earthquake treats exactly as observed, we reassign

the above-threshold label at random *within* each earthquake and re-estimate, 500 times. The null respects the assignment structure and is well behaved, with a mean of −0.0000 and a 95th percentile of +0.0035; the observed +0.0080 exceeds every one of the 500 draws, so the one-sided p is below 0.002. This is a permutation test conditional on the matched sample, not a relocation of footprints, since which villages cross MMI≥VI is set by geology and distance rather than by a lottery. It answers whether an effect this large could arise by chance given how many villages each earthquake struck. Second, our split fixes conflict history at the village level and then keeps every wave in the window, so a village whose first episode occurs at exposure contributes its later waves as well. The convention of Bazzi and Blattman (2014) defines the risk set observation by observation, keeping peaceful waves and the first episode only; that removes 1,634 of 125,290 village-waves and leaves the estimate slightly larger and less precise (+0.0084 against +0.0080), with the earthquake-level aggregation still excluding zero (+0.0088, $p = 0.048$) and the wild bootstrap weakening ($p = 0.132$). We keep the fuller sample as the reported one because the restriction is itself a function of realized outcomes, and report the risk-set version alongside it in Section S15, which collects every inference choice for this margin. Neither check overturns the sign or the magnitude; both widen the interval.

Table S20: Onset and Reversion in the Matched Within-Earthquake Design

	Estimate	(SE)	p	Observations
Panel A. Matched within-earthquake difference-in-differences				
No prior conflict episode	$+0.0080^{***}$	(0.0023)	0.0005	125,290
Prior conflict episode	-0.0251	(0.0178)	0.1595	3,490
Pooled, village-clustered	-0.0184^{***}	(0.0040)	0.0000	128,780
Pooled, earthquake-clustered	-0.0184^{*}	(0.0095)	0.0664	23 events
Pooled, *warga*	-0.0060^{*}	(0.0031)	0.0522	128,780
Panel B. Onset rates, villages with no pre-exposure conflict				
Comparison villages, post	0.0217			71,380
Treated villages, pre	0.0000			4,455
Treated villages, post	0.0282			6,700
Memo: villages with prior conflict, comparison post	0.0777			1,377
Panel C. Onset margin at the earthquake level				
Treated-village-weighted mean	+0.0080	(0.0029)	0.013	23 events
bootstrap 95% CI	[+0.0018, +0.0133]			
Equal-event mean	+0.0101		0.237	13 of 23 positive
Leave-one-earthquake-out range	[+0.0061, +0.0089]			

Panel A compares villages shaken at MMI≥VI with villages that felt the same earthquake at sub-damaging intensity (MMI in $[4, 6)$), restricting to cells that contain at least one treated and one comparison village; 23 earthquakes have matched support. All specifications absorb stack-by-village and stack-by-wave fixed effects. Standard errors are clustered at the village level except in the earthquake-clustered row, which clusters on the earthquake and reports the number of clusters in place of observations. The split in the first two rows is by whether the village reported a communal-conflict episode in the pre-exposure waves inside the matching window, at most six years of look-back rather than a complete history; Table S21 rebuilds the split from every observed pre-earthquake wave and the margin survives at +0.0076. The first row is the onset margin and the second is the margin on which mean reversion can operate. Panel B reports raw rates on the same matched sample and is descriptive. Panel C estimates the onset effect separately for each earthquake in the matched design and aggregates across them, so the unit of inference is the earthquake rather than the village; the bootstrap resamples whole earthquakes. Panel C rests on 23 matched earthquakes, and its matching cell is coarser than that of Panel A: it is the earthquake by last-pre-wave conflict cell only, without the population tercile, road and police strata, so its sample is larger and the two panels are not estimated on identical rows. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Soil conditions, measured independently of ShakeMap. The site-conditions check reported above works inside the ShakeMap product, and that is its limit: the distributed peak-ground-acceleration grid is already corrected for shear-wave velocity by USGS, so the residual it isolates is a conversion-and-felt-report residual rather than soil. Indonesia's 2022 provincial Vs30 grids give an independent measure, and we join them to village polygons by area-weighted zonal statistics, covering 99.9 percent of the village-earthquake units in this design (Table S25; the join is documented in Section S15).

Three things follow, and they do not favour the margin. First, exposure is not balanced on soil: treated villages sit on *harder* ground than their comparison villages inside the same earthquake, by 49 m/s (standard error 2, $p < 0.001$), a standardised difference of +0.34. Crossing MMI VI on firmer ground requires being closer to the rupture, so treatment status and proximity are bound together in a way the design does not break. Second, absorbing Vs30 deciles interacted with earthquake and wave cuts the estimate from +0.0076 to +0.0051 ($p = 0.033$); this is an informative test rather than a vacuous one, since 27 percent of the treatment variance survives the absorption. Third, and most telling, the association is confined to hard ground. Interacting exposure with an indicator for below-median Vs30, the main effect is +0.0158 (standard error 0.0032) and the interaction is −0.0159 (0.0044, $p < 0.001$), so the effect on soft soil is −0.0001 ($p = 0.98$), which is zero. A mechanism running through ground shaking should if anything be stronger where soil amplifies it. Finding the entire association on firm ground, and none of it on soft ground, points away from shaking and towards whatever else distinguishes villages built on harder terrain. We report this as a third reason to set the margin aside, alongside the timing placebo below.

Panel B settles what that means, because it puts the gap on one sample and one regression wave by wave. The treated-minus-comparison gap is +0.0145 one wave *before* the earthquake, +0.0038 in the wave of exposure, +0.0073 at t_{+1} and +0.0149 at t_{+2}. The change from t_{-1} to the exposure wave is −0.0107 (standard error 0.0065, $p = 0.101$), and within 30 km it is −0.0187 ($p = 0.040$). The earthquake does not widen the gap it inherits. On this comparison the gap is largest before the shock arrives and smallest in the wave the shock hits.

We take the implication plainly. The onset margin as we estimate it is not separable from a level difference between villages that were going to be shaken and villages that were not, and the design cannot deliver it as a causal quantity. We report the estimate, the tests that support it, and this test that does not, and we do not ask the reader to weigh the first two against the third.

Two things bound what the falsification implies, and neither rescues the estimate. The main onset arm excludes every village that reports conflict in any pre-exposure wave, so the villages driving this placebo are removed from the headline sample by construction; the placebo therefore shows that treated and comparison villages differ in onset propensity, not that the surviving comparison is biased by that amount. And the test needs three pre-exposure waves, so it runs on the cohorts with the longest pre-periods rather than on the full matched sample. But the assumption the onset design requires is that villages above and below the threshold in the same earthquake had comparable onset risk absent the extra shaking, and on the villages where that assumption can be checked it fails. A reader should weigh the onset result accordingly, and we say so here rather than in a referee report.

That last point can be pushed further than we first thought, though not as far as observing site conditions directly. ShakeMap intensity is converted from recorded ground motion and blended with felt reports, so the intensity implied by peak acceleration alone departs from the distributed value by a conversion-and-felt-report residual, with a median of 0.19 and a standard deviation of 0.51 intensity units, available for 89 percent of village-earthquake cells. That residual is not a clean measure of site amplification, because the distributed acceleration grid is already site-corrected. What conditioning on it does deliver is a comparison holding recorded ground motion and the intensity conversion fixed, which is the part of the objection these data can reach. Table S24 reports it.

The margin holds under every version of that test. Rebuilding exposure from the PGA-implied intensity rather than from ShakeMap gives +0.0054 ($p = 0.020$), so the result is not an artifact of the blending. Comparing villages inside deciles of instrumental acceleration, so that treated and comparison villages felt the same recorded ground motion in the same earthquake, gives +0.0120 ($p = 0.013$) on 6.9 percent of the treatment variance. Absorbing deciles of the amplification proxy

itself gives +0.0089 ($p = 0.005$), and conditioning on it directly leaves +0.0072 ($p = 0.020$) with an amplification interaction of −0.0011 (standard error 0.0081), indistinguishable from zero: the effect does not vary with how much the site amplified the shaking. Within the intersection of both sets of deciles the estimate is +0.0097 with $p = 0.07$.

That pattern is informative about which absorption was doing what. Holding physical ground motion and site response fixed leaves the margin intact and, if anything, larger. Interacting epicentral distance with post-exposure status also leaves it. What removes it is absorbing earthquake-by-distance-bin effects, which is a far stronger restriction and leaves 6.9 percent of the treatment variance. Elevation and slope remain unobserved and these proxies do not substitute for them; shear-wave velocity is observed and is reported separately above. What the exercise rules out is narrower and worth stating exactly: the estimate is not an artifact of how intensity was converted from ground motion, and it does not depend on villages whose intensity was pushed up relative to their recorded acceleration.

What follows is the rest of the battery for this margin: the longer look-back that reclassifies the two arms, the local-geography calipers and their power calculation, the checks that hold recorded ground motion and the intensity conversion fixed, and the version under the counterfactual the composition contrast is held to. None of it rescues the margin, and we report it so that the failure above is read against the full attempt rather than a selected part of it.

What the longer look-back changes. The correction does more for the continuation arm than for the onset arm. Because the short-window flag placed genuinely conflict-experienced villages in the onset arm, it was diluting the arm it removed them from: under the full observed history the continuation estimate moves from −0.0251 (standard error 0.0178), which does not clear conventional levels, to −0.0581 (0.0181, $p = 0.001$). The first-time margin barely moves, from +0.0080 to +0.0076. The definition that makes the paper's title accurate therefore also sharpens the contrast the paper is about, which is the opposite of what a fragile split would do.

Because the two margins are estimated on separate samples, the paper has so

far juxtaposed them rather than tested them against each other. Stacking the two arms while letting the earthquake-by-wave effects remain arm-specific, so each arm keeps its own counterfactual, the difference between the continuation and first-time margins is −0.0335: +0.0084 (standard error 0.0028) on the first-time arm against −0.0251 (0.0176) on the continuation arm, both read off the stacked sample (Section S15, Panel C). Under the window definition it clears only the 10 percent level, with $p = 0.061$ under village clustering, $p = 0.108$ under earthquake clustering and $p = 0.160$ under a wild cluster bootstrap. We therefore do not rest the state-dependence claim on this version.

Under the full observed history the same contrast is better identified, for the reason just given: the window definition was mixing conflict-experienced villages into the onset arm and shrinking the gap it was being asked to detect. The difference is −0.0692, significant under village clustering ($p < 0.001$), under earthquake clustering ($p = 0.005$), and under the wild cluster bootstrap ($p = 0.026$), which is the test the window definition did not pass. We therefore state the state dependence as a result rather than as a juxtaposition, with one boundary attached: it is the *difference* between the margins that clears the most demanding inference, while the onset level on its own remains marginal there ($p = 0.21$ on the risk-set sample). The claim the evidence supports is that the two margins respond differently, not that either level is established under every inference the paper reports. Panel D of Section S15 also reports the split without matching on the pre-period outcome, which is the step that would most plausibly manufacture reversion: the first-time margin is essentially unchanged at +0.0070 and the continuation decline is *larger*, not smaller, at −0.0311, so the matching attenuates that arm rather than creating it.

The pre-determined capacity moderator of Section 7.2, baseline mobile-phone signal, predicts the average decline (Table 6) but does not significantly moderate first-time onset (−0.0040, standard error 0.0060). We report that null for completeness and draw no support from it. A null on a margin this design cannot identify is not evidence for the framework, and the interval is in any case wide enough to contain moderation of a size that would matter (Section S15).

Table S21: Onset Classified Over the Full Observed Conflict History
Rebuilding the history flag from every pre-earthquake PODES wave leaves the first-time margin intact and sharpens the contrast between margins.

	Window history (paper baseline)	Full observed history	Full observed, same province
	(1)	(2)	(3)
First exposure, no prior conflict	0.0080***	0.0076***	0.0088***
village clustered	(0.0023)	(0.0023)	(0.0025)
p	0.001	0.001	0.000
earthquake clustered	(0.0029)	(0.0027)	(0.0031)
p	0.011	0.010	0.008
Earthquake-level aggregation	—	0.0082	—
block bootstrap over events		(0.0026)	
p		0.017	
earthquakes (positive)		23 (16)	
Counterfactual rate, post	0.0217	0.0216	—
Observations	125,290	122,521	82,712
Memo: continuation arm	-0.0251	-0.0581***	—
	(0.0178)	(0.0181)	
Difference, continuation − onset	—	-0.0692	—
village clustered p		0.000	
earthquake clustered p		0.005	
wild cluster bootstrap p		0.026	

Column (1) classifies a village as previously peaceful using only the pre-exposure waves inside the matching window, at most two waves. Columns (2) and (3) use every PODES wave observed before the earthquake, which for the later cohorts extends the look-back from six years to as much as nineteen. Matching cells are rebuilt under each definition so the split and the comparison set use the same history variable. PODES begins in 2005, so this is the full *observed* history rather than a complete one: episodes before 2005 are unobserved in any available source. Column (3) additionally requires treated and comparison villages to share a province. The two clustered p-values on the difference come from the full matched sample; the wild cluster bootstrap is run on the risk-set-restricted stacked sample, because `boottest` cannot take two absorbed fixed effects, and on that sample the difference is -0.0698 with an analytic earthquake-clustered p of 0.010. The onset level's own bootstrap p of 0.175 is from that same risk-set sample, where the onset arm is +0.0078 (standard error 0.0044). $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

How local is the comparison? One feature of the onset arm has to be confronted rather than noted. Conditioning on villages with no recorded conflict fixes the pre-exposure outcome at zero by construction, so the pre-period carries almost none of the identifying content and what must be believed is closer to a cross-sectional statement: that villages above and below the damage threshold within the same earthquake would have faced comparable onset risk. Realised shaking is not assigned by lottery. It falls with distance from the rupture and varies with topography and site conditions, and the matching set out above balances population and distance to the district capital and comes close on road access and a police post, standardised differences of 0.15 and 0.20, neither of which determines MMI. The geometry is not reassuring on its face: treated villages in the matched sample sit a median of 49 km from the epicentre against 149 km for their comparison villages, and only 61 percent of treated villages have any comparison village within 30 km.

Table S22 therefore rebuilds the comparison set to be local in space, in two ways, and reports everything it finds. Restricting to treated villages with a comparison village within 30, 20 and 10 km leaves the estimate at +0.0043, +0.0031 and +0.0027; absorbing local geography instead of restricting the sample leaves +0.0058 under earthquake-by-province-by-wave effects, +0.0046 under earthquake-by-25 km-cell-by-wave effects, and between +0.0009 and +0.0041 once epicentral distance bins are absorbed. Every one of these is positive. Only one clears conventional levels, the earthquake-by-province specification at $p = 0.029$; the three distance calipers do not. We state the split plainly: the calipers do not establish the effect on their own, and the one specification that keeps the full sample while absorbing province does.

They also do not overturn it, and the reason is worth separating from wishful reading. The tightest specifications remove almost all of the variation they are being asked to use: the share of treatment variance surviving absorption falls from 31 percent under the baseline to 6.9, 6.5 and 5.0 percent once distance bins, province and bearing sector are absorbed, and to 4.5 percent under the 25 km cell. The confidence interval of every specification in Panel A contains the baseline estimate of +0.0072. What the exercise establishes is that the local data cannot

distinguish the baseline effect from zero, not that they favour zero.

Panel B separates the two things a distance caliper does at once, since it changes both the comparison group and which treated villages enter the estimand. Keeping only treated villages with local support but leaving the comparison set unrestricted gives +0.0034 at 30 km and +0.0042 at 20 km, so almost the whole attenuation is a change in *which villages are compared*, not in *what they are compared against*: localising the counterfactual on top of that moves the estimate by −0.0009 and +0.0011. At 10 km the pattern reverses, with the composition step worth +0.0004 and the counterfactual step worth +0.0045, but that is the smallest and least precise sample and we do not lean on it.

Panel C applies the same restrictions to two cleaner objects. On the definition-stable *warga* outcome the margin is +0.0075 without a caliper and +0.0051 within 30 km, still significant at 5 percent on a tighter standard error than the composite achieves. The narrow intensity contrast, MMI $[6, 7)$ against $[5, 6)$, gives +0.0090 without a caliper, and that sample is geometrically far better matched than the full one, with a median distance to the nearest opposite-status village of 24 km against 86 km. Within 30 km it falls to +0.0035.

Two facts decide how the null results in Panel A should be read, and Table S23 reports both. The first is that the geographic restriction does what it is meant to do. In the matched same-earthquake sample treated villages sit on average 66 km from the epicentre against 171 km for their comparison villages, a standardised difference of −1.28; requiring a shared province narrows that to −0.89, and requiring a comparison village within 30 km closes it to −0.03. On that dimension the local sample is balanced. The other covariates are close: within 30 km the largest standardised difference is +0.15 on the police-post indicator, just above the conventional 0.10 threshold and well inside the 0.25 the matching targets. The same-province column is no better on that dimension, at +0.17, which is one reason we do not treat it as strictly dominating the others. The restriction is therefore doing the work the objection asks of it, not merely shrinking the sample.

The second is what those samples could have detected. At 80 percent power and a 5 percent test, the smallest effect the baseline could reject the null against is 0.0064, which is 0.30 of the counterfactual onset rate. Within 30 km that minimum

rises to 0.0088, or 0.47 of the base rate, and within 10 km to 0.0138, or 0.63. The headline effect of +0.0076 is *below* the minimum detectable effect of every caliper specification. Those specifications could not have rejected the null against an effect of the size we estimate even had it been exactly true, so their failure to reject carries almost no information about whether the effect is there. That is the precise sense in which we read them as imprecise rather than as evidence of absence, and it is a statement about power that can be checked rather than an appeal to it.

Table S22: The First-Time Margin Under Local Spatial Comparison
Positive throughout; significant only where the full sample is kept and province is absorbed, not under any distance caliper, and the confidence interval of every specification contains the baseline.

	Estimate	(SE)	p	Residual D variance
Panel A. First-time margin under increasingly local comparison sets				
Baseline: same earthquake × wave	0.0076***	(0.0023)	0.001	0.314
and same province	0.0094***	(0.0025)	0.000	—
Nearest opposite-status village within				
30 km	0.0043	(0.0031)	0.168	0.262
20 km	0.0031	(0.0035)	0.372	0.261
10 km	0.0027	(0.0049)	0.588	0.250
Absorbing local geography instead of restricting the sample				
Earthquake × province	0.0058**	(0.0027)	0.029	0.213
Earthquake × 25 km cell	0.0046	(0.0065)	0.484	0.045
Earthquake × 20 km distance bin	0.0041	(0.0048)	0.386	0.069
and province	0.0009	(0.0049)	0.851	0.065
and 45° bearing sector	0.0019	(0.0058)	0.742	0.050
Panel B. What the caliper changes: estimand composition or counterfactual				
30 km: (A) full sample	0.0076			
(B) local support, all controls	0.0034	(0.0027)	0.199	
(C) same treated, local controls	0.0043	(0.0031)	0.168	
A − B, estimand composition	0.0042			
B − C, localising the counterfactual	-0.0009			
20 km: (A) full sample	0.0076			
(B) local support, all controls	0.0042	(0.0030)	0.155	
(C) same treated, local controls	0.0031	(0.0035)	0.372	
A − B, estimand composition	0.0034			
B − C, localising the counterfactual	0.0011			
10 km: (A) full sample	0.0076			
(B) local support, all controls	0.0072	(0.0041)	0.081	
(C) same treated, local controls	0.0027	(0.0049)	0.588	
A − B, estimand composition	0.0004			
B − C, localising the counterfactual	0.0045			
Panel C. The same restrictions on cleaner objects				
Warga onset, no caliper	0.0075***	(0.0019)	0.000	
within 30 km	0.0051**	(0.0025)	0.043	
within 20 km	0.0049*	(0.0028)	0.081	
MMI $[6, 7)$ against $[5, 6)$, no caliper	0.0090***	(0.0030)	0.003	
within 30 km	0.0039	(0.0034)	0.245	

Table S23: Balance and Detectable Effects in the Local Comparison
The distance caliper closes the geographic gap it is meant to close, and no caliper specification could have detected an effect the size of the headline.

	Same earthquake		and same province		and within 30 km	
	Treated	Std. diff.	Treated	Std. diff.	Treated	Std. diff.
Panel A. Pre-exposure balance, treated against comparison						
Log population	7.81	0.002	7.86	-0.046	7.89	0.097
Paved road	0.76	0.145	0.76	0.092	0.77	0.098
Police post	0.18	0.195	0.18	0.173	0.17	0.153
Distance to district capital (km)	33.81	-0.064	35.08	-0.003	33.18	-0.060
Distance to epicentre (km)	66.31	-1.276	63.47	-0.887	59.15	-0.028
Mean MMI	6.73	3.983	6.65	3.586	6.47	2.403
Treated villages	2,420		2,099		1,485	
Comparison villages	24,436		15,327		3,548	

Standardised difference is the treated-minus-comparison gap divided by the pooled standard deviation, evaluated at the last pre-exposure wave on villages with no conflict recorded in any earlier wave. Values below 0.10 in absolute value are conventionally read as balanced. Distance to the epicentre and mean MMI are the two variables the matching does not target, so they show what the geographic restrictions actually buy.

Panel B. What each specification could detect, 80 percent power

	Estimate	(SE)	MDE	MDE ÷ base rate
Same earthquake × wave	0.0076	(0.0023)	0.0064	0.30
and same province	0.0094	(0.0025)	0.0071	0.33
within 30 km	0.0043	(0.0031)	0.0088	0.47
within 20 km	0.0031	(0.0035)	0.0099	0.50
within 10 km	0.0027	(0.0049)	0.0138	0.63
Earthquake × province × wave	0.0058	(0.0027)	0.0075	0.35

The minimum detectable effect is 2.802 times the standard error, the smallest true effect a two-sided 5 percent test would reject the null against 80 percent of the time. The last column expresses it as a multiple of the counterfactual onset rate in the same sample, so a value of 1 means the specification could only detect an effect that doubles onset. It is the quantity that separates an imprecise estimate from an absent one.

Table S24: Holding Recorded Ground Motion and the Intensity Conversion Fixed
The first-time margin survives exposure built from instrumental acceleration and survives absorbing the conversion residual.

	Estimate	(SE)	p
Exposure built from instrumental PGA, not ShakeMap MMI	0.0054**	(0.0023)	0.020
MMI threshold within deciles of instrumental PGA	0.0120**	(0.0048)	0.013
Conditioning on the conversion residual	0.0072**	(0.0031)	0.020
within deciles of the conversion residual	0.0089***	(0.0032)	0.005
within PGA × residual deciles	0.0097*	(0.0054)	0.071

All rows are the first-time margin on villages with no conflict recorded in any earlier wave, with stack-by-village and stack-by-wave effects and village-clustered standard errors. ShakeMap intensity is not a direct measurement: it is converted from recorded ground motion through a ground-motion-to-intensity equation and blended with felt reports. The intensity implied by peak acceleration alone therefore differs from the distributed value by a conversion-and-felt-report residual. That residual is not a clean measure of site amplification, since the distributed acceleration grid is already site-corrected, but conditioning on it does hold recorded ground motion and the intensity conversion fixed, which is the part of the site-conditions objection these data can address. Rows two to five interact that proxy, log PGA and epicentral distance with post-exposure status, so the MMI threshold has to work on top of the physical ground motion actually recorded. Vs30 and terrain are still unobserved; this is the closest the available data come to holding site conditions fixed.

Table S25: Shear-Wave Velocity: Balance, Absorption and Soil Interaction
Treated villages sit on harder ground, the margin survives absorbing Vs30 deciles, and the entire association is confined to firm soil.

	Estimate	(SE)	p	N
Panel A. Is exposure confounded with soil?				
Standardised difference in Vs30	0.343			
Within-earthquake gap (m/s)	49.28	(2.19)	0.000	
Panel B. First-time margin, soil absorbed				
Baseline (earthquake × wave)	0.00763	(0.00230)	0.001	122,517
Within Vs30 deciles	0.00510	(0.00240)	0.033	122,492
Panel C. Does the effect differ by soil?				
Exposure × soft soil	-0.01593	(0.00439)	0.000	

Vs30 is taken from the 2022 provincial grids, a source independent of ShakeMap, joined to village polygons by area-weighted zonal statistics; coverage is 99.9 percent of villages in the exposure file. Soft soil is Vs30 below the sample median. Panel A compares treated and comparison villages within the same earthquake. Panel B absorbs Vs30 deciles interacted with earthquake and wave, so the estimate uses only variation between villages on comparable ground. Standard errors are clustered at the village level.

Onset under the same counterfactual. The composition contrast weakens when the province-specific counterfactual is imposed at the earthquake level (Section 7.1), and consistency requires asking the same of the onset margin rather than only of the finding it happens to damage. The onset design is already within earthquake, since it compares villages shaken at MMI≥VI with villages that felt the same shock at sub-damaging intensity, so province-specific trajectories are largely differenced out by construction. We remove them explicitly in two ways (Table S26). Requiring treated and comparison villages to share a province leaves the estimate at +0.0099 (standard error 0.0026) against a baseline of +0.0080, still significant at 5 percent when the earthquake rather than the village is the cluster ($p = 0.010$), on 83,357 village-waves across 23 matched earthquakes. Adding province-by-wave effects on top of the stack-by-village and stack-by-wave effects the design already carries weakens the estimate to +0.0042, marginal under village

clustering ($p = 0.098$) and not distinguishable from zero when the earthquake is the cluster ($p = 0.27$). However, that specification absorbs most of the variation left after matching. We report both rather than choose. These are different restrictions and they give different answers, so we state the comparison narrowly. Requiring a shared province, which addresses the confound directly, leaves the onset estimate intact. Absorbing province-by-wave effects on top of a design that is already within earthquake does not, and under that specification the onset estimate is not distinguishable from zero. An earlier draft led with this margin on the strength of the matched estimate and the inference checks that accompany it. It does not clear the restrictions the composition contrast clears, and that is why the paper no longer rests on it.

Table S26: Onset under the Preferred Counterfactual

	Estimate	(SE)	p
Matched within-earthquake design (baseline)	$+0.0080^{***}$	(0.0023)	0.001
adding province×wave effects	$+0.0042^{*}$	(0.0025)	0.098
earthquake-clustered	+0.0042	(0.0037)	0.267
Matching restricted to the same province	$+0.0099^{***}$	(0.0026)	0.000
earthquake-clustered	$+0.0099^{***}$	(0.0035)	0.010
Observations, baseline and province×wave	125,290		
Observations, same-province matching	83,357 across 23 earthquakes		

The onset estimate is the effect of first exposure at MMI≥VI on the probability that a village with no prior communal-conflict episode reports one, against villages that felt the same earthquake at sub-damaging intensity. The design is already within-earthquake, so province-specific trajectories are largely differenced out by construction; this table asks what remains when they are removed explicitly. Adding province-by-wave effects on top of the stack-by-village and stack-by-wave effects the design already carries absorbs much of the remaining identifying variation and the estimate weakens. Restricting the matched cells so that treated and comparison villages share a province is the more informative test of the same concern, and it leaves the estimate essentially unchanged, including when the earthquake rather than the village is the cluster. Standard errors are clustered at the village level except where stated. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S15 Additional Robustness Detail

Province trends: absorbed versus conditioned on

Province trends can enter as saturated fixed effects, which a regression-based event study absorbs directly, or as covariates inside the nuisance models that the Callaway–Sant'Anna estimator fits cell by cell. The first route leaves the result intact: under village and province-by-wave effects the Sun–Abraham pre-exposure leads are jointly insignificant ($\chi^2(5) = 4.67$, $p = 0.46$) and the response is -0.0155 (standard error 0.0058) in the wave of exposure and -0.0183 (0.0066) in the wave after. The second route does not. With province indicators in the doubly-robust nuisance models the estimator loses group-time cells to overlap failure, its degrees of freedom fall from 15 to 10, and the standard error on *warga* inflates sevenfold, from 0.0034 to 0.0235. Under regression adjustment, which does not fit a propensity model, the cells survive and the composite pre-trend test stops rejecting ($p = 0.054$), but the effect attenuates to -0.0064 (0.0047) and *warga* goes to zero ($+0.0002$, $p = 0.96$), with its own pre-trend still rejecting ($p = 0.044$).

We report this rather than choose the specification whose pre-trend passes. Two readings are consistent with it and our data do not separate them: either much of the level decline is carried by differences in provincial trajectories, or conditioning on thirty-two province and island indicators inside each group-time cell absorbs treatment variation along with confounding variation. The degrees-of-freedom loss and the sevenfold standard-error inflation are the signature of the second, but they do not rule out the first. Table S27 sets the two readings out side by side.

Table S27: Province-Specific Trends: Absorbed as Fixed Effects versus Conditioned On as Covariates

	Composite		*Warga*	
Handling of province-specific trends	ATT	Pre-trend	ATT	Pre-trend
	(SE)	χ^2 (p)	(SE)	χ^2 (p)
CS(2021), wave effects, baseline covariates	-0.0177^{***}	42.9 (0.000)	—	—
	(0.0047)	df 15		
+ island-region indicators (doubly robust)	-0.0190^{**}	37.5 (0.001)	-0.0106^{**}	28.8 (0.017)
	(0.0089)	df 15	(0.0046)	df 15
+ province indicators (doubly robust)	-0.0711^{*}	50.0 (0.000)	—	—
	(0.0411)	df 10		
+ province indicators (regression adj.)	-0.0088^{*}	24.7 (0.053)	-0.0018	25.5 (0.044)
	(0.0051)	df 15	(0.0040)	df 15
Sun–Abraham, village + province × wave FE	-0.0176^{***}	4.7 (0.46)	—	—
wave of exposure	(0.0061)	df 5		
one wave after	-0.0206^{***} (0.0070)			
Observations	225,372		225,372	
doubly-robust rows with province indicators	219,457		219,457	

All rows use the same panel, window-aligned first exposure, not-yet-treated controls and the always-treated cohort excluded. The first four rows differ only in how province-specific trajectories enter the Callaway–SantÁnna nuisance models: not at all, as island-region indicators, or as province indicators, and under either the doubly-robust or the regression-adjustment implementation. The doubly-robust rows with province indicators lose group-time cells to overlap failure, which is why their degrees of freedom fall and their standard errors inflate. The final row absorbs the same trends as saturated fixed effects instead of conditioning on them as covariates. Pre-trend is the joint Wald test that all pre-treatment coefficients are zero. Standard errors are clustered at the village level. N is the number of village-waves entering the estimator. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Exposure measurement: imputed village-waves and archive coverage

Measured versus imputed exposure. A village-year with no matching ShakeMap record is assigned zero shaking, which is correct when no earthquake reached the village and incorrect when one did but the event has no footprint. Across the estimation sample 36.5 percent of village-waves carry an imputed rather than a measured first lag, and the measured share rises from 43.5 percent in 2005 to 91.1 percent in 2018. Restricting to measured village-waves leaves the composite at -0.0153 ($p = 0.01$) against the full-sample -0.0165; restricting instead to waves from 2014 onward roughly doubles it to -0.0310 ($p < 0.001$), with *warga* significant there as well (-0.0124, $p = 0.04$). We do not read the second as a clean measurement experiment, since restricting to later waves also changes the cohorts, the period and the post-exposure window, and we stop short of asserting the sign of the bias because missingness correlated with era, geography or event size need not attenuate. What the checks establish is that the estimate survives when the imputed village-waves are removed.

Earthquakes absent from the ShakeMap archive. The previous check treats missingness at the village-year level. A separate question is whether the archive itself omits earthquakes. It does: 1,951 catalogue events at $M \geq 5$ have no ShakeMap product, so they cannot enter the exposure measure at all, and a village one of them shook at MMI VI would be recorded as unexposed. To bound what that can cost, we fit an intensity-prediction equation on the events that do have coverage, regressing each event's highest village MMI on magnitude and log hypocentral distance to the nearest village, and apply it to the missing events. The fit is close ($R^2 = 0.825$ on 1,786 events, residual standard deviation 0.476 MMI units) and the highest point prediction among the missing events is 5.95, below the threshold. We do not read that as zero, because a point prediction below VI leaves the residual free to carry an event above it: applying the empirical residual distribution to every missing event, about 19 of the 1,951 are expected to have truly reached MMI VI somewhere. What matters is whether those events could have exposed a village the design treats as unexposed. Taking the 122 events whose

95 percent bound crosses VI and the radius at which it crosses, 469 in-sample villages fall inside, but 374 of them are already exposed by events the archive does contain, 243 in the excluded always-treated cohort and 131 in identified cohorts, mostly aftershocks of recorded mainshocks in Palu and Lombok. Only 95 villages, 0.22 percent of the sample, are never exposed under the design's own definition and so could be misclassified. Dropping them leaves the composite at -0.0150 (standard error 0.0040); assigning each a first exposure at the PODES wave the missed event implies, which moves them into the treated cohorts and is the adverse direction, attenuates it to -0.0142, four percent, still significant at one percent (Table S28). Two limits are worth stating. The exercise bounds catalogue events without a ShakeMap, not events missing from the catalogue itself; and the equation is fitted on events that are on average larger than the missing ones, which biases its predictions for them upward if the U.S. Geological Survey produces ShakeMaps partly on felt reports, so the bound is conservative in the direction that matters. Refitting on the post-2013 era alone, when coverage is near-complete, moves the coefficients negligibly.

Table S28: Earthquakes Absent from the ShakeMap Archive

	Any communal		*Warga*	
	Estimate	(SE)	Estimate	(SE)
Panel A. How much exposure the archive can be missing				
Events at $M \geq 5$ with no ShakeMap	1,951			
Events fitting the intensity equation	1,789 ($R^2 = 0.825$, residual s.d. 0.477)			
Highest point prediction among them	MMI 5.94			
Expected number truly reaching MMI VI	19.3			
Events whose 95% bound reaches VI	122			
villages in radius, never exposed	95 (0.22 percent of the sample)			
villages in radius, already exposed	373 (243 always-treated, 130 identified cohorts)			
Panel B. The estimate under the worst case				
Baseline, preferred specification	-0.0165^{***}	(0.0041)	-0.0042	(0.0033)
Dropping the flagged villages	-0.0167^{***}	(0.0041)	-0.0044	(0.0033)
Assigning them the implied cohort	-0.0158^{***}	(0.0042)	-0.0037	(0.0033)
Observations	225,372 (baseline and recode), 224,742 (dropped)			

Panel A asks whether earthquakes absent from ShakeMap could have produced strong shaking the exposure measure never records. The intensity equation is fitted on the events that do have coverage, regressing each event's highest village MMI on magnitude and log hypocentral distance to the nearest village: $\text{MMI}_{\max} = +2.106 + 1.119\,M - 1.028 \ln R$. A point prediction below VI does not mean no event crosses VI, so the expected count applies the empirical residual distribution to every missing event rather than reading the point predictions alone. For the events whose 95 percent bound does cross VI we take the radius at which it crosses and split the villages inside by whether the design already treats them, using windowed first exposure. Only the never-exposed villages can be misclassified; the rest are already exposed by events the archive does contain. Panel B re-estimates the preferred specification with the never-exposed flagged villages dropped and, separately, with each assigned a first exposure at the PODES wave the missed event implies, which puts them into the treated cohorts and is the adverse direction. Standard errors are clustered at the village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

This appendix reports the full results for the robustness exercises summarized in Section 6.2.

S15.1 Catalogue Start Year

The absorbing cohort is defined by ever-exposure, so it depends on how far back the earthquake catalogue reaches. Table S29 rebuilds the first-exposure cohort from three start years and re-estimates both level specifications. The result is robust in sign and significance throughout, but the magnitude attenuates monotonically as the catalogue shortens: shortening it to 2000 moves 1,373 villages out of the always-treated cohort and into the comparison group, and shrinks the estimate by 15 percent under TWFE and 18 percent under Callaway–Sant'Anna. We retain the 1990 start, which is the only one under which cohort assignment is not contaminated by unobserved pre-panel exposure, and report the shorter windows rather than resolve the question by assertion.

Table S29: Sensitivity to the Earthquake Catalogue Start Year

Catalogue start year	Absorbing TWFE (SE)	Callaway–Sant'Anna (SE)	N (CS)	Always-treated villages dropped
1990 (baseline)	-0.0202^{***}	-0.0177^{***}	225,372	9,856
	(0.0033)	(0.0047)		
1995	-0.0193^{***}	-0.0150^{***}	227,045	9,263
	(0.0033)	(0.0043)		
2000	-0.0172^{***}	-0.0143^{***}	231,049	8,483
	(0.0032)	(0.0042)		

Each row rebuilds the first-exposure cohort using only earthquakes from the stated year onward, then re-estimates the absorbing TWFE (baseline controls interacted with wave, village and wave fixed effects) and the benchmark Callaway–Sant'Anna specification. Shortening the catalogue moves villages shaken in the 1990s out of the always-treated cohort and into the comparison group, which raises the estimation sample and attenuates the estimate by 15 percent (TWFE) and 18 percent (CS). All six estimates are significant at the 1 percent level. Standard errors are clustered at the village level. The 1990 start is retained because it is the only one under which cohort assignment is not contaminated by unobserved pre-panel exposure; the attenuation under shorter catalogues is reported rather than resolved.

S15.2 Sensitivity Under the Preferred Specification

The sensitivity analysis reported above is anchored to the wave-fixed-effects panel, which is the specification the parallel-trends rule of Section 5.1 does not select. That matters here more than it usually would, because the relative-magnitudes yardstick is computed *from* the pre-treatment leads: a specification whose pre-period is noisier mechanically permits larger hypothetical violations and so breaks down sooner. Re-running the analysis on the preferred specification is therefore not a robustness check but a correction of the anchor.

Under province-by-wave effects the largest deviation between consecutive pre-treatment leads falls to 0.0118, which is 0.66 times the effect being defended, against 0.037 and 2.2 times under wave effects. The identified set accordingly excludes zero through $\bar{M} = 0.25$ and admits zero at $\bar{M} = 0.50$ (Table E.1), where the earlier analysis broke down between $\bar{M} = 0$ and $\bar{M} = 0.25$. The result is still not robust to large violations, and we do not claim otherwise; what changes is that the estimate now survives a violation a quarter as large as the worst movement observed in the pre-period, rather than failing at any positive violation at all.

S15.3 Sign and Monotonicity Restrictions, and Why They Do Not Rescue the Level

The relative-magnitudes class treats a differential trend of either sign as equally admissible. Rambachan and Roth (2023) show that a researcher willing to sign the trend, or to assume it moves monotonically, faces a smaller feasible set and so obtains a larger breakdown value. Because our estimate is negative and the pre-period under wave effects drifts upward, a reader may reasonably ask whether that restriction is what the level result needs. We ran it rather than speculate. It is not available in the Stata implementation, which offers only the unrestricted relative-magnitudes and smoothness classes, so the estimates below come from the authors' R code on the same event-study vectors.

Table S30 reports the breakdown value under each restriction, for the three communal outcomes, and in both the controlled and uncontrolled variants of the preferred specification. Three readings follow, and none of them helps.

The restriction that appears to help is degenerate. Imposing $\delta \geq 0$, or requiring δ to increase, carries the composite past $\bar{M} = 64$, and the same holds for the gate once the controls are dropped. That is not robustness. 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 $\bar{M}$ rises: in the controlled variant the composite set is $[-0.051, -0.002]$ at $\bar{M} = 1$, $[-0.125, -0.004]$ at $\bar{M} = 8$, and unchanged at $[-0.125, -0.004]$ at $\bar{M} = 64$, the lower end resting on the solver's grid rather than on anything economic. The upper end never approaches zero at any violation we can specify, so the sign is delivered by the assumption and not by the data, and the breakdown statistic ceases to carry information. We decline to quote it.

The restriction in the opposite direction behaves as it should, and costs a little: signing the trend negative lowers the composite from 0.47 to 0.44 and the gate from 0.44 to 0.39.

The definition-stable series cannot be helped by any of them. Every restricted class breaks down at $\bar{M} = 0$ for *warga*, because its first post-exposure coefficient, -0.0056 with a standard error of 0.0048 in the uncontrolled variant, is not separated from zero even under exact parallel trends. There is no sign for a restriction to protect.

One inconsistency in our own battery surfaced in the course of this exercise and is worth recording, because it is visible in the published bounds. The composite sensitivity analysis conditions on the baseline-by-wave controls while the gate analysis does not. The two columns of Table S30 report both variants for all three outcomes so that the comparison is like for like; the substantive reading is unchanged either way.

Table S30: Breakdown $\bar{M}$ Under Sign and Monotonicity Restrictions
Takeaway: the restriction that appears to rescue the level fixes the sign by assumption, the restriction in the other direction costs a little, and the definition-stable outcome cannot be rescued by either.

	Breakdown $\bar{M}$	
Restriction on the differential trend	With controls	No controls
Subtype composite		
Δ^{RM}, unrestricted	0.50	0.47
+ bias restricted, $\delta \geq 0$	≥ 64	≥ 64
+ bias restricted, $\delta \leq 0$	0.47	0.44
+ δ increasing	≥ 64	≥ 64
+ δ decreasing	0.45	0.44
Mass-brawl gate		
Δ^{RM}, unrestricted	0.27	0.44
+ bias restricted, $\delta \geq 0$	0.25	≥ 64
+ bias restricted, $\delta \leq 0$	0.23	0.39
+ δ increasing	0.25	≥ 64
+ δ decreasing	0.23	0.39
Warga *subtype*		
Δ^{RM}, unrestricted	0	0
+ bias restricted, $\delta \geq 0$	0	0
+ bias restricted, $\delta \leq 0$	0	0
+ δ increasing	0	0
+ δ decreasing	0	0

Notes: Breakdown $\bar{M}$ is the largest relative-magnitudes bound at which the 95 percent identified set for the first post-exposure effect still excludes zero, located by bisection to a tolerance of 0.02 after doubling the search ceiling from 2 up to 64. Event studies use the preferred specification (village and province × wave fixed effects, always-treated cohort excluded); the two columns differ only in whether the baseline × wave controls enter, and the no-controls column reproduces the published gate bound exactly. Sign and monotonicity restrictions follow Rambachan and Roth (2023); they are not available in the Stata implementation and are computed in R. An entry of ≥ 64 does not mean the estimate is robust to a violation sixty-four times the largest pre-period deviation. It means the restriction alone fixes the sign, for the reason given in the text, so the breakdown statistic is uninformative in those rows.

S15.4 Covariate Overlap

The doubly-robust estimator reweights control villages by their estimated propensity to be treated, so it requires that the treated propensity distribution be contained in the control one. Table S31 reports the standard diagnostics cohort by cohort. Support is not a binding problem: in no cohort, and on none of the four boundary anchors examined, does more than half a percent of controls sit above the largest treated propensity, against the five percent threshold at which support violations are normally flagged.

Weight concentration is the milder concern. In three of the twenty-two cohort-anchor cells the most heavily weighted one percent of controls carries more than a fifth of the total control weight, the 2014 cohort most visibly, where only 37 villages are treated. Those are the cells in which the estimator leans hardest on a small number of comparison villages. We do not drop them, since the leave-one-cohort-out check of Section 6.2 shows that no single cohort carries the result, but we report the concentration rather than let the doubly-robust machinery obscure it.

Table S31: Covariate Overlap by Cohort

Cohort wave	Treated villages	Controls	Max treated propensity	Controls above treated support	Max weight	Weight in top 1%
2008	880	32,615	0.084	0.02%	0.14	2.6%
2011	1,145	31,470	0.077	0.03%	0.09	2.2%
2014	37	31,433	0.175	0.02%	0.70	24.9%
2018	203	31,230	0.028	0.12%	0.04	4.3%
2021	721	30,509	0.156	0.25%	1.77	7.9%
2024	221	30,288	0.030	0.32%	0.18	5.5%

Overlap diagnostics for the doubly-robust Callaway–Sant'Anna estimator, cohort by cohort. The propensity score is fitted on the baseline covariates within each comparison. "Controls above treated support" is the share of control villages whose propensity exceeds the largest value observed among the treated, the standard support violation; it never exceeds 5 percent in any cohort, on any of the four boundary anchors examined (22 cohort-anchor cells in total). "Weight in top 1%" is the share of total control weight carried by the most heavily weighted one percent of controls; it exceeds 20 percent in three of those cells, which is where the doubly-robust estimator leans hardest on a small number of comparison villages. No cohort is dropped on overlap grounds.

S15.5 Event-Level Inference: Robustness Detail

What bounds the equal-event null

Two things bound that null. The twenty-village cutoff is a choice: lowering it leaves the treated-weighted average between -0.0242 and -0.0254 and rejecting at 1 percent at every cutoff, while the equal-event mean swings from -0.031 ($p = 0.02$) at five villages through -0.025 ($p = 0.07$) at ten to -0.004 ($p = 0.78$) at two (Appendix Table S55), which reads as an estimand that is poorly determined rather than as an absent effect. And the equal weighting masks a split the framework predicts, though we note the reading is post hoc: among the six earthquakes with the largest on-land damaging footprints the mean effect is -0.038, with five of six negative at 10 percent, while among the thirteen smaller-footprint events it is -0.004. What makes an event large here is footprint, not magnitude; the two largest-magnitude events in the window are deep offshore ruptures with small

on-land footprints and near-zero effects. The pooled evidence is carried by the large-footprint earthquakes, and we regard that as the honest ceiling on what the staggered result can claim. Table S32 collects every aggregation and inference choice for this design.

This appendix expands the earthquake-level inference summarized in Section S4.6. Because control villages are reused across stacks, we cluster two-way on earthquake event and original village; this leaves the standard error essentially unchanged (SE $= 0.0101$, $p = 0.030$), which tells us that the binding dimension for inference is the number of distinct earthquakes rather than the reuse of controls.[2] A more direct check settles the reuse question: rebuilding the design so that each control village belongs to exactly one earthquake, by drawing controls only from never-treated villages and randomly assigning each to a single shock, removes the reuse entirely and leaves the estimates essentially unchanged (treated-village-weighted -0.0222, earthquake-level bootstrap $p = 0.02$). Because the stacked design already draws controls from the whole hazard zone rather than from the immediate neighbourhood of each shock, this random partition splits the same national pool rather than substituting distant controls for local ones. Nor does the result depend on a single lucky allocation: repeating the random partition 300 times, the pooled estimate is tightly distributed around -0.017 (standard deviation 0.001, 90 percent of draws between -0.019 and -0.015, every draw negative). Reuse is therefore not what makes the estimate marginal; the small number of distinct earthquakes is. Those 19 are genuinely independent shocks rather than one rupture counted many times: under a standard aftershock window (epicentres within 100 km and origin times within 180 days) they fall in separate seismic episodes, so treating them as independent units for inference is appropriate.

On the large-versus-small split of the 19 event-specific estimates: among the six events with the largest first-treated cohorts (at least 150 villages each, and up to 824) the mean effect is -0.038, five of six are negative, four of them at the 5 percent level and the fifth at 10 percent ($t = -1.83$), with the strongest suppression reaching -0.098. We are equally candid that the sixth is positive, though far

[2]With only 19 clusters on the event dimension the two-way variance matrix is itself imprecisely estimated and required the Cameron, Gelbach, and Miller (2011) non-positive-definite adjustment, so we read it only as corroborating the event-clustered inference rather than as a sharper test.

Table S32: Event-Level Inference: All Estimands Side by Side

The effect is negative under every weighting, inference level, and control group, though its precision varies: strongest in the matched design and weakest under the equal-event mean.

	Estimate	Inference		p	Events
		SE	95% CI		
Panel A. Global staggered design (treated vs never/not-yet-treated controls)					
Pooled, variance-weighted (village clustered)	−0.0257	0.0040	—	<0.001	19
same, earthquake clustered	−0.0257	0.0105	—	0.03	19
Treated-village-weighted (event bootstrap)	−0.0268	0.0105	[−0.050, −0.009]	0.01	19
Equal-event mean	−0.0155	0.0150	—	0.32	19
Disjoint controls, treated-wtd (event boot.)	−0.0279	0.0108	[−0.051, −0.009]	0.01	19
Panel B. Same-earthquake local design (treated vs felt-controls, MMI ∈ [4, 6))					
Unmatched pooled (village clustered)	−0.0174	0.0040	—	<0.001	24
same, earthquake clustered	−0.0174	0.0098	—	0.09	24
Unmatched equal-event mean	−0.0143	0.0087	—	0.11	24
Matched pooled (earthquake clustered)	−0.0189	0.0098	—	0.07	24
Matched treated-village-weighted (event boot.)	−0.0205	0.0110	[−0.044, −0.003]	0.02	23

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

from significant (+0.004), and that the concentration in large-footprint events means six earthquakes carry most of what the pooled estimate reports. Among the thirteen smaller events the mean is close to zero (−0.004), seven of thirteen negative, indistinguishable from noise.

S15.6 Spatial Displacement: Ring-Design Detail

This appendix expands the displacement test of Section S1.5, and Table S33 reports it in full.

Two non-displacement readings fit the negative ring coefficients. These ring villages typically felt the same earthquakes at sub-VI intensities, and we cannot rule out that felt shaking itself moves conflict, since the direct low-intensity contrast could not be constructed on this catalogue (Appendix S15.18); but the ring effect is not merely each village's own sub-VI dose, since conditioning on own continuous shaking leaves it essentially unchanged (0–25 km −0.0076; 50–100 km −0.0082, both $p < 0.01$). And part reflects regional co-movement: with province×wave fixed effects the ring coefficients attenuate, the nearest staying negative and the middle becoming a precise zero (+0.0017, $p = 0.69$), none significantly positive. The ring coefficients could in principle be aggregated into a village-weighted total spatial effect, but we do not report or lean on such a magnitude: the ring contributions use village and wave, not province×wave, fixed effects, and partly reflect the felt-shaking channel, so the total is best read as the spatial extent of suppression, not a precise displacement estimate. This complements the spatial-lag check of Section S4.8: neither produces the positive sign displacement would require. Amarasinghe et al.'s own estimates by disaster type cut both ways and we report both directions. Earthquakes, the hazard closest to ours, show the largest immediate local decline among the six hazards they separate, which is consistent with the place-boundedness reading, though that coefficient is imprecise (−0.023, standard error 0.021). The outward spillover in the same column is positive and significant (0.029, standard error 0.014), and it is the sharper form of the displacement concern: for the hazard most like ours, their design does detect the relocation that our rings do not. We do not resolve that disagreement by appeal to precision. The difference we would point to is the conflict technology rather

Table S33: Spatial Displacement: Conflict in Never-Treated Villages by Proximity to Treated Villages
Takeaway: no statistically significant rise in conflict near treated villages, so the local decline is not simply relocated.

	(1) Village+wave FE (village cl.)	(2) Village+wave FE (kabupaten cl.)	(3) Village+prov. x wave FE (kabupaten cl.)
Within 0–25 km	-0.0097***	-0.0097**	-0.0021
	(0.0029)	(0.0042)	(0.0055)
25–50 km	-0.0046**	-0.0046	0.0017
	(0.0023)	(0.0034)	(0.0042)
50–100 km	-0.0094***	-0.0094***	-0.0047
	(0.0017)	(0.0029)	(0.0033)
Constant	0.0312***	0.0312***	0.0278***
	(0.0010)	(0.0014)	(0.0019)
Baseline conflict mean	0.026	0.026	0.026
Observations	211,813	211,813	211,813

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

than the hazard. Their combatants are mobile and can reallocate across districts when a shock lowers the local prize; communal conflict is between residents of the same village and does not travel with them.

S15.7 Same-Earthquake Local Design: Matched Detail

Onset and the capacity moderator

The split also lets us report how the capacity moderator behaves across the two arms, though it does not put the framework at risk in the way an earlier draft of this appendix claimed. Section 3 holds that continuation is capacity-elastic while onset is trigger-elastic, which would imply that the pre-determined moderator predicting the average decline should not moderate first-time conflict. Interacting first exposure with baseline mobile signal inside the matched design, the onset interaction is -0.0040 (standard error 0.0060, $p = 0.51$), against -0.0205 (0.0074) for the same moderator on the aggregate outcome under the preferred specification (Table 6, column 2). We do not read the onset null as confirmation. Section 8 shows that this design cannot identify the onset level at all, and a null estimated on a margin the design cannot identify carries no evidential weight either way; the interval is wide enough to contain moderation of a size that would matter. On the continuation side the interaction is -0.0279 (0.0360) and points in the predicted direction, turning a positive effect among villages without baseline signal into a negative one among villages with it, but with 559 treated pre-exposure cells the relevant support is too small to distinguish that pattern from noise, and we do not read it as evidence either. The moderator evidence that the paper does rely on is the aggregate one, and it is reported in Section 7.2.

This appendix expands the same-earthquake local design of Section 8. Two features of that design bound what it can show. First, it holds the *earthquake* and the *survey wave* fixed, but not region, soil, topography, or market access, since a ShakeMap footprint can span very different terrain; it is a same-shock comparison, not a matched-neighbour comparison. Second, the control villages also felt the earthquake, so the design estimates the incremental effect of crossing the damage threshold relative to felt shaking, not the effect of an earthquake

relative to none. The composite estimate is −0.0153, significant at 1 percent under kabupaten clustering, with a within-pre-period placebo of $p = 0.06$ that does not clear conventional levels but is not a clean pass either. To probe mean reversion, exact-matching each treated village, within its earthquake, to felt-controls that share its last-pre-wave conflict value, its full pre-period conflict path, and coarsened geography (population tercile, paved road, police post) retains 99 percent of treated cells, so the restriction has little bite, and the matched composite is similar, if slightly larger, than the unmatched one (−0.0175; earthquake-clustered $p = 0.07$, and −0.0196 with $p = 0.02$ under treated-village weighting with the event bootstrap); a narrow intensity band, treated at MMI $[6, 7)$ against controls at $[5, 6)$, gives −0.009 ($p = 0.02$). Because *warga* is rarer still, the local design does not resolve the definition-stable outcome any more cleanly than the pooled estimator does. Appendix Table S34 reports the local comparison in full.

Inference and the risk set for the first-time margin

Table S35 collects every inference choice and every sample definition we have run for the first-time margin, including the ones that weaken the result. Panel A shows that the pooled matched estimate is stable in magnitude across sample definitions and that what changes is the interval: village clustering treats each of some 125,000 village-waves as an independent draw and is not defensible here, earthquake clustering with 24 groups is the honest analytic choice, a wild cluster bootstrap is the most conservative resampling scheme, and the within-earthquake randomisation test is the one that conditions on the assignment structure rather than resampling from it. Panel B reports the aggregation we prefer, which estimates one effect per earthquake and resamples earthquakes, and which excludes zero under both sample definitions. The equal-event mean is larger but imprecise under both, the same pattern the level and composition results show. Panel C tests the two margins against each other rather than merely juxtaposing them. Panel D removes the matching on the pre-period outcome, the step most likely to manufacture reversion, and shows that it attenuates the continuation arm rather than producing it.

Two caveats belong with the table. The risk-set column conditions on realized

Table S34: Local Comparison: Villages Shaken Above vs Below the Damage Threshold by the Same Earthquake

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

Outcome	Village cluster	Earthquake-event cluster (24)	Kabupaten cluster
Composite (any communal)	-0.0174^{***}	-0.0174^{*}	-0.0174^{***}
	(0.0040)	(0.0098)	(0.0058)
Warga (definition-consistent)	-0.0060^{*}	-0.0060	-0.0060^{*}
	(0.0032)	(0.0069)	(0.0036)
Pre-trend diagnostics (village-clustered)			
Pre-lead (t_{-2}, base t_{-1})	composite $+0.014$ ($p = 0.06$); *warga* $+0.006$ ($p = 0.27$)		
Last-pre-wave baseline only	composite -0.0126^{***} ($p = 0.01$); *warga* -0.0044 ($p = 0.25$)		
Pre-window placebo (t_{-1} vs t_{-2})	composite $p = 0.06$; *warga* $p = 0.20$		

Notes: Each of 24 earthquakes forms a stack in which treated villages (mean MMI≥VI) are compared to villages that felt the *same* earthquake at sub-damaging intensity (mean MMI in $[4, 6)$) and are not themselves MMI≥VI-treated as of that event (a small share cross the threshold in a later wave of the window; excluding them moves the estimate to -0.0171, immaterially), in event-relative time (a $[-2, +2]$-wave window), with event×village and event×wave fixed effects. This holds the earthquake and the survey wave fixed, but not region, soil, or topography, since treated and felt-control villages differ; it identifies the incremental effect of shaking crossing the damage threshold within a common earthquake. $N = 147,905$ stacked village-waves (5,063 treated and 52,448 felt-control pre-treatment cells). The same coefficient is shown under three clustering choices. The composite effect is significant at the 1 percent level under village clustering and at the 1 percent level under kabupaten clustering, and at the 10 percent level under the most demanding 24-cluster earthquake-level inference, and it survives the pre-trend diagnostics only in part: the composite pre-lead is $+0.015$ ($p = 0.03$) and the pre-window placebo $p = 0.06$, neither of which is a clean pass, alongside a last-pre-wave-only baseline that leaves it at -0.0126 ($p = 0.01$). The definition-consistent *warga* estimate is negative and significant under village clustering, but we do *not* read it as overturning its pooled CS(2021) null: it is imprecise under event-level inference and falls to -0.0044 ($p = 0.25$) once only the last pre-wave anchors the baseline, so its village-clustered significance is sensitive to the pre-period. Across the 24 event-specific local estimates, 14 are negative (mean -0.014), though the equal-weight average remains imprecise. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

outcomes by construction, since which waves survive depends on when a village's first episode occurs; that is inherent to a hazard risk set and is the convention we are following, not a defect we can design away. And the wild cluster bootstrap is known to be conservative when cluster sizes are as unequal as ours, with some earthquakes contributing a handful of treated villages and others thousands, so we do not read its failure to reject as evidence of no effect any more than we would read a non-rejection anywhere else in the paper.

S15.8 Site Conditions: Shear-Wave Velocity

This appendix documents the construction behind the soil check of Section 8 and reports its full output. The reading of the results is given there and is not repeated here.

Shear-wave velocity in the upper thirty metres, Vs30, comes from Indonesia's 2022 provincial grids, a source constructed independently of the ShakeMap intensities that define exposure. We join the grids to village polygons by area-weighted zonal statistics. Sentinel values outside the physically meaningful range of 50 to 3,000 m/s are masked before the join, and the join aborts rather than writing a file if any non-finite value survives, so the coverage figure is not inflated by placeholder codes. The result covers 99.9 percent of the village-earthquake units in the matched design, which is why the check does not turn on which villages are observed. Soft soil throughout is Vs30 below the sample median.

The balance test, the estimate with Vs30 deciles interacted with earthquake and wave absorbed, and the interaction with soft soil are tabulated in the main text, together with the share of treatment variance surviving the decile absorption, which is what decides whether the restriction is informative or merely destructive of variation.

S15.9 Spatial Randomization: Full Output

Table S36 reports the spatial randomisation test in full. Each of the 500 draws relocates every earthquake's footprint to a randomly chosen never-treated village and re-estimates the level effect, so the null respects the geographic structure of exposure rather than permuting treatment at random. Two features deserve

Table S35: First-Time Conflict: Inference Choices and Risk-Set Definition
Takeaway: the sign and magnitude are stable; the first-time margin clears conventional thresholds under the design's own unit of inference and is marginal under the most conservative alternative.

	All pre-quake-peaceful village-waves	Risk set: waves up to first episode
Panel A. Pooled matched estimate, alternative inference		
Estimate	0.0080	0.0084
Village clustered	(0.0023)	(0.0028)
p	0.001	0.003
Earthquake clustered	(0.0029)	(0.0039)
p	0.011	0.044
Wild cluster bootstrap p	0.048	0.132
95% CI	[-0.0002, 0.0129]	[-0.0027, 0.0152]
Within-earthquake randomization p	<0.002	—
null 95th percentile	0.0035	—
Observations	125,290	123,656
Panel B. Earthquake-level aggregation, block bootstrap		
Treated-village-weighted mean	0.0080	0.0088
Block bootstrap SE	(0.0029)	(0.0038)
p	0.013	0.048
95% CI	[0.0018, 0.0133]	[0.0003, 0.0145]
Leave-one-earthquake-out	[0.0061, 0.0089]	[0.0055, 0.0101]
Equal-event mean	0.0101	0.0085
p	0.237	0.329
Earthquakes (positive)	23 (13)	23 (14)
Panel C. First-time versus continuation margin, difference		
First-time margin	0.0084 (0.0028)	
Continuation margin	-0.0251 (0.0176)	
Difference	-0.0335	
Village clustered p	0.061	
Earthquake clustered p	0.108	
Wild cluster bootstrap p	0.160	
Panel D. No matching at all, split by pre-exposure conflict		
First-time margin	0.0070 (0.0023), p =0.002	
Continuation margin	-0.0311 (0.0160), p =0.052	

Notes: Outcome is an indicator for any reported communal conflict. Column 1 keeps every wave in the matching window for village-earthquake pairs with no conflict in the pre-exposure waves; column 2 keeps peaceful waves and the wave of the first episode only, following the risk-set convention of Bazzi and Blattman (2014). Panels A, C, and D match exactly within earthquake on last-pre-wave conflict, the full pre-period conflict path, population tercile, paved road, and police post; Panel B uses the coarser earthquake by last-pre-wave conflict cell of the earthquake-level aggregation, so that its first column reproduces the figure reported in the text. Wild cluster bootstrap uses Webb weights and 9,999 replications, clustered on earthquake. Panel B resamples earthquakes, 2,000 replications. Panel C stacks the two arms with arm-specific earthquake by wave effects, so the reported coefficients equal those of the separate regressions.

attention beyond the p-value. One placebo draw is at least as negative as the actual estimate, which is why the one-sided p is 0.004 rather than the 1/501 a clean sweep would give. And the dispersion of the placebo distribution is 2.1 times the village-clustered standard error, which says directly that clustering on the village understates uncertainty when treatment is assigned by a few dozen earthquakes. That ratio is the reason the paper reports earthquake-level inference alongside village clustering throughout.

Table S36: Spatial Randomization Inference
500 relocations of every earthquake footprint to a random never-treated village.

Placebo draws (B)	500
Actual estimate	-0.0199
Clustered SE (village)	0.0033
Placebo mean	0.0007
Placebo sd	0.0069
Placebo minimum	-0.0203
Placebo sd / clustered SE	2.10
Draws at least as negative (k)	1
One-sided $p = (k+1)/(B+1)$	0.0040
Recentered: draws deviating at least as far (k)	9
Recentered two-sided $p = (k+1)/(B+1)$	0.0200

S15.10 Region-Type Check: Descriptive Detail

This appendix expands the region-type check of Section 6.2. Of the 43,351 in-sample villages, 6,123 (14 percent) ever report communal conflict in any of the seven waves. That figure is not in tension with the statement that nine in ten exposed villages had no prior conflict: the 14 percent counts a village as conflict-affected if it reports an episode at any point between 2005 and 2024, before or after exposure, while the nine in ten is measured only over the one or two pre-exposure waves inside the matching window of the local design. Within the seven-wave measure we find the highest rates in the known communal-conflict provinces (Maluku, 55 percent of villages; North and Central Sulawesi, 20 to 32 percent; East Nusa Tenggara, 30 percent) but substantial numbers also in populous Java (roughly 10 to 15 percent); Aceh, whose conflict was separatist rather than communal, is low (5 percent), a sign the outcome captures communal rather than armed-separatist violence. Adding

province×wave fixed effects attenuates the main absorbing effect from −0.0199 to −0.0136 (Table S37); the within-province effect remains significant at 1 percent under village clustering ($p = 0.001$) and at 5 percent under kabupaten clustering ($p = 0.04$), and the *warga* margin does not survive province trends (−0.0025, $p = 0.44$).

Table S37: Region-Type Check: Main Effect under Province-Specific Trends
Adding province×wave fixed effects attenuates the effect by about a third; a within-province component persists but is smaller and less robust.

Specification	$\widehat{D}_{\text{treat}}$	SE	p
Village + wave FE (baseline)	−0.0199***	0.0033	< 0.001
+ Province × wave FE	−0.0136***	0.0040	0.001
same, kabupaten-clustered	−0.0136	0.0064	0.04
+ Province × wave FE, *warga* outcome	−0.0025	0.0032	0.44

Notes: Absorbing first-exposure treatment on the in-sample native-2005 panel ($N = 303,457$), no controls, no conditioning on any realized outcome. The baseline row carries village and wave fixed effects; the remaining rows add province×wave fixed effects, allowing each of Indonesia's provinces its own conflict trend, so the effect is identified from within-province variation across villages. Village fixed effects already absorb region *levels*; this tests region-specific *trends*. Of the 43,351 in-sample villages, 6,123 (14 percent) ever report communal conflict; these are patterned by region, with the highest rates in the known communal-conflict provinces (Maluku, North and Central Sulawesi, and East Nusa Tenggara) but substantial numbers also in populous Java; Aceh, whose conflict was separatist rather than communal, is comparatively low, a reassuring sign that the outcome captures communal rather than armed-separatist violence. Under province×wave trends the effect changes to −0.0136, staying significant at the 5 percent level under village clustering; it is significant under the more conservative kabupaten clustering, and the definition-stable *warga* margin does not survive. The decline is therefore partly regional, with a smaller within-province component surviving. Because these fixed effects absorb most of the regional variation from which communal conflict is identified, this specification is a more demanding within-province estimate. ** $p < 0.05$, *** $p < 0.01$.

S15.11 Treatment-Outcome Timing: Full Audit

This appendix expands the treatment-outcome timing check of Section 6.2. Because PODES conflict at wave t covers the twelve months before May enumeration, an

earthquake that falls *inside* that window gives a partial-exposure $t = 0$ outcome, mixing pre- and post-shock months, while a shock that predates the window gives a clean, fully-post outcome. We audit every earthquake by exact date (Table S38, Panel A): among the 19 stacked earthquakes, 73.7 percent are clean-post and five are partial, and every major earthquake that carries the effect (Padang 2009, Pidie Jaya 2016, Lombok and Palu 2018) is clean-post. In the broader 44-event universe, which also counts the always-treated 2005-wave events, the partial-exposure share is dominated by the mega-events of that era (the December 2004 Aceh tsunami and March 2005 Nias earthquake), cohorts we already treat separately. Restricting the estimate to clean-post cohorts, which drops 54,628 partial-exposure village-waves, does not weaken the result: if anything it is slightly larger (-0.0206 versus -0.0199, Panel B, a within-noise move consistent with attenuation of the partial-exposure $t = 0$ outcome), and the definition-stable *warga* outcome is significant at 1 percent (-0.0071, standard error 0.0027). This restriction excludes the five partial cohorts among the stacked earthquakes rather than re-timing them to a later clean wave, so it changes cohort composition; re-timing is the stricter alternative we do not implement.

S15.12 Benchmark TWFE: Distributed-Lag Estimates

Table S39 reports the two-way fixed-effects benchmark with the distributed lags, on the same panel as the estimators in the paper. It is the conventional specification against which the heterogeneity-robust estimators are read, and it sits here rather than in the paper because the paper does not rest on it.

S15.13 Goodman–Bacon Decomposition

Figure S3 plots the decomposition discussed in the paper: the weight each 2×2 comparison receives against the estimate it contributes, on the sample the preferred specification uses.

Table S38: Treatment-Outcome Timing: Clean-Post versus Partial Exposure
Ninety percent of stacked-design villages are clean-post, and restricting to clean-post cohorts strengthens the effect.

Panel A. Onset-wave exposure classification (share of earthquakes)		
	Clean-post	Partial-exposure
All events (≥20 first-treated, 45 total)	75.6% (34 events)	24.4% (11 events)
The 19 stacked earthquakes	73.7% (14 events)	26.3% (5 events)
The *all events* universe is the 45 earthquakes that first expose at least 20 sample villages, a broader set than the stacked-design universe used elsewhere because it also counts the always-treated 2005-wave events. Partial exposure is dominated by the 2005-wave mega-events (the December 2004 Aceh tsunami and March 2005 Nias earthquake, together 8,663 villages), whose outcome window straddles the shock; these are the early cohorts the paper already treats separately.		
Panel B. Absorbing effect under clean-post timing		
	Composite	*Warga*
Full sample (benchmark)	-0.0199^{***}	
	(0.0033)	
Drop 2005-wave Aceh/Nias mega-events	-0.0198^{***}	
	(0.0033)	
Clean-post cohorts only	-0.0206^{***}	-0.0071^{***}
	(0.0035)	(0.0027)
Observations (full / clean-post)	303,457 / 250,320	250,320

Notes: Absorbing first-exposure treatment, village and wave fixed effects, village-clustered standard errors. A village's onset wave is *partial-exposure* if its earthquake falls inside the wave's 12-month conflict-reporting window (May$[t-1]$, May$[t]$], so the $t = 0$ outcome mixes pre- and post-shock months; *clean-post* if the shock predates the window. Restricting to clean-post cohorts drops 53,137 partial-exposure village-waves and, if anything, leaves the composite effect slightly larger (-0.0206 vs -0.0199), a within-noise move consistent with attenuation of the partial-exposure $t = 0$; the definition-stable *warga* outcome becomes significant at 1 percent ($p = 0.009$) once timing is clean. This restriction excludes the partial cohorts rather than re-timing them to a later clean wave, so it changes cohort composition. The full per-earthquake timing audit (date, pre-wave, first fully-post wave, months to the outcome window) is in the replication archive. * $p < 0.1$, *** $p < 0.01$.

Table S39: Benchmark TWFE: Survey-Window-Aligned MMI Exposure and Communal Conflict
2005-boundary panel, medium and high hazard zones

	Outcome: any communal conflict (binary)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ever exposed MMI≥VI (absorbing)	-0.0199***	-0.0217***						
	(0.0033)	(0.0034)						
MMI (max, window $t-1$)			0.0001					
			(0.0002)					
MMI≥VI (contemporaneous)								-0.0012
								(0.0017)
MMI≥VI (window $t-1$)				-0.0055**	-0.0056**			-0.0056**
				(0.0026)	(0.0026)			(0.0026)
MMI≥VI (window $t-2$)					-0.0229***			
					(0.0035)			
MMI≥VI (window $t-3$)					-0.0010			
					(0.0018)			
MMI≥V (window $t-1$)						0.0026		
						(0.0018)		
MMI≥VII (window $t-1$)							-0.0057	
							(0.0035)	
Observations	303,457	292,712	292,712	292,712	292,712	292,712	292,712	292,712
Adj. R-squared	0.073	0.072	0.071	0.071	0.071	0.071	0.071	0.071

Standard errors clustered at village level in parentheses. Columns (1) and (2) report the primary absorbing specification: a village is treated from the first survey wave following its first MMI≥VI exposure onward. Columns (3) to (8) report episodic exposure in a given window. Earthquake exposure is spatially clustered, so village-level clustering may understate uncertainty. Re-clustering at the kabupaten level (350 clusters) leaves every point estimate unchanged and gives, for column (2), SE = 0.0065, t = -3.36, p = 0.001; for column (4), SE = 0.0034, t = -1.64, p = 0.102. The absorbing estimate is the one that survives this and province-level clustering; see the clustering robustness table. All specifications include village and survey-wave fixed effects, plus seven baseline village characteristics interacted with wave dummies: household electricity, paved road, internet access, police post, health centre, primary schools, and distance to the district capital. Baseline values are taken from PODES 2003 (enumerated August 2002), before the sample period, so they cannot themselves respond to exposure. Lagged rather than baseline covariates would be bad controls: exposure raises primary schools and paved roads and lowers police presence (see the mechanism table on village characteristics). Exposure windows are May-to-May 12-month blocks aligned to PODES enumeration; window t-1 lies fully before the conflict reference window. * p<0.1, ** p<0.05, *** p<0.01

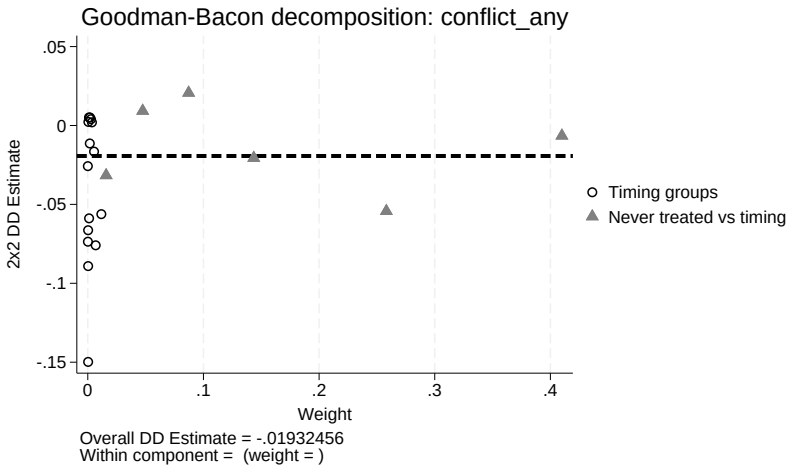


Figure S3: Goodman-Bacon Decomposition of TWFE Absorbing Estimate
Note: x-axis = weight on each 2×2 DiD; y-axis = DiD coefficient. Dot size proportional to weight. 96.2% of weight from never-treated vs. timing comparisons and 3.8% from the forbidden timing vs. timing comparisons, which return −0.0376 against −0.0186 for the clean comparisons. Sample: the 2005-boundary panel with the always-treated cohort dropped, as in the preferred specification. D_{treat} is absorbing, window-aligned first exposure, balanced panel $N = 232{,}778$. Source: Goodman-Bacon (2021).

S15.14 Episodic Estimand: dCdH Non-Absorbing Treatment

Table S40 reports the de Chaisemartin and D'Haultfoeuille (2026) non-absorbing interval-treatment estimate ($D_{interval}$) summarized in Section 6. Panel B reports the placebo (pre-trend) coefficients; the first is individually positive and significant, so we report it rather than rely on the joint test alone.

Table S40: dCdH Non-Absorbing Treatment Estimate
2005-boundary panel, medium and high hazard zones

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

Notes: Outcome: binary indicator for any communal conflict (`conflict_any`). Treatment: $D_{\text{interval},it} = 1$ if any earthquake with MMI≥VI occurred in [enum($t-1$) + 1mo, enum(t)−1yr−1mo]; non-absorbing (can switch 0→1→0). Source: USGS ShakeMap event-level rasters; PODES 2005–2024 (7 waves). Sample: KRB Medium/High seismic hazard, native 2005 village boundaries (nb2005). Column (1) TWFE: `reghdfe` with baseline covariates; SE clustered at village level. Column (2) dCdH: `did_multiplegt_dyn` (de Chaisemartin & D'Haultfœuille, 2026), Effect_ℓ = ATT for switchers ℓ waves after first switch; Placebo_ℓ tests parallel trends ℓ waves before first switch. † Controls omitted in dCdH: the estimator's switching design compares consecutive waves, and all Effect_1 comparisons require wave 2005 as baseline. TWFE with controls (-0.013^{***}) is the controls-adjusted benchmark. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

S15.15 Dropping the G6 Cohort and Aceh Province

The G6 cohort is the only positive-ATT cohort (Section 6), an artifact of a floor effect in post-conflict Aceh, where the last pre-treatment conflict rate was near zero. Table S41 confirms that this cohort neither drives the level estimate nor holds it back appreciably: dropping G6 leaves the composite ATT at −0.0203 against −0.0176 on the full sample, dropping Aceh province entirely gives −0.0185, and dropping both gives −0.0201, each significant at the 1 percent level. The average pre-trend lead remains insignificant in every restriction, and the definition-stable *warga* outcome moves in the same direction, already significant at 5 percent on the full sample (−0.0074) and strengthening once G6 is dropped (−0.0091).

Table S41: Robustness: Dropping the G6 Cohort and Aceh Province
Baseline: nb2005, window-aligned first exposure, CS(2021) with baseline controls. Takeaway: the headline estimate is stable when the positive-outlier G6/Aceh cohort is removed, so the outlier neither drives nor materially attenuates the result.

	Composite (*conflict_any*)		*Warga*	
Specification	CS(2021) ATT	Pre_avg (p)	CS(2021) ATT	N
Main (full sample)	−0.0176***	−0.0084 (0.28)	−0.0074**	225,372
Drop G6 cohort	−0.0203***	−0.0095 (0.26)	−0.0091**	224,084
Drop Aceh province	−0.0185***	−0.0088 (0.28)	−0.0074*	215,446
Drop Aceh + G6	−0.0201***	−0.0094 (0.27)	−0.0086**	214,977

Each row re-estimates the benchmark CS(2021) specification (doubly-robust weighting, not-yet-treated controls, `long2`, baseline covariates) on a restricted sample. G6 is the Aceh Pidie Jaya cohort, the only cohort with a positive ATT; Aceh province is identified by the first two digits of the village identifier. Pre_avg is the aggregate pre-treatment event-study coefficient (a conditional parallel-trends test). Standard errors are analytical, based on the Callaway–SantÁnna influence function and clustered at the village level. N is the number of village-waves entering the estimator. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S15.16 Balancing on Pre-Treatment Growth Slopes

Appendix S8 shows that villages later struck by earthquakes trend differentially before exposure on several margins, so the preferred specification relies on conditional rather than unconditional parallel trends. Because conditioning on covariate *levels*

cannot by itself remove a differential *slope*, we add each village's pre-treatment growth in log population and log households (measured over 2005–2008 and held time-invariant) directly to the control set. Population growth itself is not a source of concern: staggered-treated villages grew by 5.95 log points against 6.27 for the never-treated, a difference that is small, of the opposite sign to the one the concern anticipates, and statistically indistinguishable from zero ($p = 0.49$; households 10.30 against 10.24, $p = 0.91$). Conditioning on these slopes leaves the estimate essentially unchanged (Table S42): the composite ATT is -0.0165 versus the level estimate -0.0177, with an insignificant average pre-trend lead (-0.0076, $p = 0.33$), and the definition-stable *warga* outcome is likewise stable.

S15.17 Cohort and Event Dependence: Leave-One-Out

Table S43 documents the earthquakes behind the staggered variation, the input to the stacked event design of Section S4.6: 19 distinct events, each first exposing at least 20 in-sample villages, spread across 2006–2023 and from Sumatra to Maluku.

Table S42: Robustness: Conditioning on Pre-Treatment Growth Slopes
2005-boundary panel, window-aligned first exposure, Callaway–Sant'Anna

	Composite (*conflict_any*)		*Warga*	
Control set	CS(2021) ATT	Pre_avg (p)	CS(2021) ATT	N
Baseline controls (PODES 2003 levels)	-0.0177^{***}	-0.0077 (0.33)	-0.0075^{**}	225,372
+ pre-treatment growth slopes	-0.0176^{***}	-0.0076 (0.33)	-0.0074^{**}	225,365

The second row adds each village's pre-treatment growth in ln(population) and ln(households), measured 2005–2008 and held time-invariant, to the baseline control set (available for all 43,351 villages). Conditioning on these slopes directly is a regression adjustment for differential pre-treatment trajectories; it addresses the same concern as reweighting or matching on trends, though it is not equivalent to entropy balancing or coarsened exact matching. The estimate is Estimates are Callaway–Sant'Anna (2021) group-time average treatment effects, doubly-robust with not-yet-treated controls, with analytical influence-function standard errors clustered at the village level. essentially unchanged. Pre_avg is the aggregate pre-treatment event-study coefficient. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S43: Earthquakes Behind the Staggered Variation (Event Audit)
Takeaway: the identifying variation comes from 19 distinct earthquakes, not a handful.

Earthquake (region)	Date	Max MMI (village)	First-treated villages	Onset PODES wave
Pundong	May 2006	7.5	314	2008
Bukittinggi	Mar 2007	6.8	157	2008
Sungai Penuh	Sep 2007	6.9	94	2008
Gorontalo	Nov 2008	7.0	113	2011
Sarangani	Feb 2009	7.6	63	2011
Banjar	Sep 2009	6.5	85	2011
Pariaman	Sep 2009	8.1	824	2011
Reuleuet	Dec 2016	7.5	126	2018
Labuan Lombok	Aug 2018	8.1	353	2021
Palu	Sep 2018	8.2	252	2021
Ambon	Sep 2019	6.7	56	2021
Sukabumi	Nov 2022	8.3	149	2024
7 further events (≥20 villages each)		—	211	2008–2024
37 smaller events (<20 villages each)		—	198	2008–2024
Total (19 events, 2008–2024 onset)			**2,995**	

Notes: Each village is assigned to the earliest earthquake exposing it to mean MMI≥VI over 2000–2024. The table lists the 12 largest events (selected by first-treated-village count, shown here in chronological order), 7 further events with at least 20 first-treated villages, and 37 smaller events; the 19 events with at least 20 villages account for 2797 of the 2995 staggered-treated villages (2008–2024 onset). Region labels are auto-derived from the nearest named place in the USGS ShakeMap catalog and are approximate; all counts, dates, MMI values, and onset waves are exact. Source: USGS ShakeMap event rasters, native 2005 boundaries.

Is the result driven by one cohort or one earthquake? Group-level ATTs are heterogeneous (Section 6): the largest cohort effects are the more recent G7 and the Padang 2009 cohort G4, and G6 is positive in the no-controls specification. Table S44 answers the dependence question directly, re-estimating the no-controls CS(2021) ATT while dropping each treatment cohort one at a time and re-weighting the six cohort ATTs equally instead of by cohort share. This is a leave-one-cohort-out check, not a clean leave-one-event-out check: because a single first-exposure cohort can bundle more than one earthquake (the 2018 Lombok and Palu events

both fall in G7), dropping a cohort can remove more than one event, and isolating individual events would require the earthquake-level design we discuss as an extension. What the exercise does establish is that no single cohort drives the sign.

Table S44: Leave-One-Cohort-Out
2005-boundary panel, no controls so that all six cohorts are estimable. Takeaway: every re-estimate stays negative, so no single cohort drives the result.

Sample	ATT	SE	p	N
Full sample (no-controls benchmark)	-0.0175^{***}	(0.0044)	0.000	232,778
Leave-one-cohort-out				
Drop G3 (2008)	-0.0156^{***}	(0.0053)	0.003	228,102
Drop G4 (2011; Padang cohort)	-0.0250^{***}	(0.0059)	0.000	224,763
Drop G5 (2014)	-0.0160^{***}	(0.0044)	0.000	232,519
Drop G6 (2018; positive cohort)	-0.0201^{***}	(0.0046)	0.000	231,357
Drop G7 (2021)	-0.0127^{***}	(0.0048)	0.008	227,731
Drop G8 (2024)	-0.0178^{***}	(0.0045)	0.000	231,231
Cohort-share weighted average	-0.0202^{***}	(0.0045)	0.000	—
Equal-weight cohort average	-0.0318^{***}	(0.0122)	0.009	—

Notes: CS(2021) simple ATT (Callaway–Sant'Anna, doubly-robust, not-yet-treated controls, wave-ordinal time), outcome: any communal conflict, no controls so that all six cohorts are estimable; the with-controls headline ATT is reported in Table S1 and the no-controls full-sample benchmark here is -0.0175^{***}. Each row drops one treatment cohort at a time (the treated villages of that cohort are removed). This is a leave-one-cohort check, not a clean leave-one-event check: a single first-exposure cohort can bundle more than one earthquake (the 2018 Lombok and Palu events both fall in G7), so dropping a cohort can remove more than one event. The last two rows separate two ways of averaging the same six cohort ATT(g)s. The cohort-share weighted average is csdid_estat group's GAverage row, which weights each cohort by its share of treated units. The equal-weight average gives every cohort weight 1/6 and is computed from the cohort ATT vector and its covariance matrix, with a delta-method standard error; it is the row that speaks to whether the pooled estimate is a cohort-weighting artifact. SE clustered at village level. * p<0.1, ** p<0.05, *** p<0.01.

Every estimate is negative, ranging from -0.0127 to -0.0250 around the no-controls full-sample benchmark of -0.0175. Dropping the large Padang cohort (G4) *strengthens* the estimate to -0.0250 ($p < 0.001$), and dropping the positive G6 cohort strengthens it to -0.0201; the weakest drop is G7, at -0.0127, still

significant at 1 percent. The last two rows separate the two ways of averaging the same six cohort effects. Weighting each cohort by its share of treated units gives -0.0202, while giving every cohort weight $1/6$ gives -0.0318 (standard error 0.0122, $p = 0.009$). The equal-weight average is the one that speaks to cohort-share weighting, and it is larger in magnitude than the pooled estimate rather than smaller, so the result is not a weighting artifact. No single cohort overturns the finding, and none carries it.

S15.18 Placebo Tests and Randomization Inference

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

First, spatial randomisation inference. Because earthquakes strike spatially clustered sets of villages rather than villages drawn independently, a design-based placebo must relocate the shocks while preserving their geography. For each of 500 placebo draws we move every earthquake footprint to a randomly drawn never-treated village and assign placebo treatment to the nearest never-treated villages, preserving each footprint's village count and compactness (relocated footprints occasionally overlap, truncating a draw by about 15 percent on average), with the earthquake's true onset timing, and re-estimate the absorbing coefficient on the never-treated sample so that the true suppression cannot leak into the placebo control pool. Drawing the placebo centre from the villages themselves rather than uniformly over the map is what makes the placebo footprints resemble real ones: real footprints have a median mean-radius of 13.7 kilometres and placebo footprints drawn this way a median of 3.0 kilometres, against 170 kilometres under uniform-in-area placement, which spreads a placebo shock over an area no earthquake in the sample covers. The null's standard deviation (0.0069) is 2.1 times the village-clustered standard error (0.0033): the spatial null concedes that the uncertainty from a small number of large footprints is greater than village clustering implies, which is the few-shocks concern made quantitative, and the actual estimate clears it anyway. The null is centred essentially at zero (mean $+0.0007$). The actual estimate (-0.0199) lies below all but one placebo draw; with $k = 1$ of $B = 500$ draws at least as negative, the randomisation-inference

convention $p = (k + 1)/(B + 1)$ gives a one-sided $p = 0.004$. The result does not rest on the shift: recentering the null on its own mean, the actual deviation is exceeded by only 9 of the 500 draws (recentred two-sided $p = 0.020$; Table S36). No random spatial placement of the shocks reproduces the observed suppression.

The same exercise applied to the compositional divergence rather than to the level is weaker, and we report it rather than leave the asymmetry: relocating footprints against the village-level theft-minus-*warga* divergence puts the actual divergence of +0.058 at 1.7 standard deviations of the placebo distribution, with 21 of 500 draws at least as large, a one-sided $p = 0.044$. Both the level and the divergence clear 5 percent on this test, the divergence by the narrower margin.

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

Third, we run a wild cluster bootstrap for the deliberately conservative province-level clustering (28 clusters), using Webb weights and 9,999 replications (Roodman et al., 2019): the bootstrap p-value is 0.030 for the two-way fixed-effects composite estimate and 0.215 for the definition-stable *warga* outcome, so under this most demanding few-cluster inference the composite clears 5 percent while the *warga* arm does not. A wild-cluster bootstrap is not available for the CS(2021) estimator, so its analogous conservative figure is the province-clustered influence function reported in Section 6.2 ($p = 0.043$), not a bootstrap (Table S11). This is the single most demanding test the estimate faces; every less extreme clustering level, village and kabupaten, keeps the composite effect significant.

Fourth, a low-intensity exposure check, which we are no longer able to run. The design constructs pseudo first-exposure cohorts from events with mean MMI in the felt-but-non-damaging $[3, 5)$ band, restricted to villages that never experience any event of MMI V or above. That restriction was workable when the event catalogue began in 2000. With the catalogue extended back to 1990, as Section 4

explains it must be for the absorbing cohort assignment to be correct, almost every hazard-zone village has at some point reached MMI V, and the felt-but-undamaged comparison group collapses to 791 village-waves. Estimates on that sample are uninformative (-0.0018, $p = 0.91$ under two-way fixed effects). We therefore withdraw the low-intensity check rather than report it on a catalogue vintage inconsistent with the rest of the paper, and we make no claim about whether felt-but-non-damaging shaking moves conflict. The question of whether the effect is specific to the MMI VI damage threshold is instead addressed by the within-earthquake band comparison of Appendix S15.7, which contrasts villages at MMI $[6, 7)$ with villages at $[5, 6)$ struck by the same event and does not depend on the catalogue start year.

S15.19 Alternative MMI Thresholds

Table S45: Robustness: Alternative Shaking Thresholds
Takeaway: the effect is negative and significant under the absorbing design at all three thresholds, but the ladder is not monotone. The TWFE estimate is largest at MMI≥VI, and the CS estimate at MMI≥VII is imprecise because that cohort is small. We draw no dose-response conclusion from this table.

	(1) MMI≥V (weak)	(2) MMI≥VI (strong, main)	(3) MMI≥VII (very strong)
TWFE D_{treat} (absorbing)	-0.0069^{***}	-0.0213^{***}	-0.0174^{***}
	(0.0021)	(0.0034)	(0.0035)
CS(2021) overall ATT	-0.0117^{**}	-0.0177^{***}	-0.0054
	(0.0058)	(0.0047)	(0.0047)
Observations	165,767	225,372	269,171
Ever-treated villages	7,671	2,995	1,436
Village + wave FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

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

Table S45 shows results for MMI≥V (broader, more frequent treatment) and MMI≥VII (narrower, more destructive). Under the absorbing first-exposure design, MMI≥V yields -0.0069^{***} and MMI≥VII yields -0.0174^{***}, both consistent in sign with the level estimate but neither larger than the -0.0213^{***} at MMI≥VI. Signs are negative and significant at all three thresholds, so the result is not an artifact of the MMI VI cutoff.

S15.20 Exposure Construction: Raster-to-Village Aggregation

The treatment assigns each village the *mean* MMI across raster pixels intersecting its polygon, and a referee may reasonably ask whether the result depends on that aggregation rule: a mean can dilute localised strong shaking in large polygons, and a max can overstate it. We re-ran the zonal statistics for all 1,423 event ShakeMaps under four aggregation rules (pixel mean, pixel max, pixel median, and the raster value at the village's representative point), rebuilt the window-aligned first-exposure cohorts under each, and re-estimated the design (Table S46). As a further check, we drop the largest 5 percent of villages by polygon area (above 56 km^2), where within-polygon aggregation matters most. The estimates are stable across every construction: TWFE ranges from -0.0073 to -0.0209, all significant at the 1 percent level, and the CS(2021) ATT from -0.0059 to -0.0169, negative throughout and significant in four of the five constructions, the pixel-maximum rule being the exception at 10 percent. This panel assigns cohorts from its own windowed construction rather than from the production first-exposure file, so its levels sit below the benchmark and the comparison to read is the spread across rules, not any single row. Measurement error in the raster-to-village mapping is not driving the result.

Table S46: Robustness: Alternative Raster-to-Village Exposure Constructions
Outcome: any communal conflict; 2005-boundary panel, window-aligned absorbing first exposure, no covariates

Exposure construction (≥VI)	TWFE	(SE)	CS(2021) ATT	(SE)
Pixel mean within boundary	-0.0110^{***}	(0.0025)	-0.0100^{***}	(0.0035)
Pixel max within boundary	-0.0073^{***}	(0.0023)	-0.0059^{*}	(0.0032)
Pixel median within boundary	-0.0107^{***}	(0.0025)	-0.0096^{***}	(0.0035)
Value at representative point	-0.0101^{***}	(0.0024)	-0.0086^{**}	(0.0035)
Drop largest 5% of villages by area	-0.0224^{***}	(0.0035)	-0.0200^{***}	(0.0046)

Zonal statistics re-run over all 1,423 event ShakeMaps at native 2005 boundaries (1.78 million village-event cells, zero events missing). Each row rebuilds first-exposure cohorts from the stated aggregation rule and re-estimates the absorbing TWFE and CS(2021) specifications without covariates. TWFE: village and wave fixed effects, village-clustered SEs, $N = 220,248$–$277,200$. CS(2021): doubly-robust weighting, not-yet-treated controls, `long2`, analytical SEs. This panel assigns cohorts from its own windowed construction rather than from the production first-exposure file, so its levels are not directly comparable to the headline estimates; the informative content is the *spread* across aggregation rules. Across those rules the TWFE estimate lies between -0.0073 and -0.0224 and the CS estimate between -0.0059 and -0.0200, so the sign and broad magnitude are not an artifact of the raster-to-village aggregation rule. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S15.21 SUTVA: Spatial Spillovers

Earthquakes do not stop at village boundaries, so never-treated control villages located near treated villages could be contaminated by spillovers, biasing the never-treated counterfactual that carries most of the identification weight (Section 6). We test this two ways using village-centroid distances and a KD-tree neighbour search. First, we add each village's neighbours' treated share within 10, 25, and 50 km as a spatial lag alongside the direct treatment indicator. The direct effect survives but attenuates rather than strengthens: the absorbing -0.0217 becomes -0.0139 at 10 km, -0.0173 at 25 km and -0.0190 at 50 km, significant at 1 percent at every radius. The spillover coefficient is itself negative and significant at conventional levels at all three radii (-0.0096, -0.0066 and -0.0054 respectively), which says that never-treated villages with more treated neighbours report *less* conflict, not more.

That sign matters for how the check should be read, and it does not go the way the earlier draft of this section claimed. A negative spillover means the never-treated pool nearest the treated villages moves in the same direction as the treated villages themselves, so the control group is not clean of the shock and the unadjusted contrast understates rather than overstates the level effect. We therefore report the lag-adjusted estimates as the conservative version of the level result, and we do not claim that spillover contamination has been ruled out. Table S47 reports all five columns.

Notes to Table S47: Column (1) is the absorbing specification without a neighbour term. Columns (2) to (5) add the share of a village's neighbours within the stated radius that are already exposed, using village-centroid distances and a KD-tree neighbour search. All columns include village and wave fixed effects and the baseline covariates interacted with wave dummies, and cluster standard errors at the village. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Second, we drop never-treated villages that have any treated neighbour within 25 or 50 km (a "donut" around treated villages) and re-estimate: the direct TWFE estimate is essentially unchanged, and CS(2021) on the 25 km donut sample yields ATT -0.0057 (negative but imprecise, $p = 0.236$) with an average-lead pre-trend that does not reject (Pre_avg $+0.0034$, $p = 0.559$). The composite joint Wald test and the covariate-free anchors remain the specifications in which a pre-trend signal appears. Excluding potentially contaminated controls does not overturn the estimate, the pattern we would expect if spillover contamination were not a material threat to identification here.

Table S48 summarizes three further checks on the absorbing specification under wave fixed effects. Province × Wave fixed effects, which the rule of Section 5.1 selects as the preferred specification, move the coefficient from -0.0213^{***} to -0.0165^{***}; it remains significant at 1 percent, so the result is not an artifact of province-level shocks coinciding with earthquake timing. The balanced-panel restriction (villages present in all seven waves) leaves the estimate unchanged, because the absorbing design's sample is already balanced; no observations are dropped by the restriction. Always-treated villages are excluded from the estimation sample by construction under the window-aligned first-exposure design (any village already exposed at the 2005 survey baseline is dropped from g), so there is no separate ablation to report

Table S47: Direct Effect and Neighbour Exposure at Four Radii
The direct effect survives at every radius but attenuates.

	Panel A: any communal conflict, absorbing exposure				
	(1) No neighbour control	(2) 10 km	(3) 25 km	(4) 50 km	(5) 100 km
Ever exposed MMI≥VI (absorbing)	-0.0217***	-0.0139***	-0.0173***	-0.0190***	-0.0186***
	(0.0034)	(0.0047)	(0.0037)	(0.0035)	(0.0034)
Neighbour treated within 10 km		-0.0096**			
		(0.0041)			
Neighbour treated within 25 km			-0.0065***		
			(0.0024)		
Neighbour treated within 50 km				-0.0054***	
				(0.0019)	
Neighbour treated within 100 km					-0.0111***
					(0.0015)
Baseline mean	0.0256	0.0256	0.0256	0.0256	0.0256
Observations	292,712	292,712	292,712	292,712	292,712
	Panel B: any communal conflict, episodic exposure				
	(1) No neighbour control	(2) 10 km	(3) 25 km	(4) 50 km	(5) 100 km
MMI≥VI (window $t-1$)	-0.0055**	-0.0158***	-0.0127***	-0.0092***	-0.0056**
	(0.0026)	(0.0060)	(0.0038)	(0.0029)	(0.0027)
Neighbour treated within 10 km		0.0106*			
		(0.0054)			
Neighbour treated within 25 km			0.0075**		
			(0.0029)		
Neighbour treated within 50 km				0.0042**	
				(0.0017)	
Neighbour treated within 100 km					0.0002
					(0.0011)
Baseline mean	0.0256	0.0256	0.0256	0.0256	0.0256
Observations	292,712	292,712	292,712	292,712	292,712
	Panel C: inter-village conflict, absorbing exposure (waves 3–8)				
	(1) No neighbour control	(2) 10 km	(3) 25 km	(4) 50 km	(5) 100 km
Ever exposed MMI≥VI (absorbing)	-0.0191***	-0.0213***	-0.0190***	-0.0197***	-0.0180***
	(0.0032)	(0.0043)	(0.0035)	(0.0033)	(0.0032)
Neighbour treated within 10 km		0.0028			
		(0.0036)			
Neighbour treated within 25 km			-0.0002		
			(0.0020)		
Neighbour treated within 50 km				0.0013	
				(0.0016)	
Neighbour treated within 100 km					-0.0056***
					(0.0014)
Baseline mean	0.0126	0.0126	0.0126	0.0126	0.0126
Observations	250,896	250,896	250,896	250,896	250,896
	Panel D: inter-village conflict, episodic exposure (waves 3–8)				
	(1) No neighbour control	(2) 10 km	(3) 25 km	(4) 50 km	(5) 100 km
MMI≥VI (window $t-1$)	-0.0027*	-0.0116***	-0.0084***	-0.0066***	-0.0038**
	(0.0016)	(0.0044)	(0.0027)	(0.0019)	(0.0017)
Neighbour treated within 10 km		0.0091**			
		(0.0041)			
Neighbour treated within 25 km			0.0059***		
			(0.0022)		
Neighbour treated within 50 km				0.0044***	
				(0.0012)	
Neighbour treated within 100 km					0.0013*
					(0.0008)
Baseline mean	0.0126	0.0126	0.0126	0.0126	0.0126
Observations	250,896	250,896	250,896	250,896	250,896

for that concern. Finally, a forward-lag placebo that leads D_{treat} by one wave finds a lead coefficient of -0.0029 ($t = -0.53$, $p = 0.60$), against a contemporaneous effect of -0.0226^{***} estimated jointly with it. The lead is a clean null, so the design shows no sign of anticipatory adjustment on this margin. That is not the same as a clean pre-period: the joint Wald test rejects for the composite while *warga* passes (Section S4.5), and we lean on the definition-stable outcome and the composition contrast accordingly. The HonestDiD bounds in Section 6.2 quantify exactly how much residual violation of this kind the estimate can absorb before the sign is no longer identified.

S15.22 Additional Checks: Coefficient Stability and Spatial Inference

Coefficient stability under Oster (2019), computed for the absorbing window-aligned treatment ($\hat{\beta} = -0.0213$, $R^2 = 0.206$ with full controls): adding the full control set moves the coefficient *away* from zero, from $+0.0008$ with an R^2 of essentially zero to -0.0213. At Oster's recommended bound $R^2_{\max} = 1.3 \times R^2 = 0.267$ the implied $\delta = -2.98$ for the composite outcome (and $\delta = -0.23$ at the extreme bound $R^2_{\max} = 1$). A negative δ means that proportional selection on unobservables would have to run opposite in sign to the observed selection on covariates in order to explain the effect away: the observables, if anything, pull the coefficient away from zero rather than toward it. Like a large positive δ, this sign pattern indicates that the estimate is not an artifact of the kind of selection the observables reveal.

Conley (Conley, 1999) spatial-HAC standard errors, computed with a Bartlett kernel via a scalable KD-tree implementation on the subset of villages with matched centroids (N = 190,279, about 84 percent of the panel; underlying coefficient -0.0269), give SE = 0.0080–0.0095 across distance cutoffs of 50–150 km ($t \approx -2.8$ to -3.3; $p < 0.01$ at every cutoff, both outcomes). Alternative clustering brackets these: kabupaten-level clustering keeps the estimate significant, while province-level clustering (few clusters) is the conservative bound, at which a wild-cluster bootstrap still rejects at 5 percent for the two-way fixed-effects estimate ($p = 0.030$), as does the pooled CS(2021) ATT under province clustering ($p = 0.043$). The ordering (village-clustered < Conley < kabupaten < province) is exactly as spatial

Table S48: Additional Robustness Checks: nb2005 Baseline (Primary)

Specification	$\hat{\beta}$	SE	N	Note
Panel A: Province × Wave Fixed Effects				
Village + wave FE (baseline)	-0.0213^{***}	(0.0034)	225,372	Headline specification
Village + province×wave FE	-0.0165^{***}	(0.0041)	225,372	Absorbs province-level shocks
Panel B: Balanced Panel				
Unbalanced (baseline)	-0.0213^{***}	(0.0034)	225,372	All villages
Balanced (7 waves)	-0.0213^{***}	(0.0034)	225,372	0 obs dropped for balance
Panel C: Always-Treated Villages				
Under the absorbing windowed-first-exposure design, villages already exposed at the survey baseline ($g\_wave = 2$, i.e., first exposure at or before the 2005 wave) are excluded from g by construction, so there is no separate always-treated group left to drop as an ablation; this concern is structural to the design rather than a further robustness check.				
Panel D: Forward-Lag Placebo (No Anticipation)				
D_{treat} (current)	-0.0226^{***}	(0.0042)	225,372	Headline treatment, jointly estimated with lead
$F1\_D_{\text{treat}}$ (future, placebo)	-0.0029	(0.0055)	225,372	$t = -0.53$, $p = 0.599$ (see notes)

Notes: Baseline: absorbing two-way fixed-effects LPM with village and wave fixed effects, matching Table S57. Outcome: any communal conflict. Sample: KRB Medium/High seismic hazard zones, nb2005 windowed first-exposure design. Panel A replaces wave FE with province×wave FE to absorb regional time trends. Panel D leads D_{treat} by one wave to test for anticipation. Standard errors clustered at village level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

correlation predicts, and the effect remains significant under the distance-based correction appropriate for a localised physical shock, though it is sensitive to the most conservative few-cluster inference.

S15.23 Rare-Outcome Functional Form: Conditional Logit

Communal conflict is a rare binary outcome (2.6 percent pooled), so a linear probability model could in principle misbehave near the boundary. We re-estimate the absorbing design with a conditional (fixed-effects) logit, which conditions out the village effects without the incidental parameters problem and uses only villages whose outcome varies over time. The effect holds in the odds metric: first exposure reduces the odds of reported communal conflict by about a quarter without controls (odds ratio 0.75, $\hat{\beta} = -0.29$, cluster-robust $z = -3.14$, $N = 38{,}871$ observations in 5,553 switching villages) and by about a third with the standard baseline covariates (odds ratio 0.62, $z = -4.5$). The definition-stable *warga* outcome is a null without controls (odds ratio 0.93, $z = -0.6$) and negative with controls (0.74, $z = -2.2$), so the *warga* margin again carries its signal only conditionally, which we disclose. A fixed-effects Poisson model, an alternative rare-outcome specification, yields an incidence-rate ratio of 0.79 ($z = -2.9$). The LPM point estimates are therefore not an artifact of the linear functional form; the nonlinear models agree on sign.

S15.24 Robustness to the Control Set: Bad-Control and Baseline-Covariate Checks

Table S49 accompanies Section S4.8 and re-estimates the level ATT under alternative covariate sets, confirming that the result does not depend on the potentially endogenous infrastructure controls.

Table S49: Robustness to the Control Set: Bad-Control and Baseline-Covariate Checks
Outcome: any communal conflict (and warga*); 2005-boundary panel, window-aligned absorbing first exposure*

	Composite (*conflict_any*)			*Warga*	
Control set	CS(2021) ATT	(SE)	Pre_avg (p)	CS(2021) ATT	Observations
No covariates (anchor)	-0.0175^{***}	(0.0044)	+0.0107 (0.01)	-0.0084^{**}	232,778
Baseline PODES 2003 levels (headline)	-0.0177^{***}	(0.0047)	+0.0159 (0.00)	-0.0075^{**}	225,372
Baseline 2005 levels	-0.0177^{***}	(0.0047)	+0.0151 (0.00)	-0.0088^{**}	210,294
Lagged set, paved road dropped	-0.0155^{**}	(0.0061)	−0.0121 (0.07)	−0.0042	189,703
Geographic only (distance)	-0.0161^{***}	(0.0056)	−0.0082 (0.17)	−0.0051	191,749

Each row re-estimates the benchmark Callaway–Sant'Anna specification with a different covariate set on the same panel. The headline row uses PODES 2003 baseline levels, enumerated August 2002 and therefore predetermined with respect to every exposure in the sample. The composite estimate is stable across all five sets, from −0.0177 to −0.0155; what varies is the pre-treatment average, which is positive and significant at the 5 percent level under the three level sets and negative under the two lagged sets, where it reaches −0.0121 ($p = 0.07$) when the paved-road control is dropped and −0.0082 ($p = 0.17$) for the geographic set. Estimates are Callaway–Sant'Anna (2021) group-time average treatment effects, doubly-robust with not-yet-treated controls, with analytical influence-function standard errors clustered at the village level. Sample sizes differ because lagged covariates are unavailable in the first wave. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S16 Heterogeneity in the Conflict Response

Having examined the village-level inputs of the contact and coordination channels directly, we ask a complementary question: is the conflict-reducing effect larger where those channels have more room to operate? Table S50 reports interaction regressions that test whether the treatment effect varies with village proxies for these channels. Each column interacts the absorbing treatment with a standardized pre-treatment village characteristic, using the headline design (village and wave fixed effects, the baseline covariate set, and standard errors clustered at the village level).

Table S50: Mechanism Heterogeneity: Interaction Regressions

	(1) Monitoring	(2) Contact disruption	(3) State capacity
D_{treat}	-0.0210***	-0.0207***	-0.0215***
	(0.0035)	(0.0034)	(0.0035)
D_{treat} × Police (std.)	-0.0005		
	(0.0029)		
D_{treat} × Road access (std.)		-0.0041	
		(0.0034)	
D_{treat} × Dist. to capital (std.)			0.0033
			(0.0023)
Observations	225,372	225,372	225,372
Adj. R^2	0.073	0.073	0.073

Standard errors clustered at village level in parentheses. Absorbing windowed first-exposure treatment; village and wave fixed effects throughout. Controls are the 2002 baseline covariates interacted with wave, not lagged covariates, and each column omits the baseline level of its own moderator so that the same variable never enters twice. Moderators are the 2002 baseline values, standardized to mean zero and unit standard deviation. Baseline conflict incidence: 2.8%. * p<0.1, ** p<0.05, *** p<0.01

The interactions in Table S50 do not sharpen the channels. On the composite outcome the conflict-reducing effect does not vary significantly with pre-existing police presence (interaction -0.0005, SE $= 0.0029$), with paved-road access (-0.0041, SE $= 0.0034$), or with distance to the district capital ($+0.0033$, SE $= 0.0023$). A police interaction does appear on the definition-stable *warga* outcome, at -0.0048 (SE $= 0.0024$, significant at 5 percent), while remaining absent on the composite (-0.0024, SE $= 0.0030$; Appendix Table S53); it is specific to that outcome and to village-level clustering, so we do not build the monitoring channel on it. We read the road null as a null on this moderator, not as evidence against the contact

channel. The relevant margin for that channel is the communication infrastructure whose proxies decline after exposure, not road access, which typically stays passable after moderate shaking.

The fourth prediction of Section 3 concerns group structure. Because communal violence is fought *between* groups, the contact and mobilisation inputs it requires are more available where the local population is more polarised, so the shock's disruption of those inputs should reduce communal conflict more where polarisation is higher. Interacting the absorbing treatment with pre-treatment kabupaten *religious* polarisation (Esteban–Ray, from 2000 census religious shares; Montalvo and Reynal-Querol 2005, Esteban, Mayoral, and Ray 2012) bears this out: a one-standard-deviation increase deepens the effect by 1.6 percentage points (-0.0164). Because the moderator varies at the kabupaten level, we cluster the standard errors there; the interaction remains significant at the 5 percent level ($t = -2.50$, $p = 0.013$, 253 kabupaten clusters). This is the dimension along which Indonesia's major communal episodes have historically run (Maluku, Poso), and it is where the polarisation theory of conflict predicts the moderator should bind: polarisation, rather than fractionalization, drives conflict intensity when the contested prize is *public*, such as religious or cultural dominance and local political control, rather than private and divisible, and when group cohesion is high (Esteban, Mayoral, and Ray, 2012; Ray and Esteban, 2017). Ethnic polarisation (Bazzi and Gudgeon, 2021) points the same way in the continuous interaction but is not robust and we do not lean on it, consistent with religion proxying this public-prize axis of Indonesian communal contestation more cleanly than ethnicity. We do not, however, read this as isolating polarisation from fractionalization. For religious divisions in Indonesia, where a large majority faces smaller minorities, the two indices are nearly collinear and a joint horse-race cannot separate them, with both moderating the effect at almost identical magnitude (interaction -0.0164 and -0.0154, both significant at 5 percent). The ethnic counterparts are weaker but not uniformly absent: the ethnic polarisation interaction is insignificant (-0.0123) while ethnic fractionalization is -0.0150, significant at 5 percent (Appendix Table S54). We therefore read the gradient as evidence of an inter-group group-structure channel in the sense of Esteban, Mayoral, and Ray (2012), without attributing it to polarisation rather

than fractionalization specifically. Of the five moderators we examine (police, road access, distance to the district capital, ethnic and religious polarisation), only religious polarisation moderates the composite at conventional levels once the standard errors are clustered at the kabupaten (Appendix Table S54); the police, road and distance interactions do not move it (Table S50). These patterns are suggestive at best. The moderators are measured at the kabupaten level and moderate a village-level outcome, and, as Section S4.4 sets out, the religious gradient is not separately identified from the amount of conflict a place starts with: high-polarisation kabupaten have 2.3 times the baseline conflict rate, and once that scale is controlled the gradient either loses significance or does not survive a proportional specification. We therefore do not read these interactions as evidence for the contact-and-mobilisation account, and for the same reason we do not lean on the religious gradient against the reporting concern: a reporting shock proportional to each village's underlying rate would itself vary with polarisation through that correlation.

One further exercise speaks to the definition-stable *warga* outcome, whose pooled staggered estimate is negative but imprecise and is therefore weak evidence on its own. Because an average effect can mask heterogeneity, we ask whether the *warga* response is concentrated where the inter-group mechanism should bite. The pattern is there, subject to the identification caveat just stated. Interacting first exposure with pre-treatment religious polarisation, the effect on *warga* deepens with polarisation (interaction -0.0069, standard error 0.0037, significant at 10 percent), the same sign as the gradient we estimate for the composite (-0.0164, standard error 0.0066). On the rebuilt panel the monitoring interaction does not fall away on *warga* as an earlier draft reported: it is -0.0048 (standard error 0.0024) there against -0.0024 (0.0030) on the composite (Appendix Table S53). Both moderators therefore enter on the definition-stable measure, and neither is separately identified from the amount of conflict a place starts with, for the reason given in Appendix S21. We report the pattern and do not lean on it.

S17 Communication Infrastructure: Mobile Signal and Street Lighting

S17.1 Confounded moderators

It is not a decomposition, and one competing reading has to be priced. Baseline signal correlates with remoteness and development, so the interaction could be capturing either. Adding pre-exposure paved-road access, police-post presence and distance to the district capital as further interactions leaves the signal interaction at −0.0165 (standard error 0.0096) on the composite, which puts it at 10 percent significance, and at −0.0126 (0.0079) on *warga*, which no longer clears conventional levels. Neither the police-post interaction (−0.0044 and −0.0028) nor the distance interaction (+0.0027 and −0.0023) is distinguishable from zero. The road interaction is significant on *warga* (−0.0157, standard error 0.0052) but not on the composite (−0.0112, standard error 0.0071), so more than one infrastructure input is doing work on at least one outcome and we do not claim signal is the only one. The full horse race is reported in the manuscript.

Two outcomes from the communication-infrastructure battery in the capacity-input discussion of the manuscript are reported here. The absence of any mobile signal is essentially unchanged after exposure (+0.0013, $p = 0.83$), and so is the presence of any street lighting (+0.0003, $p = 0.97$). For mobile signal the aggregate pre-treatment event-study mean is large and significant (+0.0267, $p < 0.001$), so treated and comparison villages were already diverging before first exposure. The street-lighting pre-trend, by contrast, is small and insignificant (+0.0058, $p = 0.48$); we report it here because it belongs to the same battery, not because its pre-period fails. Neither outcome moves, so neither speaks to the capacity reading in either direction. The street-lighting series does at least pass its own pre-trend test, unlike the mobile-signal series, so its null is informative about street lighting in a way the mobile-signal null is not. An earlier draft read a positive street-lighting coefficient as post-disaster reconstruction; the estimate on the rebuilt panel is a precise zero and that reading no longer applies. Table S51 reports both outcomes in full.

Table S51: Communication-Infrastructure Outcomes
2005-boundary panel, window-aligned absorbing first exposure, no controls

Outcome	Treated pre-mean	CS(2021) ATT	(SE)	p	Pre_avg (p)		Observations
No mobile signal	0.111	+0.0013	(0.0062)	0.833	+0.0267	(0.000)	192,085
Any street lighting	0.729	+0.0003	(0.0082)	0.974	+0.0058	(0.483)	192,085

No mobile signal is the indicator that a village has no mobile-phone reception. Street lighting is built from PODES items r502a and r502b, whose 2018 redesign produces a level break; its own pre-treatment aggregate is insignificant. Callaway–Sant'Anna, doubly-robust, not-yet-treated controls, no covariates. Standard errors are analytical, based on the Callaway–Sant'Anna influence function and clustered at the village level; N is the number of village-waves entering the estimator. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S18 Social Capital: Sports Groups and Village Institutions

Four outcomes from the social-capital battery in Section S12 are reported here. None of the four moves. The presence of any active sports-activity group is flat (-0.0036, $p = 0.50$), as is *karang taruna* presence ($+0.0006$, $p = 0.96$), the sports-group count (-0.0149, $p = 0.75$) and the inverse-hyperbolic-sine of total village community institutions in the 2018–2024 panel (-0.0076, $p = 0.76$). Two of the four carry significant aggregate pre-treatment means, the sports count ($+0.1137$, $p < 0.001$) and total institutions ($+0.2108$, $p = 0.001$), so for those two treated and comparison villages were already diverging before first exposure; the two null outcomes have clean pre-periods. These outcomes move in directions consistent with either organisational disruption or measurement breaks across the 2018 questionnaire redesign, but their pre-trends caution against causal interpretation. Table S52 reports the four outcomes in full.

Table S52: Social-Capital Outcomes with Failed Pre-Trends
Baseline: 2005-boundary panel, window-aligned absorbing first exposure, no controls

Outcome	Treated pre-mean	CS(2021) ATT	(SE)	p	Pre_avg (p)		Observations
Sports-activity group (any)	0.942	−0.0036	(0.0053)	0.500	+0.0002	(0.949)	192,085
Sports groups (count)	3.742	−0.0149	(0.0477)	0.754	+0.1137	(0.000)	192,085
Karang taruna (any)	0.857	+0.0006	(0.0114)	0.957	−0.0101	(0.114)	78,119
Village institutions (ihs)	3.298	−0.0076	(0.0244)	0.755	+0.2108	(0.001)	78,119

Social-capital outcomes from PODES. The sports-group count and the village-institution index both fail their own pre-treatment test, so neither is read as evidence either way; they are reported because the mechanism discussion refers to them. Callaway–Sant'Anna, doubly-robust, not-yet-treated controls, no covariates. Standard errors are analytical, based on the Callaway–Sant'Anna influence function and clustered at the village level; N is the number of village-waves entering the estimator. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

S19 Heterogeneity on the Definition-Stable Outcome and Group Structure

This appendix reports two heterogeneity exercises discussed in Section S4.4. Table S53 re-estimates the two surviving composite moderators, police monitoring and religious polarisation, on the definition-stable *warga* outcome alongside the composite outcome: the *warga* response deepens with religious polarisation and, on the rebuilt panel, also with baseline police presence, so the imprecise pooled *warga* effect reflects heterogeneity rather than absence of an effect. Table S54 interacts first exposure with ethnic and religious polarisation *and* fractionalization, separately and in a religious horse-race. Religious group structure moderates the effect while ethnic group structure does not, but for religion the polarisation and fractionalization indices are nearly collinear and the horse-race cannot separate them, so the moderation is read as a group-structure channel rather than as polarisation specifically.

Table S53: Composite versus Definition-Stable Outcome by Moderator (2005-Boundary Panel)

	Police (village cluster)		Relig. polar. (kab. cluster)	
	(1)	(2)	(3)	(4)
	Composite	Warga	Composite	Warga
D_{treat}	-0.0209***	-0.0065**	-0.0265***	-0.0117***
	(0.0035)	(0.0027)	(0.0066)	(0.0039)
$D_{\text{treat}} \times$ Police (std.)	-0.0024	-0.0048**		
	(0.0030)	(0.0024)		
$D_{\text{treat}} \times$ Relig. polar. (std.)			-0.0164**	-0.0069*
			(0.0066)	(0.0037)
Observations	225,372	225,372	198,702	198,702
Adj. R^2	0.073	0.057	0.070	0.055

Standard errors in parentheses, clustered at village level (police) or kabupaten level (religious polarization). Absorbing windowed first-exposure treatment; village and wave fixed effects throughout. Controls are the 2002 baseline covariates interacted with wave, not lagged covariates. Moderators are pre-determined and standardized to mean zero and unit standard deviation. * p<0.1, ** p<0.05, *** p<0.01

Table S54: Polarization versus Fractionalization, Ethnic and Religious (2005-Boundary Panel)

	(1) Eth. P	(2) Eth. F	(3) Relig. P	(4) Relig. F	(5) Relig. race
D_{treat}	-0.0323***	-0.0302***	-0.0265***	-0.0264***	-0.0273***
	(0.0083)	(0.0075)	(0.0066)	(0.0066)	(0.0069)
× Ethnic polarization (std.)	-0.0123				
	(0.0081)				
× Ethnic fractionalization (std.)		-0.0150**			
		(0.0065)			
× Religious polarization (std.)			-0.0164**		-0.1560*
			(0.0066)		(0.0931)
× Religious fractionalization (std.)				-0.0154**	0.1374
				(0.0064)	(0.0902)
Observations	197,106	197,106	198,702	198,702	198,702
Kabupaten clusters	235	235	253	253	253
Adj. R^2	0.067	0.068	0.070	0.070	0.070

Kabupaten-clustered standard errors in parentheses. Absorbing windowed first-exposure treatment; village and wave fixed effects throughout. Controls are the 2002 baseline covariates interacted with wave, not lagged covariates. Moderators are pre-determined and standardized to mean zero and unit standard deviation. Column 5 is a religious horse-race with polarization and fractionalization jointly. * p<0.1, ** p<0.05, *** p<0.01

S20 SNPK/NVMS: Why It Cannot Corroborate the Village-Level Result

ACLED's Indonesia coverage begins in 2015 and so cannot reach the earthquakes that identify most main-specification cohorts, including the 2009 Padang event behind the largest treated cohort. The SNPK/NVMS incident database (Barron, Jaffrey, and Varshney, 2016), a newspaper-based violence series independent of PODES, covers 2005–2014 and is the one source whose window includes those cohorts. We attempted the check and report here why it does not deliver one.

Kabupaten-level analysis follows Bazzi and Gudgeon (2021), who use the same database at the same geographic level. Kabupaten codes are harmonized using a single fixed (2014-vintage) BPS crosswalk. SNPK's own kabupaten codes are recorded in one fixed later vintage across all incident-years rather than the code administratively valid at the time of each incident: a post-split code already appears in the 2005 incident file, years before that split existed administratively. Matching against the 2014-vintage crosswalk achieves a 99.9 percent match rate versus 97.0 percent under a naive year-specific match. The sample runs 2005–2014 and is restricted to the 18 provinces with continuous SNPK coverage over the period, since SNPK expanded to national coverage in 2012. Treatment is kabupaten-level absorbing first exposure to MMI≥VI from event dates; 37 always-treated kabupaten (first exposure before 2005) are excluded, leaving 77 ever-treated kabupaten. Kecamatan-level harmonisation, which would match the main design's geography, was not feasible: the source MFD kecamatan crosswalk ships only as an archive format not extractable in this environment.

Restricting to the 18 coverage-constant provinces does not hold recorded incidence fixed. Between 2005–2009 and 2010–2014 the share of never-treated kabupaten recording at least one identity conflict rises from 0.269 to 0.440, a factor of 1.63; among ever-treated kabupaten it rises from 0.086 to 0.190, a factor of 2.21. Both groups rise, and they rise by similar proportions. What differs is the base: never-treated kabupaten enter the window with roughly three times the recorded incidence of ever-treated ones.

That difference alone is enough to generate a negative estimate in the absence

of any treatment effect. A common proportional increase in reporting raises a high-base group by more percentage points than a low-base group, and a difference-in-differences estimated in levels reads the resulting gap as a decline among the treated. The signature is visible across outcomes. The implied level change is -0.039 for any social conflict, whose control-to-treated base ratio is 1.60; -0.054 for vigilante violence, at a base ratio of 1.77; and -0.067 for identity conflict, at a base ratio of 3.14. The estimated effect is ordered by the base gap rather than by any feature of the outcome itself.

Estimating the same specification in proportional rather than level form settles the question. Replacing the linear probability model with Poisson on incident counts, holding the kabupaten and year fixed effects and the kabupaten-level clustering unchanged, reverses every sign. Identity conflict moves from -0.0818 (SE 0.0239) in levels to $+0.514$ (SE 0.452) in proportional form, any social conflict from -0.006 (SE 0.020) to $+0.293$ (SE 0.231), and vigilante violence from -0.005 (SE 0.034) to $+0.189$ (SE 0.383). No proportional estimate is distinguishable from zero, and none points in the direction the level estimates suggest. An estimate that changes sign with the scale on which it is measured is not evidence about behaviour.

The contrast with the main analysis is the point of reporting this at all. The village-level composition result is estimated on PODES, a census whose enumeration protocol does not expand mid-sample, and it survives the identical test: the outcome-interacted specification re-estimated as Poisson preserves both the sign of each response and the ranking between communal violence and property crime. SNPK does not survive it, because what moves over the SNPK window is how much violence gets recorded rather than how much occurs. That the reporting technology rather than the underlying behaviour can drive the result is not a concern peculiar to us. Studying Indonesian crime over the same period, Hardiman (2025) finds that newspaper-reported incidents show no relationship with income shocks while survey-based victimisation over the same shocks does, which is the same divergence between a media-driven series and an enumeration that we encounter here.

West Sumatra, home of the Padang cohort, lies outside the 18 coverage-constant

provinces in any case, and extending the panel to include it does not recover the check. Across 2005–2014 the twelve Padang-cohort kabupaten record identity conflict in only two years, one kabupaten in 2006 and two in 2012, and the broader conflict measure sits at one kabupaten in every year except 2012 and 2014, when it reaches three. That is almost no variation from which to identify anything, and what little there is falls in years unrelated to the September 2009 earthquake.

We therefore claim no SNPK corroboration of the village-level result. The gap this leaves is real and we state it plainly: no independent violence series covers the cohorts that identify the main estimates, and the village-level result rests on PODES together with the internal-validity evidence assembled above.

S21 Polarization Heterogeneity in Full

Having examined the village-level inputs directly, we ask whether the effect is larger where those channels have more room to operate. Interacting first exposure with pre-treatment religious polarisation deepens the effect by 1.6 percentage points per standard deviation (-0.0164, standard error 0.0066, kabupaten clustering), and a gradient of the same sign appears on the definition-consistent *warga* outcome (-0.0069, standard error 0.0037). The monitoring interaction, by contrast, is absent on the composite (-0.0024, standard error 0.0030) and present on *warga* (-0.0048, standard error 0.0024; Appendix Table S53), so the two moderators do not separate the inter-group reading from a generic restoration of order as cleanly as an earlier draft claimed.

We then apply to this gradient the same scale test that governs the rest of the paper, and it does not pass. Religious polarisation is not orthogonal to how much conflict a place has to begin with. Baseline conflict rises from 2.1 percent in the least polarised tercile of kabupaten to 4.8 percent in the most polarised, a ratio of 2.3, and the kabupaten-level correlation between baseline conflict and religious polarisation is 0.37 (0.39 within the estimation sample). A response that is common in proportional terms therefore delivers a larger *absolute* decline where polarisation is high, with no separate role for polarisation at all. Three tests bound how much of the gradient this accounts for. Adding an interaction of exposure

with the leave-one-out kabupaten baseline conflict rate retains 40 percent of the polarisation gradient and removes its significance (-0.0070, standard error 0.0048), with the baseline-conflict interaction itself entering at -0.0141 (standard error 0.0053). Replacing that linear control with a flexible set of baseline-conflict tercile interactions, which imposes no functional form on the scale channel, leaves the polarisation gradient at -0.0141 (standard error 0.0058). And estimating the interaction proportionally, as Poisson, gives -0.0675 with a 95 percent confidence interval of $[-0.269, 0.134]$, which does not reject a common proportional response.

The gradient is therefore not separately identified from the amount of conflict a place starts with: whether it survives depends on how baseline conflict is controlled for, and the proportional specification cannot distinguish it from nothing. We report it because it is the natural test of the mechanism, and we report its failure because the same standard governs the external-validation exercise we discard in Appendix S20. We do not read it as evidence for the contact-and-mobilisation account. That account rests on the composition contrast, which does survive the proportional test, and on the direct *gotong royong* measurement. Appendix S16 gives the interaction tables, the polarisation-versus-fractionalization collinearity, and the weaker ethnic counterparts in full.

S22 Material Relocated from the Paper's Appendix

The sections that follow were carried in the paper's own appendix in earlier drafts. They are reported in full here rather than in the paper because none of them is load-bearing for the result: each answers a question a referee may reasonably raise, and each is available in full if raised, but none of them is needed to read the paper's argument. Nothing has been cut; the text, tables and figures are unchanged, and the paper cross-references them by their numbers here.

S23 Additional Data Figures

This section collects the descriptive figures the paper refers to but does not display, together with Figure S7, which plots the eight estimates of the estimator comparison as a forest plot. The table in the paper gives the same numbers; the figure is the faster way to see that the sign and rough magnitude do not turn on the estimator.

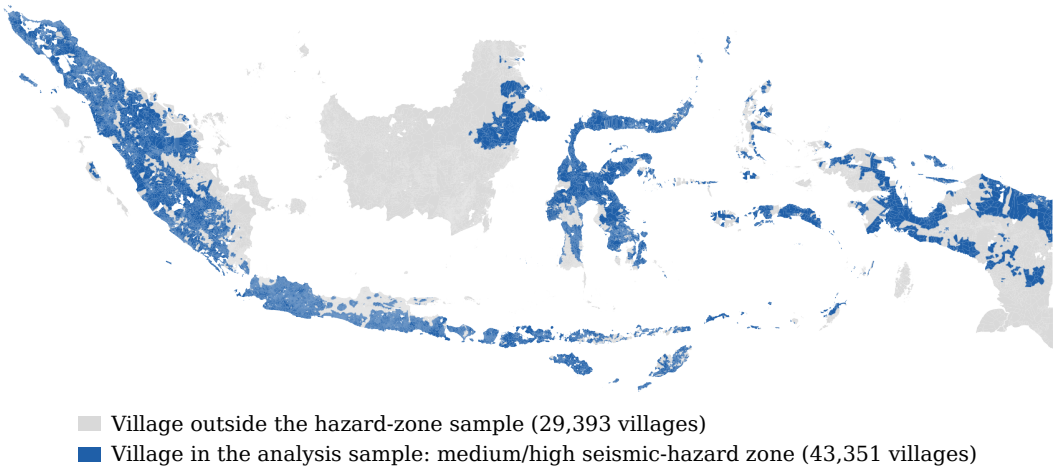


Figure S4: Analysis Sample: Medium/High Seismic-Hazard Zones
Note: Each village is drawn as its constant 2005 boundary polygon. Blue: the analysis sample, villages in the medium/high seismic-hazard zone (43,351 villages); light grey: the rest of Indonesia, outside the sample (30,001 villages). Hazard zones from the official seismic-hazard map of the Center for Volcanology and Geological Hazard Mitigation (PVMBG); village polygons from Statistics Indonesia (BPS).

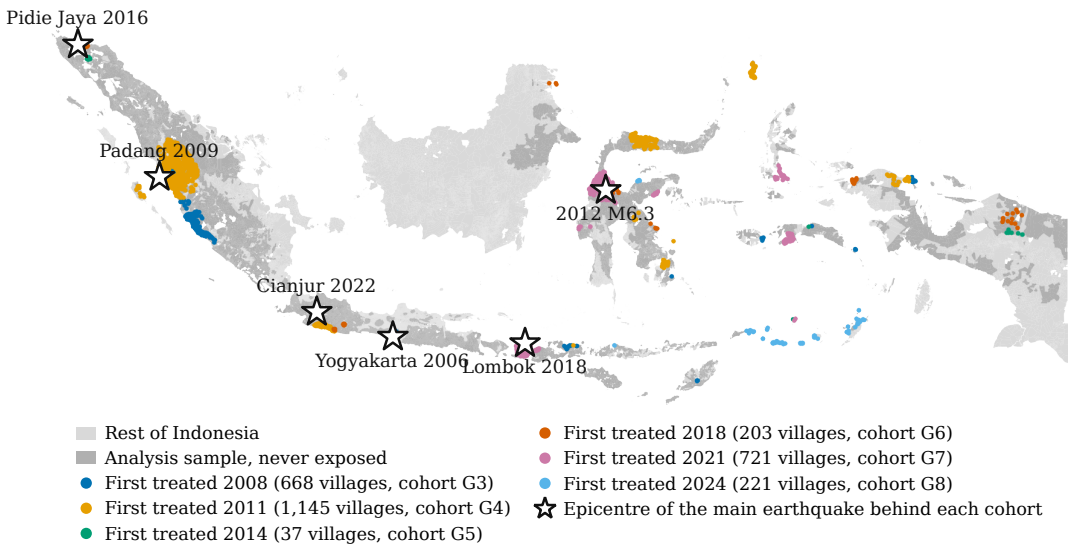


Figure S5: Ever-Treated Villages by First-Exposure Cohort, with the Main Earthquake Behind Each Cohort

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

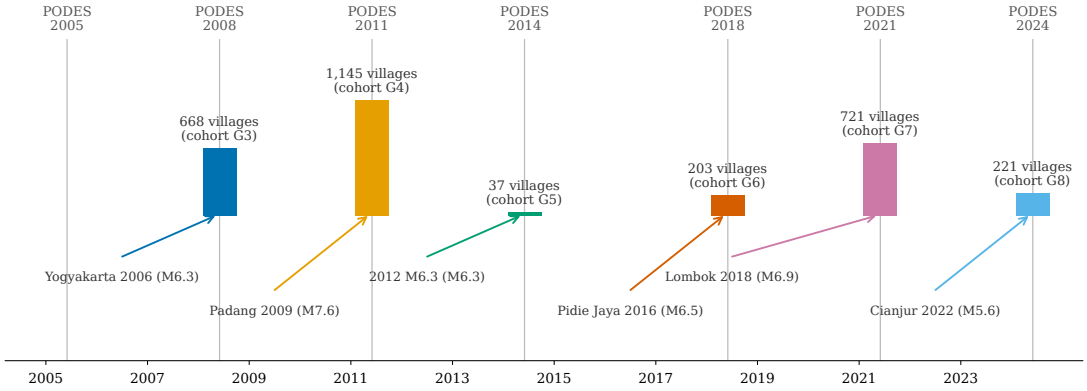


Figure S6: Treatment Cohorts on the Survey Calendar

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

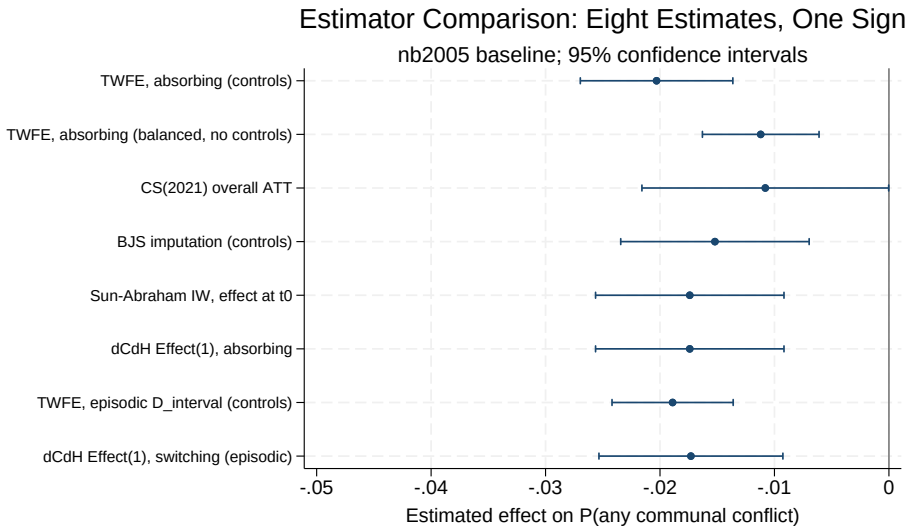


Figure S7: Estimator Comparison: Eight Estimates, One Sign
Note: Point estimates and 95% confidence intervals for the eight specifications in Table S57. Outcome: any communal conflict. Sample: the 2005-boundary panel, window-aligned exposure timing. The two TWFE rows are estimated with baseline covariates ($N = 225{,}372$) and on the balanced no-controls sample ($N = 232{,}778$); neither is the no-controls full-panel estimate of -0.0199 used as the reference point in the robustness discussion, which retains the always-treated cohort ($N = 303{,}457$). The takeaway is not the exact equality of coefficients across estimators but the stability of the sign and magnitude: all eight estimates are negative and significant at the 1 percent level, and lie in a narrow band of roughly -1.3 to -2.0 percentage points, so the result is not an artifact of a single estimator or treatment definition. These are not eight independent tests, since they draw on the same data and identifying variation; the point is robustness to estimator choice, not independent replication.

S23.1 Event-Size Threshold and the Two Event-Level Estimands

The stacked event design of Section S4.6 keeps earthquakes with at least twenty in-sample first-treated villages, which yields 19 events. The catalogue contains forty-seven earthquakes with at least two such villages. Because each stack is built on its own controls (Wing, Freedman, and Hollingsworth, 2024), an earthquake's own difference-in-differences estimate does not depend on which other earthquakes are included, so the table below is computed from a single set of per-event estimates re-aggregated at each cutoff.

Table S55: Event-Level Estimands by Minimum Event Size
The treated-village-weighted average is unmoved by the cutoff; the equal-event mean is not.
Note: Each row re-aggregates the same per-earthquake difference-in-differences estimates over the earthquakes that meet the stated minimum number of in-sample first-treated villages. Outcome is the composite communal-conflict indicator. Each per-event estimate uses village and wave fixed effects on an event-relative window of $[-2, +2]$ against never-treated and not-yet-treated controls. The p-value for the treated-village-weighted average comes from an earthquake block bootstrap that resamples whole events (9,999 draws, seed 42); the p-value for the equal-event mean is a two-sided t-test on the event-specific coefficients. The twenty-village row reproduces the estimates reported in Section S15.

	Sample		Treated-village-weighted		Equal-event mean	
Minimum treated villages	Events	Villages	Estimate	p	Estimate	p
≥ 20 (paper)	19	2797	−0.0268	0.005	−0.0155	0.315
≥ 10	28	2918	−0.0276	0.002	−0.0258	0.065
≥ 5	33	2947	−0.0280	0.003	−0.0311	0.021
≥ 2	47	2986	−0.0267	0.004	−0.0048	0.764

S24 Composition by Cohort

The crime module is fielded in all seven PODES waves, so the theft and conflict effects are identified on the same seven-wave grid and by the same cohorts. This appendix reports both outcomes cohort by cohort on the common sample, so the reader can see which cohorts carry the contrast.

Crime outcomes are estimated without covariates, so their pre-trend diagnostics are unconditional. Adding the standard baseline covariates leaves the theft point estimate similar (+0.062, $p < 0.001$) but, by shortening the pre-period, causes it to fail the pre-trend test; we therefore prefer the no-controls specification, whose pre-trend also fails (−0.043, $p = 0.04$). The role theft plays does not depend on that test: rejecting a uniform collapse in reporting capacity requires only that the same respondent, in the same villages and the same recall window, recorded more theft while not recording more communal conflict. An earlier version of this passage attributed the cohort weighting of the theft effect to the crime module being absent in 2011 and 2014. That was wrong: the module is fielded in all seven waves, and those two were missing from our extraction rather than from the survey. With them restored, every event-time cell draws on the full cohort set, so it is worth asking whether the contrast is a cohort artifact. It is not, though the support is narrower than a uniform pattern. Estimating both outcomes cohort by cohort on the common sample (Appendix Table S56), communal conflict falls while theft

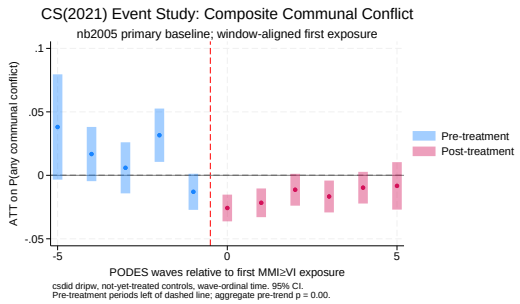
rises in the 2008, 2011 and 2021 cohorts, which together hold 84 percent of the treated villages in this sample. The three exceptions are the smallest cohorts: 2014 (37 villages), 2018 (predominantly Aceh, where communal conflict rises from a near-zero floor, as Table S1 and its discussion already report), and 2024 (observed at one post-exposure wave only). The contrast is therefore carried by the large cohorts rather than holding uniformly across all of them.

Table S56: Communal Conflict and Theft, Cohort by Cohort
Common crime-module sample; never-treated comparison; village and wave fixed effects

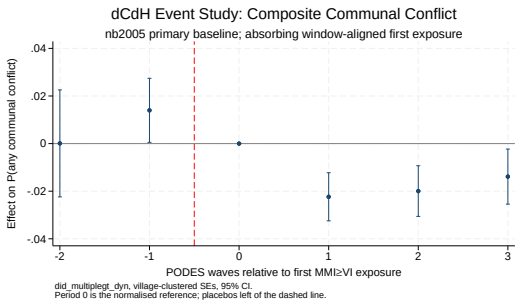
Cohort	Villages	Communal		Theft	
		ATT	(SE)	ATT	(SE)
G3 (2008)	549	−0.0156	(0.0085)	0.0616	(0.0223)
G4 (2011)	980	−0.0091	(0.0058)	0.1199	(0.0145)
G5 (2014)	37	0.0312	(0.0292)	−0.0376	(0.0627)
G6 (2018)	170	0.0208	(0.0087)	0.0797	(0.0283)
G7 (2021)	694	−0.0373	(0.0076)	0.0479	(0.0156)
G8 (2024)	219	−0.0058	(0.0152)	−0.0393	(0.0322)

Each cohort is estimated separately on the common crime-module sample against never-treated villages, with village and wave fixed effects and village-clustered standard errors. The always-treated cohort is omitted: it is treated in every wave, so its within-village coefficient is not defined. Communal falls while theft rises in G3, G4 and G7, which together hold 2,223 of the 2,649 treated villages in this sample (84 percent). The exceptions are the smallest cohorts: G5 (37 villages), G6 (predominantly Aceh, where communal conflict rises from a near-zero floor), and G8 (the 2024 cohort, observed at one post-exposure wave only).

S25 Event Study Figures



(a) CS(2021), 2005-boundary panel



(b) dCdH, 2005-boundary panel

Figure S8: Event Study, 2005-boundary panel: Callaway–Sant'Anna and de Chaisemartin–D'Haultfoeuille

Note: Dynamic effects in event time, waves relative to first MMI≥VI exposure, with pointwise 95 percent confidence intervals. Outcome: any communal conflict. Panel (a) is Callaway–Sant'Anna (2021), doubly-robust, not-yet-treated controls, `long2` convention, $N = 225{,}372$. It is the conventional benchmark on the wave-fixed-effects panel, where its joint pre-trend test rejects, and it is shown here beside the episodic estimator; the main-text event study of Figure 1 instead uses Sun–Abraham under province-by-wave effects, for the reasons given in that caption. Panel (b) is the de Chaisemartin–D'Haultfoeuille non-absorbing interval estimator on the same panel. Under wave fixed effects the average pre-treatment lead is insignificant for both outcomes while the joint Wald test rejects; the preferred province-by-wave specification passes both (Section 6.2).

S26 Fixed-Effect Structures, Estimator Comparison, and Decay

This appendix collects the four tables that Sections 5.1 and S4.7 argue from: the full estimator comparison, the inter-village and inter-group results under province-by-wave effects, the progression of the estimate across fixed-effect structures, and the decay profile against months since the most recent damage-level shaking.

Table S57: Estimator Comparison: Effect of Earthquake Exposure on Communal Conflict (nb2005 Baseline)

	Estimate	SE	Controls	N
Panel A: Absorbing first exposure ($D_{treat} = \mathbf{1}\{t \geq g\}$, window-aligned)				
TWFE (D_{treat})	-0.0213^{***}	(0.0034)	Yes	225,372
TWFE (D_{treat}, balanced)	-0.0193^{***}	(0.0033)	No	232,778
CS(2021) overall ATT	-0.0177^{***}	(0.0047)	Yes	225,372
BJS imputation (Borusyak et al.)	-0.0157^{***}	(0.0037)	Yes	225,372
Sun–Abraham IW, effect at t_0	-0.0223^{***}	(0.0051)	No	232,778
dCdH, Effect_1 (absorbing)	-0.0224^{***}	(0.0052)	No	190,366
Panel B: Episodic exposure ($D_{interval}$, non-absorbing, can switch from 0 to 1 and back to 0)				
TWFE (D_{interval})	-0.0140^{***}	(0.0026)	Yes	292,712
dCdH, Effect_1 (switching)	-0.0150^{***}	(0.0040)	No	249,237
Ever-treated villages (absorbing, estimable)			2,995	
Switcher villages (episodic, Effect_1)			3,868	
Pre-trend checks			CS Pre_avg n.s.; SA t_{-2} lead $+0.0138^{**}$; dCdH joint placebo n.s.	
Observations	225,372	232,778	225,372	225,372

Notes: Outcome: any communal conflict. All rows use the survey-window-aligned exposure timing (first PODES wave enumerated after the event). Point estimates differ across rows because of sample (balanced vs. unbalanced), controls (baseline-by-wave covariates where feasible; no-controls rows use specifications for which the estimator's own first-step regression is infeasible with controls at the baseline wave), and estimand (overall ATT vs. effect at t_0 vs. switcher effect). CS(2021): doubly-robust, not-yet-treated control, analytical SEs. BJS: Borusyak, Jaravel & Spiess imputation estimator, efficient under homogeneity but requiring parallel trends in all periods (a stronger assumption than CS, which requires it only post-treatment). Sun–Abraham: interaction-weighted, never-treated control cohort. dCdH: de Chaisemartin & D'Haultfœuille. SE clustered at village level. The Sun–Abraham and absorbing dCdH rows round to the same estimate and standard error at four decimals by coincidence, not by duplication: the underlying values are -0.022349 (SE 0.005139) and -0.022354 (SE 0.005158) respectively, on different samples. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Table S58: Strong Shaking and Communal Conflict: Province × Wave Fixed Effects, Always-Treated Cohort Excluded

	(1) Any communal	(2) Mass-brawl gate	(3) Inter-village	(4) Inter-group
Strong shaking (MMI ≥ VI)	-0.0154***	-0.0173***	-0.0168***	-0.0035
	(0.0040)	(0.0042)	(0.0037)	(0.0032)
Pre-exposure mean	0.0624	0.0670	0.0371	0.0372
Effect (% of pre-exposure mean)	-24.7	-25.8	-45.3	-9.3
Pre-trend joint p	0.610	0.554	0.187	0.257
Observations	232,778	232,778	199,524	232,778

Standard errors clustered at village level in parentheses. All columns include village and province × wave fixed effects. The always-treated cohort (villages already exposed when the panel opens, 9,856 of 43,351) is excluded, which removes the units that have no pre-period; earlier-treated staggered cohorts can still serve as controls for later ones, as the regional-trends discussion in the main text reports. Pre-trend joint p tests leads at -4, -3 and -2 against zero. * p<0.1, ** p<0.05, *** p<0.01

Table S59: Inter-Village Conflict Across Fixed-Effect Structures

	(1) Wave FE	(2) Prov × wave	(3) + drop always	(4) Kab × wave
Strong shaking (MMI ≥ VI)	-0.0165***	-0.0165***	-0.0168***	-0.0095**
	(0.0031)	(0.0037)	(0.0037)	(0.0044)
Observations	260,106	260,106	199,524	199,494

Outcome is inter-village communal conflict. Village fixed effects throughout. Column (3) is the preferred specification. Column (4) absorbs 62 percent of the remaining treatment variance (6.40 to 2.43 percent); its standard-error inflation of 1.55 tracks the variance loss (predicted 1.62) and its interval contains column (3), so it is uninformative rather than contradictory. * p<0.1, ** p<0.05, *** p<0.01

Table S60: Decay of the Effect with Time Since Shaking, Earthquake Fixed Effects

	(1) Any communal	(2) Mass-brawl gate	(3) Inter-village
0–12 months	-0.0029	-0.0009	-0.0105**
	(0.0031)	(0.0033)	(0.0051)
12–24	0.0008	0.0020	-0.0018
	(0.0039)	(0.0044)	(0.0026)
24–36	-0.0094	0.0047	-0.0096*
	(0.0084)	(0.0093)	(0.0057)
36–48	-0.0053	-0.0038	-0.0064***
	(0.0034)	(0.0036)	(0.0024)
48–72	-0.0002	0.0008	-0.0008
	(0.0036)	(0.0039)	(0.0029)
72–120	0.0032	0.0041*	0.0015
	(0.0023)	(0.0025)	(0.0016)
Decay test p (recent = distant)	0.2642	0.5557	0.0121
Observations	303,456	303,456	260,105

Months elapsed between the most recent MMI $\geq$ VI shaking and the survey. Village, province $\times$ wave, and EARTHQUAKE fixed effects throughout. Event fixed effects are essential rather than optional: without them each lag window is dominated by a different earthquake in a different wave, and the profile looks flat for reasons of composition rather than timing. Identification is therefore within-exposed and across elapsed time, not exposed versus unexposed. The 120+ month bin is collinear with the event effects and is absorbed. Decay test compares the two most recent bins against the two most distant. * p<0.1, ** p<0.05, *** p<0.01

Data Availability

This is a restricted-access reproduction package, and we state precisely what can and cannot be shared.

Confidential and not redistributable. The village-level PODES microdata and the BPS village-boundary product (*Peta Desa*) are confidential products of BPS-Statistics Indonesia and may not be redistributed by us. Researchers obtain them directly from BPS under its data-access terms. Because the derived analysis panels carry BPS village identifiers, they are withheld for the same reason and must be rebuilt from source.

Public or licensed, obtained from the provider. USGS ShakeMap MMI and PGA rasters are public domain; the PVMBG/ESDM seismic-hazard zone maps and the SNPK/NVMS incident database are public; ACLED and IPUMS International require each user to register and build their own extract under the providers' terms. Exact sources and download instructions for each are given in the deposit.

Provided in full. All analysis code is included: the programs that clean the raw sources, build the village panel, and produce every table and figure in the paper and appendix. A researcher with BPS access can rebuild the panel from the code and reproduce every reported number, and each number is traceable through the deposit's verification file to the program and log that produced it.

The replication 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.

S27 Staggered DiD: Group-Level ATTs

S27.1 Cohort heterogeneity

Group-level heterogeneity. No single cohort drives the effect: five of six estimable cohort ATTs are negative, and the one positive cohort, G6, is predominantly Aceh villages beginning from a near-zero floor less than a decade past the Helsinki peace agreement. Dropping G6, or Aceh altogether, leaves the estimate unchanged or larger (-0.0203^{***}; Appendix Tables S61 and S15).

Table S61: Callaway–Sant'Anna Group-Level ATT(g)
Takeaway: five of the six cohorts have negative ATTs; the only positive cohort is G6, the Aceh Pidie Jaya cohort, whose pre-treatment conflict rate sits at a post-conflict floor and can only revert upward.

Cohort	First treated wave	ATT	SE	z	p
Panel A: Composite (any communal conflict)					
G3	Wave 3 (2008)	−0.0151	0.0082	−1.85	0.065
G4	Wave 4 (2011)	−0.0121	0.0070	−1.73	0.084
G5	Wave 5 (2014)	−0.1281	0.0704	−1.82	0.069
G6	Wave 6 (2018)	+0.0323	0.0121	+2.66	0.008
G7	Wave 7 (2021)	−0.0527	0.0118	−4.47	0.000
G8	Wave 8 (2024)	−0.0244	0.0174	−1.40	0.160
Simple ATT	(all cohorts)	−0.0177	0.0047	−3.81	0.000
Panel B: Warga *(definition-consistent)*					
G3	Wave 3 (2008)	−0.0067	0.0070	−0.96	0.336
G4	Wave 4 (2011)	−0.0056	0.0054	−1.05	0.293
G5	Wave 5 (2014)	−0.0846	0.0569	−1.49	0.137
G6	Wave 6 (2018)	+0.0252	0.0097	+2.60	0.009
G7	Wave 7 (2021)	−0.0202	0.0088	−2.30	0.022
G8	Wave 8 (2024)	−0.0182	0.0166	−1.10	0.271
Simple ATT	(all cohorts)	−0.0075	0.0037	−2.04	0.042

Group-time ATTs from Callaway–Sant'Anna with baseline PODES 2003 covariates, doubly-robust weighting and not-yet-treated controls. G6, the Aceh Pidie Jaya cohort, is the only positive-ATT group; Supplementary Material Section S7 shows that dropping it leaves the pooled estimate essentially unchanged. G5 contains 37 treated villages and its estimate is correspondingly imprecise.

References

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

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.

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

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

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

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.

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

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

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

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

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.

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

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

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.

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

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

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

de Chaisemartin, Clement and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9):2964–2996.

———. 2026. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *Review of Economics and Statistics* 108 (4):863–880.

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2018. "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns." *American Economic Journal: Applied Economics* 10 (2):202–233.

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

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

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

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

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.

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

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

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.

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

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

Olden, Andreas and Jarle Møen. 2022. "The Triple Difference Estimator." *The Econometrics Journal* 25 (3):531–553.

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

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

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

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

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

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

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

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.

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

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

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.

Wing, Coady, Seth M. Freedman, and Alex Hollingsworth. 2024. "Stacked Difference-in-Differences." NBER Working Paper 32054, National Bureau of Economic Research.

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