Believing without Belonging?
Natural Disasters and the Margins of Religiosity*

Michal Brzezinski†

August 30, 2026

Abstract

Religion is often modelled as insurance against shocks. Do natural disasters make people more religious? I decompose the question into belief, participation, connectedness, and private coping, within person, for 143,648 adults in 22 countries and Hong Kong (Global Flourishing Study) linked to 145 disasters by distance to the affected footprint. Preregistered pooled tests detect no average effect on any margin, and the coping null is precise: the interval rules out changes beyond 0.02 SD at every registered distance threshold. Exploratory evidence suggests disasters recompose religiosity rather than raise it. Between the first two survey waves attendance falls where disruption bites—about 0.025 on a 1–5 scale, heavily in one flood catastrophe—while belief in God is about 0.7 percentage points higher one wave after first exposure among those not exposed again; exposed believing attenders are 1.75 percentage points likelier to stop attending while *still believing*. The pattern fits a higher price of participation, not weaker demand for religion.

1 Introduction

That adversity turns people toward God is one of the oldest claims in social science. Individuals are more religious when an earthquake has recently struck close by, and the difference persists into the next generation of migrants (Bentzen, 2019); Google searches for prayer surged when COVID-19 arrived (Bentzen, 2021); and religious communities historically spread fastest where agricultural risk was highest (Ager and Ciccone, 2018). The canonical economic reading is insurance: religion supplies material support, social ties, and psychological comfort exactly when formal insurance fails (Chen, 2010; Auriol et al., 2020). The psychological reading, and the one to which the disaster evidence is usually attributed, is religious coping—turning to faith to make sense of events that cannot be controlled or explained (Pargament et al., 2000). Yet almost all of this evidence compares different people or places, it treats "religiosity" as a single quantity that shocks push up or down, and it rarely measures coping itself.

Religiosity is not a single quantity. Believing in God, attending services, and turning to religion for comfort are different goods with different production technologies: attendance requires time,

*Preliminary version. Data: Global Flourishing Study Waves 1–3 and restricted geolocation, accessed under GFS Registry preregistration (gmc5p) with University of Warsaw ethics approval (23/2026); disaster events from GDACS. The data providers had no role in the analysis and no right to review results prior to dissemination. The author received no external funding for this research and has no financial relationships or other conflicts of interest to disclose. No output reports coordinates or any geographic cell with fewer than ten respondents. Replication code accompanies the paper.

†University of Warsaw. Email: mbrzezinski@wne.uw.edu.pl.

physical access, and a functioning congregation; belief requires none of these. A disaster that destroys roads and gathering places while raising the demand for meaning should therefore move these margins in *opposite* directions—a composition effect that the standard "more or less religion" question cannot detect, and that cross-sectional comparisons across more and less disaster-prone places cannot identify. This paper asks whether natural disasters change religiosity *within the same person*, and on which margin, using the Global Flourishing Study (GFS)—a panel following 207,919 adults in 22 countries and Hong Kong, 143,648 of whom are interviewed at least twice and form the estimation sample (Padgett et al., 2025; VanderWeele et al., 2025)—linked to 145 major geolocated natural disasters (tropical cyclones, earthquakes, floods, wildfires, volcanic eruptions, and droughts) occurring between each respondent's own interviews.[1] Exposure is measured in two ways: at the public administrative region, and—under restricted access governed by a preregistration—as the distance between each respondent's location, rounded to the nearest degree, and each disaster's footprint, the mapped area the event affected. Identification is within-person throughout: with two waves, first differences with country fixed effects; with three, individual and country-by-wave fixed effects. In both cases I compare individuals whose location was reached by a disaster between their interviews with compatriots whose location was not, netting out every time-invariant trait—including how disaster-prone a person's home region permanently is, which is precisely what confounds comparisons across people and places.

I state the preregistered answer first. The confirmatory test—the within-person effect of exposure within 100 km, the registered primary radius, pooled over all three waves—detects no average effect on any margin: attendance −0.007 (SE 0.005), belief +0.005 (0.006, under the pooled convention explained in Section 4), private coping −0.004 (0.006), none with a false-discovery q below 0.4, and design-based randomization inference agrees (Table 1). Under the registration's own decision rules, the confirmatory hypotheses are not supported. For coping the null is precise: the confidence interval excludes average responses beyond ±0.02 standard deviations at the region, at every registered distance threshold, in the 0–50 km ring, and—in an exploratory split—inside the footprint itself. Only the outer rings, which carry far fewer respondents, are less precise.

The exploratory anatomy beneath that null occupies the rest of the paper, and its pieces fit together. Attendance falls where and when disruption bites (Section 5). In the Waves 1–2 window, exposure within 100 km lowers attendance by 0.016–0.020 standard deviations, about 0.025 points on the 1–5 scale. The decline is concentrated three to six months after the event and recovers within the window; it is larger within 100 km of the footprint than at the region, and it grows monotonically with baseline financial insecurity: the decline is 0.032 larger in the most insecure third of respondents than in the most secure ($p = 0.007$). It is also carried by floods, heavily by one: the 2024 Kenyan floods account for nearly half the magnitude, and the estimate is sign-stable but insignificant without them (−0.011, $p = 0.12$). Event-level two-way clustering leaves the window estimate intact ($p = 0.03$). This is a real but concentrated disruption, not a universal law.

Belief moves later, and in the opposite direction. It is the exploratory result this paper tests hardest, and the one those tests leave least settled. The rise appears in three designs. In the Waves 1–2 window, cut by months since the event, belief peaks six to twelve months out (+0.050 under the regional measure). In the three-wave event study it is +0.019 one wave after first exposure. And in the cohort never exposed again—where a delayed response is not mixed with the effect of a

[1] Each is an Orange- or Red-alert event in the Global Disaster Alert and Coordination System (GDACS), a cooperation framework of the United Nations and the European Commission (https://www.gdacs.org); Section 2 describes the event list and how each disaster's affected area is mapped.

second shock—it is +0.026 ($p = 0.02$), about 0.7 percentage points more believers; the difference between this cohort and the re-exposed one is imprecise, and the single testable pre-exposure period shows no movement, though with one pre-period that check is weak. Unlike attendance, belief is indifferent to distance—as strong across the affected region as within 100 km—consistent with conviction responding to an event's salience rather than to physical closeness. The contrast suggests the paper's central pattern: the same shock that empties pews strengthens conviction—"believing without belonging," in Davie's phrase (Davie, 1994). Two marginal coefficients cannot establish that joint state, so I estimate the transition directly. Take those who at Wave 1 both believed and attended: by Wave 2 each has either stayed, stopped attending while still believing, stopped believing while still attending, or left both. Exposed respondents are 1.75 percentage points likelier to reach the second of these, and estimating all four shows the movement goes there rather than into either form of belief loss. Two qualifications travel with the result: the test was not registered, and it is a Waves 1–2 phenomenon. A third is a caution about reading it. Belief almost never changes between waves in these data, so believers who stop attending remain believers nearly mechanically—the transition is largely the attendance exit seen from the joint state, and what the four-way split adds is the direction it takes. Section 5 carries all three in full.

The coping null is the paper's strongest positive claim—a *precise* null on the very channel the literature invokes—so it draws the objection every null draws: that counting everyone within a wide radius as exposed diluted whatever response there was. The restricted geolocation answers that objection on its own terms. If dilution were the explanation, the estimate should grow as exposure sharpens from the whole region to the footprint itself. It does not grow; it stays flat. And the same sharpened measure does move other things—attendance down, self-rated mental health and national belonging slightly up—so exposure is registering in people's lives. What the null covers is bounded by what the data record: where a disaster struck, not whom it harmed. Injury, bereavement, and displacement go unobserved, and most people within 100 km of a footprint escape them, so this is a statement about ordinary residents of affected areas, not about disaster victims. For them, proximity does not detectably raise private religious coping.

Both responses concentrate among respondents with prior religious attachment—the affiliated and those raised attending weekly—while the never-religious show neither. That gradient is consistent with disasters reactivating religious capital rather than creating it, though no single between-group difference is precisely estimated. A simple household-production reading organizes the whole pattern (Azzi and Ehrenberg, 1975; Iannaccone, 1992). Attendance is the time- and infrastructure-intensive margin of religiosity; a local disaster raises its price—damaged buildings, displaced clergy, cleanup demands on time—just as the repeal of blue laws raised the relative price of church-going in Gruber and Hungerman (2008). Belief is produced without market inputs and responds instead to the demand side: mortality salience and the search for meaning that adversity triggers (Pargament et al., 2000; Vail et al., 2010; Norenzayan and Shariff, 2008). A shock that raises the price of the congregational margin while raising the demand for meaning organizes what I find: less attendance, more belief. Two things do not fit. The same demand story implies a rise in private coping, which is also produced without market inputs; that was the registered prediction, and it is precisely the one the data rule out—suggesting conviction and religious practice do not respond as a single good, which is the paper's own thesis applied to the demand side. And among those never exposed again, attendance is still depressed a wave later, longer than damaged buildings and cleanup time would explain.

Two adjacent literatures link religion and risk. One asks where instrumental religious practice

takes hold in the first place: Espín-Sánchez et al. (2026) show that the timing structure of environmental risk—specifically, whether the chance of rain rises the longer a dry spell runs—determines where rainmaking rituals are persuasive enough to endure over long horizons, and Butinda et al. (2023) find that anti-theft rituals raise sellers' inventory and profits only among those who believe in the ritual's efficacy. Both concern the long-run selection of practices and the role of prior belief, not the short-run response of an established believer to a shock. The other treats religion as an input rather than an outcome: Hasan et al. (2024) find that pre-existing church membership predicts stronger local economic recovery after the 2005 hurricane season, and Bentzen et al. (2026) trace how religion in government draws new attenders, concentrated within a denominational in-group. This paper runs the other way: religiosity is the outcome, and the question is what a disaster does to it, within the same person, over one to two years.

My contribution to the economics of religion is threefold. First, identification and scale: the literature's evidence on shocks and religiosity is predominantly cross-sectional or aggregate—across districts (Bentzen, 2019), historical counties (Ager and Ciccone, 2018; Belloc et al., 2016), or survivors of a single event (Aslam et al., 2022)—and the closest within-person antecedent follows one earthquake in one country (Sibley and Bulbulia, 2012). I estimate within-person responses to more than a hundred disasters across 22 countries and Hong Kong. Second, measurement and decomposition: I measure private religious coping directly—the channel to which the disaster response is usually attributed, and one that shock studies rarely record—and bound its average response near zero. Measuring belief, participation, and connectedness in the same panel then yields exploratory evidence that the "disasters increase religiosity" conclusion depends on which margin one measures—a belief-in-God measure would find religion rising, an attendance measure would find it falling, and both would be right. Third, dynamics: the transitory participation dip against the delayed belief response, though outside the registered set, distinguishes a disruption mechanism from a demand mechanism in a way level comparisons cannot.

A preregistration disciplines the restricted-data analysis: the hypotheses, the 100-km primary treatment, the outcome families, test sidedness, and the moderator set were registered before the sensitive data were accessed, and everything outside them is labeled as such. The design also passes the checks that within-person identification makes testable: disaster exposure does not predict panel attrition (retention is 62.9% among the exposed against 63.0% among the unexposed), treated and untreated respondents are balanced at baseline (all gaps below 0.03 standard deviations), and disasters that concluded in the year *before* the first interview do not predict subsequent within-person changes.

Section 2 describes the panel, the disaster data, and the exposure construction, public and restricted; Section 3 sets out the design and its threats; Section 4 gives the registered answer and the proximity test that resolves the coping null; Section 5 presents the margin decomposition, its dynamics, and the joint-state transition behind the title; Section 6 works through the threats to each claim; Section 7 concludes.

2 Data and exposure construction

Panel. I use the Global Flourishing Study (GFS), a longitudinal survey that follows the same individuals annually across 22 countries and Hong Kong (VanderWeele et al., 2025; Padgett et al., 2025). The panel contains 207,919 first-wave respondents. Of these, 128,868 complete the second annual wave and 111,486 the third: 96,684 second-wave completers (75%) who are re-interviewed, plus

14,802 who skipped the second wave and returned. Respondents observed at least twice—143,648 in all—form the estimation sample. The third wave was obtained under the study's registry access after the public Waves 1–2 analysis was completed.

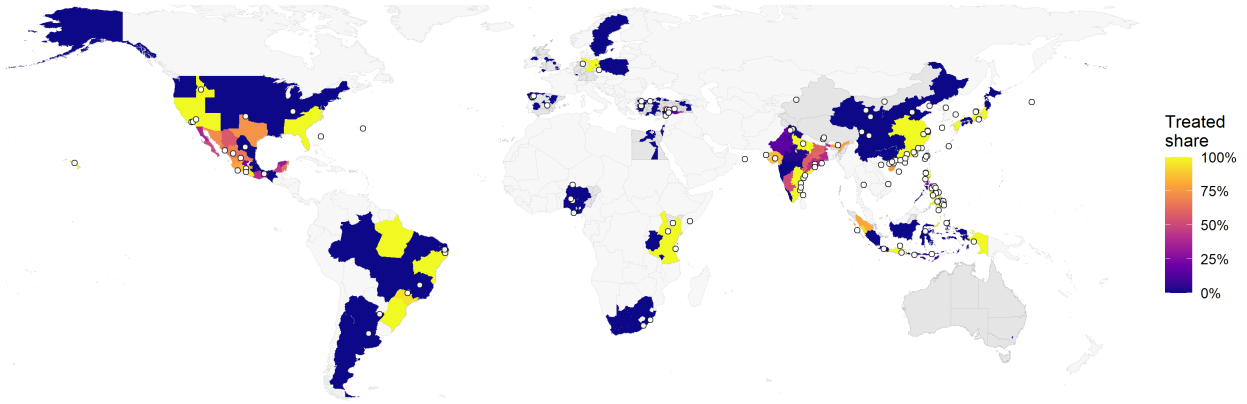
Religiosity outcomes. I study four outcome groups. *Private coping*: drawing strength or comfort from one's religion or spirituality, frequency of prayer or meditation, and feeling loved or cared for by God; the preregistered coping index averages the standardized values of these three components and is rescaled to unit variance. *Connectedness*: how often the respondent feels connected to a religion or form of spirituality. *Institutional participation*: frequency of religious service attendance. *Belief*: belief in God or a spiritual force, and belief in life after death. The third-wave codebook confirms every item's wording and scale are unchanged across waves, so within-person changes reflect respondents rather than the questionnaire. All ordinal items are reverse-coded in the raw data; I reorient them so that higher values denote more religiosity and standardize within country. Two standardization conventions are reported side by side: the registered one, within country *and* wave, which is the headline throughout, and a pooled alternative, one mean and standard deviation per country across all observed waves, which keeps within-person changes comparable over time. The resulting estimates differ by at most 0.0012 standard deviations. Belief is quoted under the pooled convention because the registered rule is undefined for it in one country, as Section 4 explains.

Exposure window. The survey records the date of every interview. The interval between consecutive interviews—a respondent-specific window of roughly twelve to eighteen months—defines the period over which I measure disaster exposure for the wave that closes it. For the minority who skip the second wave and return in the third (14,802 respondents), the window spans the full gap. The primary specification does not distinguish these long windows, so its estimate averages over windows of unequal length; the balanced three-wave check in Section 6, which drops the returners, is the corresponding robustness. Fieldwork timing differs sharply across countries, so windows are built from each person's own dates rather than a country-constant interval.

Disaster exposure. I draw major natural-disaster events from the Global Disaster Alert and Coordination System (GDACS), retaining all Orange- and Red-alert events—the system's two highest severity tiers—between January 2022 and January 2026; with third-wave fieldwork ending in December 2025, the inventory covers every respondent's last interval in full. The Waves 1–2 windows draw on 101 events in the study countries: 36 tropical cyclones, 28 earthquakes, 25 floods, 7 wildfires, and 5 volcanic events. Extending coverage through the third wave adds 44 more (145 in total); the new events are appended without re-querying the earlier ones, so adding the third wave could not alter the original Waves 1–2 exposure assignment. For each event I retrieve the affected-area polygons—for cyclones, the union of timestamped wind-field footprints along the track—rather than relying on a single centroid, which lets a storm that crosses many provinces register exposure in all of them. I intersect these footprints with first-level administrative boundaries and link them to GFS public regions, which are reported at the state or province level for most countries. A respondent is exposed if a qualifying event's footprint reaches their baseline region and the event's active period overlaps their personal window. Across the Waves 1–2 windows this yields 44,764 exposed respondents, distributed across cyclones (18,413), floods (16,338), and earthquakes (9,584); hazard categories overlap, and wildfire and volcanic exposure are not tabulated separately, so the components need not sum to the total. Distinguished by geographic reach, 36,922 respondents experienced a

local disaster (an event reaching at most half of their country's matched regions) and 7,842 only a whole-country shock. Treated and untreated regions coexist within most countries (Figure 1); Online Appendix A gives the breakdown of treated respondents by country and hazard.

Figure 1: Disaster exposure across the study countries



Note: Administrative regions of the 22 study countries and Hong Kong are shaded by the share of respondents treated; the remaining land is drawn in outline for geographic context only and carries no data. Gray regions lie inside a study country but could not be matched by name to a survey region, so they carry no data. White points mark the centroids of all 145 GDACS Orange- and Red-alert events, one marker per event regardless of hazard type; the hazard composition is in Online Appendix A by country and in Online Appendix J by outcome margin. The map shows that treated and untreated regions coexist within most countries, which is the variation the within-country comparison uses. Online Appendix A reproduces the map at full-page size.

Restricted geolocation. The analyses in Sections 4–6 use the GFS restricted geolocation, obtained under an accepted GFS Registry preregistration.[2] The preregistration binds the confirmatory analysis to the hypotheses and specifications registered in advance; everything beyond them is labeled as such. Each respondent's location is recorded at each annual interview as the median point of their interviewing cluster—the small local area within which face-to-face fieldwork is organized—or, in the web-survey countries, the centroid of their postal code; both are rounded to the nearest whole degree, a deliberate disclosure protection of roughly 100 km resolution. The rounded location is the respondent's *grid cell*: everyone whose true location rounds to the same point shares a cell. I take the rounded point as the location estimate and measure the geodesic distance from it to each event's affected-area polygon (zero if the point lies inside the footprint; events reported only as a centroid enter as a 50-km buffer around it, here and in the regional assignment). Exposure at each interval uses the location reported at the interval's *start*, so that moving in response to a disaster cannot reassign exposure. Coverage is near-complete: 99.98% of respondents have usable baseline coordinates. Exactly one recorded location is plainly erroneous—it lies more than 100 km outside its own country—and is excluded. The primary treatment, fixed in the preregistration, is an indicator for any qualifying footprint within 100 km during the interval. The preregistered robustness set

[2]With ethics approval from the University of Warsaw. Published output contains no coordinates and no statistic computed on fewer than ten respondents in any grid cell; analysis files in the project repository carry only binned distances, and the restricted files are held in access-logged storage, per the ethics protocol.

varies it along each dimension fixed in advance: wider and narrower radii (50, 150, and 250 km), non-overlapping distance rings in place of nested thresholds, exposure graded by the events' severity scores, separate indicators for each hazard type, and bands for months since the event. The whole-degree rounding can displace any location by up to roughly 50 km, so I construct no distance band narrower than that: the data cannot tell such distances apart. Few respondents move: 3.6% change grid cell across their second-wave interval and 3.3% across their third (4,664 and 3,656 respondents; Online Appendix D reports the mover checks).

3 Empirical strategy

Specification. With two waves the within-person estimator is a first difference. Let R_{ic} denote a standardized religiosity outcome for individual i in country c. I estimate

$$\Delta R_{ic} = \beta D_{ic}^{\text{local}} + \delta D_{ic}^{\text{natl}} + \lambda_c + \varepsilon_{ic}, \tag{1}$$

where ΔR_{ic} is the Waves 1–2 change, D_{ic}^{local} indicates that a *local* disaster—one reaching at most half of the country's matched regions—was active in respondent i's baseline region during their personal between-interview window, D_{ic}^{natl} indicates exposure to a whole-country shock, and λ_c are country fixed effects, which in a differenced equation absorb any common national shift between the waves. Differencing removes every time-invariant individual trait: childhood religion, denomination, personality, and the fixed disaster-proneness of the respondent's location, the confound that complicates cross-sectional comparisons of religiosity across hazard-prone places (Bentzen, 2019). The coefficient of interest is β, identified from comparing exposed and unexposed respondents *within the same country*. That comparison needs unaffected compatriots, which is what whole-country shocks—events reaching more than half of their country's matched regions; 18 of the 128 country-events, including the late-2023 Kenyan floods—mostly eliminate. Ten of the eighteen reach every matched region, so no same-country control exists at all. The remaining eight reach 52–96% of matched regions, and the few regions left untouched sit beside a disaster that struck most of their country—plausibly affected through markets, family networks, and national attention—so they are contaminated rather than clean controls. D^{natl} therefore serves only to keep nationally exposed respondents out of the control group, and I attach no interpretation to δ: its estimate flips sign depending on how exposure is coded—regional or proximate, retained or dropped national rows—so it measures nothing stable. Standard errors are clustered at the region level in this two-wave design.

Three waves and proximate exposure. With the third wave and the restricted geolocation, the preregistered primary specification generalizes (1) to a two-way fixed-effects panel in levels,

$$R_{ict} = \alpha_i + \lambda_{c\times t} + \beta D_{i,t-1\to t}^{100} + \delta D_{i,t-1\to t}^{\text{natl}} + \varepsilon_{ict}, \tag{2}$$

where α_i are individual fixed effects, $\lambda_{c\times t}$ country-by-wave effects, and D^{100} indicates a qualifying footprint within 100 km of the respondent's start-of-interval cell during the personal window ending at wave t; with two waves and regional exposure, (2) is identical to (1)—the two give the same estimates—so the bridge table (Table 3) can vary the exposure measure while everything else stays fixed. Because D^{100} records exposure only in the interval that closes at t, β measures the within-interval response. Responses that arrive later, or persist from earlier exposure, are not separated out here; the dynamic designs below do that, and the pooled null does not by itself rule them

out. Standard errors are clustered on the respondent's grid cell, with region-clustered variants reported for comparability. The preregistration fixes one-sided tests for the directional hypotheses (coping upward, attendance downward), two-sided tests elsewhere, and Benjamini–Hochberg false-discovery q-values within the primary and secondary outcome families.

Those two directions are where the household-production reading enters the design itself, not just the interpretation. Attendance needs time, travel, and a functioning congregation; private coping needs none of these. A disaster that raises the price of attending while raising the demand for meaning should therefore push attendance down and coping up—and the registration committed to exactly those directions in advance. Separating the margins is therefore not a robustness exercise—a single index would net the two effects against each other and report their difference as the answer.

Dynamics require more care than a two-way fixed-effects event study would give them. Respondents are first exposed at different waves, so treatment is staggered, and if the response differs across cohorts—plausible here, since the events themselves differ in severity and hazard—the conventional estimator weights some cohort-period effects negatively and can return a coefficient outside the range of every underlying effect. Event-study dynamics therefore use the interaction-weighted estimator of Sun and Abraham (2021) around the first exposed wave, with never-exposed respondents as the clean control group, and the Borusyak et al. (2024) imputation estimator as robustness; the two rest on different assumptions about which comparisons identify the effect, so their disagreement is informative rather than a nuisance, and Section 6 reports it. A confirmatory dynamic claim requires the pre-exposure lead to be jointly insignificant. Design-based inference asks how unusual the real estimate is among fake geographies: within each country, respondents' grid cells are reshuffled among them one-for-one, exposure is rebuilt from the shuffled locations, and (2) is re-estimated on each of 1,000 such draws.

Identification. The identifying assumption is that, conditional on individual and country-by-wave fixed effects, whether a disaster reaches a respondent's location during their interview window is unrelated to the within-person trajectory their religiosity would otherwise have followed. Formally, strict exogeneity: $\mathbb{E}\big[\varepsilon_{ict} \mid \{D^{100}_{i,s-1\to s}\}_{s=2}^{T}, \alpha_i, \lambda_{c\times t}\big] = 0$—the time-varying disturbance must be unrelated to the respondent's entire exposure path, not only to current exposure. The *level* of religiosity may correlate with exposure freely; the individual effect absorbs it.

Two features make that assumption plausible here rather than merely convenient. First, treatment is the intersection of two processes that are separately determined and neither of which responds to an individual's religiosity. Whether a qualifying event occurs, when, and how far its footprint extends is geophysical and meteorological; when a respondent is interviewed is set by the survey's fieldwork calendar, which differs sharply across countries and is fixed long before any particular event. A respondent is treated when these two happen to overlap. Second, exposure is assigned from the location reported at the *start* of the interval, so a respondent who moves in response to a disaster cannot be reassigned by that move—an endogeneity that a contemporaneous location measure would admit.

What the assumption does rule out is subtler and worth naming. It fails if the scheduling of fieldwork within a country correlates with regional religiosity trends in a way that lines up with when events strike—for instance if a survey firm systematically visits more devout provinces in the same months that a monsoon arrives. Country-by-wave effects absorb any common national timing, so the concern is specifically differential, within-country, region-by-season scheduling. It is not directly testable, but it has implications that are: exposed and unexposed respondents should look alike at

baseline, exposure should not predict who stays in the panel, and events that concluded *before* a window opened should not predict changes within it. Section 6 reports all three.

Because differencing absorbs fixed regional disaster risk, β is the effect of a region being *reached* by an event during the window. It is an intention-to-treat magnitude, and plausibly smaller than the effect of direct exposure would be, since many respondents coded as exposed were far from the impact zone or interviewed long after the event; I do not press that into a formal bound, because misclassification, spillovers, and dynamics could each break the ordering. Two limits on what β means follow. First, it is *proximity* to an event rather than personal loss—Section 6 finds no average rise in hardship among the exposed. Second, if belief responds to region-wide salience, respondents just outside the exposure radius are partially treated and the within-country contrast is conservative for the belief margin; the stacked design's 250-km clean-control rule (Section 5) addresses precisely this.

Threats and diagnostics. Four concerns are testable, and Section 6 reports each test. *Selective attrition*: the design conditions on Wave 2 completion, so disaster-correlated dropout would bias β; I test whether exposure predicts retention using a counterfactual window for all Wave 1 respondents, running from each respondent's first interview to that date plus their country's median between-wave gap, so that it requires no second interview date. *Selection on levels or trends*: I test baseline balance on demographics and religiosity, and I run a placebo assigning local disasters that *concluded* in the year before the Wave 1 interview, estimated on the subsample whose Wave 1 interview is late enough for a full year of event coverage; such events should not predict subsequent within-person changes. The registration's term is "placebo exposure"; strictly it is a carryover check on real prior events—a null shows limited persistence into the window—rather than a test of untreated parallel trends. *Few-cluster inference*: identifying variation is spatially clustered—regional in the two-wave design, cell-level in the proximate one—so alongside clustered standard errors I report randomization inference matched to each design: permuted region-level event histories within countries for the two-wave window, and the cell-location scheme described above for the registered specification, each with full re-estimation on 1,000 draws. *Representativeness*: I re-estimate with the GFS longitudinal weights, which restore national representativeness, and with inverse-probability attrition weights—each retained respondent weighted by the inverse of their estimated probability of staying in the panel, estimated from Wave-1 characteristics—so that stayers stand in for those who left.

4 The registered answer

Table 1 accounts for every registered hypothesis against its primary test before any exploratory result appears. The registered primary analysis—equation (2), pooled over all three waves with the 100-km footprint indicator—detects no average effect on any margin: attendance −0.007 (SE 0.005; one-sided $p = 0.08$), the coping index −0.004 (0.006), and belief +0.005 (0.006, under the pooled yardstick—the registered one cannot be computed for belief in Egypt, as explained below). Correcting for the five outcomes of the registered primary family (attendance, the coping index, its comfort and prayer components, and connectedness), the smallest false-discovery q is 0.41—attendance, the only outcome whose sided p is small—and randomization inference over 1,000 within-country permutations of cell locations agrees ($p = 0.26$, 0.49, and 0.48). The registered interaction batteries—36 preregistered contrasts interacting exposure with personal moderators (age, gender, financial strain, loneliness, baseline religiosity) and country context (religiosity, state capacity, traditional values),

for attendance and the coping index—do no better: their smallest Benjamini–Hochberg q is 0.12. Under the preregistration's own decision rules, H1–H4 are not supported, and only the belief-lag component of H5—the prediction that belief responds with a delay—survives as suggestive (Section 5).

Table 2 asks whether the answer depends on implementation: it re-estimates the primary analysis under each defensible reading of the registration. Two choices separate the readings. The registration standardizes outcomes within country *and* wave, whereas the pooled alternative uses one mean and standard deviation per country. And the registration drops whole-country *shocks* from the within-country estimates—which the exposure construction already does, since events reaching more than half of a country's matched regions never generate a treated observation—while the affected respondent–intervals remain in the sample with a national-shock control, which is what makes the registered contrast between local and national exposure estimable at all. Column (1) implements that reading and is the headline throughout. It differs from the pooled yardstick in column (3) by at most 0.0012 standard deviations, and for the primary family the design-based p-values move by less than 0.02. Belief's design-based p moves more (0.56 against 0.48), because the registered convention is not equivalent for that outcome.[3] Columns (2) and (4) apply a stricter reading still, dropping the national-event respondent–intervals outright; under it the pooled attendance estimate moves from -0.007 to $+0.000$ ($p = 0.53$), which makes the null cleaner rather than weaker.

These are precise nulls, not uninformative ones. For the preregistered coping index the 95% interval excludes average responses beyond ±0.02 standard deviations at the affected region, at every registered distance threshold, and in the innermost ring: $[-0.011, +0.008]$ at the region, $[-0.014, +0.008]$ within 50 km, $[-0.015, +0.007]$ within 100 km, $[-0.018, +0.003]$ within 150 km, $[-0.014, +0.007]$ within 250 km, and $[-0.016, +0.008]$ in the 0–50 km ring. An exploratory refinement—not among the registered thresholds—takes the limit case, respondents whose own grid cell lies inside the event's mapped affected area: there too the interval is $[-0.013, +0.011]$. Only the outer rings—the mutually exclusive bands beyond 50 km, each carrying far fewer respondents—are less precise. For attendance the interval excludes average declines larger than 0.017 standard deviations—about 0.025 points on the 1–5 attendance scale. The ±0.02 boundary is the interval the data delivered rather than a preregistered equivalence margin, so two anchors give it content. In raw units, 0.02 standard deviations of the yes/no comfort item is about 0.8 percentage points, and of the 1–5 prayer scale about 0.02 points. And in the registered upward direction the exclusion is tighter: neither the affected region, nor any registered threshold, nor the 0–50 km ring admits a rise of even +0.01 standard deviations—smaller than the responses this panel does detect, the Waves 1–2 attendance decline (0.016–0.020 in magnitude) and the clean-cohort belief rise (+0.026).

That precision is what makes the coping null informative rather than merely negative, because the obvious way to explain the null away was dilution. When only the public regional geography was available, two readings survived: a true null, or a real effect watered down—attenuated—by a yes/no region-level measure that counts respondents hundreds of kilometres from the impact zone as exposed. The restricted geolocation was requested to tell the two apart, and it does. Measuring exposure by distance is the dilution story's own test: sharpening the measure should make the estimate grow. It does not. The preregistered coping index is -0.004 (SE 0.006) at 100 km, -0.004

[3] Belief entered the registration as a secondary outcome: the primary hypotheses concerned coping and attendance. In Egypt at Waves 2 and 3 everyone gives the same belief answer, so the within-country-wave standard deviation is zero and the registered z-score cannot be computed; the 9,251 Egyptian belief observations drop from that regression. Belief is therefore quoted under the pooled convention wherever it appears.

Table 1: Preregistration accounting: every registered hypothesis against its primary test

Hypothesis	Primary test	Estimate [95% CI]	p	q	RI p	Decision	Exploratory follow-up
H1: private coping rises	Coping index, registered primary, one-sided > 0	-0.004 $[-0.015, +0.007]$	0.77	0.98	0.49	**Not supported** (precise null)	Null at every radius and inside footprints
H2: attendance falls short-run	Attendance, registered primary, one-sided < 0	-0.007 $[-0.017, +0.003]$	0.08	0.41	0.26	**Not supported**	A decline appears in the exploratory Waves 1–2 window, 0–6 months after events; concentrated in floods, in Kenya, and among the financially insecure (Section 5)
H3: context amplifies (religious context, collective shocks)	16 registered context interactions	min $p = 0.005$ (non-monotone)	—	min $q = 0.12$	—	**Not supported**	Whole-country events enter only as controls (Section 3)
H4: coping concentrates (attached, insecure, older)	10 registered coping interactions	min $p = 0.023$	—	min $q = 0.28$	—	**Not supported after FDR**	Only the age contrast points the predicted way before correction ($p = 0.45$); the who-responds gradients are in Section 5
H5a: coping responds fast, then decays	Coping timing bands; earliest (0–3 month) band	-0.024 (SE 0.011)	0.03	—	—	**Contradicted** (early *dip*)	Returns to zero after three months; comfort estimate coincides (-0.024, $p = 0.03$)
H5b: belief responds slowly, persists	Event-study rel+1; pre-trend gate	$+0.019$ (SE 0.008)	0.02	—	—	**Suggestive**	Belief rises one wave later in the cohort never exposed again: 0.026 ($p = 0.022$; design-based $p = 0.04$); the clean-vs-re-exposed gap is imprecise ($p = 0.09$); pre-exposure leads show nothing ($p =$ 0.29–0.86)

Note: The pooled two-way fixed-effects specification over all three waves with the 100 km footprint indicator is the registered primary analysis. The H1–H2 rows report its estimates with outcomes standardized within country×wave, as the registration specifies; they equal column (1) of Table 2. p-values are one-sided where registered (H1 upper, H2 lower); q-values are Benjamini–Hochberg false-discovery-rate (FDR) adjustments within the registered five-outcome primary family, computed on the same sided p-values the table displays; RI p-values come from 1,000 within-country permutations of cell locations with full re-estimation under the same convention. Registered interaction batteries are summarized by their smallest p and BH q across 36 contrasts. Exploratory follow-ups (right-most notes; Sections 5–6) are labeled throughout and none enters a confirmatory claim. The Wave 1–2 estimates that motivated the registration predate it and are therefore exploratory by construction.

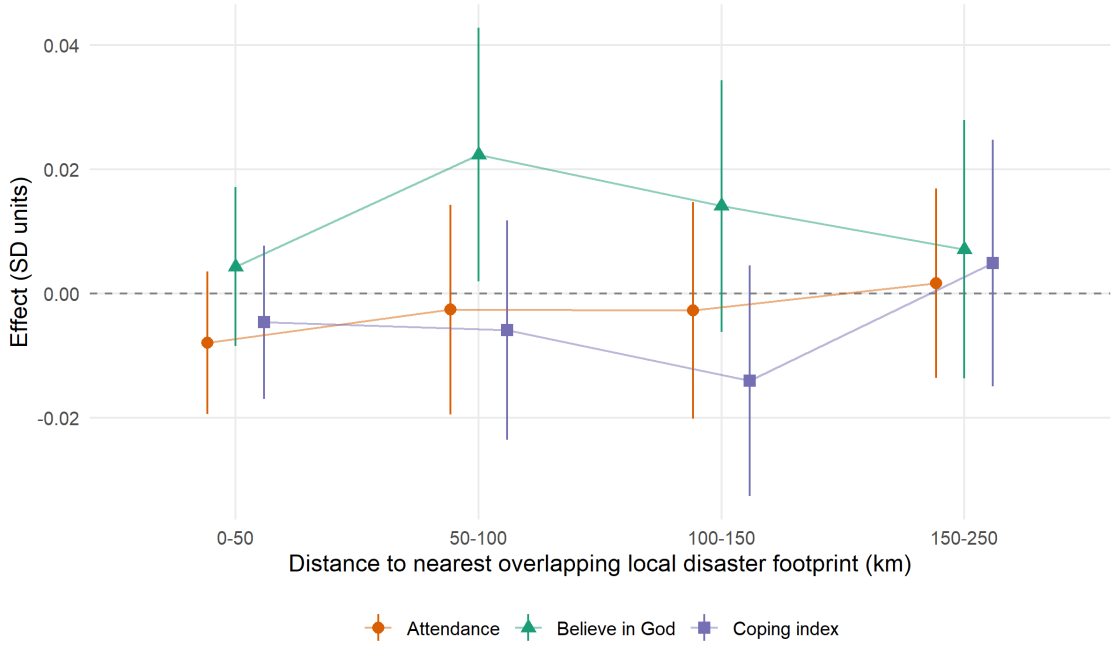
Table 2: Preregistered within-person estimates: the registered specification and its alternatives

	(1) **Registered**	(2) Registered strict	(3) Pooled-z (former)	(4) Pooled-z strict	95% CI of (1)	RI p
Attendance	−0.0069	+0.0003	−0.0068	+0.0018	[−0.017, +0.003]	0.26
	(0.0050)	(0.0049)	(0.0052)	(0.0049)		
Coping index	−0.0042	+0.0009	−0.0044	+0.0010	[−0.015, +0.007]	0.49
	(0.0055)	(0.0055)	(0.0056)	(0.0056)		
Comfort	−0.0045	+0.0007	−0.0052	+0.0002	[−0.016, +0.007]	—
	(0.0057)	(0.0062)	(0.0058)	(0.0063)		
Prayer	+0.0020	−0.0011	+0.0026	−0.0001	[−0.010, +0.014]	—
	(0.0060)	(0.0063)	(0.0060)	(0.0063)		
Connectedness	+0.0001	−0.0052	+0.0009	−0.0042	[−0.011, +0.011]	—
	(0.0055)	(0.0063)	(0.0056)	(0.0064)		
Secondary outcome						
Believe in God	+0.0042	+0.0048	+0.0054	+0.0066	[−0.007, +0.015]	0.56
	(0.0057)	(0.0063)	(0.0059)	(0.0063)		
Observations (coping index)	383,897	341,982	383,897	341,982		
Observations (belief)	320,876	282,581	330,127	291,832		
National-event rows	retained	excluded	retained	excluded		
Standardization	ctry×wave	ctry×wave	ctry	ctry		

Note: Within-person effect of a qualifying disaster footprint within 100 km during the interval, on each standardized outcome; individual and country×wave fixed effects; standard errors clustered on the respondent's grid cell in parentheses. Column (1) is the registered specification and the paper's headline: outcomes standardized within country×wave, as the registration states, with whole-country shocks entering as a control rather than as local exposure (Reading A). Columns differ from their neighbour in one element at a time — (2) adds the stricter reading that also drops respondent–intervals overlapping a national event, (3) replaces the yardstick with one mean and standard deviation per country pooling waves, (4) applies both changes. "95% CI of (1)" gives the registered interval; "RI p" is a design-based p-value from 1,000 permutations of grid-cell locations within country, with full re-estimation on each draw, computed on the same country×wave standardized outcomes as column (1); it is run for the three outcomes whose nulls the text leans on—attendance, the coping index, and belief—while the index's comfort and prayer components and connectedness are not separately permuted. Registered sidedness is one-sided for the coping family (upper) and attendance (lower) and two-sided otherwise; Benjamini–Hochberg q-values are computed on those sided p-values within each column's primary family, and none falls below 0.41 (the minimum, attendance in column (1)). Counts are shown for the coping index only; the other primary outcomes differ solely through item non-response, ranging from 382,098 to 383,897 in column (1). Belief is reported below the rule because the country×wave z-score is undefined for Egypt at Waves 2 and 3, where the item has zero variance: those 9,251 rows leave columns (1) and (2), so belief is quoted in the text under the pooled convention, column (3).

(0.006) in the 0–50 km ring, and −0.001 (0.006) inside the footprint (Figure 2). These distance categories are built from the rounded grid cells, not true locations: the cell centre stands in for the respondent, so any stated distance can be off by up to roughly 50 km, and Section 6 reports a simulation that bounds the consequences. Narrower bands also change *who* is treated, not only how well exposure is measured, so the flat profile weighs against a large uniform coping response rather than by itself ruling out attenuation in the regional design. What proximate exposure *does* move—attendance down, self-rated mental health and national belonging slightly up—suggests the treatment is real and the median nearby respondent bears little personal loss. The data therefore support a specific claim: for the typical resident of an affected area—a bystander to the event rather than its victim—proximity does not detectably raise private religious coping.

Figure 2: Religiosity response by distance to the nearest disaster footprint



Note: Within-person two-way fixed-effects estimates on mutually exclusive distance rings, relative to respondents with no overlapping local event within 250 km; three waves pooled; whole-country shocks controlled; 95% confidence intervals, grid-cell-clustered standard errors. Distance is measured from the respondent's 0-decimal cell to the event's affected-area polygon (zero inside the footprint).

One registered prediction is contradicted outright. H5a held that coping would respond fastest immediately after exposure. Instead both the coping index and its comfort component *dip* in the first three months (−0.024 and −0.024, two-sided $p = 0.03$; the registered test looked only for a rise, so against it this dip scores p close to one) before returning to zero within the window (Table 1; the within-window profile, where the dip is visible directly, is in Online Appendix E). The sign is wrong for the coping hypothesis, not merely the magnitude.

Everything that follows is the exploratory anatomy of this null, labelled as such. The Waves 1–2 results that motivated the registration predate it and are exploratory by construction. The dynamic and proximity analyses are registered specifications, but the registration made their confirmatory status conditional on the pooled primary test rejecting first—and it did not. The summary is not that disasters move religiosity and the pooled test missed it; it is that any average response is small,

and the interesting structure sits in margins, distances, and horizons.

5 The anatomy beneath the null

Nothing in this section is confirmatory. The Waves 1–2 estimates predate the registration; the dynamic and joint-state analyses are registered designs that the decision rule of Section 4 demotes to exploratory; the rest are post-hoc constructions, labelled where they appear. What makes the pattern worth reporting is that the pieces are mutually consistent and point at a mechanism the pooled averages cannot see: a shock that raises the price of participation while raising the demand for meaning should move the margins in opposite directions, and it does.

5.1 Participation falls where disruption bites

Between the first two waves, exposure within 100 km lowers attendance by 0.016–0.020 standard deviations—about 0.025 points on the 1–5 attendance scale (equation (1) with the 100-km indicator; Table 3)—and the effect is patterned in the way a disruption account predicts. It is larger within 100 km of the footprint than at the level of the affected region, concentrated three to six months after the event and recovered by the end of the window, and it grows monotonically with baseline financial insecurity—the decline is 0.032 larger in the most insecure third of respondents than in the most secure ($p = 0.007$; Online Appendix H). By hazard it concentrates in floods: -0.029 ($p = 0.02$), with cyclones, earthquakes, and wildfires individually near zero. The split is descriptive—no cross-hazard contrast is tested—but floods are where infrastructure damage and cleanup demands on time would put a disruption effect. Online Appendix J reports the same battery for every outcome margin, including the earthquake column that bears on Bentzen (2019), taken up at the end of this section. Section 6 reports how much of this rides on a single flood system.

Table 3 is the bridge from the public regional design to the restricted-geolocation one. Its first two column pairs hold the matched Waves 1–2 sample fixed while sharpening only the exposure measure: the clean attenuation comparison, under which the attendance estimate grows rather than shrinks. The third pair widens to the full geocoded sample, and the fourth pools all three waves, arriving at the registered null of Section 4.

5.2 Belief moves later, and in the other direction

The third wave turns the within-window timing profile into an event study around first proximate exposure. Cohorts are defined by the first wave whose interval contained a local disaster within 100 km, never-exposed respondents are the control group, and the reference period is the wave before first exposure. The single available lead—Wave 1 for the Wave-3-exposed cohort—is the registered pre-trend check: joint p-values run from 0.29 to 0.86 across outcomes, which is read as showing no detectable pre-movement rather than as establishing parallel trends. For Wave-2 skippers the interval closing at Wave 3 spans both gaps, so an event from the earlier calendar window enters their history at first exposure rather than one wave later; they are a tenth of the panel, and the pooled estimates survive the balanced-panel restriction that drops them (Section 6).

The two margins are mirror images. Attendance falls on impact (-0.014, $p = 0.05$) and belief rises with a lag: $+0.011$ ($p = 0.10$) on impact, growing to $+0.019$ ($p = 0.02$) one wave later (full paths in Online Appendix E). The timing bands—months since the most recent event, now over

Table 3: From regional to proximate exposure: the attenuation test

	Region, W1–W2 (matched sample)		≤100 km, W1–W2 (same sample)		≤100 km, W1–W2 (full geocoded)		≤100 km, three waves (TWFE)	
Comfort from religion	0.000	(0.008)	−0.009	(0.009)	−0.007	(0.009)	−0.005	(0.006)
Prayer/meditation	−0.012	(0.009)	−0.009	(0.010)	−0.009	(0.010)	0.003	(0.006)
Connectedness	−0.007	(0.009)	−0.012	(0.009)	−0.012	(0.009)	0.001	(0.006)
Attendance	−0.015**	(0.007)	−0.019**	(0.009)	−0.020**	(0.008)	−0.007	(0.005)
Coping index (prereg.)	0.000	(0.009)	−0.007	(0.010)	−0.006	(0.009)	−0.004	(0.006)
Believe in God†	0.017**	(0.008)	0.009	(0.009)	0.007	(0.009)	0.005	(0.006)

Note: Each cell is a separate regression of the standardized outcome on the exposure indicator and the whole-country control, with individual and country-by-wave fixed effects. Column pair 1: the public regional exposure measure (baseline administrative region), Waves 1–2, on the region-matched sample (respondents whose survey region links to a map boundary), region-clustered standard errors. Pairs 2–3: any disaster footprint within 100 km of the respondent's grid cell at the interval start, grid-cell-clustered standard errors, on the matched and the full geocoded samples. Pair 4: all three waves pooled. The preregistered coping index averages the within-country z-scores of comfort, prayer, and feeling loved by God, rescaled to unit variance. †Secondary outcome. Benjamini–Hochberg q-values across the five primary outcomes in the pooled column, computed on the registration's sided p-values, are all above 0.47. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

all three waves—agree in shape, belief peaking at six to twelve months (+0.017, $p = 0.03$; the corresponding Waves 1–2 regional band, quoted in the introduction, is larger at +0.050). Unlike attendance, belief is indifferent to distance—as strong across the affected region as within 100 km—consistent with conviction responding to an event's salience rather than to physical closeness.

The response is also item-specific, and that deserves naming rather than silence. Belief is measured with two items; only belief in God moves. Belief in life after death is flat in the pooled test (−0.004, SE 0.007) and across every hazard in the all-outcomes battery (Online Appendix J). If proximity genuinely raised religious conviction, both indicators of the construct might be expected to respond; that only one does suggests the two items tap different dimensions of belief rather than one underlying conviction. A separate caveat attaches to the moving item itself: belief in God sits near its ceiling in much of the sample, so the estimated response is drawn from the thin margin with room to move. The "believing" in this paper's title rests on the one item that moves, and the claim should be read with that scope.

The complication is re-exposure. Of the first-exposed cohort observed at Wave 3, 42% (9,743 of 23,169) is exposed *again* before that interview, so the "one wave later" coefficient mixes a delayed response to the first shock with the immediate effect of a second. Splitting it is the paper's central piece of evidence and its principal complication at once. For belief the delayed rise sits in the clean cohort: +0.026 ($p = 0.022$; randomization inference with exposure and cohorts recomputed on each of 1,000 draws, $p = 0.040$)—about 0.7 percentage points—against −0.003 among the re-exposed, and +0.033 ($p = 0.05$) in the registered stacked design. The split itself is imprecise: the clean-minus-re-exposed difference is +0.019 ($p = 0.09$), or +0.025 ($p = 0.04$) under the stacked design. (Here and below, quoted differences are interaction contrasts from joint models, not arithmetic gaps between the cohort coefficients.)

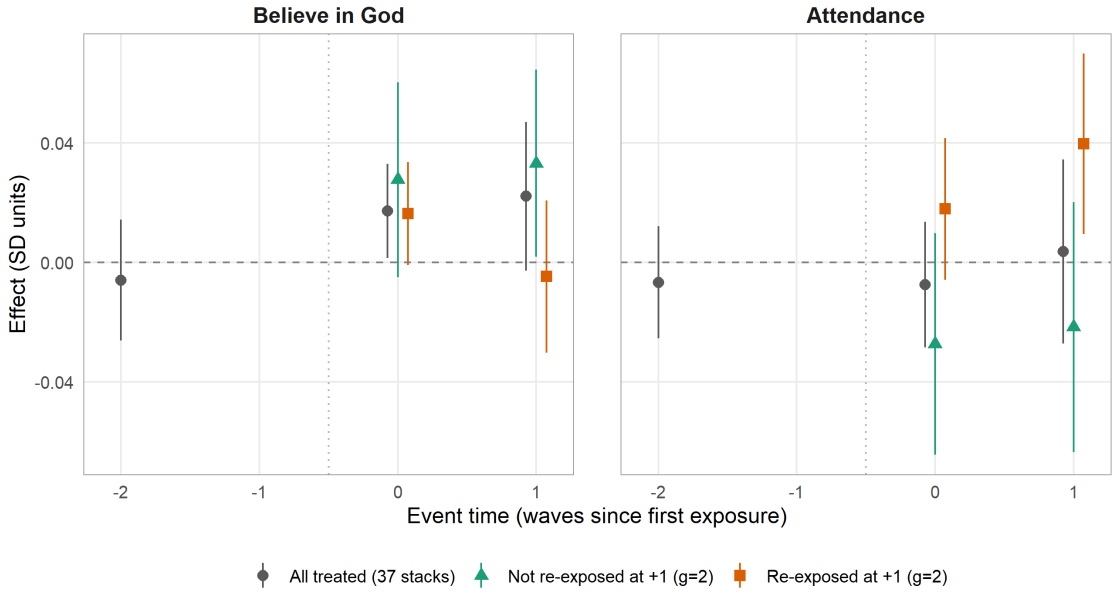
For attendance the split changes the interpretation rather than the magnitude. The pooled "recovery" averages two different paths: the clean cohort remains *depressed* one wave later (−0.024, $p = 0.08$, after −0.030 on impact), while the re-exposed show a significantly *positive* coefficient (+0.028, $p = 0.005$; difference −0.036, $p = 0.008$). What looked like recovery is partly composition. Re-exposure is not randomly assigned—chronically hazard-prone places select into it—so the split

describes what the pooled coefficient mixes rather than identifying the causal effect of a second exposure. A distributed-lag specification separating current from lagged exposure shows the same belief ordering (lagged +0.012 against contemporaneous +0.007, neither individually significant).

The registration named a stacked design around large discrete events; it could not be run at the first implementation and is executed here post hoc, so its results carry exploratory status despite the registered specification. It builds 37 event-by-country stacks from local events with at least 200 treated respondent–intervals within 100 km, using as controls same-country respondents with no local event within 250 km in any interval, and absorbs stack-by-respondent and stack-by-wave fixed effects with standard errors clustered two ways on the index event and the grid cell. Figure 3 shows the resulting event-time paths for the full sample and for the two re-exposure arms separately. It reproduces the ordering: attendance depressed at impact, belief rising afterwards and concentrated in the arm not exposed again. One difference from the event study: here, with large events only, belief is already positive at impact—consistent with bigger disasters moving salience faster—so the sharp zero-then-rise lag is the event study's result; the stacked design shows the weaker ordering of higher-later.

Table 5, in Section 6, collects the checks that constrain the belief result, and they constrain it more than any other in the paper: the magnitude is estimator-dependent and worst-case attrition bounds include zero.

Figure 3: Stacked event design: event-time paths



Note: Pooled stacked-event estimates across 37 event×country stacks (local events with at least 200 treated respondent–intervals within 100 km; controls are same-country respondents with no local event within 250 km in any interval); stack×respondent and stack×wave fixed effects; reference period is event time −1, omitted by construction; 95% confidence intervals, standard errors clustered two ways, on the stack's event and on the respondent's grid cell. Event time is waves since the cohort's first exposure, so with three waves the axis spans −2 to +1: time −2 is contributed only by the Wave-3-exposed cohort and time +1 only by the Wave-2-exposed cohort. The split series divide the Wave-2 first-exposed cohort by re-exposure before Wave 3 and draw on the 30 (not re-exposed) and 25 (re-exposed) stacks containing such cohorts; a stack can contain both arms, so the two counts overlap and exceed 37. Exploratory; implements the registration's stacked design post hoc (Online Appendix B).

The timing is what makes the pattern informative. Two stories could explain a disaster changing religiosity. In the first—*disruption*—nothing changes in what people want: services are simply hard to reach for a while, so participation drops and then returns. In the second—*demand*—the disaster changes what people want from religion. The pattern speaks to both: the early, fading attendance dip looks like disruption, and the later, persistent belief rise looks like demand. The fit is not perfect: among those never exposed again, attendance stays depressed a wave later, longer than damaged buildings and cleanup time explain, so the exit from participation is not purely practical. But a study comparing *levels* of religiosity across people or places at one point in time—which is most of the literature—could not have separated the two stories at all; only timing can.

5.3 Why separating the margins is the result

Separating the margins is not a presentational choice; collapsing them erases the finding. A conventional single-index study of these same data—the average of standardized attendance, belief, and prayer, run through the same 100-km specifications—finds nothing anywhere (Online Appendix C): −0.000 (SE 0.005) pooled over three waves, −0.010 (0.008) in the Waves 1–2 window, flat dynamics. A five-item composite and its first principal component behave the same way, the principal component tilting mildly negative one wave after exposure (−0.017, $p = 0.05$) because belief, the one rising margin, is the item it weights least. A prayer-only design—the closest analogue to the Google-search prayer measures of Bentzen (2021)—is equally flat (+0.003 pooled, −0.009 in Waves 1–2). Of the four registered outcome groups, two respond, and with opposite signs—participation down, belief in God up; private coping and connectedness are flat throughout, and within belief the life-after-death item never moves. Opposite-signed margins cancel inside any one number, which is why a study measuring religiosity as a single quantity can report a null, a rise, or a fall from the same underlying behaviour depending on which items it happens to average.

5.4 Who responds

If disasters work by reactivating existing religious attachments rather than creating new ones, both responses should concentrate among respondents who already hold religious capital (Azzi and Ehrenberg, 1975; Iannaccone, 1992). The subgroup estimates line up. The attendance decline is detected only among the religiously attached: −0.024 ($p = 0.01$) among respondents affiliated at baseline against −0.002 among the unaffiliated, and −0.021 ($p = 0.04$) among those who attended weekly at age twelve against −0.003 among those who never did. Belief shows the same gradient—+0.024 ($p = 0.02$) among the weekly-raised and +0.023 ($p = 0.01$) among the affiliated, against small and insignificant estimates for the never-religious (plotted in Online Appendix H).

The qualification is that the formal tests are weaker than the pattern. Estimated from a fully saturated model with country-by-group fixed effects, the between-group differences carry the expected signs but are imprecise (affiliation: $p = 0.07$ for attendance, $p = 0.10$ for belief; childhood attendance: $p = 0.22$ and $p = 0.35$), and across the 36 registered interaction contrasts nothing survives false-discovery control. This is a consistent pattern, not a sharply established interaction. One result runs the other way and is worth recording: among respondents who never attended as children, religious coping *declines* after exposure (−0.024, $p = 0.01$), a hint that shocks polarize rather than uniformly increase religiosity—though with twenty-one group-by-outcome cells a single contrast carries little weight. What the pattern weighs against is an insurance account cast in terms of

new adherents: among the never-religious, neither belief nor attendance detectably moves—though conversion itself, a transition into affiliation, is not the tested outcome.

5.5 Believing without belonging

Two marginal coefficients moving in opposite directions do not establish that the same people hold the joint state. Estimating the transition directly: between Waves 1 and 2, exposed believing attenders are 1.75 percentage points likelier to stop attending while continuing to believe ($p = 0.0003$), and a decomposition over all four destinations—stay, stop attending while believing, stop believing while attending, leave both—places essentially the whole movement there rather than in belief loss: the four effects sum to zero by construction, and the two belief-losing destinations are individually negligible.

Three sample qualifications travel with it, all in Table 4. The test was not registered. It is a Waves 1–2 phenomenon: the same estimate between Waves 2 and 3 is −0.002 ($p = 0.74$), and the two windows differ significantly ($p = 0.007$), so this is not a standing regularity of the panel. And under the stricter sample that drops national-event intervals it falls to +0.007 ($p = 0.08$). It survives survey weighting (+0.013, $p = 0.01$) and an ordinal definition of attendance decline (+0.011, $p = 0.04$), but weakens at a weekly attendance threshold (+0.010, $p = 0.09$). Because belief is sticky in these data, much of the transition is the attendance exit restated rather than an independent fact about belief; the decomposition is what shows the movement goes to the believing-without-attending cell rather than out of belief altogether. A final caveat is identification, not sampling: conditioning on baseline believing attenders selects a risk set, and the change-score protection against fixed differences in transition propensity does not carry over—exposure must be as good as random within that baseline state—so these estimates read as descriptive. The complete battery—all four destinations, alternative thresholds, and the cross-window test—is in Online Appendix G.

5.6 What the pattern amounts to

A household-production reading organizes it. Attendance is the time- and infrastructure-intensive margin: a local disaster raises its price through damaged buildings, displaced clergy, and cleanup demands on time, much as the repeal of blue laws raised the relative price of church-going in Gruber and Hungerman (2008). Belief is produced without market inputs and responds to the demand side—mortality salience and the search for meaning that adversity triggers. A shock that shifts demand up while raising the price of the congregational margin implies more belief and less attendance among those who already hold the stock of religious capital, which is what the exploratory evidence shows. Two pieces do not fit, and the paper has already shown both: the same demand logic implied a rise in private coping—the registered prediction—and Section 4 rules it out with the paper's tightest interval, suggesting conviction and religious practice do not respond as a single good; and the clean cohort's attendance stays depressed a wave later, longer than a purely price-based story explains (Section 5.2). The hazard split offers the demand story no support in either direction: in the all-outcomes battery the belief column's near-significant cell is cyclones (+0.010, $p = 0.06$), while the earthquake coefficient is larger but noisier (+0.012, SE 0.014), and no cross-hazard contrast is estimated—inconclusive rather than adverse (Online Appendix J).

The two halves also cohere as a package, not only as separate estimates. Across the same 1,000 within-country cell permutations used throughout—both statistics recomputed on each draw, so their dependence is preserved—43% of permuted geographies produce an opposite-signed pair on

Table 4: The exploratory results and the checks that most constrain them

Exploratory result	Estimate	Decisive check	Result
Attendance falls, Waves 1–2 within 100 km	−0.020	excluding Kenya	−0.011 ($p = 0.12$)
Belief rises one wave after first exposure	+0.019	imputation estimator; Lee bounds	+0.009 ($p = 0.43$); bounds cross zero
Coping dips in the first three months	−0.024	direction registered under H5a	*opposite* sign (two-sided $p = 0.03$)
Believing attenders stop attending, Waves 1–2	+0.018	the same test, Waves 2–3	−0.002 ($p = 0.74$)
		window equality	−0.020 ($p = 0.007$)
		strict national-event sample	+0.007 ($p = 0.080$)

Note: Each row pairs an exploratory result with the check that constrains it most, rather than with the checks it passes. The attendance decline is sign-stable but loses significance once the 2024 Kenyan floods are removed. The belief response is positive under the event-study and stacked designs but small and imprecise under imputation, and worst-case attrition bounds include zero (Table 5). A joint design-based test treats the two headline rows as a package: across the same 1,000 within-country cell permutations, 43% of draws produce an opposite-signed attendance–belief pair—two noisy nulls often do—but a pair as extreme as the one observed almost never arises (joint randomization $p = 0.003$; marginal one-sided $p = 0.015$ and 0.038). The coping dip in the first three months after exposure is listed because it runs *against* the registered prediction, which had private coping rising fastest immediately after the event; the comfort component behaves the same way (−0.024, $p = 0.03$). The joint-state transition is a Waves 1–2 phenomenon: it does not appear between Waves 2 and 3, and the two windows differ significantly. The transition row is in probability units (a 1.75-percentage-point increase); the other rows are in standard-deviation units. None of these results carries a confirmatory claim: the first, second, and fourth were not registered, and the third is the registered H5a specification read against its registered direction.

the two margins; two noisy nulls often do. But the observed pair is jointly extreme: counting draws in which the attendance estimate is at least as negative *and* the belief estimate at least as positive as observed gives a joint design-based p of 0.003, against marginal one-sided values of 0.015 for the Waves 1–2 attendance decline and 0.038 for the clean-cohort belief rise. The joint test is exploratory, like everything in this section, and it inherits the permutation scheme's exchangeability premise (Section 6), but it answers a specific objection: the recomposition is not two unrelated accidents of specification.

The summary is deliberately narrow. Averaged over everyone within reach of a disaster, religiosity does not move, and for private coping that null is precise. Beneath it, participation and conviction appear to move on different schedules and in opposite directions, concentrated among the already-religious and, for participation, in one flood catastrophe. That is a hypothesis with unusually detailed support, not a finding the registration licenses.

6 Credibility

This section is organized by threat rather than as a catalog of robustness checks. Each subsection names an objection that would overturn a claim made above, reports the evidence that bears on it, and states what survives. Four of the checks return adverse results, and they narrow the claims they touch.

6.1 Does coarse location manufacture the null?

The location measure is rounded to whole degrees, so an indicator built from it misclassifies some respondents. If that noise were driving the coping null, the estimate would grow as the measure sharpened. It does not: the coping index is flat from the affected region down to the interior of the footprint (Figure 2), and the confidence interval excludes responses beyond ±0.02 standard deviations at every registered distance threshold, though not in the sparser outer rings. Two further checks bound the problem directly. A simulation re-draws each location with a random error of up to roughly 50 km—the size the rounding could be off—and re-estimates everything on each of 50 such datasets. The results barely depend on where the error lands: the pooled attendance estimate reaches conventional significance in only 6% of draws, and the clean-cohort belief result stays significant in 98%. Under that perturbation model, rounding-scale positional error is unlikely to have generated the belief finding, and does not rescue the attendance one. The caveat is that narrower bands change *who* is treated as well as how well treatment is measured, so the flat profile weighs against a large uniform response rather than mechanically excluding attenuation. Online Appendix D reports the mover and location-timing checks the third wave makes possible; the registered-radius intervals are listed in Section 4, and the within-window timing profile is in Online Appendix E.

6.2 Is the inference too optimistic?

Two objections apply. The first is that clustering on the respondent's grid cell understates dependence when exposure is spatially correlated: one disaster reaches many cells at once, so if outcomes co-move across cells, the clustered standard errors are too small. Randomization inference probes it without relying on an asymptotic approximation, by permuting grid-cell locations within country, rebuilding exposure from each respondent's own interview dates, and re-estimating the full model

on 1,000 draws. The design-based p-values are 0.26, 0.49, and 0.48 for attendance, coping, and belief—the last under the pooled convention, as throughout—agreeing with the clustered ones. The permutation null treats cell labels as exchangeable within country, which residence and hazard geography do not guarantee, so the exercise is a stringent geographic stress test rather than exact design-based inference; on that reading, the null is not a clustering artifact. The Waves 1–2 estimate gets the parallel treatment: clustering simultaneously on events and grid cells leaves it intact ($p = 0.03$). The second objection is multiplicity. It is met by not claiming: within the registered primary family the smallest Benjamini–Hochberg q, computed on the registered-sidedness p-values, is 0.41; across the 36 registered interaction contrasts the smallest is 0.12; and every result outside the registered set is labelled exploratory and carries no confirmatory weight. Online Appendix C details the permutation scheme and reports it under both standardization conventions; Online Appendix H gives the complete moderator battery.

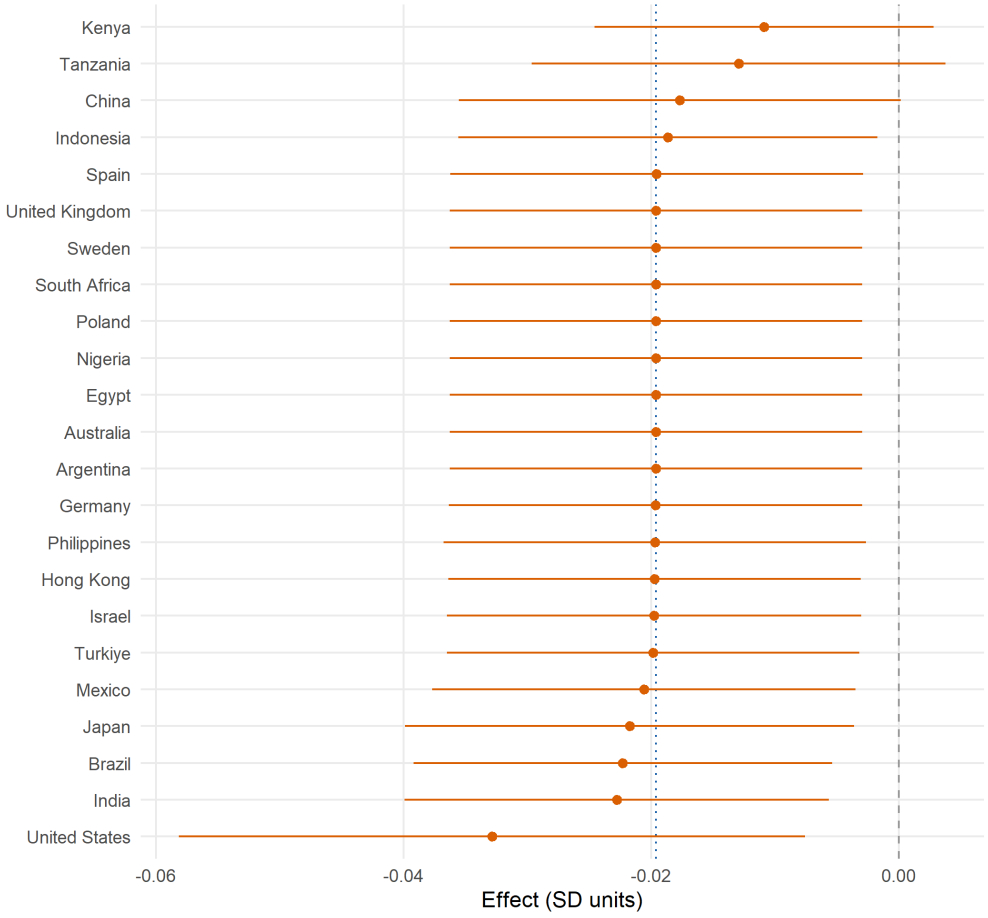
6.3 Is the attendance decline one country?

Largely, yes, and the paper says so. The Waves 1–2 attendance decline is −0.020 standard deviations; the April–May 2024 Kenyan floods account for close to half of that magnitude, and the attribution is to the event, not merely to the country: every locally treated Kenyan respondent in that window (1,622) is treated by that one sub-national flood, which reached 4 of Kenya's 41 matched regions. Excluding Kenya leaves −0.011 (SE 0.007, $p = 0.12$)—the same sign, no longer distinguishable from zero (Figure 4). Tanzania, where the late-2023 arm of the same East African flood system is locally coded, is the second most influential exclusion (−0.013, SE 0.009, $p = 0.13$); no other single country moves the estimate past $p = 0.10$. The stricter reading that drops national-event intervals likewise halves the estimate, mechanically so here: the separate, whole-country Kenyan floods of late 2023 overlap 99.6% of Kenyan intervals in the window, so that rule removes essentially all Kenyan respondents, treated and control alike. Online Appendix F gives the per-country estimates with a random-effects summary—an average that lets the true effect differ by country. Leave-one-country-out holds the sign throughout, so this is concentration rather than a single influential outlier. The claim the evidence supports is a real but concentrated disruption, not a universal law.

6.4 How settled is the delayed belief result?

Less than any other exploratory result, and Table 5 is arranged so a reader can see why. The response is positive in every specification that retains the cohort structure: +0.019 ($p = 0.02$) in the Sun–Abraham event study; +0.022 ($p = 0.09$) in the registered stacked design overall, rising to +0.033 ($p = 0.05$) among those never exposed again and near zero among the re-exposed; and randomization inference on the clean cohort returns $p = 0.04$. The practical concern is that the magnitude is not stable across conventional estimators: imputation returns +0.009 ($p = 0.43$) and equal-weighted stacking +0.001 ($p = 0.97$), so the result depends on how event-time cells are weighted, and neither of those is a worst-case construction. Attrition adds the formal bound. Reweighting Wave-3 completers to the Wave-2 population leaves +0.021 ($p = 0.06$), but Lee bounds—which do assume the worst about who leaves the panel—run from −0.039 to +0.074 and therefore include zero. Finally, only one pre-treatment lead is available; it shows no detectable pre-movement ($p = 0.60$), but a single available lead is a weak test and cannot exclude gradual anticipation. The defensible statement is suggestive rather than established: in the specifications that keep the cohort structure, belief moves later than participation and in the opposite direction; its magnitude—and, under the estimators that

Figure 4: Leave-one-country-out influence: attendance, Waves 1–2, 100 km exposure



Note: Each point re-estimates the exploratory Waves 1–2 100-km attendance specification excluding one country; the dotted vertical line is the full-sample estimate (−0.020); 95% confidence intervals, grid-cell-clustered standard errors. All estimates remain negative; excluding Kenya moves the estimate to −0.011 ($p = 0.12$), and excluding Tanzania to −0.013 ($p = 0.13$).

reweight event-time cells, the movement itself—is not pinned down. Online Appendix E reports the event-study path and the within-window profile behind them.

Table 5: The delayed belief response against the checks that could overturn it

Check	Estimate	p	What it tests
Sun–Abraham event study	+0.019 (0.008)	0.019	baseline
never re-exposed	+0.026 (0.011)	0.022	delayed response, clean reading
Registered stacked design	+0.022 (0.013)	0.090	the registered dynamic design
never re-exposed	+0.033 (0.016)	0.047	
re-exposed	−0.005 (0.013)	0.723	re-exposure selection
equal-weighted stacks	+0.001 (0.035)	0.969	weighting across stacks
Imputation estimator	+0.009 (0.011)	0.430	estimator dependence
Attrition-reweighted	+0.021 (0.011)	0.062	differential Wave-3 retention
Lee bounds (clean cohort)	[−0.039, +0.074]	—	worst-case attrition
Randomization inference (clean)	—	0.040	design-based inference
Single pre-treatment lead	—	0.604	pre-movement

Note: Every row estimates the change in belief in God one wave after first proximate exposure, except where stated. Standard errors in parentheses; p two-sided. The response is positive in every specification that retains the cohort structure, but its magnitude and precision are *not* stable: the imputation estimator and equal-weighted stacking both return small, imprecise effects, and the Lee bounds—which assume the worst about who leaves the panel—cross zero. The re-exposed cohort shows nothing, which is the pattern a delayed consolidation predicts, but the formal clean-versus-re-exposed difference is itself imprecise ($p = 0.09$). All rows are exploratory: the registration made their confirmatory status conditional on the pooled primary test rejecting, and it did not. The single pre-treatment lead is a weak check and cannot rule out gradual pre-movement.

6.5 Does the design itself hold?

The checks within-person identification makes testable all pass, though checks of this kind constrain the identifying assumption rather than establish it: none tests untreated outcome trends directly. Disaster exposure does not predict panel attrition: retention is 62.9% among the exposed against 63.0% among the unexposed, and exposure predicts third-wave retention only marginally, with reweighting changing nothing. Treated and untreated respondents are balanced on baseline religiosity and demographics, every gap below 0.03 standard deviations. Locally concluded disasters from the year *before* the first interview do not significantly predict subsequent within-person changes on the responsive margins, though the attendance cell is imprecise (−0.019, SE 0.012)—and this is a carryover check on real prior events, not a parallel-trends test, so it weighs against location-linked persistence rather than ruling it out. Requiring a balanced three-wave panel leaves attendance at −0.003 ($p = 0.71$), and the estimates survive alternative radii, severity-graded exposure, and region-clustered standard errors.

A null on every religious margin invites the objection that the treatment is simply too weak to register—that a footprint reaching a respondent's location is not an event in their life. Non-religious outcomes weigh against it. Proximate exposure slightly raises self-rated mental health and national belonging, and attendance falls while participation in non-religious groups does not (+0.005, SE 0.006), so the exposure measure is picking up something respondents experienced. Estimating the two participation responses jointly, the religious-minus-secular difference is −0.012 (SE 0.007, $p = 0.10$)—correctly signed for a congregation-specific disruption, though itself imprecise, so the

specificity claim is suggestive rather than established. The treatment registers; the religious response to it is small. Online Appendix I reports these outcomes in full.

6.6 What the design cannot answer

Two boundaries matter for interpretation. First, exposure is proximity, not loss: the treatment records that a qualifying footprint reached the respondent's location, not that the respondent was injured, bereaved, or displaced. Most people within 100 km of a footprint are bystanders, so these are intention-to-treat estimates over a population in which the directly harmed are a minority, and nothing here licenses a claim about disaster *victims*. Second, whole-country catastrophes mostly erase the within-country comparison: ten of the eighteen national events reach every matched region in their country, leaving no unaffected compatriot to compare against. The contrast that can still be estimated flips sign across exposure codings, so national events enter the models only as controls, with no effect claimed for them. The estimand is the effect of a nearby disaster on an ordinary resident of an affected area, relative to a compatriot whose area was not reached.

7 Conclusion

Whether adversity strengthens religion may be the wrong question, twice over. The preregistered answer is that nearby disasters do not detectably change religiosity on average: the pooled confirmatory tests are null on every margin, and precisely so for private coping, where the interval excludes average responses beyond ±0.02 standard deviations at every registered distance threshold and inside the footprint itself. That null is not an artifact of coarse geographic coding. It survives the test its rival explanation proposed—sharpening exposure from the region to the footprint leaves the estimate flat, where the dilution story required it to grow.

The question's premise is what fails. Religiosity is not one quantity moving one way, and the exploratory anatomy beneath the null suggests it recomposes rather than rises. In a panel that follows the same individuals in 22 countries and Hong Kong through 145 major disasters, local events appear to push people temporarily *out* of religious institutions while pulling them *toward* religious conviction, both concentrated among those with prior religious attachment. A single-margin study of these same data could mislead in either direction—a belief-in-God item alone would show religion strengthening, attendance alone would show it weakening—and a composite of them shows nothing at all.

That composition speaks to how religion functions as insurance (Chen, 2010; Ager and Ciccone, 2018; Auriol et al., 2020). The margin through which religious communities deliver material and social support is participation, and participation is what a local disaster disrupts, at the moment this support would be most valuable. The margin that survives is belief, whose production needs no infrastructure. Because both responses concentrate among those who already hold religious capital, shocks appear to deepen existing attachments rather than create new ones—reactivation rather than conversion—though the subgroup contrasts behind that reading are individually imprecise.

Three limits bound all of this. The estimand is proximity, not personal loss, so nothing here describes what disasters do to the people they harm directly. The participation response is concentrated in one East African flood system. And every part of the recomposition story is exploratory—the registered tests are null, and the paper's discipline is to say so rather than promote what the data merely suggest. What would settle the story is exposure graded by within-footprint intensity, which

can distinguish severity where distance cannot, and a horizon long enough to see whether the belief response is a lasting shift in the stock of believers or the echo of a single shock.

References

Ager, Philipp and Antonio Ciccone, "Agricultural Risk and the Spread of Religious Communities," *Journal of the European Economic Association*, 2018, *16* (4), 1021–1068.

Aslam, Maqsood, Laurent Weill, and Hira Iqbal, "Do Mountains Move Faith?," *Economic Modelling*, 2022, *117*, 106077.

Auriol, Emmanuelle, Julie Lassébie, Amma Panin, Eva Raiber, and Paul Seabright, "God Insures those Who Pay? Formal Insurance and Religious Offerings in Ghana," *The Quarterly Journal of Economics*, 2020, *135* (4), 1799–1848.

Azzi, Corry and Ronald Ehrenberg, "Household Allocation of Time and Church Attendance," *Journal of Political Economy*, 1975, *83* (1), 27–56.

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

Bentzen, Jeanet Sinding, "Acts of God? Religiosity and Natural Disasters across Subnational World Districts," *The Economic Journal*, 2019, *129* (622), 2295–2321.

_ , "In Crisis, We Pray: Religiosity and the COVID-19 Pandemic," *Journal of Economic Behavior and Organization*, 2021, *192*, 541–583.

_ **, Alessandro Pizzigolotto, and Lena Lindbjerg Sperling**, "Divine Policy: The Impact of Religion in Government," *American Economic Journal: Applied Economics*, 2026, *18* (1), 195–247.

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

Butinda, Lewis Dunia, Aimable Amani Lameke, Nathan Nunn, Max Posch, and Raúl Sánchez de la Sierra, "On the Importance of African Traditional Religion for Economic Behavior," Working Paper 31430, National Bureau of Economic Research, Cambridge, MA July 2023. Working paper; no journal version as of August 2026.

Chen, Daniel L., "Club Goods and Group Identity: Evidence from Islamic Resurgence during the Indonesian Financial Crisis," *Journal of Political Economy*, 2010, *118* (2), 300–354.

Davie, Grace, *Religion in Britain since 1945: Believing without Belonging*, Oxford: Blackwell, 1994.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Nicholas Ryan, "Praying for Rain," *The Quarterly Journal of Economics*, 2026, *141* (3), 2363–2422.

Gruber, Jonathan and Daniel M. Hungerman, "The Church versus the Mall: What Happens When Religion Faces Increased Secular Competition?," *The Quarterly Journal of Economics*, 2008, *123* (2), 831–862.

Hasan, Iftekhar, Stefano Manfredonia, and Felix Noth, "Church Membership and Economic Recovery: Evidence from the 2005 Hurricane Season," *The Economic Journal*, 2024, *134* (664), 3306–3332.

Iannaccone, Laurence R., "Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives," *Journal of Political Economy*, 1992, *100* (2), 271–291.

Norenzayan, Ara and Azim F. Shariff, "The Origin and Evolution of Religious Prosociality," *Science*, 2008, *322* (5898), 58–62.

Padgett, R. Noah, Richard G. Cowden, Manas Chattopadhyay, Ying Han, John Honohan, Zacc Ritter, Rajesh Srinivasan, Byron R. Johnson, and Tyler J. VanderWeele, "Survey Sampling Design in Wave 1 of the Global Flourishing Study," *European Journal of Epidemiology*, 2025, *40* (4), 391–406.

Pargament, Kenneth I., Harold G. Koenig, and Lisa M. Perez, "The Many Methods of Religious Coping: Development and Initial Validation of the RCOPE," *Journal of Clinical Psychology*, 2000, *56* (4), 519–543.

Sibley, Chris G. and Joseph Bulbulia, "Faith after an Earthquake: A Longitudinal Study of Religion and Perceived Health before and after the 2011 Christchurch New Zealand Earthquake," *PLoS ONE*, 2012, *7* (12), e49648.

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

Vail, Kenneth E., Zachary K. Rothschild, Dave R. Weise, Sheldon Solomon, Tom Pyszczynski, and Jeff Greenberg, "A Terror Management Analysis of the Psychological Functions of Religion," *Personality and Social Psychology Review*, 2010, *14* (1), 84–94.

VanderWeele, Tyler J., Byron R. Johnson, Piotr T. Bialowolski, Rebecca Bonhag, Matt Bradshaw, Thomas Breedlove, Brendan Case, Ying Chen, Zhuo Job Chen, Victor Counted et al., "The Global Flourishing Study: Study Profile and Initial Results on Flourishing," *Nature Mental Health*, 2025, *3* (6), 636–653. Paper has 47 authors; first ten listed.

Online Supplementary Appendix to "Believing without Belonging? Natural Disasters and the Margins of Religiosity"

Michal Brzezinski*

August 29, 2026

This appendix reports the material supporting the paper. Sections, tables and figures numbered with a leading letter are in this appendix; anything referred to as "the paper's" is in the main document. No output here reports coordinates or any statistic computed on fewer than ten respondents in a geographic cell.

Contents

A Data construction and the restricted-geography protocol **2**

B Registration, amendment, and implementation **2**
B.1 A registered test of the insurance mechanism 4
B.2 Deviations from the preregistration 5

C The primary specification and additional inference **6**

D Distance, treatment measurement, and location error **7**

E Dynamic estimators, re-exposure, and attrition **8**

F Country and event influence **8**

G The joint-state transition **9**

H Moderators and religious capital **11**

I Mechanisms and secular outcomes **13**

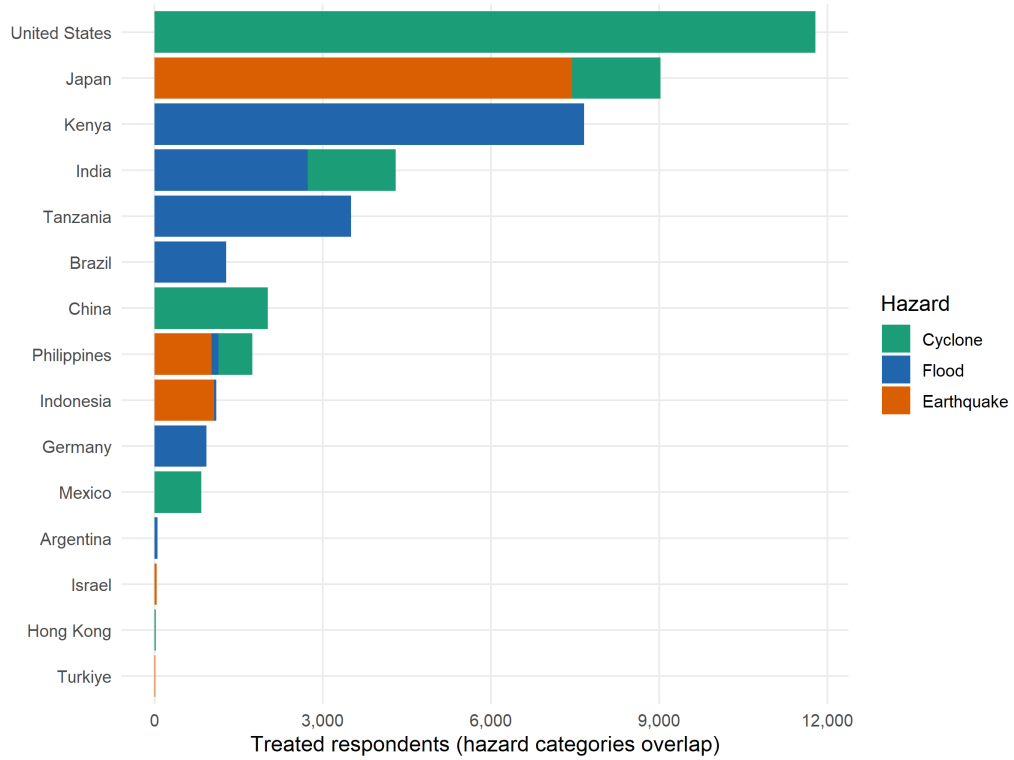
J Hazards, radii, and remaining diagnostics **14**

*University of Warsaw. Email: mbrzezinski@wne.uw.edu.pl.

A Data construction and the restricted-geography protocol

This section records what the main text's data section compresses. Exposure is built from GDACS Orange- and Red-alert events between January 2022 and January 2026, intersected with administrative boundaries and linked to GFS public regions; the restricted analyses replace that regional assignment with the geodesic distance from each respondent's whole-degree cell to the event's affected-area polygon. Figure A1 shows how treated respondents distribute across countries and hazards, which is the composition every pooled estimate averages over. Throughout, no output reports coordinates and no statistic is computed on fewer than ten respondents in a geographic cell; the sensitive files are held in access-logged storage outside the project tree, and analysis files inside it carry only binned distances.

Figure A1: Treated respondents by country and hazard



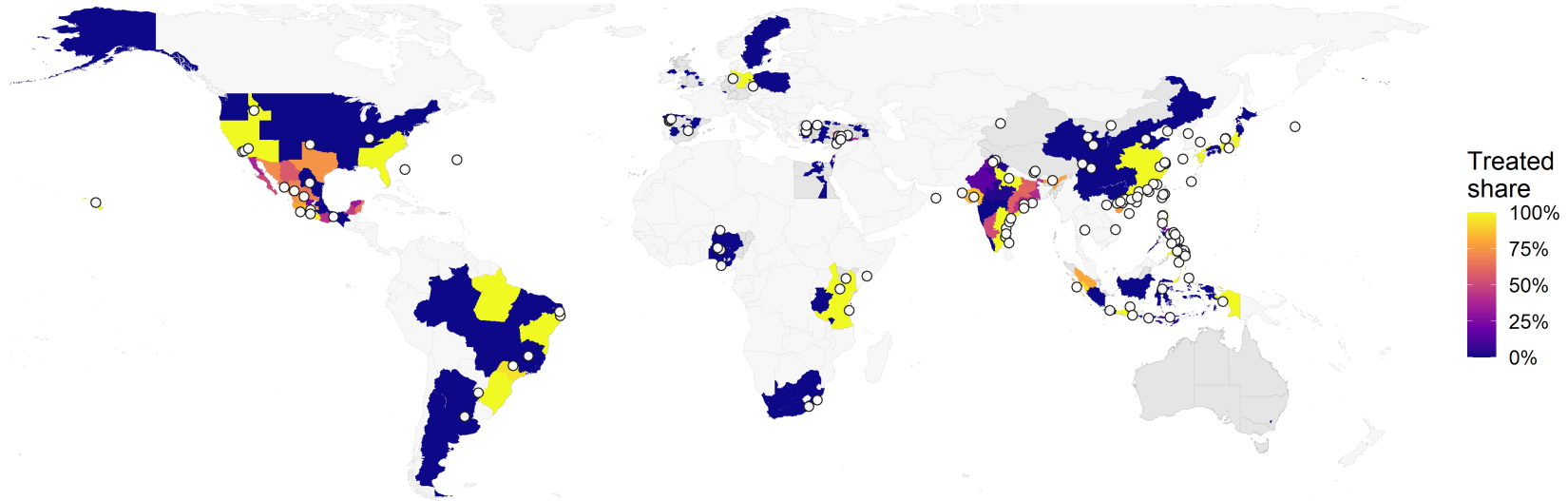
Note: Second-wave completers in a region reached by a GDACS Orange- or Red-alert event during their Waves 1–2 window. Hazard categories overlap.

The paper's exposure map is reproduced here at full-page size, rotated, because at the width of a text page the administrative detail and the individual event locations are difficult to resolve. Nothing in it differs from the version in the paper except scale.

B Registration, amendment, and implementation

This section collects the full registration record: the amendment registered and tested after the exploratory country splits, and the complete list of implementation choices that deviate from or

Figure A2: Disaster exposure across the study countries (full-page reproduction)



Note: As the exposure map in the paper's data section, enlarged. Administrative regions of the 22 study countries and Hong Kong are shaded by the share of respondents treated; the remaining land is outline only and carries no data, as do the gray regions, which lie inside a study country but could not be matched by name to a survey region. White points mark the centroids of all 145 GDACS Orange- and Red-alert events, one marker per event regardless of hazard type. Treated and untreated regions coexist within most countries, which is the variation the within-country comparison uses.

extend the registered text.

B.1 A registered test of the insurance mechanism

The country-level inequality patterns collected in Appendix H came from more than thirty country-level contrasts run beyond the original registration, so I registered them properly before testing them again. An amendment to the preregistration, submitted on 20 July 2026—returned to draft on 24 July for a clarification of its justification field alone, with no substantive section reopened, and accepted by the GFS Registry on 30 July—fixed two hypotheses, their exact measures, the models, a two-test multiplicity correction, and a decision rule in advance. H-A1 predicts that the attendance disruption is larger below the median of a respondent's own country's baseline equivalized household-income distribution; H-A2 predicts that this gradient is concentrated in less-unequal countries. The registered measure upgrades the bracket rank used above by converting each bracket to the midpoint of its country-specific range and dividing by the square root of household size. It is available for 90.5% of respondents, with South Africa worst at 20.4% missing.

Neither hypothesis is supported. The registered individual-level interaction is +0.014 (SE 0.011; 95% CI $[-0.009, +0.036]$; $q = 0.23$), which carries the predicted sign but does not approach the registered threshold; because the interval excludes moderation larger than 0.05 standard deviations in either direction, this is a precise null rather than an uninformative one. The cross-level triple interaction is −0.034 (SE 0.027; 95% CI $[-0.087, +0.019]$; $q = 0.23$), also correctly signed and also short of the threshold, but its interval is too wide to rule out moderation of the size at issue, so it is inconclusive—the outcome the amendment's power note anticipated for a median split of 22 countries and Hong Kong. One qualification attaches to both: because H-A1 and H-A2 were selected after inspecting those thirty-plus contrasts and are evaluated on the same panel, the amendment fixes the specifications, sidedness, and decision rule in advance but cannot restore full prospectivity—its two-test correction governs the amendment's own family, not the earlier search—so these are specification-fixed follow-up tests rather than selection-adjusted confirmation. The correct summary is therefore asymmetric: H-A1 is a precise null in the interval sense, excluding individual-level moderation beyond 0.05 standard deviations, H-A2 is uninformative, and the insurance mechanism accordingly remains open rather than closed by these tests.

The descriptive pattern nonetheless survives the better measure and sharpens: the exposure effect among below-median respondents in less-unequal countries is −0.054 ($p = 0.003$), against −0.002, −0.015, and +0.002 in the other three cells. That single cell is not evidence of moderation—the interaction is the test of moderation, and it fails twice—and the registered robustness set shows why restraint is warranted. Excluding Kenya and Japan, the two countries flagged as influential, collapses the individual-level term to +0.002 ($p = 0.86$) and reverses the triple interaction to +0.015 ($p = 0.69$). Leave-one-country-out holds both signs, so it is the pair that matters, not either country alone. The unequivalized variant (+0.010) and a continuous rank in place of the median split (+0.015) agree with the registered estimate and with each other. I report the failure as registered: it does not license substituting the correlated GDP-per-capita or state-capacity splits, which remain descriptive, not confirmatory.

B.2 Deviations from the preregistration

Eight implementation choices deviate from or extend the registered text; each is disclosed where it matters in the body and collected here. The fourth is closed as of this version: the registered stacked design is now implemented and reported alongside the original estimator.

1. **Whole-country shocks.** The registration drops whole-country *shocks* from the within-country estimates. The exposure construction already does this—a nationally flagged event never generates a treated observation—and the affected respondent–intervals are retained with a control whose coefficient is never interpreted, which is what makes the registered local-versus-national contrast estimable at all. A stricter reading that also drops those respondent–intervals from the sample is reported alongside: it halves the Waves 1–2 attendance estimate (-0.010, $p = 0.16$), because the whole-country late-2023 Kenyan floods overlap nearly every Kenyan interval in that window—removing those respondents removes, with them, the sub-national April–May 2024 flood exposure that supplies Kenya's local treatment—and it moves the pooled estimate to +0.000 (the paper's specification table, column 2).

2. **Coping index.** The registered index averages comfort, prayer, and feeling loved by God (unit variance); it is the confirmatory object throughout. The earlier two-component index (comfort, prayer) is retained only for comparability with the Waves 1–2 results that predate the registration.

3. **Standardization.** *This deviation is now closed for the primary family.* The registration specified standardization within country–wave. Earlier versions of this paper standardized within country using moments pooled across all observed waves—so that within-person changes stayed comparable over time, and because rescaling within each country–wave lets a later wave rescale identical within-person changes—and reported that pooled rule as the headline. The registered country–wave rule is now the headline for the registered primary family, and the two conventions are reported side by side in the paper's primary table. They differ by at most 0.0012 SD, flip no sign, and move the design-based p-values by less than 0.02. What remains is a genuine limitation of the registered rule rather than a choice: it is undefined where an item has zero variance in a country–wave cell. That occurs once—belief in God is invariant in Egypt at Waves 2 and 3—so 9,251 rows leave the columns that use it, and belief, a secondary outcome, continues to be quoted under the pooled convention wherever it appears. Verification checks reproduce the original Waves 1–2 estimates exactly under the two-wave version of the pooled rule—a continuity check with the earlier public-data analysis rather than a replication of the registered rule.

4. **Event-study estimator.** The registration described stacked designs around large discrete events with Callaway and Sant'Anna (2021) as robustness; the implementation first used the interaction-weighted first-exposure-cohort estimator of Sun and Abraham (2021) with a never-exposed control group and Borusyak et al. (2024) imputation as robustness (software constraints precluded the Callaway–Sant'Anna implementation). The registered stacked design is now implemented as well—37 event×country stacks with 250-km clean controls, reported in the paper's event-study section and the paper's stacked event-study figure—

closing this deviation post hoc: registered in design, post hoc in construction. The Sun–Abraham estimator remains the expositional lead for continuity with the two-wave analysis.

5. **Distance bands.** Bands below 50 km were not constructed: the 0-decimal coordinate rounding makes them unmeasurable. The inside-footprint split reported in the text is exploratory.

6. **Revised restricted-geolocation file.** The Wave-3 release supplied all three waves of restricted geolocation in a single file superseding the earlier one. Rebuilding exposure from it reproduced every existing value for every existing respondent–interval with one exception: one Spanish respondent's Wave-2 coordinates were withdrawn between releases, which removes their single interval and, with it, their baseline row. No treated count changes. The amendment's disclosed exploratory estimates were recomputed against the rebuilt panel and all reproduce at the precision disclosed, except two that shift in the fourth decimal—the country-inequality split excluding Kenya and Japan, from +0.0403 to +0.0404, and the motivating triple interaction, from −0.0265 to −0.0266. The registration was not amended for these, since the figures are labeled exploratory there and the shifts change no sign, verdict, or claim.

7. **Multiplicity in the amendment family.** The amendment specifies a sharpened Benjamini–Hochberg correction within its two-test family; I apply the standard procedure used elsewhere in the paper, which is weakly more conservative. Both q-values are near 0.23, far from the threshold, so the choice does not matter here.

8. **Zero-income respondents.** The amendment's validity window for the income bracket code excludes the explicit zero-income categories, so 8,612 respondents reporting no household income fall outside the moderator. They are substantively zero rather than missing; I follow the registered rule as written and record the exclusion here.

C The primary specification and additional inference

The paper's primary table reports the registered specification beside three alternatives that each differ from their neighbour in one element. Two details sit behind it. First, the registered country-by-wave standardization is undefined wherever an item has zero variance in a country-wave cell; this occurs once, for belief in God in Egypt at Waves 2 and 3, and removes 9,251 rows from the columns that use the registered yardstick. Belief is therefore quoted under the pooled convention throughout the paper, and the two conventions are reported side by side so the difference is visible rather than asserted.

Second, the design-based inference reported in that table permutes each respondent's grid-cell label within their country—reshuffling the country's observed cells among its respondents one-for-one—rebuilds 100-km exposure from the respondent's own interview dates at the permuted location, and re-estimates the full model on each of 1,000 draws. The permutation is run separately under both standardization conventions on the same draw sequence, so the two sets of p-values are directly comparable; they differ by less than 0.02 for the primary family, and by 0.08 for belief, where the registered convention drops the Egypt zero-variance rows and the two columns are therefore not estimated on the same sample.

Table A1 reports the single-index counterfactual quoted in the main text: the same data collapsed into conventional composites and run through the paper's three lenses. Nothing in it approaches significance except the principal component's mild negative tilt one wave after exposure, which arises because that combination gives the smallest weight to belief, the one rising margin.

Table A1: The single-index counterfactual: composites of the same data

	Pooled, three waves	Waves 1–2	One wave after exposure
Three-item average (attendance, belief, prayer)	−0.000 (0.005)	−0.010 (0.008)	−0.006 (0.008)
Five-item average	−0.001 (0.005)	−0.013 (0.008)	−0.011 (0.008)
First principal component (five items)	−0.001 (0.005)	−0.014 (0.009)	−0.017 (0.009)

Note: Each composite is built exactly as the paper's outcomes are: the mean of the listed within-country z-scores, rescaled to unit variance (the principal component is sign-oriented so belief enters positively). Columns re-run the paper's three lenses on each composite: the registered pooled specification, the Waves 1–2 window, and the event-study coefficient one wave after first exposure; standard errors clustered on the respondent's grid cell in parentheses. The prayer-only benchmark quoted in the text is the prayer row of the paper's main outcome tables. Exploratory throughout.

D Distance, treatment measurement, and location error

The Wave-3 coordinates make two checks possible that were not before. The first addresses a specific threat to the re-exposure split of the paper's event-study section: a respondent who moved away between Waves 2 and 3 is mechanically less likely to be counted as re-exposed, so the "clean" arm could be enriched with movers and the belief result could reflect relocation rather than the absence of a second shock. The arms do differ—3.05% of clean respondents changed grid cell against 2.37% of the re-exposed ($p = 0.002$)—but the enrichment is far too small to carry a coefficient. Dropping the 1,953 flagged movers from the stacked sample moves the clean-cohort belief estimate one wave after first exposure by 0.0002, from +0.033 to +0.033, and leaves the re-exposed arm near zero. Both checks in this section are exploratory.

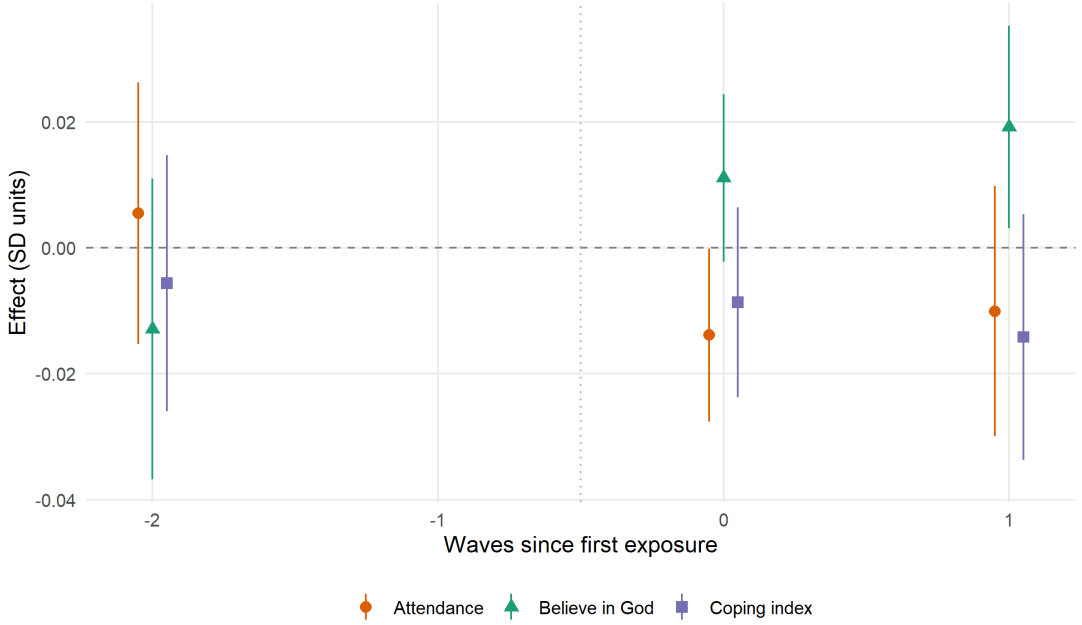
The second check reverses the registered timing convention. Exposure is assigned at each interval's start so that moving after a disaster cannot reassign it; with coordinates at both ends of every interval, I can ask what the mirror-image choice would give. The two codings agree for 99.3% of second-wave and 99.5% of third-wave intervals, and the differences run in both directions—585 second-wave intervals switch from unexposed to exposed, 387 the other way. Recoding only the third wave leaves every estimate within 0.0003 of its registered value; recoding all intervals leaves attendance at −0.007, belief at +0.004, and coping at −0.001. The timing convention matters for the interpretation of the design but not, in these data, for the estimates.

The bridge from the public regional indicator to the 100-km footprint indicator—the attenuation test on the matched Waves 1–2 respondents—is the exposure-measure table in Section 5 of the main text.

E Dynamic estimators, re-exposure, and attrition

The main text reports the stacked design and the checks that constrain the delayed belief result. Two further views of the same dynamics are collected here. Figure A3 gives the Sun–Abraham event study around first proximate exposure, with never-exposed respondents as controls and the wave before first exposure as the reference period; it is the expositional lead in the two-wave analysis and is reported here for continuity with it. Figure A4 shows the within-window profile that the event study generalizes—the response by time elapsed since the most recent local disaster—where the early coping dip that contradicts the registered H5a prediction is visible directly.

Figure A3: Event study around first proximate exposure

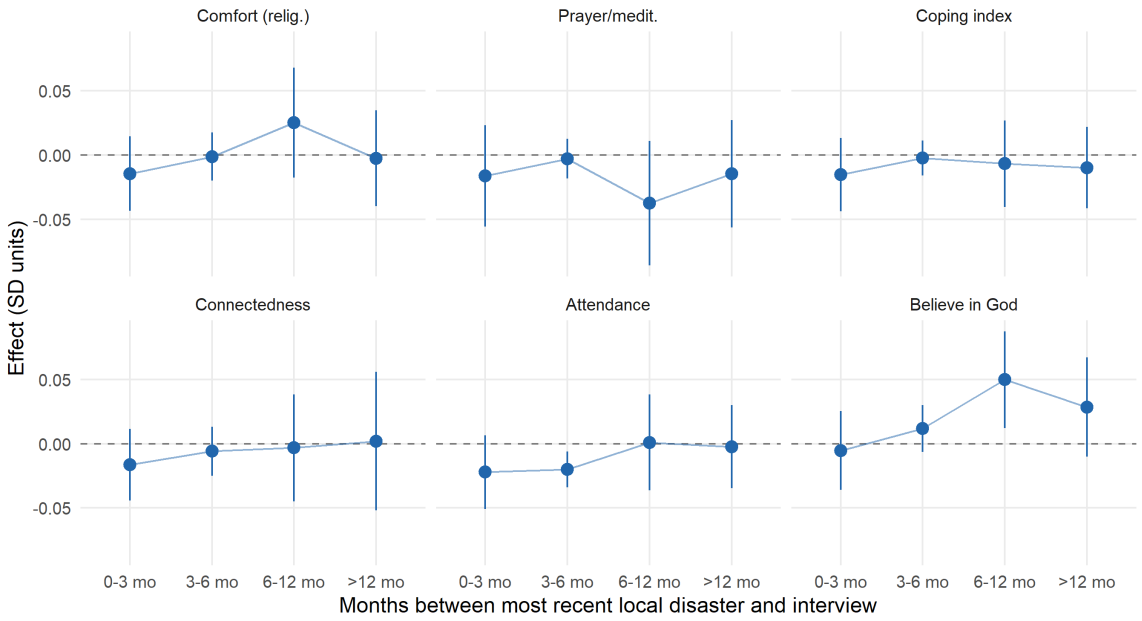


Note: Sun–Abraham interaction-weighted estimates by wave relative to the first interval with a local disaster footprint within 100 km; never-exposed respondents are the control group and the wave before first exposure the reference period; 95% confidence intervals, grid-cell-clustered standard errors. The single available lead—a weak, limited-power check—shows no detectable differential pre-movement (joint p-values 0.29–0.86 across outcomes).

F Country and event influence

Figure A5 re-estimates the Waves 1–2 100-km specification country by country (twelve countries have at least ten treated respondents in the window) and summarizes with DerSimonian–Laird random effects—an average that lets the true effect differ by country. For attendance the random-effects mean is −0.019 (SE 0.013), matching the pooled estimate, with real heterogeneity ($I^2 =$ 66%: two-thirds of the cross-country dispersion is genuine rather than sampling noise; Q-test $p = 0.001$) concentrated exactly where the main text says it is: the East African floods (Kenya −0.114, Tanzania −0.085) against near-zero estimates in the United States and Japan. For belief

Figure A4: Religiosity response by time since the most recent local disaster



Note: Mutually exclusive bands defined by the months between the most recent local event and the Wave 2 interview; the omitted category is respondents with no local exposure. 95% confidence intervals, region-clustered standard errors. Waves 1–2; exploratory, predates the registration.

the Waves 1–2 forest is flat (random-effects −0.002, SE 0.006, $I^2 = 33\%$)—as it should be, since the belief pattern is a delayed dynamic (the paper's event-study section), not a Waves 1–2 level shift. The pooled estimates are averages over visibly heterogeneous settings; no single-country reading survives this plot, and none is claimed.

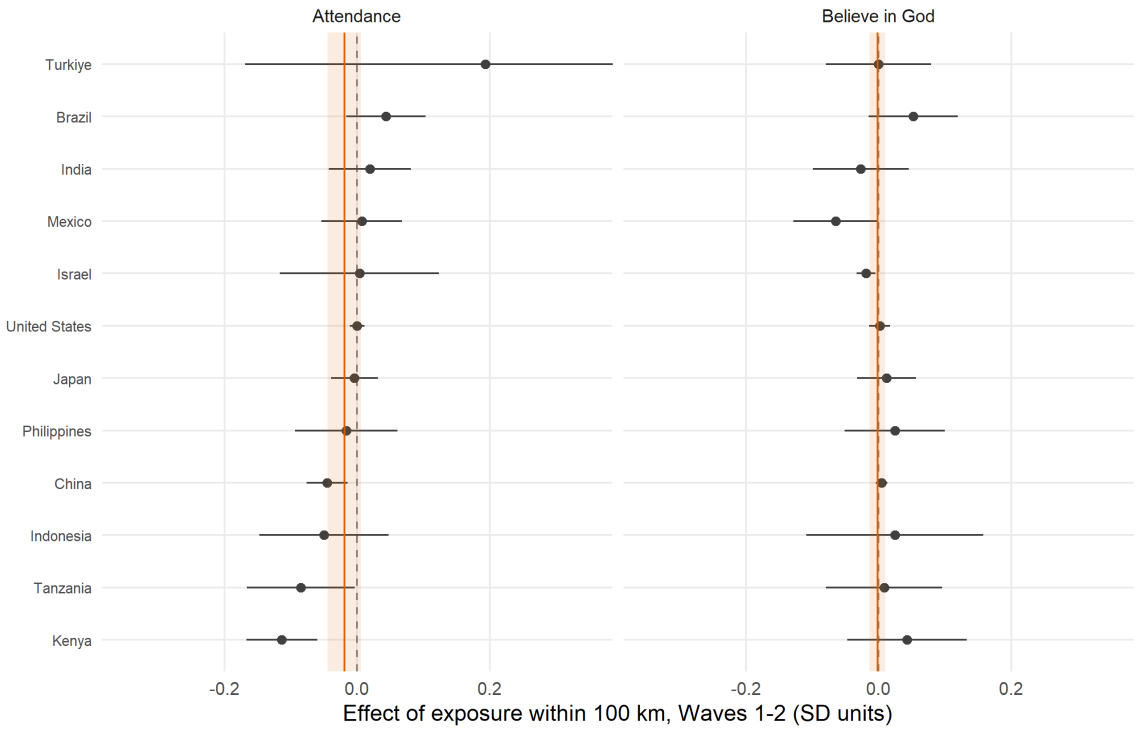
Figure A6 collects the robustness of the two identified margins across the weighting, sample, and treatment-definition variants described in the main text's credibility section, so that part of the battery can be read at once rather than in prose; the distance profiles are in the paper's registered-answer section and the clustering variants in its credibility section.

G The joint-state transition

The main text reports the transition to believing without attending together with the checks that most constrain it. The full battery behind those checks is as follows. The four mutually exclusive destinations—remaining a believing attender, moving to believing without attending, keeping attendance while losing belief, and leaving both—are estimated jointly on the pooled panel, and their effects sum to zero by construction, which the code asserts rather than assumes. The movement is concentrated in the second destination; the two belief-losing destinations are individually negligible (−0.001 and +0.001).

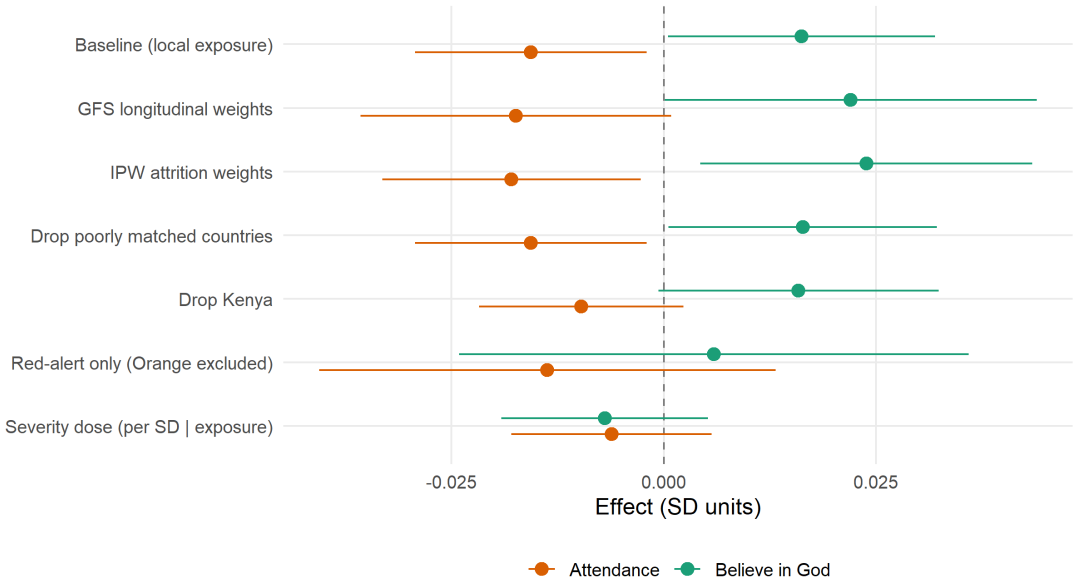
The result is estimated at the registered monthly-or-more attendance threshold (+0.012 on the pooled panel) and re-estimated at a weekly threshold (+0.010, $p = 0.09$), under an ordinal definition requiring at least a one-category attendance decline (+0.011, $p = 0.04$), and with the survey's longitudinal weights (+0.013, $p = 0.01$); a multinomial-logit specification reported as

Figure A5: Per-country estimates with a random-effects summary (Waves 1–2, 100 km)



Note: Per-country estimates of exposure within 100 km on the Waves 1–2 panel with individual and wave fixed effects; 95% confidence intervals, grid-cell-clustered standard errors; countries with at least ten treated respondents (excluded for fewer: Argentina, Australia, Egypt, Germany, Hong Kong, Nigeria, Poland, South Africa, Spain, Sweden, United Kingdom). The orange line and band are the DerSimonian–Laird random-effects mean and its 95% interval. The horizontal axis is truncated at ±0.35; Türkiye's wide interval extends beyond it. Exploratory.

Figure A6: Robustness of the two identified margins



Note: Local-disaster effects on attendance and belief in God across weighting, sample, and treatment-definition variants; 95% confidence intervals, region-clustered standard errors. The severity row is the effect of a one-standard-deviation increase in summed event-severity scores, conditional on exposure.

average marginal effects gives +0.015 for the believing-without-attending destination. The cross-window comparison is the decisive check: the same estimate between Waves 2 and 3 is −0.002 ($p = 0.74$), and a formal test rejects equality of the two windows ($p = 0.007$). That is why the main text describes the transition as a Waves 1–2 phenomenon rather than a standing feature of the panel.

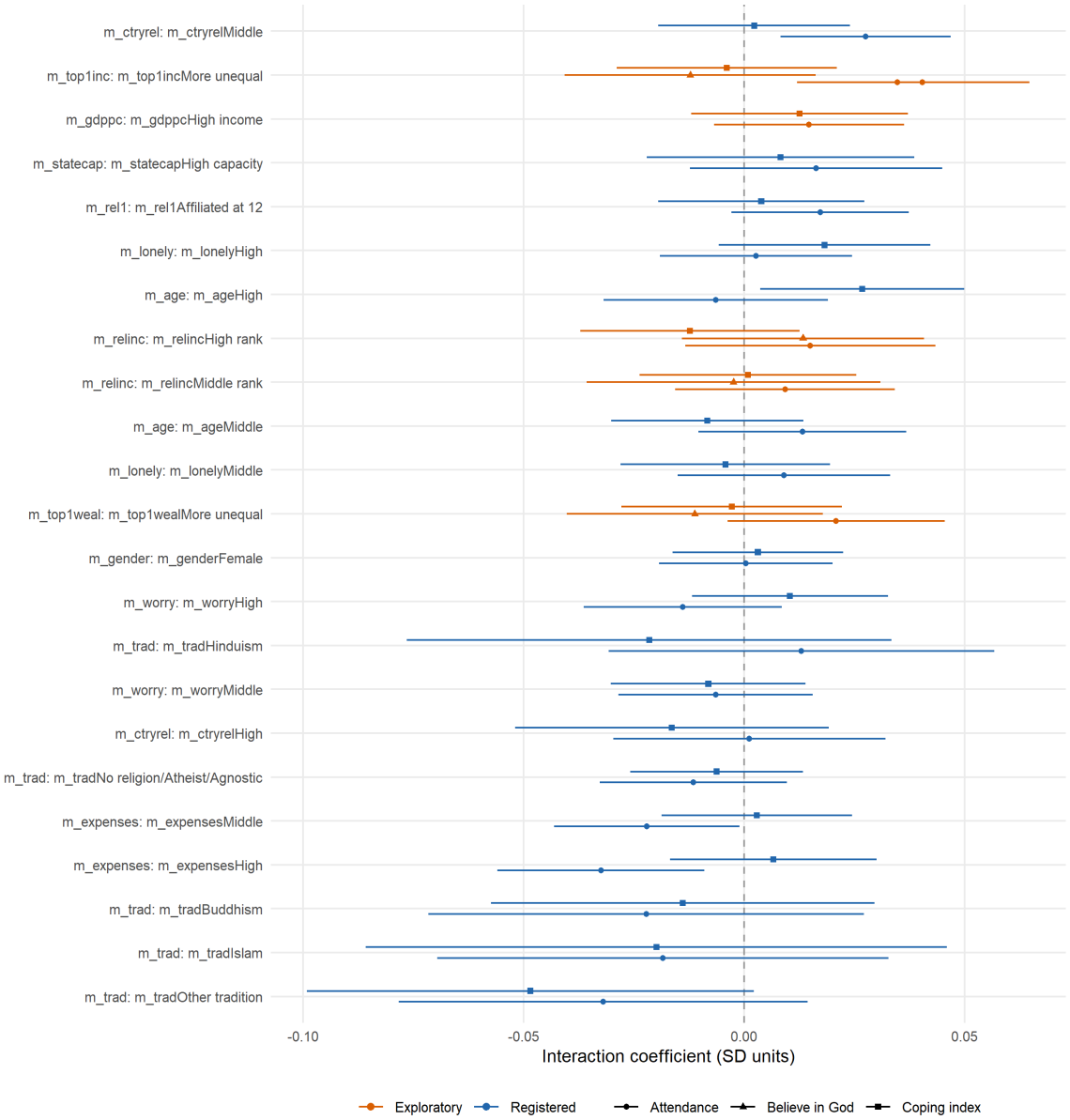
H Moderators and religious capital

Figure A7 plots the complete interaction-contrast battery—the 36 registered contrasts of the pre-registered moderator battery and the 15 exploratory contrasts added afterwards—with 95% confidence intervals. (The thirty-plus country-level splits that motivated the income amendment of Appendix B are country subsamples, not interaction contrasts, and are accounted for there.) Across the registered battery the smallest Benjamini–Hochberg q is 0.12; nothing survives false-discovery-rate control at conventional thresholds, and the main text accordingly reports patterns, not findings.

The exploratory country-level splits are reported here for completeness. A split by GDP per capita at purchasing-power parity behaves like the state-capacity split in the main text (−0.015 in the poorer half of countries versus −0.000 in the richer; difference $p = 0.18$); with a rank correlation of 0.87 between the two country splits—only Argentina, China, India, and Türkiye change sides—income and institutional capacity cannot be separated across 22 countries and Hong Kong, and both echo the individual-level insecurity gradient.

Motivated by evidence that inequality raises religiosity (Solt et al., 2011), I also split countries

Figure A7: The interaction-contrast battery



Note: Each point is the interaction of the 100-km exposure indicator with a moderator level, relative to that moderator's reference group, from the saturated specification; blue points are preregistered contrasts, orange points exploratory ones; 95% confidence intervals, grid-cell-clustered standard errors. Registered contrasts: smallest Benjamini–Hochberg $q = 0.12$.

by the top-1% share of post-tax disposable income and of net personal wealth (WID.world, 2024 values; World Inequality Database, 2026; the post-tax disposable series is the post-tax concept available for all 22 countries and Hong Kong). The result runs in the unexpected direction: the attendance disruption is concentrated in the *less*-unequal half (−0.031 versus +0.004; interaction +0.035, $p = 0.003$), a contrast that strengthens rather than dissolves when Kenya and Japan, the two largest treated cells, are excluded (+0.040, $p = 0.001$); the wealth split points the same way (+0.021, $p = 0.10$), and neither coping nor belief shows inequality heterogeneity. One speculative reconciliation with Solt et al. (2011) is insurance: where inequality is high and safety nets thin, the club value of participation may hold congregations together precisely when disasters disrupt them. With over thirty country-level contrasts examined beyond the preregistration, this is a hypothesis for designed future work, not a finding—and Appendix B reports what happened when that designed test was registered and run.

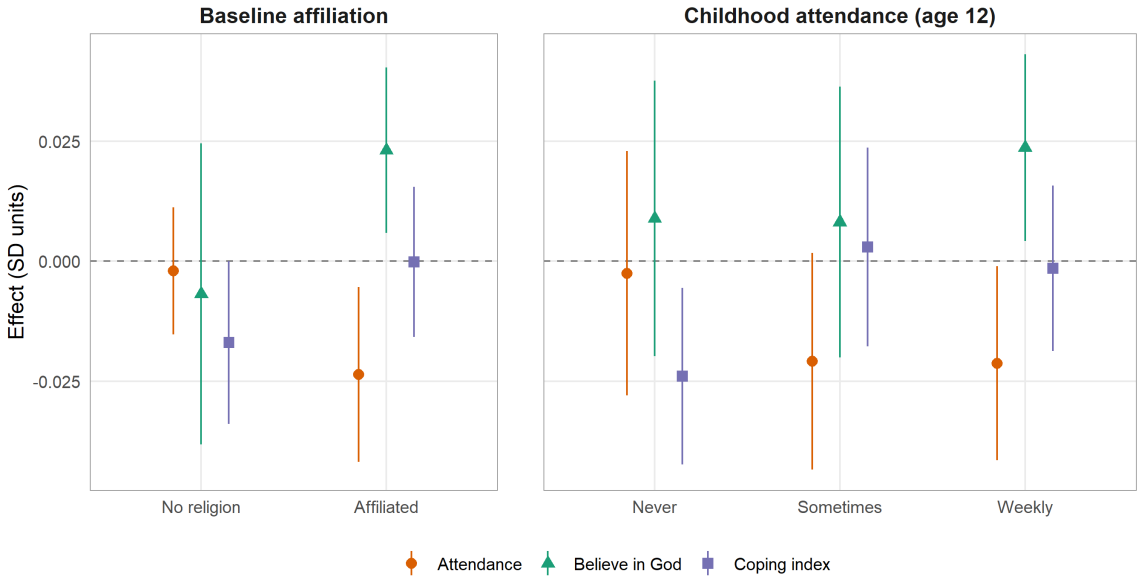
An individual-level version uses each respondent's position in their own country's baseline household-income distribution: the within-country rank of the GFS income bracket (currency- and bracket-scheme-free by construction), cut into within-country tertiles—usable for 91% of respondents, with no country missing more than 21%, though brackets measure unequivalized household income. It points the same way, more weakly: the attendance disruption declines monotonically in relative rank (−0.017 bottom tertile, −0.008 middle, −0.002 top; differences individually insignificant), and splitting below-median respondents by country inequality reproduces the macro pattern (below-median in *less*-unequal countries: −0.046, $p = 0.01$; in more-unequal ones: −0.000), though the formal three-way interaction is imprecise ($p = 0.32$). Relative position moderates less sharply than felt financial strain, and the position-by-inequality pattern is exploratory.

Figure A8 shows the subgroup estimates by baseline religious attachment that the main text summarizes: the attendance and belief responses appear among the affiliated and the weekly-raised and are absent among the never-religious, with between-group tests that carry the expected signs but remain imprecise.

I Mechanisms and secular outcomes

A null on every religious margin raises the question whether the treatment registers in respondents' lives at all. The non-religious outcomes speak to it. Proximate exposure moves self-rated mental health (+0.017, SE 0.006, $p = 0.007$) and national belonging (+0.015, SE 0.007, $p = 0.02$) in the pooled three-wave specification, so the exposure measure is capturing something respondents experienced. Both movements run upward: in a sample where the directly harmed are a small minority, a slight average improvement is consistent with the solidarity and mutual aid that follow disasters, and the argument here uses only that the measure moves, not the direction. The Waves 1–2 regional benchmarks for the mechanism battery (Table A5) are noisier and individually insignificant, which is the attenuation ordering the sharper measure predicts. Figure A9 makes the sharper comparison: attendance at religious services falls while participation in non-religious groups does not (+0.005, SE 0.006)—the pattern a congregation-specific disruption predicts, though the formal religious-minus-secular contrast is −0.012 (SE 0.007, $p = 0.10$), correctly signed but imprecise, so the location claim is suggestive rather than established. These outcomes are reported as mechanism descriptors; none enters a confirmatory claim, and the

Figure A8: Local-disaster effects by baseline religious attachment



Note: Group-specific first-difference estimates of local-disaster exposure, by baseline religious affiliation (left) and by religious service attendance at age 12 (right); 95% confidence intervals, region-clustered standard errors. Waves 1–2; exploratory, predates the registration.

mechanism family carries the same multiple-testing caveat as the moderator battery. The underlying first-difference estimates are tabulated with the Waves 1–2 benchmarks in Appendix J.

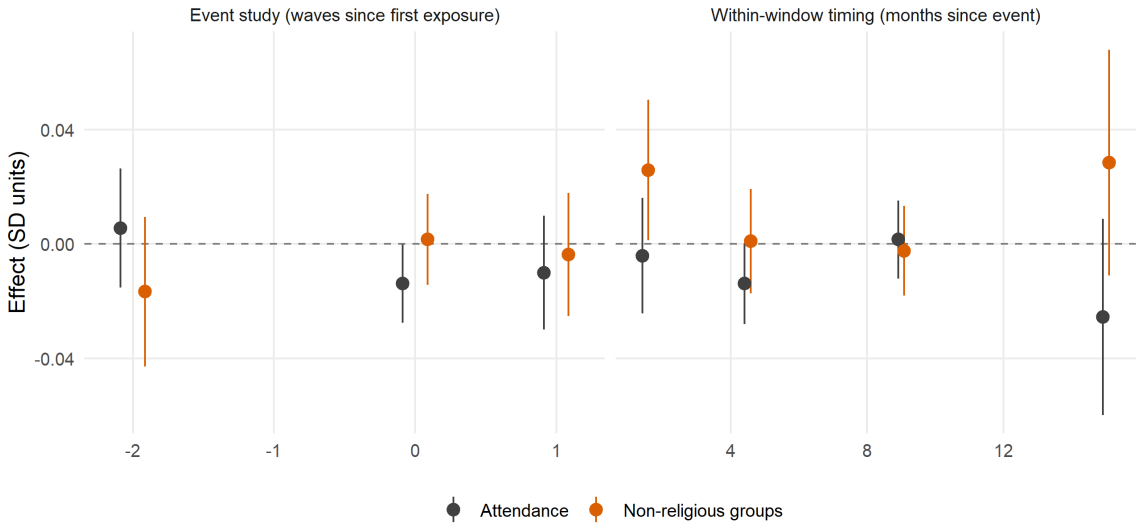
J Hazards, radii, and remaining diagnostics

Hazard-specific exposure indicators belong to the preregistered robustness set. The main text reports the attendance and coping rows of that battery; Table A2 reports it for every outcome margin, each estimated in a single specification that enters the four hazard indicators jointly alongside the national-event control.

Two features are worth drawing out. First, the attendance disruption is a flood phenomenon (-0.029, $p = 0.02$), with cyclones, earthquakes and wildfires individually near zero—consistent with participation falling where infrastructure damage and cleanup demands on time actually bind, rather than wherever a disaster occurs.

Second, the earthquake column speaks directly to Bentzen (2019), the closest published benchmark. That study finds individuals more religious when an earthquake has recently struck near them, and attributes the response to religious coping; it compares districts within countries, measures belief- and identity-type items, and does not observe religious coping directly. Here, within person and on the same hazard, belief in God is +0.012 (SE 0.014; 95% confidence interval $[-0.016, +0.040]$). The point estimate runs in the direction that study reports, but the interval is wide enough that these data neither confirm nor exclude an effect of the reported size, and I claim no earthquake-specific replication. The preregistered coping index, by contrast, is -0.014 (SE 0.012), with an upper confidence limit near +0.009 standard deviations: the chan-

Figure A9: Attendance versus non-religious group participation



Note: Religious service attendance and participation in non-religious groups (an identical five-point frequency scale, reoriented so higher means more participation). Left: Sun–Abraham event-study paths around first exposure. Right: within-window timing bands, plotted at band midpoints in months since the most recent local event within 100 km. 95% confidence intervals, grid-cell-clustered standard errors. Exploratory.

nel to which the response has been attributed is bounded close to zero on the same hazard, in the same sample, under the same specification. Because the two designs standardize different instruments on different scales, this bounds the estimand measured *here*; it is not a claim that the published coefficient is excluded. A further feature of the belief column deserves note: its near-significant cell is cyclones (+0.010, $p = 0.06$), while the earthquake coefficient is larger but much noisier (+0.012, SE 0.014) and no cross-hazard contrast is estimated. The hazard pattern for belief is therefore inconclusive, and it should not be read as mechanism evidence in either direction.

Figure A10 gives the same hazard decomposition on the Waves 1–2 first-difference design with regional exposure, and the three tables that follow report the Waves 1–2 estimates in full: the local-disaster specification that motivated the registration, the pooled any-disaster benchmark against which the attenuation test is read, and the complete robustness battery for both. All three predate the registration and are exploratory by construction.

Table A2: Hazard-specific within-person estimates at 100 km, all margins

	Cyclone	Flood	Earthquake	Wildfire
Believe in God	+0.0104*	−0.0092	+0.0122	−0.0137
	(0.0055)	(0.0145)	(0.0143)	(0.0139)
Life after death	+0.0024	−0.0239	−0.0006	+0.0113
	(0.0057)	(0.0164)	(0.0140)	(0.0074)
Coping index (registered 3-item)	+0.0052	−0.0313*	−0.0140	−0.0311**
	(0.0055)	(0.0169)	(0.0119)	(0.0122)
Religious comfort	+0.0058	−0.0381**	−0.0091	−0.0129
	(0.0058)	(0.0156)	(0.0136)	(0.0164)
Prayer/meditation	+0.0048	−0.0035	+0.0057	−0.0059
	(0.0059)	(0.0152)	(0.0136)	(0.0146)
Connectedness	−0.0010	+0.0067	−0.0018	−0.0126
	(0.0054)	(0.0142)	(0.0116)	(0.0145)
Attendance	+0.0048	−0.0291**	−0.0018	−0.0045
	(0.0045)	(0.0127)	(0.0124)	(0.0064)

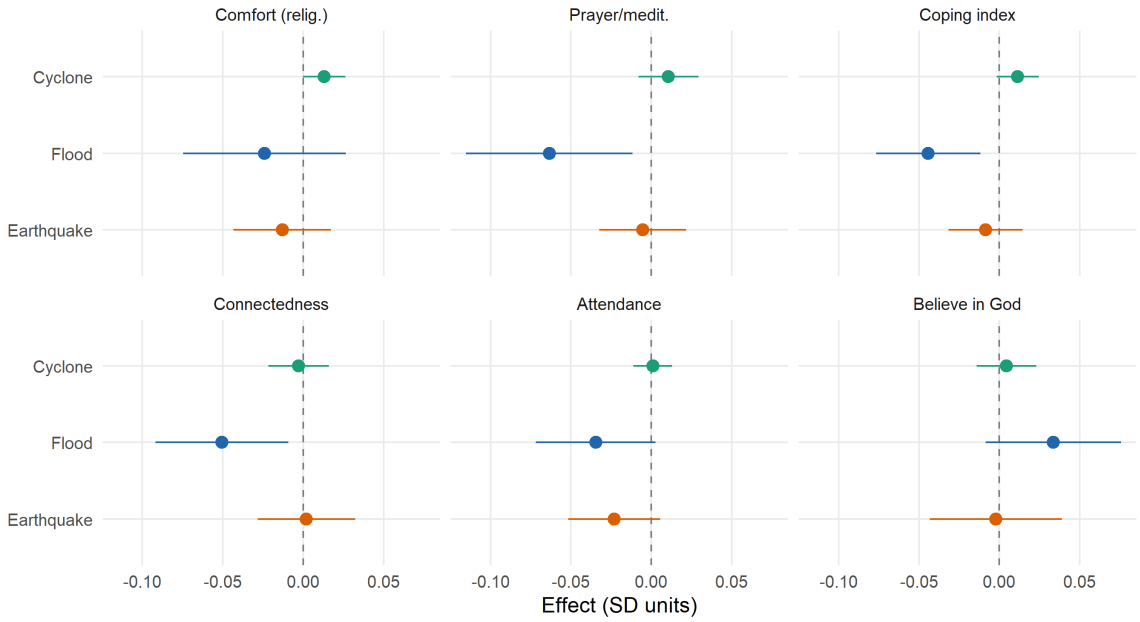
Note: Within-person change in each standardized outcome on indicators for cyclone, flood, earthquake and wildfire exposure within 100 km during the interval (hazard categories overlap, so a respondent can carry more than one indicator), estimated jointly with the national-event control and individual and country-by-wave fixed effects; grid-cell-clustered standard errors in parentheses. Hazard-specific indicators are part of the preregistered robustness set; the decomposition across outcome margins is reported here for completeness. Two-sided p; the registration assigns sidedness to the pooled exposure test, not to this decomposition. With 28 cells the isolated significant entries carry the multiple-testing caveat of Appendix H and are not treated as findings. Earthquake-exposed respondents: 13,732.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3: Local-disaster exposure and within-person change in religiosity (Waves 1–2; exploratory, predates the registration)

	Local disaster		Whole-country shock		
	Estimate	(Std. error)	Estimate	(Std. error)	N
Religious comfort	0.000	(0.008)	0.014	(0.028)	113,230
Prayer/meditation	−0.012	(0.009)	−0.055*	(0.029)	113,353
Coping index	−0.006	(0.007)	−0.020	(0.021)	113,996
Connectedness	−0.007	(0.009)	−0.024	(0.024)	113,516
Attendance	−0.016**	(0.007)	0.000	(0.028)	113,109
Believe in God	0.016**	(0.008)	−0.005	(0.032)	94,824
Life after death	−0.003	(0.010)	−0.069**	(0.030)	72,689

Note: Each row is one regression of the within-person change in the standardized outcome on an indicator for exposure to a local (sub-national) GDACS Orange- or Red-alert disaster whose active period overlaps the respondent's personal Wave 1–2 window, and an indicator for whole-country shocks (events reaching more than half of a country's matched regions), with country fixed effects. Whole-country shocks leave no contemporaneous geographically unaffected comparison within their country; their coefficients are identified off respondent-specific window overlap and partial geographic reach, and are reported as contrasts, not causal effects. The coping index is the two-item (comfort, prayer) version used throughout the Waves 1–2 analysis; the registered three-item index postdates these results (Appendix B). Standard errors clustered by region. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Figure A10: Hazard-specific first-difference estimates (pooled exposure)



Note: Within-person change in each standardized outcome on indicators for cyclone, flood, and earthquake exposure; 95% confidence intervals, standard errors clustered by region. Waves 1–2; exploratory, predates the registration.

Table A4: Pooled any-disaster exposure: the attenuation benchmark (Waves 1–2; exploratory, predates the registration)

	Estimate	(Std. error)
Religious comfort	0.000	(0.008)
Prayer/meditation	−0.011	(0.009)
Connectedness	−0.012	(0.009)
Attendance	−0.009	(0.006)
Coping index	−0.006	(0.006)
Believe in God	0.016**	(0.008)
Life after death	−0.006	(0.009)

Note: Each row is a separate regression of the within-person change in the standardized outcome on an indicator for any GDACS Orange- or Red-alert disaster in the respondent's baseline region during their Wave 1–2 window, with country fixed effects. The coping index is the two-item (comfort, prayer) version used throughout the Waves 1–2 analysis (Appendix B). Standard errors clustered by region. $N \approx 113{,}000$. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A5: Mechanism outcomes: first-difference estimates of pooled disaster exposure (Waves 1–2; exploratory, predates the registration)

	Estimate	(Std. error)
Worry: monthly expenses	0.007	(0.011)
Worry: safety/food/housing	0.009	(0.011)
Loneliness	−0.002	(0.009)
Mental health	0.006	(0.009)
Depressed	0.000	(0.010)
Anxious	0.001	(0.011)
Donated	0.008	(0.011)
Helped stranger	−0.002	(0.012)
Volunteered	0.010	(0.011)
Non-religious group participation	−0.008	(0.009)

Note: Within-person change in each standardized mechanism outcome on the any-disaster indicator, country fixed effects, region-clustered standard errors. Outcomes keep their original survey coding, and coding direction varies across items, so the table is read for precision—no cell approaches significance—rather than for signs. $N \approx 113{,}000$. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A6: Robustness of the local-disaster estimates (Waves 1–2; exploratory, predates the registration)

	Attendance		Believe in God		Coping index	
Baseline (local exposure)	−0.016**	(0.007)	0.016**	(0.008)	−0.006	(0.007)
GFS longitudinal weights	−0.017*	(0.009)	0.022**	(0.011)	−0.005	(0.008)
Attrition-reweighted (inverse probability)	−0.018**	(0.008)	0.024**	(0.010)	−0.004	(0.007)
Drop poorly matched countries	−0.016**	(0.007)	0.016**	(0.008)	−0.006	(0.007)
Drop Kenya	−0.010	(0.006)	0.016*	(0.008)	−0.004	(0.007)
Red-alert only (Orange-only-treated excluded)	−0.014	(0.014)	0.006	(0.015)	−0.002	(0.013)
Severity dose (per SD, conditional on exposure)	−0.006	(0.006)	−0.007	(0.006)	−0.001	(0.005)
Placebo: pre-window disasters	−0.019	(0.012)	−0.007	(0.016)	0.005	(0.011)

Note: Each cell is a separate regression of the within-person change in the standardized outcome, with country fixed effects and region-clustered standard errors. Treatment is exposure to a local (sub-national) GDACS Orange- or Red-alert disaster whose active period overlaps the respondent's personal between-interview window; whole-country shocks (events reaching more than half of a country's matched regions) enter separately in every specification. The poorly matched countries are Australia, Germany, Hong Kong, Spain, and the United Kingdom, where region-name matching between the survey and the boundary data is weakest. The severity row reports the effect of a one-standard-deviation increase in summed event-severity scores, conditional on local exposure. The Red-alert row excludes Orange-only-treated respondents. The placebo row assigns local disasters that concluded in the year before the Wave 1 interview, on the subsample with a full year of event coverage. Randomization inference (1,000 within-country permutations of region event histories, re-estimating the full model): attendance $p = 0.074$, belief $p = 0.078$. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

References

Bentzen, Jeanet Sinding, "Acts of God? Religiosity and Natural Disasters across Subnational World Districts," *The Economic Journal*, 2019, *129* (622), 2295–2321.

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

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

Solt, Frederick, Philip Habel, and J. Tobin Grant, "Economic Inequality, Relative Power, and Religiosity," *Social Science Quarterly*, 2011, *92* (2), 447–465.

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

World Inequality Database, "WID.world: World Inequality Database," https://wid.world 2026. Series sdiinc992j and shweal992j (top 1% shares, p99p100), accessed 15 July 2026 via the Stata `wid` package.