

# Can Mental Health Programs Have Social Externalities? Experimental Evidence from Myanmar\*

Nicolas Cerkez

Alexandra Hartman

Jonathan L. Weigel

Kyaw Zay Ya

June 20, 2026

## Abstract

Non-specialist group-based psychosocial programs are often promoted in conflict settings where access to mental health professionals is limited. While the theory of change underlying such programs is a reduction of mental illness among the individual participants, their group structure may also transform the social fabric of the community. This paper studies one such program in war-torn Myanmar. To help disentangle the importance of the group from the curriculum, we randomly invite participants to either a group-based psychosocial skills intervention or a group activity intervention involving cooking, crafting, or gardening. Although neither intervention measurably improved mental health, both increased social cohesion—an effect we attribute to the group structure. The skills treatment also shifted the locus of perceived responsibility for community service provision from the individual to the local leader, consistent with the curriculum’s emphasis on help-seeking. The results illustrate how targeted psychosocial programs can have broader impacts on social cohesion and local governance in fragile settings.

*Keywords: Mental Health, Social Cohesion, Local Governance*

*JEL Codes: D12, D91, I12, I15, I31, O12*

---

\*Cerkez: University of Oxford, nicolas.cerkez@qeh.ox.ac.uk; Hartman: University College London, alexandra.hartman@ucl.ac.uk; Weigel: UC Berkeley, jweigel@berkeley.edu; Ya: Swiss Tropical and Public Health Institute, University of Basel, dr.kyawzayya87@gmail.com. We thank Matthew Basilico, Gharad Bryan, Giuseppe Raviola, Lauren Young and seminar participants at the IFS, Nuffield College, the Klosters Workshop, EPSA, and the CSAE conference for helpful comments and suggestions as well as Fengzhe (Alain) Liu for providing excellent research assistance. This research was funded by ELRHA and supported by Innovations for Poverty Action, Myanmar. The project was registered at AEARCTR-0012552 and AEARCTR-0015406 (see Appendix A.3 for details) and received human subjects approval through UCL’s IRB, IPA’s IRB (15186), and CEAB’s IRB (CEAB-2022-010 and CEAB-2023-014). All errors remain our own.

# 1 Introduction

Psychosocial programs delivered by non-specialist facilitators are now common across low- and middle-income countries, where access to specialist mental health care remains rare. Governments and NGOs scale these interventions as a cost-effective response to mental distress. Yet, the evidence on their effectiveness is mixed. In recent randomized evaluations, Barker et al. (2022) report sizable improvements in mental health from a group-based curriculum, while Haushofer et al. (2020) find no detectable effects from a comparable individual psychosocial package. Moreover, there is a dearth of evidence from settings with ongoing conflict, where the burden of mental illness may be particularly large.

Many of these interventions are implemented *in groups* (Table A1), such as for example among the rural poor in Ghana (Barker et al., 2022) or among high-risk young men in urban Monrovia, Liberia (Blattman et al., 2017). Group delivery is attractive for cost and reach (WHO, 2024), but it also bundles curriculum content with social exposure. As a result, we often cannot tell whether benefits (if found) arise because participants learn skills from the curriculum, or because they meet regularly in a supportive group that creates new networks, social support, accountability, and simple friendship. Moreover, these programs’ group-based structure may reshape the social fabric of local communities, yet measuring such effects is typically beyond the scope of targeted mental health questionnaires.

This study helps fill these gaps by studying a village-level randomized controlled trial (RCT) with three treatment arms across 90 villages in war-torn Myanmar: (i) a *group skills* arm—a light-touch, group-delivered psychosocial curriculum adapted from Problem Management Plus (PM+); (ii) a *group activity* arm—regular meetings involving cooking, crafting, or gardening designed to reproduce the frequency, duration, and cooperative structure of the skills arm *without* psychosocial content; and (iii) a pure control group. In both treatment arms, sessions were held weekly for ten weeks, lasted 90 minutes each, and were led by trained non-specialist facilitators from local health organizations. Our design therefore allows us to separate (a) a *group effect* arising from repeated cooperative contact and network formation (skills and activities vs. control) from (b) a *curriculum effect*, the incremental value of psychosocial skill-building beyond mere group exposure (skills vs. activities).

In addition to clinical outcomes such as the PHQ-9 (for depression) and the GAD-7 (for anxiety), we assess whether group-delivered programs could potentially reshape social cohesion and local governance, even absent change in mental health outcomes. Regular, cooperative interaction can build trust, reciprocity, and a sense of mutual accountability. Moreover, standard curricula in this space seek to normalize emotional openness and encourage help-seeking behavior, which may alter the (high) perceived value of self-reliance in this setting and correspondingly strengthen expectations of local leaders. Motivated by these channels—and by influential evidence that group-based cognitive behavioral therapy (CBT) reduced crime and violence in Liberia through social and self-regulation pathways (Blattman et al., 2017)—we put social and civic outcomes at the core of the evaluation: perceived closeness (with in- and out-groups), trust (measured in surveys and games), and beliefs and behaviors regarding who should provide local services (e.g., education, health, redistribution) and who actually volunteers.

Compared to the control group, neither program significantly affects standard mental health measures. This null is precisely estimated: we are powered to detect effects of magnitudes on par or below those reported in prior work. Moreover, according to causal-forest analysis, there is little evidence of sub-groups benefiting despite the average null. We rule out several alternative explanations, such as mis-measurement due to stigma or to somatic presentation of mental distress.

By contrast, both interventions improved social cohesion: trust, measured through self-reports and a trust game, increased by 0.179–0.211 SDs compared to the control group; perceived closeness increased by 0.205 SDs in the group skills treatment and 0.076 SDs (not significant) in the group activity treatment. There is no statistical difference in the impact of the two programs on either component of social cohesion. Our interpretation is that the common element of both interventions—the group—drives the change. In particular, repeated cooperative interaction appears to have lowered social distance, normalized cross-group engagement, and created opportunities for trust to develop through shared activity. This is consistent with evidence that the benefits of intergroup contact largely arise from interpersonal exposure rather than structured content (Lowe, 2024).

Participants in the group skills intervention also shift away from the norm of self-reliance and instead expect more from their local leaders. Specifically, they assign greater responsibility

to local leaders (0.453 SDs) and less to themselves in domains like education, health care, and redistribution. This large effect does not appear in the group activity arm. We thus attribute it to the impact of the curriculum, which emphasized the importance of emotional openness and help-seeking. Indeed, a core objective of many psychosocial interventions is to relax the norm of self-reliance—allowing people to admit they are struggling to grow accustomed to asking others for support in difficult times. Consistent with this interpretation, participants in the group skills treatment place greater value on emotional sharing over suppression and report feeling more comfortable discussing negative emotions—such as unhappiness, stress, or depression—with friends.

If groups are potent sources of social healing, why don't they form in the absence of interventions like the one we study? Two frictions are first-order in our setting. Stigma around seeking psychosocial help can deter disclosure and participation. Consistent with this, our baseline reveals low self-reported distress, likely reflecting underreporting rather than absence of hardship. A second friction is social heterogeneity and legacies of conflict. As with many conflict-affected settings, the study area<sup>1</sup> is religiously and ethnically diverse (Buddhist and Christian communities among others) with fault lines drawn by decades of violence between armed groups and the state that reduce cross-group contact. Structured, facilitator-led groups lower both frictions by offering a dignified reason to meet and an explicit norm of respectful interaction that lowers the reputational cost of seeking support across lines of difference.

This research makes three contributions. First, we show that psychosocial programs can generate social externalities by increasing trust, closeness, and shifting the perceived responsibility for local governance. While prior work has documented non-clinical effects of psychosocial interventions, such as reductions in crime and violence among high-risk young men (Blattman et al., 2017), we provide (to our knowledge) the first experimental evidence that such programs can also strengthen social ties and reshape civic expectations in conflict-affected communities.

Second, we contribute to the literature on *group-based* psychosocial interventions in developing countries by isolating the effect of the group and the curriculum. Past work explores both individual-level (e.g., Haushofer et al., 2020; McKelway et al., 2023) and group-level interventions

---

<sup>1</sup>For the safety of our implementing partners, we withhold the name of the state and the identities of the organizations involved.

(e.g., Blattman et al., 2017; Barker et al., 2022).<sup>2</sup> Our group activity treatment arm sheds light on the mechanisms through which group-based psychosocial programs operate—specifically, how group interaction can shape and interact with the social fabric of the community.<sup>3</sup>

Third, we provide rare experimental evidence on the efficacy of a mental health intervention from a low-income setting in the midst of a civil war. In such settings, public health systems typically collapse and prove incapable of addressing the mental distress created by conflict. This is why mental health practitioners advocate non-specialist programs like the one we study (Raviola et al., 2019). Although there is evidence from post-conflict (e.g., Blattman et al., 2017), conflict-affected communities (e.g., Moya et al., 2024), and refugee settings (e.g., Hussam et al., 2022), experimental evidence from settings experiencing active civil war remains scarce (to our knowledge).<sup>4</sup> While our results offer limited grounds for optimism regarding the alleviation of individual mental distress, we view the positive impacts on social cohesion as particularly important in light of the ongoing war in Myanmar, which is one of the severest and most underfunded humanitarian crises (International Crisis Group, 2025). Simple group interaction may produce social externalities large enough to sustain community resilience in the face of large-scale violence.

## 2 Experiment and Data

### 2.1 Experimental Design

Our implementation partner—whose identity we withhold for security reasons—is a local non-governmental organization that provides health services in conflict-affected regions of Myanmar. With decades of in-country experience, it has earned the trust of actors on all sides of the conflict,

---

<sup>2</sup>See Table A1 for an overview of the literature.

<sup>3</sup>Our study also helps to reconcile some of the mixed results from past work—e.g., the positive effects in Barker et al. (2022) compared to the null effects in Haushofer et al. (2020). Because our experimental design closely mimics that of Barker et al. (2022), we can rule out differences in structure, duration, or timing as explanations for the mixed evidence. We suspect that the gender mix of the group is the key difference—and that, in particular, the mixed-gender groups we study may have inhibited some therapeutic gains due to stigma surrounding emotional vulnerability, especially among men.

<sup>4</sup>For instance, a literature review of the health literature on psychosocial interventions in conflict-affected settings is almost entirely focused on programs in refugee and internally displaced person camps (Kleber et al., 2014). The rare exceptions are small-N or non-experimental. For example, Ayoughi et al. (2012) studies a counseling RCT that finds benefits in a sample of 61 women in Afghanistan.

granting it rare access to areas such as our study site, which lies at the epicenter of the civil war between the military government and ethnic armed organizations.

Our sample consists of 90 rural communities located in areas outside the control of Myanmar’s government at the time of the research.<sup>5</sup> Communities were selected based on logistical feasibility, security considerations, and community interest. Due to the ongoing conflict, our local implementing partner identified villages that were deemed accessible and safe for enumerators and program facilitators. Staff from the local partner then visited each community and met with local leaders and the village chief to determine eligibility. Specifically, local leaders helped with recruitment by identifying people in need who could benefit from the program.<sup>6</sup> Given the extreme poverty and agrarian livelihoods of the population, our partner believed that relying on local leaders’ knowledge of need would lead to better targeting than proxy-means testing.<sup>7</sup> Those individuals identified as good candidates were then invited to register for participation, providing us with lists of interested individuals in each village. The 13 selected participants per village were those who registered first. Importantly, participant selection occurred prior to randomization and did not differ across treatment. In total, our baseline sample consists of 1,168 individuals across the 90 communities.

We implemented two stages of randomization. First, communities were randomly assigned to either group skills, group activity, or control — with 30 communities in each category. Second, within the 60 treatment communities, we randomly assigned half to receive an additional culturally specific spiritual intervention.

## 2.2 Treatments

Our study focuses on two interventions that aimed to improve mental well-being and social cohesion. Both interventions were administered to groups of 13 participants and took place in ten 90-minute sessions, administered over ten weeks. The sessions were delivered by non-

---

<sup>5</sup>We use the term community to encompass villages (the great majority) and a few larger towns that were split into a few parts to be of similar size to the modal towns in our study sample.

<sup>6</sup>Severely mentally-ill individuals require specialist care and were also not suitable for this (or similar) programs.

<sup>7</sup>Targeting based on socio-economic status is common among psychosocial programs in developing countries (Table A1). Targeting using local leader knowledge is also practiced in similar contexts such as subsidy allocation (Alatas et al., 2019; Basurto et al., 2020).

specialist facilitators from the region trained by the implementation partner.

**Group skills.** The group skills treatment was a standard psychosocial curriculum designed for delivery by trained non-specialists. It was developed jointly by the implementation partners and a Burmese psychiatrist experienced in conflict-affected communities. The curriculum extends the well-known individual Problem Management Plus (PM+) to the group format and adapts it to the context in Myanmar.

The curriculum covered mental health literacy, stress and coping, recognition of depression and anxiety, basic counseling and active listening, resilience and self-care, stress management techniques (including breathing exercises, progressive muscle relaxation, and sleep hygiene), strategies for managing negative thoughts and grounding, alcohol and substance use, structured problem-solving, and strengthening social support networks. The sessions were interactive, combining “mini-lectures” by the facilitators with case scenarios and group discussions; cases were tailored to local realities. The program emphasized skill-building, self-reflection, and community connection rather than intensive clinical therapy.

In particular, the sessions sought to cultivate emotional openness by encouraging participants to share personal experiences, express vulnerability, and recognize that discussing distress can be a source of strength rather than shame. This focus on openness was operationalized through repeated exercises in active listening, empathy, and trust-building, such as paired sharing and small-group reflection. The approach aimed to normalize emotional expression, reduce stigma, and foster collective resilience within communities.

Overall, this curriculum shares many key features with group CBT, as implemented by Barker et al. (2022), and individual PM+, as implemented by Haushofer et al. (2020), particularly in psychoeducation, stress management, structured problem-solving, and social support. The main difference from group CBT is the absence of formal cognitive restructuring. The main difference from PM+ is our broader scope (e.g., including substance use) and the integration of mindfulness-inspired grounding techniques. Compared with both, our program was lighter and framed less as therapy than as a community psychosocial support package, emphasizing flexibility, cultural tailoring, and emotional openness. Table A2 summarizes these differences as well as comparisons in terms of design and findings.

**Group activity.** The group activity was set up in the same way as the group skills program, but instead of delivering curriculum content over 10 weeks, facilitators convened a group to practice an activity determined in the first session, including cooking, handicraft, or gardening. Different groups chose different activities, so the average effect of the treatment is unlikely to reflect specific attributes of the individual activities but rather the common group participation element.<sup>8</sup> Facilitators scheduled the meeting times for the following nine sessions of equivalent time of day and duration (as compared to the skill program). Participants were not provided with any specific psychosocial skills training during the sessions. The group activity arm is thus not a placebo, in the strict sense of the term, but a deliberately constructed counterfactual designed to replicate the social structure of group therapy without its curricular content.

**Spiritual intervention.** The cross-randomized component was implemented at the start of each session for approximately 15 minutes. It consisted of two parts: first, a culturally appropriate, non-denominational “prayer for others,” in which participants wished for the well-being of others; second, a brief moment of silence / meditation. This time was in addition to, not in place of, other activities during the session. This treatment was randomized at the community level.

**Implementation.** 28 facilitators implemented the program across 60 communities. These facilitators were highly trained but non-specialists. On average, they were 26 years old, the majority were female (68%), and most (86%) were employed in the health sector. Only 21% had prior experience working on mental health programs.

## 2.3 Data and Outcomes

Baseline data collection was completed prior to implementation with all potential participants by trained enumerators during in-person interviews. Two to six months following program implementation, the same individuals completed an endline survey. In addition to basic demographic

---

<sup>8</sup>The fact that people chose the activity could independently impact their engagement (Olken, 2010; Dal Bó et al., 2010). This aspect was deemed necessary by the implementer to adapt the program and generate high participation rates across all the different communities. Moreover, ultimately the pattern of results we observe largely assuages this concern.

measures, we collected data on mental health, social cohesion, and local responsibility.<sup>9</sup>

**Mental health.** We measure mental health at baseline and endline using standard measures of depression and anxiety with the Patient Health Questionnaire 9 (PHQ-9) and Generalized Anxiety Disorder 7 (GAD-7). Both of these measurements are disorder-specific screening tools.<sup>10</sup>

We enumerated two additional mental health measures at endline. First, we measure mental distress using the Kessler Psychological Distress Scale (K6). To overcome possible stigma or other concerns given the presence of enumerators, we designed a private elicitation approach for the K6 in which participants put on a headset and answer the questions without the enumerator seeing or hearing their answers. Second, we included a physical health module, which can capture mental distress in settings with stigma and was judged a locally relevant manifestation of morbidity (Ohrnberger et al., 2017; Shavitt et al., 2016).

**Social cohesion.** We study self-reported trust in and perceived closeness to various groups, such as family, friends, and members of different religious and ethnic communities. We also included a trust game in which respondents (“Player 1”) were asked how much of an initial endowment of 2000 MMK they would send to an anonymous member drawn randomly from the whole village/community (“Player 2”), under the assumption that any amount sent would be tripled before Player 2 decided how much to return. Due to the conflict setting, the local IRB required the stakes of the game to be hypothetical. But following De Quidt et al. (2018), we randomly included a version politely recommending that respondents give more to Player 2 — and this prompt had no impact (Table A17). We therefore assume the trust game results remain robust to experimenter demand effects. Finally, to ensure comprehension, respondents completed a test question following the instructions; 83.36% and 83.80% answered correctly at baseline and endline, respectively.

---

<sup>9</sup> We originally preregistered two separate pre-analysis plans with the intention of writing two papers. We later decided to integrate the analyses into a single paper. See Appendix A.3 for details.

<sup>10</sup>By contrast, the Kessler Psychological Distress Scale (K6), the main outcome measure used in Barker et al. (2022), and the General Health Questionnaire (GHQ-12), the main outcome measure used in Haushofer et al. (2020), are nonspecific distress measures that capture broader psychological strain rather than mapping directly onto a single diagnostic category (see Table A2).

**Local responsibility.** We first ask participants to assign responsibility for the provision of various community services, such as education, dispute resolution, infrastructure, security, redistribution, and health care, to a range of actors. These include the “neighborhood watch,” village leader, central government, NGOs, international organizations, religious groups, and the respondents themselves (“myself”). The beliefs were collected only at endline. To capture behavioral consequences of these beliefs, we also collect information on participants’ involvement in community volunteering, specifically in health committees, development committees, and flood response teams.

## 2.4 Balance and Descriptive Statistics

Our sample is balanced across control and treatment groups (Table A3). The average respondent is female, aged 41, with less than primary education, identifying as Buddhist, with a monthly income around 22 USD. At baseline, the average PHQ-9 score is approximately 3 (on a 0–27 scale), and the average GAD-7 score is about 2 (on a 0–21 scale), both indicating low levels of mental distress.<sup>11</sup> Given the prevailing view among our implementation partners and other health experts in Myanmar that there is a large undiagnosed and untreated burden of mental illness in the country, as well as evidence from other settings that poverty and conflict exacerbate mental health (Ridley et al., 2020; Charlson et al., 2019), the low baseline incidence of mental illness we find presents a puzzle.<sup>12</sup>

The most likely explanation is suppressed reporting due to stigma associated with mental illness and the importance of the “saving face” norm in this context.<sup>13</sup> In cross-country assessments, self-reported mental illness is consistently lower in this region than in comparable countries (with similar levels of poverty and conflict).<sup>14</sup> To assess the role of stigma, we intro-

---

<sup>11</sup>Although they must be locally validated for clinical interpretation, a rule of thumb is that PHQ-9 and GAD-7 scores of 10 or above indicate depression or anxiety. In our sample, 4.5% and 1.2% would be classified as depressed or anxious, respectively, according to these benchmarks.

<sup>12</sup>To provide a clearer picture of respondents’ direct exposure to conflict, we asked about experiences with sexual violence, other forms of physical violence, loss of a primary residence, loss of land or a business, loss of private property, forced labor, and separation from family members. Table A4 reports the share of respondents who experienced each type of conflict-related hardship, as well as the average number of such experiences per respondent. Approximately 40% of respondents report having experienced at least one of these forms of conflict exposure.

<sup>13</sup>See, e.g., AAPP (2017); MIMU (2021); Aye et al. (2020) and Lakin et al. (2021) on stigma and mental illness in Myanmar, and Smith (2025) for evidence outside Myanmar.

<sup>14</sup>See “anxiety disorders” and “major depressive disorders” from the Global Burden of Disease Database.

duced two additional measures of mental distress: an audio-based elicitation of the K6, in which enumerators could neither see nor hear respondents’ answers, and a measure of physical health. The K6 measure, however, reveals similarly low levels of reported mental distress as the direct survey questions. This suggests that reluctance to disclose symptoms in front of enumerators is unlikely to be the primary channel through which stigma operates in this setting. The physical health module was rolled out at endline in light of evidence that mental illness can be more easily detected through physical symptoms in Burmese populations (Steel et al., 2009; Riley et al., 2017; Tay et al., 2019). This is because (a) physical health questions are less subject to stigma and reporting hesitancy, and (b) patients often express mental illness through physical symptoms (independent of reporting). According to this measure, we observe higher levels of morbidity: respondents indicate a positive answer to 4.8 physical symptoms on average out of 15 overall. Of course, this measure also captures purely physical symptoms with no link to mental illness. But the gap between it and the traditional mental health indices—4.8/15 or 0.32 for physical vs. 3/27 or 0.11 for PHQ-9 and 2/21 or 0.10 for GAD-7—combined with the consensus in the health literature about the usefulness of measuring mental illness via physical symptoms in this population, provides suggestive evidence that stigma likely leads to under-reporting of mental distress in our sample.<sup>15</sup>

How reliable then are the mental health measures we have? We validate our primary metrics in two ways.<sup>16</sup> First, we show that these very different metrics, including the physical health index, are highly correlated (Figure A1). Such covariance suggests that there is important signal (and not just pure noise) to these measures, despite their low average levels. Second, we examine how well-known drivers of mental illness, such as individual conflict exposure and poverty, co-move with reported distress (Table A5). Individuals exposed to conflict, those reporting poorer

---

(Last accessed: December 17, 2025.)

<sup>15</sup>All but 74 individuals report experiencing at least one of the 15 physical symptoms from the survey. Ordered from most to least commonly reported, the symptoms are: headache (629 respondents), pain in the arms or legs (618), exhaustion/fatigue (575), dizziness (508), back or neck pain (503), difficulty sleeping (418), stomach pain or discomfort (353), constipation (319), toothache (286), itching or skin rash (269), excessive cold/chills (217), excessive heat/fever (159), diarrhea or loose stool (130), more than usual hair loss (126), and vomiting (103).

<sup>16</sup>The main alternative to our stigma interpretation is that individuals have adapted to ongoing hardship by developing coping mechanisms and a high degree of resilience (e.g. Wood, 2003). We believe that this interpretation may apply to some respondents, but we agree with our implementation partner’s local health experts that it seems implausible that, on average, the neediest individuals in an impoverished rural region beset by civil war would be so seldom struggling with even mild depression and anxiety.

physical health, and those who consume more alcohol exhibit higher PHQ-9 and GAD-7 scores, consistent with greater depression and anxiety. Conversely, respondents who are employed or perceive themselves as wealthier report lower scores on both measures. Reassuringly, these patterns all align with expectations based on work on determinants of mental health in other settings. Ultimately, we believe our measures understate the true burden of mental illness in the population but nonetheless provide a signal of real distress. We thus do not focus on rule-of-thumb cutoffs—e.g., PHQ scores above 10—but study changes in participants’ raw scores across these different metrics.

## 2.5 Attrition and Engagement

In total, our baseline sample consists of 1,168 individuals across the 90 communities. Table A6 provides a detailed overview of our sample by treatment status, showing that attrition between baseline and endline was somewhat higher in the control group (11.28%) than in the treatment groups (6.15% and 6.67%). Differential attrition by treatment is confirmed in Table A7.

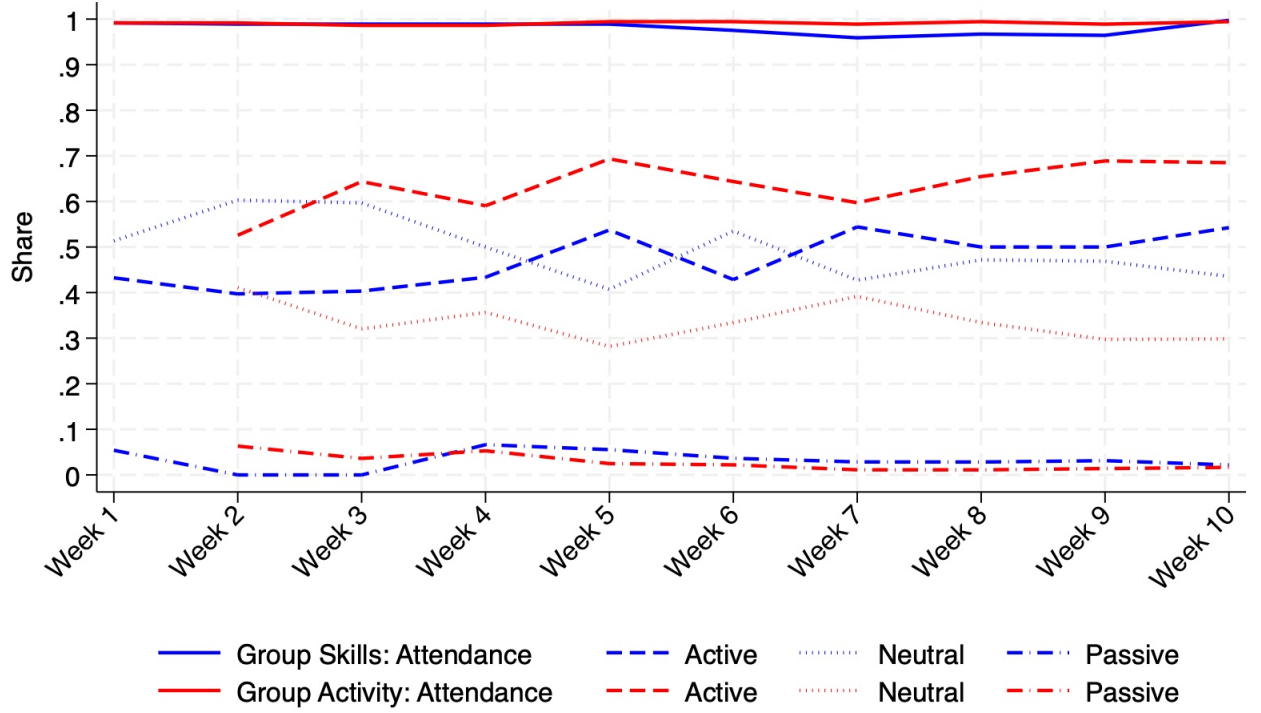
Our main specification thus restricts to respondents observed at both baseline and endline, and we report Lee bounds (Lee, 2009) to assess robustness to differential attrition (Table A10). We also show that results are qualitatively unchanged in samples that replace attriting respondents by drawing replacements from the reserve lists.<sup>17</sup>

To provide a zero-stage, we analyze program data collected by the facilitators on a weekly basis on attendance and engagement during the intervention. As Figure 1 shows, attendance was very high throughout both programs. According to the facilitators’ (admittedly biased) assessments, the great majority of participants were active or neutral (not passive) during the sessions. Participants appear to have been somewhat more active in the group activity treatment.

---

<sup>17</sup>This is not our preferred specification because the replacement protocol differed between treatment and control. In treatment, participants could be replaced during the first three weeks of the program (2% of participants in treatment were replaced in this way), while in control, participants could only be replaced at endline. This differential procedure could change the sample in ways that are endogenous to treatment. That said, the fact that robustness checks *including* the replacement respondents deliver nearly identical results suggests that attrition was either sufficiently idiosyncratic or sufficiently rare as to be essentially inconsequential. Nonetheless, we retain the most conservative sample definition (baseline only) in our main analyses.

Figure 1: Attendance and Engagement



**Notes:** The figure presents the weekly attendance and engagement shares for each treatment group, conditional on not attriting between baseline and endline. Engagement is categorized as Active, Neutral, or Passive.

### 3 Empirical Strategy

To identify the impact of our interventions, we rely on the following ANCOVA specification

$$y_{i,EL} = \alpha + \beta_1 \text{Group Skills}_i + \beta_2 \text{Group Activity}_i + \gamma_1 y_{i,BL} + \delta_s + \delta_e + \varepsilon_i,$$

where  $y_{i,EL}$  and  $y_{i,BL}$  denote the outcomes described in the previous section at endline and baseline, respectively.  $\text{Group Skills}_i$  and  $\text{Group Activity}_i$  are dummy variables indicating assignment to the psychosocial skills and psychosocial activity interventions, respectively.  $\delta_s$  and  $\delta_e$  are strata and enumerator fixed effects; standard errors are clustered at the village level.

## 4 Results

Table 1 reports our main results for our three main outcome groups: mental health, social cohesion, and local responsibility.<sup>18</sup>

Table 1: Main Results

|                                | Mental Health     |                  | Social Cohesion     |                    | Local Responsibility |                      |                    |
|--------------------------------|-------------------|------------------|---------------------|--------------------|----------------------|----------------------|--------------------|
|                                | PHQ-9             | GAD-7            | Trust Index         | Closeness Index    | Leader Responsible   | Self Responsible     | Self Volunteer     |
|                                | (1)               | (2)              | (3)                 | (4)                | (5)                  | (6)                  | (7)                |
| T1: Group Skills               | 0.023<br>(0.074)  | 0.087<br>(0.076) | 0.211***<br>(0.073) | 0.205**<br>(0.083) | 0.453***<br>(0.129)  | -0.435***<br>(0.126) | -0.152*<br>(0.084) |
| T2: Group Activity             | -0.039<br>(0.080) | 0.046<br>(0.067) | 0.179**<br>(0.075)  | 0.076<br>(0.084)   | 0.149<br>(0.115)     | -0.069<br>(0.103)    | -0.092<br>(0.076)  |
| Mean of outcome at baseline    | 0.000             | 0.000            | 0.000               | 0.000              |                      |                      | 0.000              |
| Mean of outcome in controls    |                   |                  |                     |                    | -0.205               | 0.172                |                    |
| p-value: T1 = T2               | [0.394]           | [0.538]          | [0.679]             | [0.123]            | [0.011]              | [0.005]              | [0.426]            |
| BL of outcome                  | Yes               | Yes              | Yes                 | Yes                | No                   | No                   | Yes                |
| Strata fixed effects           | Yes               | Yes              | Yes                 | Yes                | Yes                  | Yes                  | Yes                |
| Enumerator fixed effects       | Yes               | Yes              | Yes                 | Yes                | Yes                  | Yes                  | Yes                |
| SEs clustered at village-level | Yes               | Yes              | Yes                 | Yes                | Yes                  | Yes                  | Yes                |
| Clusters (villages)            | 90                | 90               | 90                  | 90                 | 90                   | 90                   | 90                 |
| Observations                   | 1075              | 1075             | 1075                | 1075               | 1019                 | 1019                 | 1075               |

**Notes:** The table reports results from OLS regressions of our main outcomes on treatment variables, including strata and enumerator fixed effects. Standard errors are clustered at the village-level. The outcomes in Columns (1) and (2) are standardized scores of the PHQ-9 and GAD-7, which measure depression and anxiety, respectively. Columns (3) and (4) present two sub-indices of social cohesion: trust and closeness (see Table A16 for details). Columns (5) and (6) capture attribution beliefs regarding the provision of community services, focusing on two entities: the “Village Leader” (Column (5)) and “Myself” (Column (6)). These outcomes are indices constructed separately for each entity across all community services (see Figure 2 for details). Column (7) reports an index based on three variables measuring respondents’ volunteering behavior (see Table A19 for details). Indices in Columns (3) to (7) are constructed in two steps: first, we average the underlying variables; second, we standardize the resulting average.

### 4.1 No Effects on Individual Mental Health

The outcomes in Columns (1) and (2) of Table 1 are standardized versions of our main mental health measures. Neither the group skills nor the group activity interventions had a significant effect on these outcomes.

We explore this null result in detail in Online Appendix A. First, we show that it is precisely estimated. The minimal detectable effect is roughly 0.2 standard deviations, which is smaller than typical effect sizes for comparable psychological interventions. It is therefore unlikely that

<sup>18</sup>Tables A8 and A9 replicate the main results using the alternative estimation samples that include replacement respondents from (i) treatment villages only and (ii) both treatment and control villages. In addition, Table A10 reports Lee bounds around the main treatment effects to assess the robustness of the results to differential attrition. Across all specifications, our main findings remain unchanged.

insufficient power explains the null result. Second, we examine whether the null masks meaningful heterogeneity. Across specifications, including interactions with conflict exposure and causal-forest analyses that estimate conditional average treatment effects, we find no evidence that any subgroup experienced mental health improvements from either program. The cross-randomized spiritual component also does not appear to moderate the effect. Third, we also find no treatment effects when considering two alternative outcomes we added to the endline survey (the audio K6 and physical health index) designed to improve measurement in the presence of stigma or somatic presentation of mental illness. Finally, a range of alternative specifications and inference methods (e.g., randomization inference) yield similar results.<sup>19</sup>

Ultimately, the evidence is most consistent with the idea that these interventions simply did not move individual mental health substantially in this sample. We discuss how this null result sheds light on the mixed evidence in past literature studying similar programs targeting general populations in low-income countries in Online Appendix A. For instance, we suggest that our results make it less likely that group size, session number or duration, or the time gap between intervention and endline are good candidate explanations for why some past studies, like Haushofer et al. (2020), find null results while others, like Barker et al. (2022), find reductions in mental illness. We suspect that a key factor explaining our null results is gender heterogeneity in a context of stigmatized mental illness reporting. Having both men and women present during group skills groups may have constrained the openness and emotional sharing that drive therapeutic change.

## 4.2 Group Interventions Strengthen Social Cohesion

However, we find positive treatment effects on social cohesion. Both group skills and group activity treatments increased self-reported trust by 0.211 and 0.179 SDs, respectively (Table 1, Column 3). In addition, the skills intervention increased perceived closeness to others by 0.205 SDs; the group activity intervention led to a statistically insignificant increase of 0.076 SDs (Column 4). That said, for both outcomes these treatment effects are statistically indistinguishable

---

<sup>19</sup>Barker et al. (2022) document significant “churn” in mental distress in Ghana: those with depression at baseline often no longer suffered at endline, and vice versa. We explore whether such churn is present in our sample and potentially masking treatment effects in Table A11. We find little evidence to suggest as much.

from one another.

We decompose the trust and closeness index into their underlying components and show further robustness tests in Tables A16 and A17, respectively. The group activity treatment causes a 0.171 SD increase in giving in the trust game, and the group skills treatment causes a 0.099 SD increase (though the latter is not statistically significant) (Table A16, Column 9). We also consider whether experimenter demand effects could explain the results. Following De Quidt et al. (2018), we examine a version of the trust game in which participants were told they would be doing the researchers a favor by sending more money to the other player. We observe no increase in giving when participants receive this prompt (Table A17, Column 3). Demand effects are unlikely to explain the estimated increases in trust.

Because both interventions increased social cohesion and we cannot reject equality of their effects, we interpret these gains as arising from their shared element: sustained participation in a small group. This interpretation aligns with our core hypothesis that repeated cooperative interaction builds trust, reciprocity, and mutual accountability, and with evidence that the benefits of intergroup contact largely stem from interpersonal exposure rather than programmatic content (Lowe, 2024).

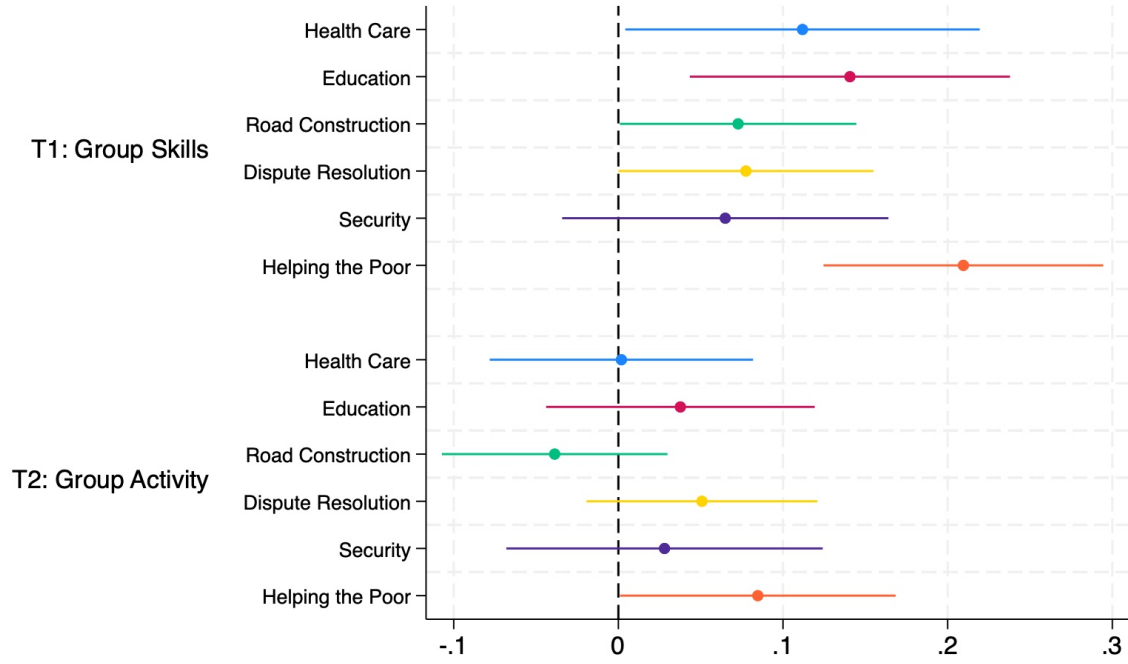
Three features of the intervention environment reinforce this mechanism. First, the repeated nature of engagement—ten weekly 90-minute meetings—created stable opportunities for interaction, allowing relationships to form and deepen over time. Second, the cooperative structure of the activities—whether cooking, handicrafts, or gardening in the group activity treatment or group discussions and case scenario analysis in the group skills treatment—required participants to collaborate and coordinate effort toward shared goals rather than mere co-presence. Third, trained facilitators helped lower barriers to participation and interaction, particularly across religious and ethnic lines, by establishing norms of inclusion, respect, and equal voice.

Together, these elements generated a dignified and predictable social space in which participants could interact repeatedly, cooperate concretely, and engage across lines of difference. The resulting gains in trust and perceived closeness—still detectable two to six months after program completion—suggest that these groups approximated the conditions of collective problem solving that underpin social cohesion and the formation of social capital more broadly (Ostrom, 1990; Miguel and Gugerty, 2005).

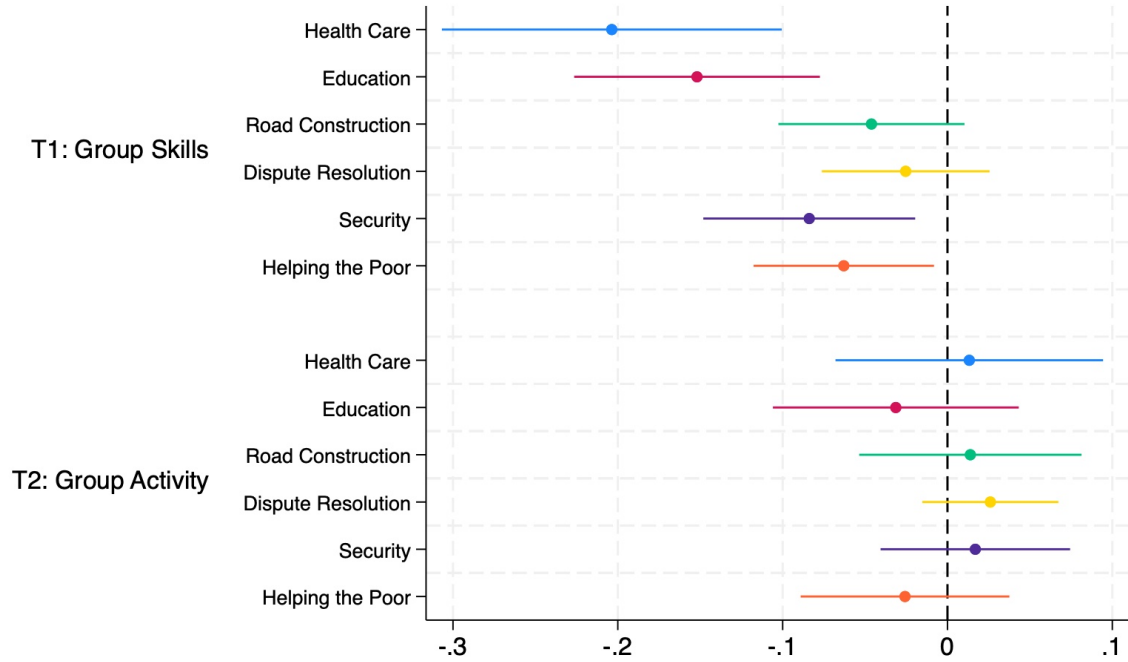
### 4.3 From Self-Reliance to Local Leadership

The group skills intervention also caused a large reduction in the importance of self-reliance and a corresponding increase in the perceived responsibilities of local leaders. Specifically, respondents in the group skills treatment see themselves as less individually responsible for upholding health care, education, and other community services. This effect is large: a 0.435 SD decline in the importance of self-reliance. Moreover, we observe an increase of similar magnitude (0.453 SDs) toward holding the village leader responsible instead. Participants could choose whom they viewed as responsible for these services from a list of potential actors—including also the neighborhood watch, the central government, NGOs, international organizations, and religious groups—but they consistently shifted from the individual to the local leader in the group skills treatment group (Table A18). This shift in perceived responsibility is apparent across many community services but even more pronounced for health, education, and helping the poor (Figure 2). In particular, the erosion of a self-reliance norm is particularly pronounced for health (Figure 2, Panel B). This is reassuring since the group skills treatment was explicitly focused on mental health. It also resonates with the specific emphasis in the curriculum on social determinants of health and the importance of asking for help from others, as we discuss further below.

Figure 2: Components of Local Responsibility



(a) Components of Leader Responsibility



(b) Components of Self Responsibility

**Notes:** The panels display coefficients from OLS regressions of various outcomes on treatment variables and strata and enumerator fixed effects. Standard errors are clustered at the village-level. The outcomes in Panel (a) (Panel (b)) are six binary variables indicating whether respondents view the village leader (themselves) as responsible for providing various community services. All treatment effects are thus expressed in percentage points. We report 90% confidence intervals.

Does participants’ behavior align with these normative shifts in whom they hold responsible for community services? Although we lack objective behavioral measures, participants in the group skills intervention are 0.152 SDs less likely to report volunteering on community committees (Table 1, Column 7). This decrease is largely driven by less engagement in local development committees—in which the local leader plays a key role—rather than other groups such as the flood response team that are headed by leaders at higher levels (Table A19).<sup>20</sup> Although this evidence is only suggestive, it is consistent with the observed shift from self-reliance toward greater expectations of community leaders.

Importantly, we observe no comparable effects for individuals assigned to the group activity intervention and can reject the null of equivalent effects across treatments. We therefore attribute these shifts in perceived responsibility for community services to the *curriculum* followed in the group skills treatment, rather than to group participation or facilitator contact per se.

What exactly is the mechanism through which the curriculum shapes the perceived importance of self-reliance and local leadership? We consider five possibilities.

First, the curriculum explicitly emphasized the importance of emotional openness and help-seeking. Participants were encouraged to share personal experiences, listen supportively, and view requesting assistance as a strength rather than a weakness. The curriculum also trained participants to distinguish between challenges that can be addressed individually and those that require support from others or community institutions. By reframing effective coping as collaborative rather than self-reliant, these lessons plausibly generalized to civic life, leading participants to view community welfare and service provision as shared responsibilities and to attribute a greater role to village leaders. Consistent with this mechanism, participants in the group skills treatment—but not the group activity treatment—exhibit a large increase in emotional openness (0.324 SD) (Table 2, Columns 1–5). Column (1) presents a standardized index combining three measures of comfort discussing difficult emotions, explored individually in Columns (3)–(5), and one measure contrasting emotional sharing with suppression (Column (2)). Each component moves in the same direction: treated participants place greater value

---

<sup>20</sup>One may wonder if the intervention simply reduced the time respondents had available for committee participation. However, this appears unlikely given that the endline survey occurred several months after the group interventions finished — although, in the absence of detailed time use data, we cannot rule this interpretation out entirely.

on sharing emotions and report greater comfort discussing unhappiness, stress, anxiety, and depression with friends.

Second, an alternative interpretation is that increased technical knowledge about mental health might have shifted responsibility attribution by highlighting structural determinants of well-being. We examine this possibility using quasi-test questions in the endline survey concerning technical definitions of depression, counseling, and grounding. Although group skills participants appear to score better in their technical mental health knowledge, the difference is not statistically significant. This is not surprising given that the curriculum focused more on building norms and habits associated with mental well-being rather than teaching technical diagnostic and therapeutic material.

Third, the group skills treatment might have reduced participants' perceived locus of control, making them feel less personally in control of their lives and more at the mercy of external forces. If so, it would have been natural to re-attribute individual responsibility for community services to local leaders. Using standard locus of control questions based on the World Values Survey model, we find a marginally significant, small positive effect in the group activity treatment, and no effect in the group skills treatment. This does not appear to explain why the group skills curriculum shifted respondents from self-reliance to demanding more from local leaders.

Fourth, it could be that the curriculum made people less altruistic toward the community and thus decreased how much they felt responsible for contributing to the common good. This might seem at odds with the social cohesion results. However, trust and closeness need not map one-to-one into personal responsibility: individuals may trust others more because they expect leaders to take charge and promote collective action. We examine a behavioral measure of altruism akin to a dictator game: participants were asked how much of their survey thank you payment they wanted to donate to a randomly selected person from the village. The group skills treatment did not significantly decrease altruism as would have been predicted by this mechanism.

A final possibility is that the group skills program itself—by exemplifying a visible form of community service provision—shifted responsibility attribution toward local leaders (who were the ones who initially endorsed the program and helped with recruitment in the village) (Berman et al., 2011; Beath et al., 2025). We view this explanation as unlikely, however, because the

group activity program was also endorsed and enabled by village leaders in exactly the same way—yet, it had no impact on the perceived responsibility for community service provision.

Taken together, the evidence points to emotional openness as the primary mechanism through which the curriculum reshaped community responsibility attribution. By redefining coping as collaborative rather than self-reliant, the group skills intervention shifted how participants think about both personal hardship and the appropriate role of local leaders in addressing collective needs.

Table 2: Mechanisms

|                                | Emotional Openness  |                     |                                                                                          |                      | Technical Knowledge |                  |                  |                  |                  |                   |                     |
|--------------------------------|---------------------|---------------------|------------------------------------------------------------------------------------------|----------------------|---------------------|------------------|------------------|------------------|------------------|-------------------|---------------------|
|                                | Sharing             |                     | How comfortable do you feel talking to your friends when you feel [...] about your life? |                      | Index               | Depression       |                  | Counseling       |                  | Grounding         |                     |
|                                | Index               | Suppressing         | unhappy                                                                                  | stressed and anxious |                     | (6)              | (7)              | (8)              | (9)              | Locus of Control  | Altruism            |
|                                | (1)                 | (2)                 | (3)                                                                                      | (4)                  | (5)                 | (6)              | (7)              | (8)              | (9)              | (10)              | (11)                |
| T1: Group Skills               | 0.324***<br>(0.081) | 0.284***<br>(0.078) | 0.051**<br>(0.024)                                                                       | 0.048**<br>(0.022)   | 0.038*<br>(0.021)   | 0.153<br>(0.094) | 0.151<br>(0.099) | 0.117<br>(0.072) | 0.039<br>(0.074) | 0.022<br>(0.040)  | -16.687<br>(35.002) |
| T2: Group Activity             | 0.052<br>(0.080)    | 0.027<br>(0.077)    | 0.055**<br>(0.023)                                                                       | 0.006<br>(0.021)     | -0.005<br>(0.021)   | 0.046<br>(0.075) | 0.036<br>(0.076) | 0.017<br>(0.070) | 0.064<br>(0.067) | 0.062*<br>(0.034) | 24.489<br>(36.482)  |
| Mean of outcome at baseline    | 0.000               | -0.064              | 0.883                                                                                    | 0.867                | 0.866               | -0.116           | -0.126           | -0.036           | -0.046           | 0.624             | 889.684             |
| Mean of outcome in controls    |                     |                     |                                                                                          |                      |                     |                  |                  |                  |                  |                   |                     |
| p-value: T1 = T2               | [0.001]             | [0.003]             | [0.844]                                                                                  | [0.057]              | [0.032]             | [0.229]          | [0.189]          | [0.245]          | [0.760]          | [0.304]           | [0.224]             |
| BL of outcome                  | Yes                 | No                  | Yes                                                                                      | Yes                  | Yes                 | No               | No               | No               | No               | Yes               | Yes                 |
| Strata fixed effects           | Yes                 | Yes                 | Yes                                                                                      | Yes                  | Yes                 | Yes              | Yes              | Yes              | Yes              | Yes               | Yes                 |
| Enumerator fixed effects       | Yes                 | Yes                 | Yes                                                                                      | Yes                  | Yes                 | Yes              | Yes              | Yes              | Yes              | Yes               | Yes                 |
| SEs clustered at village-level | Yes                 | Yes                 | Yes                                                                                      | Yes                  | Yes                 | Yes              | Yes              | Yes              | Yes              | Yes               | Yes                 |
| Clusters (villages)            | 90                  | 90                  | 90                                                                                       | 90                   | 90                  | 90               | 90               | 90               | 90               | 90                | 90                  |
| Observations                   | 1075                | 1075                | 1075                                                                                     | 1075                 | 1075                | 1075             | 1075             | 1075             | 1075             | 1075              | 1075                |

**Notes:** The table presents results from OLS regressions of various outcomes on treatment variables and strata and enumerator fixed effects. Standard errors are clustered at the village-level. Columns (1)–(5) report regressions for outcomes relating to emotional openness. Column (1) presents an index of questions in Columns (2)–(5). The index is constructed by first averaging the underlying variable and then standardizing the resulting average. Column (2) reports the share of individuals saying that sharing and talking about their sadness is better than ignoring and suppressing it. Specifically, respondents were presented with two statements. Statement 1 read “some people think it is better to ignore and suppress sadness and negative emotions” while statement 2 read “others think it is better to share and talk about their sadness with others.” After being asked which statement they agree with, we asked respondent how strongly they agree with the statement (agree very strongly, agree strongly, agree somewhat strongly). The resulting variable contains the values 2.5, 1.5, 0.5, -0.5, -1.5, -2.5, where 2.5 indicates agreeing very strongly with statement 2 and -2.5 indicates agreeing very strongly with statement 1. The outcome is standardized. Columns (3) through (5) report binary outcomes indicating whether respondents feel somewhat or very comfortable (vs. somewhat or very uncomfortable) talking to their friends. Columns (6)–(9) report regressions for outcomes relating to technical knowledge. Column (6) presents an index of questions in Columns (7)–(9). The index is constructed by first averaging the underlying variable and then standardizing the resulting average. The outcome in Column (7) is based on a question asking individuals to name, out of eight options, symptoms of depression. The options were: (i) sad mood, (ii) poor sleep, (iii) bad appetite, (iv) pain, (v) not being able to make decisions, (vi) low energy, (vii) all of the above, and (viii) none of the above. The correct answer is either listing all options of (i) to (vi) or (vii). Participants receive one point per correct symptom, yielding a score from 0 to 6. The outcome is standardized. The outcome in Column (8) is based on a question asking individuals to name, out of six options, what counseling is. The options were: (i) giving advice, (ii) active listening, (iii) helping someone to see a different point of view, (iv) lecturing, (v) all of the above, and (vi) none of the above. The correct answers are (ii) and (iii), and respondents receive one point per correct choice, without penalty for incorrect ones. The score ranges from 0 to 2 and is standardized. The outcome in Column (9) is based on a question asking individuals to name, out of six options, what grounding is. The options were: (i) exercising on the ground, (ii) planting plants in the ground (gardening), (iii) noticing what you’re thinking and feeling, (iv) slowing down your breathing, (v) all of the above, and (vi) none of the above. The correct answers are (iii) and (iv), scored similarly to Column (8), with a standardized outcome. Column (10) presents a binary outcome equal to 1 if respondents agree that “most of the success or failure in people’s lives result from their choices and actions” rather than from “forces beyond their control.” Column (11) captures the amount (in Myanmar kyat) that the respondent is willing to donate from their survey payment to a randomly selected anonymous individual in their village.

## 5 Conclusion

Non-specialist, group-based psychosocial programs are widely used in conflict-affected settings as a pragmatic response to mental distress where professional care is scarce. This paper provides evidence that their impacts can extend beyond individual mental health. In war-torn Myanmar, neither a group-based psychosocial skills program nor a structurally identical group activity program improved standard measures of mental health. Yet both increased social cohesion, measured through self-reported trust and closeness as well as behavior in a trust game. We attribute these increases to the group structure itself—repeated, facilitated, cooperative interaction—which was common across treatment arms. At the same time, the psychosocial curriculum appears to have reshaped expectations about local community service provision, weakening norms of self-reliance and increasing the perceived responsibilities of local leaders. Evidence points to the curriculum’s emphasis on emotional openness and help-seeking as one mechanism behind this result. Together, these findings show that commonly used psychosocial interventions can generate meaningful social and even political externalities, strengthening interpersonal trust and reshaping civic expectations in settings where authority, leadership, and sovereignty are contested.

Should psychosocial programs be explicitly designed with these broader social effects in mind—for example, by intentionally forming mixed-identity groups or by incorporating civic content alongside therapeutic material? Doing so may offer opportunities to strengthen social cohesion and reshape governance expectations in deeply divided settings. At the same time, such design choices involve tradeoffs. In contexts marked by stigma or sharp social cleavages, engineered mixing may suppress emotional sharing and openness—the channels through which both therapeutic and non-therapeutic benefits appear to operate. Understanding how best to navigate the potential tradeoff between emotional openness and cross-group cohesion in the design of psychosocial programming—particularly in settings with contested authority—remains an important agenda for future research.

## 6 Bibliography

- AAPP**, “Treating Common Mental Health Disorders in Burma and the Need for Comprehensive Mental Health Policies,” *Assistance Association for Political Prisoners Policy Paper*, 2017.
- Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, Ririn Purnamasari, and Matthew Wai-Poi**, “Does Elite Capture Matter? Local Elites and Targeted Welfare Programs in Indonesia,” *AEA Papers and Proceedings*, may 2019, 109, 334–339.
- Angelucci, Manuela and Daniel Bennett**, “The Economic Impact of Depression Treatment in India: Evidence from Community-based Provision of Pharmacotherapy,” *American Economic Review*, 2024, 114 (1), 169–198.
- Athey, Susan and Stefan Wager**, “Estimating Treatment Effects with Causal Forests: An Application,” *Observational studies*, 2019, 5 (2), 37–51.
- Aye, Win Thu, Aye Aye Lwin, Mark J. Siedner, Myint Myint Han, Ni Ni Soe Chit, Thaingi Maung, Moe Pwint Oo, Thiri Thandar Aung, Chan Myint Tun Soe, Kyaw Ko Ko Linn, Zin Ko Kyaw, and Aye Aye Mon**, “Mental Distress and Associated Factors Among 18–49-Year-old Citizens of Yangon, Myanmar: A Cross-Sectional Study,” *BMC Public Health*, 2020, 20 (1), 1224.
- Ayoughi, Sarah, Inge Missmahl, Roland Weierstall, and Thomas Elbert**, “Provision of Mental Health Services in Resource-poor Settings: A Randomised Trial Comparing Counselling with Routine Medical Treatment in North Afghanistan (Mazar-e-Sharif),” *BMC psychiatry*, 2012, 12 (1), 14.
- Baird, Sarah, Berk Özler, Chiara Dell’Aira, Luca Parisotto, and Danish Us-Salam**, “Therapy, Mental Health, and Human Capital Accumulation Among Adolescent Girls in Uganda,” *Journal of Development Economics*, 2025, 176, 103473.
- Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko**, “Maternal Depression, Women’s Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial,” *American Economic Review*, 2020, 110 (3), 824–859.
- Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry**, “Cognitive Behavioral Therapy Among Ghana’s Rural Poor is Effective Regardless of Baseline Mental Distress,” *American Economic Review: Insights*, 2022, 4 (4), 527–545.
- Bass, Judith, Richard Neugebauer, Kathleen F Clougherty, Helen Verdeli, Priya Wickramaratne, Lincoln Ndogoni, Liesbeth Speelman, Myrna Weissman, and Paul Bolton**, “Group Interpersonal Psychotherapy for Depression in Rural Uganda: 6-month Outcomes: Randomised Controlled Trial,” *The British Journal of Psychiatry*, 2006, 188 (6), 567–573.

- Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson**, “Decentralization and efficiency of subsidy targeting: Evidence from chiefs in rural Malawi,” *Journal of Public Economics*, may 2020, *185*, 104047.
- Beath, Andrew, Fotini Christia, and Ruben Enikolopov**, “Can Development Programs Counter Insurgencies? Evidence from a Field Experiment in Afghanistan,” *American Economic Journal: Applied Economics*, 2025, *17* (3), 80–116.
- Berman, Eli, Jacob N Shapiro, and Joseph H Felter**, “Can Hearts and Minds be Bought? The Economics of Counterinsurgency in Iraq,” *Journal of Political Economy*, 2011, *119* (4), 766–819.
- Bhat, Bhargav, Jonathan De Quidt, Johannes Haushofer, Vikram H Patel, Gautam Rao, Frank Schilbach, and Pierre-Luc P Vautrey**, “The Long-run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes,” 2022.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan**, “Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia,” *American Economic Review*, 2017, *107* (4), 1165–1206.
- , **Sebastian Chaskel, Julian C Jamison, and Margaret Sheridan**, “Cognitive Behavioral Therapy Reduces Crime and Violence Over Ten Years: Experimental Evidence,” *American Economic Review: Insights*, 2023, *5* (4), 527–545.
- Bó, Pedro Dal, Andrew Foster, and Louis Putterman**, “Institutions and Behavior: Experimental Evidence on the Effects of Democracy,” *American Economic Review*, 2010, *100* (5), 2205–2229.
- Bolton, P., C. Lee, E. E. Haroz, L. Murray, S. Dorsey, and C. Robinson**, “A Transdiagnostic Community-based Mental Health Treatment for Comorbid Disorders: Development and Outcomes of a Randomized Controlled Trial among Burmese Refugees in Thailand,” *PLoS Medicine*, 2014, *11* (11), e1001757.
- Bolton, Paul, Judith Bass, Richard Neugebauer, Helen Verdeli, Kathleen F Clougherty, Priya Wickramaratne, Liesbeth Speelman, Lincoln Ndogoni, and Myrna Weissman**, “Group Interpersonal Psychotherapy for Depression in Rural Uganda: A Randomized Controlled Trial,” *JAMA*, 2003, *289* (23), 3117–3124.
- , —, **Theresa Betancourt, Liesbeth Speelman, Grace Onyango, Kathleen F Clougherty, Richard Neugebauer, Laura Murray, and Helen Verdeli**, “Interventions for Depression Symptoms Among Adolescent Survivors of War and Displacement in Northern Uganda: A Randomized Controlled Trial,” *JAMA*, 2007, *298* (5), 519–527.
- Brown, Christina, Supreet Kaur, Geeta Kingdon, and Heather Schofield**, “Cognitive Endurance as Human Capital,” *The Quarterly Journal of Economics*, 2025, *140* (2), 943–1002.

- Bryant, Richard A, Alison Schafer, Katie S Dawson, Dorothy Anjuri, Caroline Mulili, Lincoln Ndogoni, Phiona Koyiet, Marit Sijbrandij, Jeannette Ulate, Melissa Harper Shehadeh et al.**, “Effectiveness of a Brief Behavioural Intervention on Psychological Distress Among Women with a History of Gender-based Violence in Urban Kenya: A Randomised Clinical Trial,” *PLoS Medicine*, 2017, *14* (8), e1002371.
- Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller**, “Bootstrap-based Improvements for Inference with Clustered Errors,” *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.
- Charlson, Fiona, Mark van Ommeren, Abraham Flaxman, Joseph Cornett, Harvey Whiteford, and Shekhar Saxena**, “New WHO Prevalence Estimates of Mental Disorders in Conflict Settings: A Systematic Review and Meta-analysis,” *The Lancet*, 2019, *394* (10194), 240–248.
- Chibanda, Dixon, Helen A Weiss, Ruth Verhey, Victoria Simms, Ronald Munjoma, Simbarashe Rusakaniko, Alfred Chingono, Epiphania Munetsi, Tarisai Bere, Ethel Manda et al.**, “Effect of a Primary Care-based Psychological Intervention on Symptoms of Common Mental Disorders in Zimbabwe: A Randomized Clinical Trial,” *JAMA*, 2016, *316* (24), 2618–2626.
- Cuijpers, Pim, Filip Smit, Ernst Bohlmeijer, Steven D Hollon, and Gerhard Andersson**, “Efficacy of Cognitive Behavioural Therapy and Other Psychological Treatments for Adult Depression: Meta-analytic Study of Publication Bias,” *The British Journal of Psychiatry*, 2010, *196* (3), 173–178.
- , **Matthias Berking, Gerhard Andersson, Leanne Quigley, Annet Kleiboer, and Keith S Dobson**, “A Meta-analysis of Cognitive Behavioural Therapy for Adult Depression, Alone and in Comparison with Other Treatments,” *The Canadian Journal of Psychiatry*, 2013, *58* (7), 376–385.
- Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy**, “Sex Workers, Stigma, and Self-image: Evidence from Kolkata Brothels,” *Review of Economics and Statistics*, 2022, *104* (3), 431–448.
- Haushofer, Johannes, Robert Mudida, and Jeremy P Shapiro**, “The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being,” 2020.
- Hussam, Reshmaan, Erin M Kelley, Gregory Lane, and Fatima Zahra**, “The Psychosocial Value of Employment: Evidence from a Refugee Camp,” *American Economic Review*, 2022, *112* (11), 3694–3724.
- International Crisis Group**, “Myanmar’s Dangerous Drift: Conflict, Elections and Looming Regional Détente,” July 2025, (184). Accessed 2025-12-16.

- John, Anett and Kate Orkin**, “Can Simple Psychological Interventions Increase Preventive Health Investment?,” *Journal of the European Economic Association*, 2022, 20 (3), 1001–1047.
- Kleber, Rolf, K de Jong, Jeroen Knipscheer, and N Ford**, “The Efficacy of Psychosocial Interventions for Adults in Contexts of Ongoing Man-made Violence-A Systematic Review.,” *Health*, 2014, 6, 504–516.
- Lakin, Daniel P., Lwin Lai Win, Thein Thein, Phyo Thitsar Win, and David H. Chae**, “Masculinity and Mental Health Treatment Initiation for Former Political Prisoners in Myanmar,” *BMC Public Health*, 2021, 21 (1), 2244.
- Lee, David S**, “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *The Review of Economic Studies*, 2009, 76 (3), 1071–1102.
- Lowe, Matt**, “Has Intergroup Contact Delivered?,” *Annual Review of Economics*, 2024, 17.
- Maselko, Joanna, Siham Sikander, Elizabeth L Turner, Lisa M Bates, Ikhlaq Ahmad, Najia Atif, Victoria Baranov, Sonia Bhalotra, Amina Bibi, Tayyaba Bibi et al.**, “Effectiveness of a Peer-delivered, Psychosocial Intervention on Maternal Depression and Child Development at 3 Years Postnatal: A Cluster Randomised Trial in Pakistan,” *The Lancet Psychiatry*, 2020, 7 (9), 775–787.
- McKelway, Madeline**, “Women’s Self-Efficacy and Economic Outcomes: Experimental Evidence from India,” *Available at SSRN 5158816*, 2025.
- , **Abhijit Banerjee, Erin Grela, Frank Schilbach, Miriam Sequeira, Garima Sharma, Girija Vaidyanathan, and Esther Duflo**, “Effects of Cognitive Behavioral Therapy and Cash Transfers on Older Persons Living Alone in India: A Randomized Trial,” *Annals of Internal Medicine*, 2023, 176 (5), 632–641.
- McKenzie, David and Owen Ozier**, “Why Ex-post Power Using Estimated Effect Sizes is Bad, But an Ex-post MDE is Not,” *World Bank Development Impact Blog*, 2019.
- Miguel, Edward and Mary Kay Gugerty**, “Ethnic Diversity, Social Sanctions, and Public Goods in Kenya,” *Journal of Public Economics*, 2005, 89 (11-12), 2325–2368.
- MIMU**, “Tackling Perceptions to Address Mental Health in Myanmar,” *Myanmar Information Management Unit Report*, 2021.
- Moya, Andrés, Maria José Torres, Juliana Sánchez-Ariza, Arturo Harker, Alicia Lieberman, Blasina Nino, and Vilma Reyes**, “Caregiver Mental Health and Early Childhood Development: Experimental Evidence from a Conflict-Affected Setting,” 2024.
- Ohrnberger, Julius, Eleonora Fichera, and Matt Sutton**, “The Relationship Between Physical and Mental Health: A Mediation Analysis,” *Social Science & Medicine*, 2017, 195, 42–49.

- Olken, Benjamin A**, “Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia,” *American Political Science Review*, 2010, 104 (2), 243–267.
- Orkin, Kate, Rob Garlick, Mahreen Mahmud, Richard Sedlmayr, Johannes Haushofer, and Stefan Dercon**, “Aspiring to a Better Future: Can a Simple Psychological Intervention Reduce Poverty?,” 2023.
- Ostrom, Elinor**, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.
- Patel, Vikram, Benedict Weobong, Helen A. Weiss, Arpita Anand, Bhargav Bhat, Basavraj Katti, Sona Dimidjian, Ricardo Araya, Steve D. Hollon, Michael King, Lakshmi Vijayakumar, A-La Park, David McDaid, Terry Wilson, Richard Velleman, Betty R. Kirkwood, and Christopher G. Fairburn**, “The Healthy Activity Program (HAP), A Lay Counsellor-delivered Brief Psychological Treatment for Severe Depression, in Primary Care in India: A Randomised Controlled Trial,” *The Lancet*, 2017, 389 (10065), 176–185.
- , **Helen A Weiss, Neerja Chowdhary, Sujit Naik, Shoba Pednekar, Sudipto Chatterjee, Mary De Silva, Bhargav Bhat, Ricardo Araya, Michael King et al.**, “Effectiveness of an Intervention Led by Lay Health Counselors for Depressive and Anxiety Disorders in Primary Care in Goa, India (MANAS): A Cluster Randomized Controlled Trial,” *The Lancet*, 2010, 376 (9758), 2086–2095.
- Quidt, Jonathan De, Johannes Haushofer, and Christopher Roth**, “Measuring and Bounding Experiment Demand,” *American Economic Review*, 2018, 108 (11), 3266–3302.
- Rahman, Atif, Abid Malik, Siham Sikander, Christopher Roberts, and Francis Creed**, “Cognitive Behaviour Therapy-based Intervention by Community Health Workers for Mothers with Depression and Their Infants in Rural Pakistan: A Cluster-randomised Controlled Trial,” *The Lancet*, 2008, 372 (9642), 902–909.
- , **Muhammad Naseem Khan, Syed Usman Hamdani, Anna Chiumento, Parveen Akhtar, Huma Nazir, Anum Nisar, Aqsa Masood, Iftikhar Ud Din, Nasir Ali Khan et al.**, “Effectiveness of a Brief Group Psychological Intervention for Women in a Post-conflict Setting in Pakistan: A Single-blind, Cluster, Randomised Controlled Trial,” *The Lancet*, 2019, 393 (10182), 1733–1744.
- Raviola, Giuseppe, John A Naslund, Stephanie L Smith, and Vikram Patel**, “Innovative Models in Mental Health Delivery Systems: Task Sharing Care with Non-specialist Providers to Close the Mental Health Treatment Gap,” *Current Psychiatry Reports*, 2019, 21, 1–13.
- Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel**, “Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms,” *Science*, 2020, 370 (6522), eaay0214.

- Riley, Andrew, Allison Varner, Peter Ventevogel, Md. Taimur Hasan, and C. Welton-Mitchell**, “Daily Stressors, Trauma Exposure, and Mental Health Among Stateless Rohingya Refugees in Bangladesh,” *Transcultural Psychiatry*, 2017, 54 (3), 304–331.
- Shavitt, Sharon, Young Ik Cho, Timothy P Johnson, Duo Jiang, Allyson Holbrook, and Marina Stavrakantonaki**, “Culture Moderates the Relation Between Perceived Stress, Social Support, and Mental and Physical Health,” *Journal of Cross-Cultural Psychology*, 2016, 47 (7), 956–980.
- Shreekumar, Advik and Pierre Luc Vautrey**, “Managing Emotions: The Effects of Online Mindfulness Meditation on Mental Health and Economic Behavior,” *Unpublished Manuscript, Massachusetts Institute of Technology*, 2022.
- Smith, Emma C**, “Stigma and Social Cover: A Mental Health Care Experiment in Refugee Networks,” 2025.
- Steel, Zachary, Tien Chey, Derrick Silove, Catherine Marnane, Richard A. Bryant, and Mark van Ommeren**, “Association of Torture and Other Potentially Traumatic Events with Mental Health Outcomes Among Populations Exposed to Mass Conflict and Displacement: A Systematic Review and Meta-analysis,” *JAMA*, 2009, 302 (5), 537–549.
- Tay, Alvin Kuowei, Andrew Riley, Rokayea Islam, C. Welton-Mitchell, B. Duchesne, V. Waters, A. Varner, B. Moussa, A. N. M. Mahmudul Alam, M. A. Elshazly, D. Silove, and Peter Ventevogel**, “The Culture, Mental Health and Psychosocial Well-being of Rohingya Refugees: A Systematic Review,” *Epidemiology and Psychiatric Sciences*, 2019, 28 (5), 489–494.
- Vlassopoulos, Michael, Abu Siddique, Tabassum Rahman, Debayan Pakrashi, Asad Islam, and Firoz Ahmed**, “Improving Women’s Mental Health During a Pandemic,” *American Economic Journal: Applied Economics*, 2024, 16 (2), 422–455.
- Wager, Stefan and Susan Athey**, “Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests,” *Journal of the American Statistical Association*, 2018, 113 (523), 1228–1242.
- WHO**, *Psychological Interventions Implementation Manual: Integrating Evidence-based Psychological Interventions Into Existing Services*, Geneva: World Health Organization, 2024.
- Wood, Elisabeth Jean**, *Insurgent Collective Action and Civil War in El Salvador*, Cambridge University Press, 2003.

# A Online Appendix

## A.1 The Null Effect on Mental Health

**Heterogeneity.** It is possible that our null result masks underlying heterogeneous treatment effects. To explore this, we start by looking at heterogeneous treatment effects of both programs by conflict exposure. Table A12 presents these results. We find no evidence of heterogeneous treatment effects based on self-reported conflict exposure.

To explore heterogeneity more systematically, we employ causal forests (Wager and Athey, 2018; Athey and Wager, 2019), which estimate conditional average treatment effects (CATEs) for each respondent using random forests to identify the relevant margins of heterogeneity.

Figure A2 reports the distributions of these CATEs for the PHQ-9 (Panels (a) and (c)) and the GAD-7 (Panels (b) and (d)) for the group skills (Panels (a) and (b)) and the group activity (Panels (c) and (d)) programs. All four histograms are centered around zero (and our estimates from Table 1), indicating no significant heterogeneous treatment effects.<sup>21</sup>

**Power.** Another possibility is that, given our sample size, we simply lack power to detect reasonably sized treatment effects. Our study was powered to detect effects of roughly 0.2 standard deviations on the PHQ-9 and GAD-7 scales.<sup>22</sup> Analyses of psychological interventions, including CBT and PM+, report treatment effects between 0.4 and 0.7 standard deviations (Cuijpers et al., 2010, 2013; Bryant et al., 2017), far exceeding our MDE. Therefore, it is unlikely that insufficient power explains our null findings.

Haushofer et al. (2020), who also report a null result, take this one step further by asking how likely it is that the true effect of the program is smaller than 0.2SD (the MDE). This probability is captured by the negative predictive value (NPV), i.e., the probability that no true effect of a given magnitude exists, conditional on observing a null result at a given power and significance level. The NPV is defined as  $NPV = \frac{(1-\alpha)(1-\pi)}{(1-\alpha)(1-\pi)+\beta\pi}$  where  $\alpha = 0.05$  is the

---

<sup>21</sup>Athey and Wager (2019) recommend conducting an omnibus test of heterogeneity. We implement this by estimating separate “simple” causal forests comparing T1 vs. C and T2 vs. C. The results remain unchanged, and the omnibus test fails to detect any significant heterogeneity. All forests are estimated using 10,000 trees.

<sup>22</sup>This is because, for 80% power and a 5% significance level, the minimal detectable effect equals approximately  $2.8 \times$  the standard error (McKenzie and Ozier, 2019). Hence, for the PHQ-9,  $2.8 \times 0.074 = 0.207$  and  $2.8 \times 0.080 = 0.224$ , and for the GAD-7,  $2.8 \times 0.076 = 0.213$  and  $2.8 \times 0.067 = 0.188$ .

false positive rate,  $\beta = 0.2$  the false negative rate, and  $\pi$  the prior probability of a true effect. The NPV thus depends on the prior. For instance, if we are agnostic about the program’s effectiveness ( $\pi = 0.5$ ), then  $NPV = 0.83$ , meaning that, given our null result, the posterior probability that the true effect is smaller than 0.2SD is 83%.<sup>23</sup>

**Further robustness tests.** Table A13 reports additional robustness checks. Columns (1)–(4) apply alternative inference methods. Columns (1) and (2) replicate the mental health results from Table 1, but compute inference using the wild cluster bootstrap (Cameron et al., 2008) and report the corresponding bootstrapped p-values.<sup>24</sup> Columns (3) and (4) instead present p-values based on randomization inference. In Columns (5) and (6), we return to the main specification but exclude enumerator fixed effects. As shown in the table, there is no significant effect of either program on the PHQ-9 or the GAD-7.

In the last two columns of Table A13, we examine two alternative mental health measures: the K6, elicited via audio, and the physical health index. Both outcomes are standardized. Consistent with our main results, we find no significant treatment effects of either program on these alternative measures.

At baseline, 4.5% and 1.2% of participants report being depressed or anxious, respectively. It is therefore conceivable that the program may have effects within this (relatively small) subsample. However, Table A14 shows that this is not the case. For the PHQ-9, where the affected subsample is somewhat larger, the estimated effects do move in the expected direction: individuals who were depressed at baseline experience a non-significant reduction in their depression scores.<sup>25</sup>

Finally, Table A15 indicates that the cross-randomization had no effect on the effectiveness of either program.

**Discussion** Our mental health results suggest that the program, which targeted the general population, had no discernible impact on mental disorder. What does this imply for mental

---

<sup>23</sup>As  $\pi$  increases, the NPV decreases, which is intuitive: the higher the prior belief in the program’s effectiveness, the less willing one is to interpret a null finding as evidence of no true effect.

<sup>24</sup>The wild cluster bootstrap is most useful when the number of clusters is small, which is not the case here; we report these results for completeness.

<sup>25</sup>This suggests that the interventions may hold some promise for individuals with poorer baseline mental health, but the small number of such cases limits power to detect meaningful effects.

health interventions that target the general population?

Table A2 helps answer this question by comparing our study with the two most closely comparable recent studies: Barker et al. (2022), who find that a group CBT program reduced mental disorders, and Haushofer et al. (2020), who find that an individual PM+ intervention did not.<sup>26,27</sup>

A key difference between Barker et al. (2022) and Haushofer et al. (2020) lies in program structure. The former delivered 12 weekly 90-minute sessions in gender-separated groups of 10 participants, whereas the latter offered only five individual 90-minute sessions. Our design closely mirrors Barker et al. (2022): ten 90-minute sessions in mixed-gender groups of 13 participants, suggesting that differences in group size, session length, or intensity are unlikely to explain divergent outcomes across studies.

Another distinction concerns timing. Whereas the endline in Barker et al. (2022) occurred one to three months after treatment, Haushofer et al. (2020) measured outcomes roughly 13 months later. Our survey took place two to six months post-intervention, making it unlikely that our null results are due to post-intervention fade-out.

Finally, compared to both studies, which rely primarily on nonspecific mental health measures, our use of a broader set of disorder-specific and nonspecific indicators indicates that the null finding is not driven by measurement choices.

An important contextual factor that could explain the difference with Barker et al. (2022) is gender composition and stigma. In our context, stigma around admitting psychological distress remains high and is reinforced by gender norms that discourage men—and, in mixed settings, often women as well—from emotional disclosure. Consistent with this, our baseline data show that men report significantly lower levels of depression and anxiety than women (Table A5). We therefore suspect that mixed-gender groups may have constrained the openness and emotional sharing that drive therapeutic change.

---

<sup>26</sup>We abstract away from the obvious contextual difference that our study was conducted in an ongoing conflict.

<sup>27</sup>Table A1 motivates our comparison with these two studies. Among the studies targeting general, unscreened populations, the interventions most comparable to ours are Haushofer et al. (2020) and Barker et al. (2022) because the other studies in Column (2) differ either in population (e.g., elderly individuals in McKelway et al. (2023); high-risk youth in Blattman et al. (2017, 2023)) or in contextual features such as implementation during the COVID-19 pandemic (Vlassopoulos et al., 2024).

## A.2 Appendix Tables and Figures

Table A1: Psychosocial Interventions at a Glance

|                                       | Not General Population<br>(Screened for Mental Disorders)<br>(1)                                                                                                                                                                                                                                                                                   | General Population<br>(Not Screened for Mental Disorders)<br>(2)                                                                                                                                                                                                                                                                                                       |
|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Individual-Level Interventions</b> | <i>Directly related:</i><br>Rahman et al. (2008) (CBT, THP)<br>Patel et al. (2010)<br>Bolton et al. (2014) (CETA)*<br>Chibanda et al. (2016) (Friendship Bench)<br>Bryant et al. (2017) (PM+)<br>Patel et al. (2017) (HAP)<br>Baranov et al. (2020) (CBT, THP)<br>Bhat et al. (2022) (HAP, THPP)<br>Angelucci and Bennett (2024) (Pharmacotherapy) | <i>Directly related:</i><br>Haushofer et al. (2020) (PM+)<br>McKelway et al. (2023) (CBT)<br>Vlassopoulos et al. (2024) (Telecounseling)**<br><i>Indirectly related:</i><br>Shreekumar and Vautrey (2022) (Mindfulness Meditation)<br>Brown et al. (2025) (Cognitive Practice Intervention)                                                                            |
|                                       | <i>Directly related:</i><br>Bolton et al. (2003) (IPT-G)<br>Bass et al. (2006) (IPT-G)<br>Bolton et al. (2007) (IPT-G)<br>Rahman et al. (2019) (PM+)<br>Maselko et al. (2020) (THPP+)<br>Baird et al. (2025) (IPT-G)                                                                                                                               | <i>Directly related:</i><br>Blattman et al. (2017) (CBT, STYL)<br>Barker et al. (2022) (CBT)<br>Blattman et al. (2023) (CBT, STYL)<br><i>Indirectly related:</i><br>John and Orkin (2022) (Visualization/Planning Interventions)<br>Ghosal et al. (2022) (Self-image Training Program)<br>Orkin et al. (2023) (Aspiration Workshop)<br>McKelway (2025) (Self-efficacy) |

**Notes:** This table summarizes recent psychosocial interventions relevant to our study, organized by whether the programs are delivered at the individual or group level (rows) and whether they target screened clinical populations or the general population (columns). With the exception of Maselko et al. (2020), Haushofer et al. (2020), and McKelway et al. (2023), all directly related interventions report at least some improvement in mental health outcomes—although in several cases the effects are limited to specific subgroups or fade over time. An asterisk (\*) denotes studies conducted in refugee settings; two asterisks (\*\*) denote studies implemented during the COVID-19 pandemic. No listed study involves mixed-gender group therapy sessions. Among the studies targeting general, unscreened populations, the interventions most comparable to ours are Haushofer et al. (2020) and Barker et al. (2022). Other studies in Column (2) differ either in population (e.g., elderly individuals in McKelway et al. (2023); high-risk youth in Blattman et al. (2017, 2023)) or in contextual features such as implementation during the COVID-19 pandemic (Vlassopoulos et al., 2024).

Table A2: Comparison with Barker et al. (2022) and Haushofer et al. (2020)

|                                                        | This Paper<br>(1)                                                                                                            | Barker et al. (2022)<br>(2)                                                                                                                                               | Haushofer et al. (2020)<br>(3)                                                                                                                                                                             |
|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Panel A. Study Context</b>                          |                                                                                                                              |                                                                                                                                                                           |                                                                                                                                                                                                            |
| Country                                                | Myanmar                                                                                                                      | Ghana                                                                                                                                                                     | Kenya                                                                                                                                                                                                      |
| GDP per capita (2024, PPP, \$)                         | 5,997                                                                                                                        | 7,062                                                                                                                                                                     | 6,619                                                                                                                                                                                                      |
| Urban or rural                                         | Rural                                                                                                                        | Rural                                                                                                                                                                     | Rural                                                                                                                                                                                                      |
| Ongoing conflict                                       | Yes                                                                                                                          | No                                                                                                                                                                        | No                                                                                                                                                                                                         |
| <b>Panel B. Intervention</b>                           |                                                                                                                              |                                                                                                                                                                           |                                                                                                                                                                                                            |
| Year of intervention                                   | 2023                                                                                                                         | 2016-2017                                                                                                                                                                 | 2017-2018                                                                                                                                                                                                  |
| Intervention type                                      | Adapted PM+                                                                                                                  | CBT                                                                                                                                                                       | PM+                                                                                                                                                                                                        |
| Target population                                      | General population                                                                                                           | General population                                                                                                                                                        | General population                                                                                                                                                                                         |
| Group or individual                                    | Group                                                                                                                        | Group                                                                                                                                                                     | Individual                                                                                                                                                                                                 |
| Group size                                             | 13                                                                                                                           | 10                                                                                                                                                                        | 1                                                                                                                                                                                                          |
| Group composition                                      | Mixed gender                                                                                                                 | Separated by gender                                                                                                                                                       | -                                                                                                                                                                                                          |
| Number of sessions                                     | 10                                                                                                                           | 12                                                                                                                                                                        | 5                                                                                                                                                                                                          |
| Length of each session                                 | 90 minutes                                                                                                                   | 90 minutes                                                                                                                                                                | 90 minutes                                                                                                                                                                                                 |
| Counselor characteristics                              | 28 highly-trained facilitators,<br>68% are female,<br>57% completed high school,<br>86% employed in health sector            | 37 counselors (and assistants)<br>with a Bachelor's Degree,<br>most commonly in psychology<br>or development studies                                                      | 72 Community Health Workers:<br>volunteers who had completed<br>secondary school                                                                                                                           |
| Training offered to counselors                         | Substantial classroom-based<br>and practical training                                                                        | Two weeks of classroom training,<br>one week of practice sessions<br>(delivered in communities excluded<br>from study based on size)                                      | 9 days of classroom training,<br>5 supervised training<br>sessions with clients                                                                                                                            |
| <b>Panel C. Research Design</b>                        |                                                                                                                              |                                                                                                                                                                           |                                                                                                                                                                                                            |
| Number of communities                                  | 90                                                                                                                           | 258                                                                                                                                                                       | 233                                                                                                                                                                                                        |
| Community selection criteria                           | Partner NGO selected villages<br>that were deemed safe,<br>with road access, and existing<br>ties to local leader            | District Assemblies identified<br>communities with high poverty<br>levels, road access, no existing<br>graduation programs, census verified<br>45+ compounds in community | Partner NGO selected villages in<br>which they were prepared to work,<br>Nakuru County chosen due to high<br>levels of poverty, high baseline rates<br>of poor mental health, and existing<br>NGO presence |
| Sample size                                            | 1170                                                                                                                         | 7227                                                                                                                                                                      | 5756                                                                                                                                                                                                       |
| Individuals receiving therapy                          | 390                                                                                                                          | 1290                                                                                                                                                                      | 1018                                                                                                                                                                                                       |
| Household selection criteria                           | NGO, with village leader<br>advertized program and identified<br>potential participants, who<br>were then allowed to sign up | Households in 40 poorest<br>compounds, in census                                                                                                                          | Households without brick,<br>stone, or metal walls                                                                                                                                                         |
| <b>Panel D. Results</b>                                |                                                                                                                              |                                                                                                                                                                           |                                                                                                                                                                                                            |
| Time between end of<br>intervention and endline survey | 2-6 months                                                                                                                   | 1-3 months                                                                                                                                                                | 2-23 months (median: 13 months)                                                                                                                                                                            |
| Outcome variables                                      | PHQ-9, GAD-7, K6                                                                                                             | K10                                                                                                                                                                       | GHQ-12                                                                                                                                                                                                     |
| Main Effect on Mental Health                           | No effect                                                                                                                    | Improvement                                                                                                                                                               | No effect                                                                                                                                                                                                  |

**Notes:** The table extends Appendix Table 1 from Barker et al. (2022) by including the current paper in the comparison.

Table A3: Household Balance

| Variable                                   | (1)<br>Control     | (2)<br>Group Skills | (3)<br>Group Activity | (4)<br>(1) = (2)  | (5)<br>(1) = (3)  | (6)<br>(2) = (3)    |
|--------------------------------------------|--------------------|---------------------|-----------------------|-------------------|-------------------|---------------------|
| Household size                             | 5.376<br>(2.086)   | 5.240<br>(1.962)    | 5.297<br>(2.060)      | -0.123<br>(0.204) | -0.069<br>(0.207) | -0.054<br>(0.206)   |
| Respondent is male                         | 0.465<br>(0.500)   | 0.475<br>(0.500)    | 0.442<br>(0.497)      | 0.009<br>(0.013)  | -0.022<br>(0.018) | 0.033*<br>(0.017)   |
| Age of respondent                          | 41.517<br>(14.212) | 43.205<br>(12.468)  | 39.637<br>(12.522)    | 1.769<br>(1.399)  | -1.769<br>(1.352) | 3.569***<br>(1.194) |
| Completed primary school                   | 0.310<br>(0.463)   | 0.260<br>(0.439)    | 0.325<br>(0.469)      | -0.048<br>(0.055) | 0.016<br>(0.051)  | -0.065<br>(0.052)   |
| Number of children   having children       | 3.443<br>(1.863)   | 3.399<br>(1.678)    | 3.258<br>(1.840)      | -0.036<br>(0.184) | -0.173<br>(0.183) | 0.139<br>(0.176)    |
| Respondent is Buddhist                     | 0.895<br>(0.307)   | 0.899<br>(0.302)    | 0.906<br>(0.292)      | 0.005<br>(0.039)  | 0.009<br>(0.038)  | -0.005<br>(0.037)   |
| Monthly income in 2022 USD                 | 18.872<br>(28.789) | 20.809<br>(31.337)  | 27.181<br>(145.570)   | 1.852<br>(2.969)  | 8.252<br>(8.053)  | -6.405<br>(8.102)   |
| Respondent has been exposed to conflict    | 0.448<br>(0.498)   | 0.402<br>(0.491)    | 0.396<br>(0.490)      | -0.045<br>(0.081) | -0.052<br>(0.087) | 0.007<br>(0.073)    |
| Generalized Self Efficacy Scale (10-40)    | 26.723<br>(7.090)  | 26.104<br>(7.647)   | 25.547<br>(7.220)     | -0.551<br>(1.104) | -1.151<br>(1.065) | 0.559<br>(1.069)    |
| PHQ-9 Depression score (0-27)              | 2.818<br>(3.374)   | 3.030<br>(3.035)    | 3.308<br>(3.400)      | 0.203<br>(0.363)  | 0.492<br>(0.417)  | -0.277<br>(0.372)   |
| GAD-7 Anxiety score (0-21)                 | 1.645<br>(2.266)   | 1.787<br>(2.118)    | 2.082<br>(2.410)      | 0.133<br>(0.273)  | 0.446<br>(0.313)  | -0.298<br>(0.307)   |
| Respondent feels close to immediate family | 0.916<br>(0.278)   | 0.934<br>(0.248)    | 0.893<br>(0.310)      | 0.020<br>(0.023)  | -0.022<br>(0.029) | 0.042*<br>(0.023)   |
| Respondent feels close to extended family  | 0.558<br>(0.497)   | 0.590<br>(0.492)    | 0.593<br>(0.492)      | 0.037<br>(0.044)  | 0.039<br>(0.054)  | -0.004<br>(0.047)   |
| Respondent feels close to neighbors        | 0.283<br>(0.451)   | 0.339<br>(0.474)    | 0.316<br>(0.466)      | 0.057<br>(0.049)  | 0.034<br>(0.046)  | 0.023<br>(0.040)    |
| Respondent feels close to Buddhists        | 0.324<br>(0.469)   | 0.396<br>(0.490)    | 0.365<br>(0.482)      | 0.072<br>(0.060)  | 0.042<br>(0.058)  | 0.030<br>(0.057)    |
| Respondent feels close to Christians       | 0.165<br>(0.371)   | 0.246<br>(0.431)    | 0.228<br>(0.420)      | 0.079*<br>(0.047) | 0.061<br>(0.049)  | 0.018<br>(0.051)    |
| Respondent feels close to strangers        | 0.038<br>(0.190)   | 0.066<br>(0.248)    | 0.036<br>(0.186)      | 0.029<br>(0.019)  | -0.002<br>(0.015) | 0.030*<br>(0.017)   |
| Observations                               | 346                | 366                 | 364                   | 712               | 710               | 730                 |

**Notes:** Columns (1)—(3) show sample means and standard deviations (in parentheses) for each household characteristic as measured at baseline. Columns (4)—(6) display coefficients and standard errors derived from OLS regressions of the corresponding household characteristic on a treatment dummy variable, and strata fixed effects. Standard errors are clustered at the village-level.

Table A4: Household Balance: Conflict

| Variable                                          | (1)<br>Control   | (2)<br>Group Skills | (3)<br>Group Activity | (4)<br>(1) = (2)  | (5)<br>(1) = (3)  | (6)<br>(2) = (3)  |
|---------------------------------------------------|------------------|---------------------|-----------------------|-------------------|-------------------|-------------------|
| Respondent has been exposed to conflict           | 0.448<br>(0.498) | 0.402<br>(0.491)    | 0.396<br>(0.490)      | -0.045<br>(0.081) | -0.052<br>(0.087) | 0.007<br>(0.073)  |
| Number of conflicts experienced                   | 1.092<br>(1.625) | 0.948<br>(1.640)    | 0.986<br>(1.764)      | -0.138<br>(0.270) | -0.096<br>(0.296) | -0.034<br>(0.267) |
| Respondent experienced: sexual violence           | 0.064<br>(0.244) | 0.036<br>(0.185)    | 0.044<br>(0.205)      | -0.028<br>(0.026) | -0.019<br>(0.032) | -0.008<br>(0.025) |
| Respondent experienced: other physical violence   | 0.188<br>(0.391) | 0.150<br>(0.358)    | 0.173<br>(0.379)      | -0.035<br>(0.069) | -0.012<br>(0.070) | -0.022<br>(0.061) |
| Respondent experienced: loss of primary residence | 0.142<br>(0.349) | 0.096<br>(0.294)    | 0.135<br>(0.342)      | -0.045<br>(0.049) | -0.005<br>(0.057) | -0.039<br>(0.044) |
| Respondent experienced: loss of land or business  | 0.072<br>(0.259) | 0.112<br>(0.316)    | 0.082<br>(0.275)      | 0.039<br>(0.026)  | 0.010<br>(0.026)  | 0.030<br>(0.031)  |
| Respondent experienced: loss of private property  | 0.107<br>(0.309) | 0.115<br>(0.319)    | 0.102<br>(0.303)      | 0.007<br>(0.038)  | -0.005<br>(0.037) | 0.013<br>(0.031)  |
| Respondent experienced: forced labor              | 0.246<br>(0.431) | 0.216<br>(0.412)    | 0.198<br>(0.399)      | -0.027<br>(0.064) | -0.044<br>(0.060) | 0.019<br>(0.052)  |
| Respondent experienced: separation from family    | 0.182<br>(0.386) | 0.167<br>(0.373)    | 0.137<br>(0.345)      | -0.016<br>(0.052) | -0.046<br>(0.053) | 0.030<br>(0.051)  |
| Observations                                      | 346              | 366                 | 364                   | 712               | 710               | 730               |

**Notes:** Columns (1)–(3) show sample means and standard deviations (in parentheses) for each household characteristic as measured at baseline. Columns (4)–(6) display coefficients and standard errors derived from OLS regressions of the corresponding household characteristic on a treatment dummy variable, and strata fixed effects. Standard errors are clustered at the village-level.

Table A5: Baseline Characteristics Correlate with Mental Health Measures

|                                            | PHQ-9                |                    |                  |                     |                    | GAD-7                |                     |                      |                   |                    |                     |                   |                      |                    |
|--------------------------------------------|----------------------|--------------------|------------------|---------------------|--------------------|----------------------|---------------------|----------------------|-------------------|--------------------|---------------------|-------------------|----------------------|--------------------|
|                                            | (1)                  | (2)                | (3)              | (4)                 | (5)                | (6)                  | (7)                 | (8)                  | (9)               | (10)               | (11)                | (12)              | (13)                 | (14)               |
| Respondent is male                         | -0.340***<br>(0.055) |                    |                  |                     |                    |                      |                     | -0.305***<br>(0.056) |                   |                    |                     |                   |                      |                    |
| Age of respondent                          |                      | 0.006**<br>(0.002) |                  |                     |                    |                      |                     |                      | -0.002<br>(0.002) |                    |                     |                   |                      |                    |
| CAGE score (0, no alcohol to 5, alcoholic) |                      |                    | 0.013<br>(0.023) |                     |                    |                      |                     |                      |                   | 0.049**<br>(0.022) |                     |                   |                      |                    |
| General health (1, excellent to 5, poor)   |                      |                    |                  | 0.240***<br>(0.031) |                    |                      |                     |                      |                   |                    | 0.140***<br>(0.026) |                   |                      |                    |
| Respondent is employed                     |                      |                    |                  |                     | -0.127*<br>(0.071) |                      |                     |                      |                   |                    |                     | -0.087<br>(0.072) |                      |                    |
| Perceived economic rank (1-10)             |                      |                    |                  |                     |                    | -0.049***<br>(0.015) |                     |                      |                   |                    |                     |                   | -0.037***<br>(0.013) |                    |
| Respondent has been exposed to conflict    |                      |                    |                  |                     |                    |                      | 0.255***<br>(0.071) |                      |                   |                    |                     |                   |                      | 0.172**<br>(0.074) |
| Mean of outcome at baseline                |                      |                    |                  | 0.000               |                    |                      |                     |                      |                   |                    | 0.000               |                   |                      |                    |
| Strata fixed effects                       | Yes                  | Yes                | Yes              | Yes                 | Yes                | Yes                  | Yes                 | Yes                  | Yes               | Yes                | Yes                 | Yes               | Yes                  | Yes                |
| Enumerator fixed effects                   | Yes                  | Yes                | Yes              | Yes                 | Yes                | Yes                  | Yes                 | Yes                  | Yes               | Yes                | Yes                 | Yes               | Yes                  | Yes                |
| SEs clustered at village-level             | Yes                  | Yes                | Yes              | Yes                 | Yes                | Yes                  | Yes                 | Yes                  | Yes               | Yes                | Yes                 | Yes               | Yes                  | Yes                |
| Clusters (villages)                        | 90                   | 90                 | 90               | 90                  | 90                 | 90                   | 90                  | 90                   | 90                | 90                 | 90                  | 90                | 90                   | 90                 |
| Observations                               | 1075                 | 1075               | 1071             | 1075                | 1075               | 1075                 | 1075                | 1075                 | 1075              | 1071               | 1075                | 1075              | 1075                 | 1075               |

**Notes:** The table reports baseline correlations between respondent characteristics and standardized measures of depression and anxiety. Columns (1)–(7) present results for the PHQ-9 score, and Columns (8)–(14) for the GAD-7 score. All regressions include strata and enumerator fixed effects, and standard errors are clustered at the village-level. The regressors of interest are: (i) an indicator for male respondents; (ii) respondent age (years); (iii) the CAGE score, ranging from 0 for respondents who do not drink alcohol to 5 for alcoholics; (iv) self-reported general health, ranging from 1 (“excellent”) to 5 (“poor”); (v) an indicator for employment; (vi) perceived economic rank on a scale from 1 to 10 (with 10 denoting the “best”); and (vii) an indicator for having been exposed to conflict.

Table A6: Sample Size and Attrition

|                                                  | Control<br>(1) | Group Skills<br>(2) | Group Activity<br>(3)   | Row Total/Average<br>(4) |
|--------------------------------------------------|----------------|---------------------|-------------------------|--------------------------|
| <b>Panel A. Baseline</b>                         |                |                     |                         |                          |
| Number of villages                               | 30             | 30                  | 30                      | 90                       |
| Number of participants per village               | 13             | 13                  | 13 (12 in two villages) | 13                       |
| Total number of participants                     | 390            | 390                 | 388                     | 1168                     |
| <b>Panel B. Attrition Rates</b>                  |                |                     |                         |                          |
| Attrition rate: villages                         | 0%             | 0%                  | 0%                      | 0%                       |
| Attrition rate without replacement: participants | 11.28%         | 6.15%               | 6.67%                   | 7.88%                    |
| Attrition rate with replacement: participants    | 0.00%          | 4.62%               | 4.10%                   | 2.74%                    |

**Notes:** The table describes sample sizes across treatment arms. Panel A reports the baseline sample composition. In treatment villages, participants who stopped attending the intervention during the first three weeks could be replaced by individuals from the original reserve list maintained in each village. In control villages, replacement respondents were drawn from the same reserve lists at endline to maintain comparable sample sizes across study arms. Panel B reports attrition rates with and without replacement respondents.

Table A7: Attrition

|                                | Respondent Attrits Between Baseline and Endline |                     |
|--------------------------------|-------------------------------------------------|---------------------|
|                                | (1)                                             | (2)                 |
| T1: Group Skills               | -0.051**<br>(0.024)                             | -0.038*<br>(0.021)  |
| T2: Group Activity             | -0.051*<br>(0.027)                              | -0.051**<br>(0.026) |
| Mean of outcome                | 0.079                                           |                     |
| Controls                       | No                                              | Yes                 |
| Strata fixed effects           | Yes                                             | Yes                 |
| SEs clustered at village-level | Yes                                             | Yes                 |
| Clusters (villages)            | 90                                              | 90                  |
| Observations                   | 1168                                            | 1154                |

**Notes:** The table reports OLS regressions of a dummy variable indicating whether the respondent attrited between baseline and endline on treatment indicators and strata fixed effects. Standard errors are clustered at the village-level. Column (1) includes no additional controls, while Column (2) adds three baseline covariates: the respondent's age, gender, and a dummy equal to one if the respondent completed more than primary school education.

Table A8: Main Results with Replacement Respondents from Treatment

|                                | Mental Health     |                  | Social Cohesion     |                    | Local Responsibility |                      |                    |
|--------------------------------|-------------------|------------------|---------------------|--------------------|----------------------|----------------------|--------------------|
|                                | PHQ-9             | GAD-7            | Trust Index         | Closeness Index    | Leader Responsible   | Self Responsible     | Self Volunteer     |
|                                | (1)               | (2)              | (3)                 | (4)                | (5)                  | (6)                  | (7)                |
| T1: Group Skills               | 0.029<br>(0.074)  | 0.089<br>(0.075) | 0.205***<br>(0.074) | 0.201**<br>(0.083) | 0.472***<br>(0.131)  | -0.449***<br>(0.130) | -0.155*<br>(0.084) |
| T2: Group Activity             | -0.038<br>(0.079) | 0.046<br>(0.066) | 0.185**<br>(0.075)  | 0.079<br>(0.084)   | 0.146<br>(0.116)     | -0.066<br>(0.107)    | -0.085<br>(0.076)  |
| Mean of outcome at baseline    | 0.000             | 0.000            | 0.000               | 0.000              |                      |                      | 0.000              |
| Mean of outcome in controls    |                   |                  |                     |                    | -0.205               | 0.172                |                    |
| p-value: T1 = T2               | [0.360]           | [0.503]          | [0.801]             | [0.137]            | [0.008]              | [0.005]              | [0.350]            |
| BL of outcome                  | Yes               | Yes              | Yes                 | Yes                | No                   | No                   | Yes                |
| Strata fixed effects           | Yes               | Yes              | Yes                 | Yes                | Yes                  | Yes                  | Yes                |
| Enumerator fixed effects       | Yes               | Yes              | Yes                 | Yes                | Yes                  | Yes                  | Yes                |
| SEs clustered at village-level | Yes               | Yes              | Yes                 | Yes                | Yes                  | Yes                  | Yes                |
| Clusters (villages)            | 90                | 90               | 90                  | 90                 | 90                   | 90                   | 90                 |
| Observations                   | 1091              | 1091             | 1091                | 1091               | 1035                 | 1035                 | 1091               |

**Notes:** The table replicates Table 1, but enlarges the sample to include replacement individuals in treatment groups.

Table A9: Main Results with Replacement Respondents from Treatment and Control

|                                | Mental Health     |                  | Social Cohesion    |                    | Local Responsibility |                      |                    |
|--------------------------------|-------------------|------------------|--------------------|--------------------|----------------------|----------------------|--------------------|
|                                | PHQ-9             | GAD-7            | Trust Index        | Closeness Index    | Leader Responsible   | Self Responsible     | Self Volunteer     |
|                                | (1)               | (2)              | (3)                | (4)                | (5)                  | (6)                  | (7)                |
| T1: Group Skills               | 0.022<br>(0.073)  | 0.071<br>(0.072) | 0.185**<br>(0.075) | 0.190**<br>(0.082) | 0.454***<br>(0.128)  | -0.450***<br>(0.130) | -0.156*<br>(0.083) |
| T2: Group Activity             | -0.035<br>(0.077) | 0.030<br>(0.064) | 0.174**<br>(0.076) | 0.075<br>(0.083)   | 0.129<br>(0.111)     | -0.074<br>(0.106)    | -0.097<br>(0.077)  |
| Mean of outcome at baseline    | 0.000             | 0.000            | 0.000              | 0.000              |                      |                      | 0.000              |
| Mean of outcome in controls    |                   |                  |                    |                    | -0.175               | 0.155                |                    |
| p-value: T1 = T2               | [0.431]           | [0.521]          | [0.886]            | [0.161]            | [0.008]              | [0.006]              | [0.433]            |
| BL of outcome                  | Yes               | Yes              | Yes                | Yes                | No                   | No                   | Yes                |
| Strata fixed effects           | Yes               | Yes              | Yes                | Yes                | Yes                  | Yes                  | Yes                |
| Enumerator fixed effects       | Yes               | Yes              | Yes                | Yes                | Yes                  | Yes                  | Yes                |
| SEs clustered at village-level | Yes               | Yes              | Yes                | Yes                | Yes                  | Yes                  | Yes                |
| Clusters (villages)            | 90                | 90               | 90                 | 90                 | 90                   | 90                   | 90                 |
| Observations                   | 1135              | 1135             | 1135               | 1135               | 1079                 | 1079                 | 1135               |

**Notes:** The table replicates Table 1, but enlarges the sample to include replacement individuals in treatment and control groups.



Table A11: Incidence and Transition Rates of Mental Disorders

| Endline mental distress (PHQ-9)                    |                          |                                        |                                    |                                        |                                                |                                     |              |
|----------------------------------------------------|--------------------------|----------------------------------------|------------------------------------|----------------------------------------|------------------------------------------------|-------------------------------------|--------------|
| Level of baseline mental distress, control group   | Share at baseline<br>(1) | Minimal depression<br>(N = 247)<br>(2) | Mild depression<br>(N = 84)<br>(3) | Moderate depression<br>(N = 13)<br>(4) | Moderately severe depression<br>(N = 2)<br>(5) | Severe depression<br>(N = 0)<br>(6) | Total<br>(7) |
|                                                    |                          |                                        |                                    |                                        |                                                |                                     |              |
| Panel A. Transition Matrix: PHQ-9 in Control Group |                          |                                        |                                    |                                        |                                                |                                     |              |
| (a) Minimal depression (N = 259)                   | 74.39                    | 78.38                                  | 19.31                              | 2.32                                   | 0.00                                           | 0.00                                | 100%         |
| (b) Mild depression (N = 73)                       | 21.10                    | 54.79                                  | 39.73                              | 4.11                                   | 1.37                                           | 0.00                                | 100%         |
| (c) Moderate depression (N = 9)                    | 2.60                     | 22.22                                  | 44.44                              | 33.33                                  | 0.00                                           | 0.00                                | 100%         |
| (d) Moderately severe depression (N = 4)           | 1.16                     | 25.00                                  | 25.00                              | 25.00                                  | 25.00                                          | 0.00                                | 100%         |
| (e) Severe depression (N = 1)                      | 0.29                     | 100.00                                 | 0.00                               | 0.00                                   | 0.00                                           | 0.00                                | 100%         |
| (f) Share at endline                               |                          | 71.39                                  | 24.28                              | 3.76                                   | 0.58                                           | 0.00                                |              |

| Endline anxiety (GAD-7)                            |                          |                                     |                                 |                                    |                                  |              |
|----------------------------------------------------|--------------------------|-------------------------------------|---------------------------------|------------------------------------|----------------------------------|--------------|
| Level of baseline anxiety, control group           | Share at baseline<br>(1) | Minimal anxiety<br>(N = 311)<br>(2) | Mild anxiety<br>(N = 32)<br>(3) | Moderate anxiety<br>(N = 2)<br>(4) | Severe anxiety<br>(N = 1)<br>(5) | Total<br>(6) |
|                                                    |                          |                                     |                                 |                                    |                                  |              |
| Panel B. Transition Matrix: GAD-7 in Control Group |                          |                                     |                                 |                                    |                                  |              |
| (a) Minimal anxiety (N = 308)                      | 89.02                    | 91.23                               | 8.12                            | 0.65                               | 0.00                             | 100%         |
| (b) Mild anxiety (N = 35)                          | 10.12                    | 80.00                               | 17.14                           | 0.00                               | 2.86                             | 100%         |
| (c) Moderate anxiety (N = 3)                       | 0.87                     | 66.67                               | 33.33                           | 0.00                               | 0.00                             | 100%         |
| (d) Severe anxiety (N = 0)                         | 0.00                     |                                     |                                 |                                    |                                  |              |
| (e) Share at endline                               |                          | 89.88                               | 9.25                            | 0.58                               | 0.29                             |              |

**Notes:** The table reports the incidence and transition rates of mental disorders at baseline and endline in the control group, based on the PHQ-9 and GAD-7 measures. Cells above the diagonal reflect worsened mental health, cells on the diagonal indicate no change, and cells below the diagonal represent improvements in mental health status.

Table A12: Heterogeneity by Self-reported Conflict Exposure

|                                          | PHQ-9             | GAD-7             |
|------------------------------------------|-------------------|-------------------|
|                                          | (1)               | (2)               |
| T1: Group Skills                         | -0.023<br>(0.077) | -0.031<br>(0.083) |
| T1: Group Skills x Exposed to Conflict   | 0.110<br>(0.117)  | 0.137<br>(0.131)  |
| T2: Group Activity                       | -0.058<br>(0.093) | -0.074<br>(0.093) |
| T2: Group Activity x Exposed to Conflict | 0.052<br>(0.117)  | 0.081<br>(0.116)  |
| Mean of outcome at baseline              | 0.000             | 0.000             |
| Coefficient T1 + interaction             | 0.087             | 0.106             |
| p-value: Coefficient T1 + interaction    | [0.414]           | [0.319]           |
| Coefficient T2 + interaction             | -0.006            | 0.007             |
| p-value: Coefficient T2 + interaction    | [0.957]           | [0.936]           |
| BL of outcome                            | Yes               | Yes               |
| Strata fixed effects                     | Yes               | Yes               |
| Enumerator fixed effects                 | Yes               | Yes               |
| SEs clustered at village-level           | Yes               | Yes               |
| Clusters (villages)                      | 90                | 90                |
| Observations                             | 1075              | 1075              |

**Notes:** The table reports OLS estimates of standardized PHQ-9 and GAD-7 scores, measuring depression and anxiety, respectively, on treatment indicators, their interactions with an indicator for conflict exposure, the conflict exposure indicator itself, and strata and enumerator fixed effects. Standard errors are clustered at the village-level.

Table A13: Further Robustness Tests

|                                   | PHQ-9             | GAD-7            | PHQ-9             | GAD-7            | PHQ-9             | GAD-7            | K6 Audio          | Physical Index    |
|-----------------------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|-------------------|-------------------|
|                                   | (1)               | (2)              | (3)               | (4)              | (5)               | (6)              | (7)               | (8)               |
| T1: Group Skills                  | 0.023<br>[0.748]  | 0.087<br>[0.267] | 0.023<br>[1.000]  | 0.087<br>[1.000] | 0.054<br>(0.122)  | 0.079<br>(0.146) | -0.024<br>(0.080) | -0.012<br>(0.080) |
| T2: Group Activity                | -0.039<br>[0.631] | 0.046<br>[0.485] | -0.039<br>[1.000] | 0.046<br>[1.000] | -0.036<br>(0.118) | 0.004<br>(0.156) | 0.074<br>(0.071)  | -0.021<br>(0.077) |
| Mean of outcome at baseline       | 0.000             | 0.000            | 0.000             | 0.000            | 0.000             | 0.000            |                   |                   |
| Mean of outcome in controls       |                   |                  |                   |                  |                   |                  | -0.082            | -0.051            |
| p-value: T1 = T2                  | [0.391]           | [0.534]          | [1.000]           | [1.000]          | [0.442]           | [0.645]          | [0.293]           | [0.907]           |
| BL of outcome                     | Yes               | Yes              | Yes               | Yes              | Yes               | Yes              | No                | No                |
| Strata fixed effects              | Yes               | Yes              | Yes               | Yes              | Yes               | Yes              | Yes               | Yes               |
| Enumerator fixed effects          | Yes               | Yes              | Yes               | Yes              | No                | No               | Yes               | Yes               |
| SEs clustered at village-level    | No                | No               | No                | No               | Yes               | Yes              | Yes               | Yes               |
| Wild cluster bootstrap [p-value]  | Yes               | Yes              | No                | No               | No                | No               | No                | No                |
| Randomization inference [p-value] | No                | No               | Yes               | Yes              | No                | No               | No                | No                |
| Clusters (villages)               | 90                | 90               | 90                | 90               | 90                | 90               | 90                | 90                |
| Observations                      | 1075              | 1075             | 1075              | 1075             | 1076              | 1076             | 1075              | 1074              |

**Notes:** Columns (1)—(2) and (3)—(4) replicate Columns (1) and (2) of Table 1, but report  $p$ -values based on the wild cluster bootstrap (Columns (1)—(2)) and randomization inference (Columns (3)—(4)), respectively. Columns (5) and (6) again replicate Columns (1) and (2) of Table 1, excluding enumerator fixed effects. The last two columns report OLS estimates for two additional mental health measures, controlling for treatment indicators as well as strata and enumerator fixed effects. Standard errors are clustered at the village-level. Column (7) uses the standardized K6 score as the outcome. The K6 consists of six items related to general mental health—nervousness, hopelessness, restlessness, feeling that everything is an effort, sadness, and worthlessness—each scored from 0 (“none of the time”) to 4 (“all of the time”), with total scores ranging from 0 to 24. To minimize enumerator effects and stigma, the K6 was administered via audio headset. Respondents were asked: “During the last 30 days, about how often did you feel [...]?” Column (8) reports results using a standardized physical health index, constructed from responses to 15 questions on symptoms experienced in the past 30 days (e.g., stomach pain, vomiting, joint pain, hair loss). The index is based on the first principal component from a principal component analysis and is standardized.

Table A14: Floor Effects

|                                       | PHQ-9             | Respondent is depressed | GAD-7            | Respondent is anxious |
|---------------------------------------|-------------------|-------------------------|------------------|-----------------------|
|                                       | (1)               | (2)                     | (3)              | (4)                   |
| T1: Group Skills                      | 0.024<br>(0.073)  | 0.009<br>(0.017)        | 0.078<br>(0.075) | 0.014<br>(0.010)      |
| T2: Group Activity                    | -0.013<br>(0.072) | -0.008<br>(0.016)       | 0.041<br>(0.063) | 0.007<br>(0.007)      |
| T1: Group Skills x Depressed at BL    | -0.063<br>(0.503) | -0.265<br>(0.163)       |                  |                       |
| T2: Group Activity x Depressed at BL  | -0.560<br>(0.436) | -0.207<br>(0.132)       |                  |                       |
| T1: Group Skills x Anxious at BL      |                   |                         | 1.193<br>(0.918) | -0.000<br>(0.011)     |
| T2: Group Activity x Anxious at BL    |                   |                         | 0.542<br>(0.752) | 0.000<br>(0.009)      |
| Mean of outcome at baseline           | 0.000             | 0.045                   | 0.000            | 0.012                 |
| Coefficient T1 + interaction          | -0.039            | -0.256                  | 1.272            | 0.014                 |
| p-value: Coefficient T1 + interaction | [0.938]           | [0.106]                 | [0.170]          | [0.189]               |
| Coefficient T2 + interaction          | -0.573            | -0.215                  | 0.583            | 0.007                 |
| p-value: Coefficient T2 + interaction | [0.213]           | [0.108]                 | [0.451]          | [0.412]               |
| BL of outcome                         | Yes               | Yes                     | Yes              | Yes                   |
| Strata fixed effects                  | Yes               | Yes                     | Yes              | Yes                   |
| Enumerator fixed effects              | Yes               | Yes                     | Yes              | Yes                   |
| SEs clustered at village-level        | Yes               | Yes                     | Yes              | Yes                   |
| Clusters (villages)                   | 90                | 90                      | 90               | 90                    |
| Observations                          | 1075              | 1075                    | 1075             | 1075                  |

**Notes:** The table reports OLS estimates where outcomes are regressed on treatment indicators, their interactions with baseline indicators for depression or anxiety, and the baseline indicators themselves. All specifications include strata and enumerator fixed effects, and standard errors are clustered at the village-level. The outcome variables in Columns (1) and (3) are standardized PHQ-9 and GAD-7 scores, respectively, while Columns (2) and (4) use binary indicators equal to one if the respondent is classified as depressed (PHQ-9 score  $\geq 10$ ) or anxious (GAD-7 score  $\geq 10$ ).

Table A15: Impacts of the Cross-randomization on Mental Health

|                                             | PHQ-9             |                   | GAD-7            |                   |
|---------------------------------------------|-------------------|-------------------|------------------|-------------------|
|                                             | (1)               | (2)               | (3)              | (4)               |
| T1: Group Skills                            | 0.023<br>(0.074)  | 0.087<br>(0.095)  | 0.087<br>(0.076) | 0.167*<br>(0.090) |
| T2: Group Activity                          | -0.039<br>(0.080) | 0.025<br>(0.091)  | 0.046<br>(0.067) | 0.061<br>(0.087)  |
| T1: Group Skills x Spiritual Intervention   |                   | -0.150<br>(0.122) |                  | -0.177<br>(0.122) |
| T2: Group Activity x Spiritual Intervention |                   | -0.126<br>(0.133) |                  | -0.027<br>(0.107) |
| Mean of outcome at baseline                 | 0.000             |                   | 0.000            |                   |
| p-value: T1 = T2                            | [0.394]           |                   | [0.538]          |                   |
| Coefficient T1 + interaction                |                   | -0.063            |                  | -0.010            |
| p-value: Coefficient T1 + interaction       |                   | [0.490]           |                  | [0.922]           |
| Coefficient T2 + interaction                |                   | -0.101            |                  | 0.033             |
| p-value: Coefficient T2 + interaction       |                   | [0.371]           |                  | [0.686]           |
| BL of outcome                               | Yes               | Yes               | Yes              | Yes               |
| Strata fixed effects                        | Yes               | Yes               | Yes              | Yes               |
| Enumerator fixed effects                    | Yes               | Yes               | Yes              | Yes               |
| SEs clustered at village-level              | Yes               | Yes               | Yes              | Yes               |
| Clusters (villages)                         | 90                | 90                | 90               | 90                |
| Observations                                | 1075              | 1075              | 1075             | 1075              |

**Notes:** The table reports OLS estimates of standardized PHQ-9 scores (Columns (1)—(2)) and GAD-7 scores (Columns (3)—(4)) on treatment indicators, controlling for strata and enumerator fixed effects. Standard errors are clustered at the village-level. Columns (2) and (4) additionally include an interaction between the treatment indicator and an indicator for assignment to the spiritual intervention (cross-randomization), as well as the cross-randomization indicator itself.

Table A16: Components of Social Cohesion

| Panel A.                    | Trust               | How much do you trust [...]? |                  |                    |                   |                    |                   |                   | Trust              |
|-----------------------------|---------------------|------------------------------|------------------|--------------------|-------------------|--------------------|-------------------|-------------------|--------------------|
|                             | Index               | Immediate family             | Extended family  | Neighbors          | Same religion     | Different religion | Religious leader  | Strangers         | Game               |
|                             | (1)                 | (2)                          | (3)              | (4)                | (5)               | (6)                | (7)               | (8)               | (9)                |
| T1: Group Skills            | 0.211***<br>(0.073) | 0.032**<br>(0.013)           | 0.030<br>(0.022) | 0.050**<br>(0.024) | -0.006<br>(0.018) | 0.084**<br>(0.034) | -0.008<br>(0.013) | 0.037<br>(0.030)  | 0.099<br>(0.078)   |
| T2: Group Activity          | 0.179**<br>(0.075)  | 0.014<br>(0.010)             | 0.007<br>(0.019) | 0.002<br>(0.024)   | -0.022<br>(0.015) | 0.054<br>(0.039)   | -0.016<br>(0.010) | 0.056*<br>(0.030) | 0.171**<br>(0.084) |
| Mean of outcome at baseline | 0.000               | 0.979                        | 0.909            | 0.840              | 0.931             | 0.652              | 0.962             | 0.292             | 0.000              |
| p-value: T1 = T2            | [0.679]             | [0.041]                      | [0.235]          | [0.053]            | [0.344]           | [0.391]            | [0.458]           | [0.567]           | [0.384]            |
| Clusters (villages)         | 90                  | 90                           | 90               | 90                 | 90                | 90                 | 90                | 90                | 90                 |
| Observations                | 1075                | 1075                         | 1075             | 1075               | 1075              | 1075               | 1075              | 1075              | 1075               |

| Panel B.                       | Closeness          | How close do you feel to [...]? |                   |                   |                     |                    |                   |                     |                    |
|--------------------------------|--------------------|---------------------------------|-------------------|-------------------|---------------------|--------------------|-------------------|---------------------|--------------------|
|                                | Index              | Nuclear family                  | Extended family   | Neighbor          | Same religion       | Different religion | Same ethnicity    | Different ethnicity | Strangers          |
|                                | (1)                | (2)                             | (3)               | (4)               | (5)                 | (6)                | (7)               | (8)                 | (9)                |
| T1: Group Skills               | 0.205**<br>(0.083) | -0.013<br>(0.022)               | 0.016<br>(0.040)  | 0.001<br>(0.036)  | 0.169***<br>(0.042) | 0.075**<br>(0.033) | 0.072*<br>(0.037) | 0.049**<br>(0.022)  | 0.037**<br>(0.014) |
| T2: Group Activity             | 0.076<br>(0.084)   | -0.001<br>(0.020)               | -0.014<br>(0.045) | -0.054<br>(0.033) | 0.098***<br>(0.037) | 0.074**<br>(0.030) | 0.057<br>(0.035)  | 0.005<br>(0.022)    | 0.002<br>(0.014)   |
| Mean of outcome at baseline    | 0.000              | 0.914                           | 0.581             | 0.313             | 0.376               | 0.190              | 0.388             | 0.100               | 0.046              |
| p-value: T1 = T2               | [0.123]            | [0.609]                         | [0.442]           | [0.145]           | [0.086]             | [0.979]            | [0.712]           | [0.035]             | [0.033]            |
| BL of outcome                  | Yes                | Yes                             | Yes               | Yes               | Yes                 | Yes                | Yes               | Yes                 | Yes                |
| Strata fixed effects           | Yes                | Yes                             | Yes               | Yes               | Yes                 | Yes                | Yes               | Yes                 | Yes                |
| Enumerator fixed effects       | Yes                | Yes                             | Yes               | Yes               | Yes                 | Yes                | Yes               | Yes                 | Yes                |
| SEs clustered at village-level | Yes                | Yes                             | Yes               | Yes               | Yes                 | Yes                | Yes               | Yes                 | Yes                |
| Clusters (villages)            | 90                 | 90                              | 90                | 90                | 90                  | 90                 | 90                | 90                  | 90                 |
| Observations                   | 1075               | 1075                            | 1075              | 1075              | 1061                | 1061               | 1075              | 1075                | 1075               |

**Notes:** The table reports OLS estimates of outcomes on treatment indicators, controlling for strata and enumerator fixed effects. Standard errors are clustered at the village-level. Column (1) in both panels replicates Columns (3) and (4), respectively, of Table 1. The remaining columns in each panel report the components for the indices. Outcomes in Panel A include seven indicators of how much the respondent trusts various entities—immediate family, extended family, neighbors, people of the same religion, people of different religions, religious leaders, and strangers—and a standardized measure of the amount (in Myanmar kyat) the respondent sent to another player in a standard trust game. In the game, respondents were endowed with 2,000 MK, chose an amount to send (which was tripled), and the recipient decided how much to return. Outcomes in Panel B include eight indicators of how close the respondent feels to their nuclear family, extended family, neighbors, people of the same religion, people of different religions, people of the same ethnicity, people of different ethnicities, and strangers.

Table A17: Robustness of Trust Impacts

|                                | Trust Index without Trust Game | Trust Game           |                      |
|--------------------------------|--------------------------------|----------------------|----------------------|
|                                |                                | Normal               | Incentivized         |
|                                | (1)                            | (2)                  | (3)                  |
| T1: Group Skills               | 0.186**<br>(0.082)             | 39.324<br>(30.951)   | -68.280*<br>(40.535) |
| T2: Group Activity             | 0.086<br>(0.077)               | 68.294**<br>(33.618) | 74.442*<br>(41.954)  |
| Mean of outcome at baseline    | 0.000                          | 736.645              | 1003.253             |
| p-value: T1 = T2               | [0.190]                        | [0.384]              | [0.000]              |
| BL of outcome                  | Yes                            | Yes                  | Yes                  |
| Strata fixed effects           | Yes                            | Yes                  | Yes                  |
| Enumerator fixed effects       | Yes                            | Yes                  | Yes                  |
| SEs clustered at village-level | Yes                            | Yes                  | Yes                  |
| Clusters (villages)            | 90                             | 90                   | 90                   |
| Observations                   | 1075                           | 1075                 | 1075                 |

**Notes:** The table reports OLS estimates of outcomes on treatment indicators, controlling for strata and enumerator fixed effects. Standard errors are clustered at the village-level. Column (1) reports results for a Trust Index similar to that in Column (3) of Table 1, but excluding the Trust Game component (i.e., constructed from Columns (2)–(8) of Panel A in Table A16). Column (2) reports the amount (in Myanmar kyat) out of a 2,000 MK endowment that the respondent chose to send to another player in a standard trust game; the amount sent was tripled, and the recipient decided how much to return. Column (3) examines experimenter demand effects by repeating the trust game with an added prompt informing respondents that they would be doing the researchers a favor if they sent more than they normally would.

Table A18: Other Responsible Entities

|                                | Who is responsible for providing public services? |                   |                     |                   |                  |
|--------------------------------|---------------------------------------------------|-------------------|---------------------|-------------------|------------------|
|                                | Neighborhood watch                                | Central gov.      | NGOs                | Int. orgs.        | Relig. group     |
|                                | (1)                                               | (2)               | (3)                 | (4)               | (5)              |
| T1: Group Skills               | -0.012<br>(0.073)                                 | -0.082<br>(0.138) | -0.156<br>(0.095)   | -0.101<br>(0.088) | 0.020<br>(0.069) |
| T2: Group Activity             | 0.006<br>(0.046)                                  | -0.062<br>(0.121) | -0.178**<br>(0.089) | -0.092<br>(0.106) | 0.016<br>(0.075) |
| Mean of outcome in controls    | 0.013                                             | 0.048             | 0.111               | 0.151             | 0.036            |
| p-value: T1 = T2               | [0.808]                                           | [0.863]           | [0.824]             | [0.891]           | [0.957]          |
| BL of outcome                  | No                                                | No                | No                  | No                | No               |
| Strata fixed effects           | Yes                                               | Yes               | Yes                 | Yes               | Yes              |
| Enumerator fixed effects       | Yes                                               | Yes               | Yes                 | Yes               | Yes              |
| SEs clustered at village-level | Yes                                               | Yes               | Yes                 | Yes               | Yes              |
| Clusters (villages)            | 90                                                | 90                | 90                  | 90                | 90               |
| Observations                   | 1019                                              | 1019              | 1019                | 1019              | 986              |

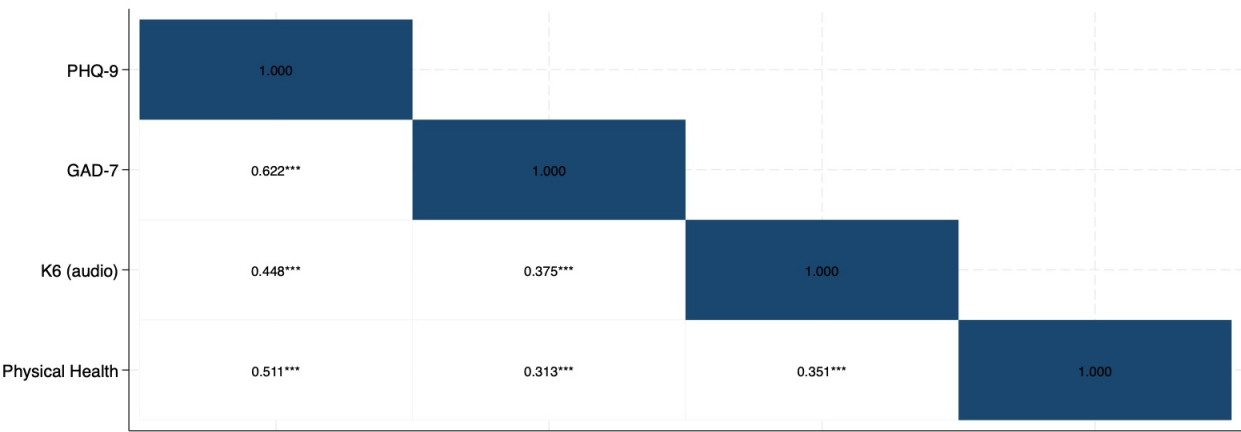
**Notes:** The table presents results from OLS regressions of various variables measuring respondents' attribution of community services provision on treatment variables and strata and enumerator fixed effects. Standard errors are clustered at the village-level. For a range of public services—health care, education, road construction, dispute resolution, security, and helping the poor—respondents were asked to indicate whether each of the five entities listed in the column headers is responsible for providing the community service. The outcomes are standardized indices, constructed separately for each entity across all community services in two steps: first, we compute the average of the underlying binary indicators; second, we standardize the resulting average.

Table A19: Components of Volunteering

|                                | Health committee   | Development committee | Flood response team |
|--------------------------------|--------------------|-----------------------|---------------------|
|                                | (1)                | (2)                   | (3)                 |
| T1: Group Skills               | -0.034<br>(0.041)  | -0.066**<br>(0.029)   | -0.044<br>(0.043)   |
| T2: Group Activity             | -0.072*<br>(0.042) | -0.020<br>(0.030)     | 0.015<br>(0.036)    |
| Mean of outcome at baseline    | 0.332              | 0.545                 | 0.187               |
| p-value: T1 = T2               | [0.285]            | [0.151]               | [0.128]             |
| BL of outcome                  | Yes                | Yes                   | Yes                 |
| Strata fixed effects           | Yes                | Yes                   | Yes                 |
| Enumerator fixed effects       | Yes                | Yes                   | Yes                 |
| SEs clustered at village-level | Yes                | Yes                   | Yes                 |
| Clusters (villages)            | 90                 | 90                    | 90                  |
| Observations                   | 1070               | 1074                  | 1064                |

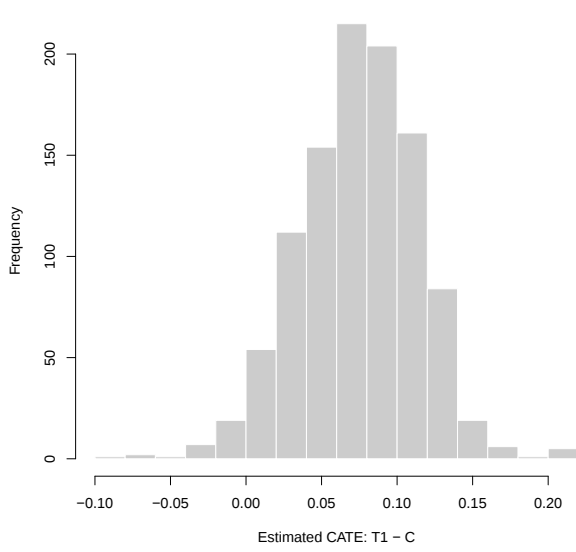
**Notes:** The table presents results from OLS regressions of various variables measuring respondents' volunteering activities on treatment variables and strata and enumerator fixed effects. Standard errors are clustered at the village-level. The outcomes in Columns (1), (2), and (3) are binary indicators for whether the respondent is voluntarily part of a (i) health committee, (ii) development committee, or (iii) flood response team, respectively.

Figure A1: Correlations of Different Mental Health Measures

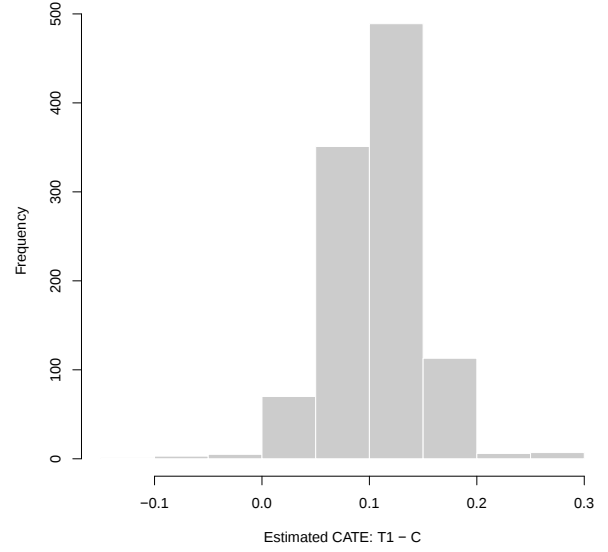


**Notes:** The figure presents raw correlations among four measures of mental disorder in the control group at endline: the PHQ-9, GAD-7, and K6 scales, administered via audio, and a physical health index.

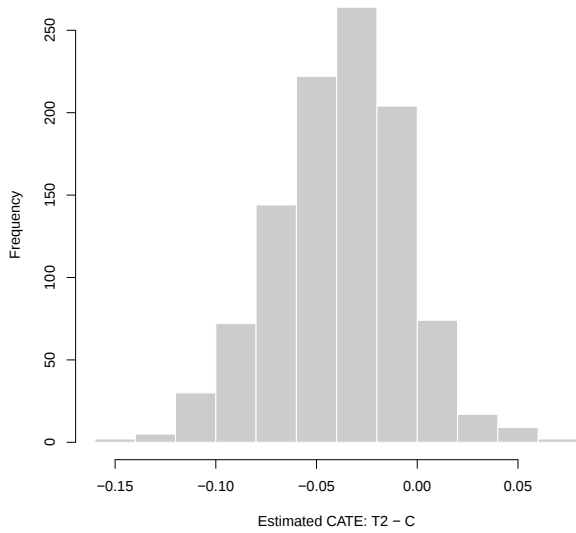
Figure A2: Causal Forests: out-of-bag CATE



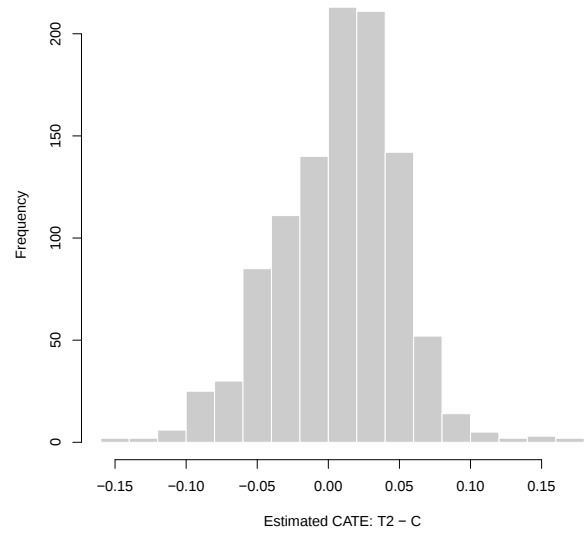
(a) PHQ-9



(b) GAD-7



(c) PHQ-9



(d) GAD-7

**Notes:** The figure presents the distributions of conditional average treatment effects estimated using multi-arm causal forests. Panels (a) and (c) show treatment effects of the group skills and group activity interventions, respectively, on the standardized PHQ-9 score, which measures depression. Panels (b) and (d) show corresponding effects on the standardized GAD-7 score, which measures anxiety.

### A.3 Explanation of Pre-analysis Plans

**Overview.** As noted in the acknowledgements, the project was preregistered under two pre-analysis plans (PAPs): AEARCTR-0012552 and AEARCTR-0015406. As footnote 9 explains, we originally intended to produce two separate papers, each with its own PAP, but subsequently decided to consolidate the analyses into a single paper. In the PAPs we call the “Group Skills” intervention “Group Counselling” and the “Group Activity” intervention “Group Solidarity.”

**AEARCTR-0012552.** This PAP covers the mental health and social cohesion results presented in the paper. Differences between the PAP and the presented analysis are:

- The PAP states that we will explore heterogeneity along four dimensions only (gender, religion, education, and conflict exposure). In the paper, we only report heterogeneity by conflict explicitly (Table A12), and instead add Causal Forests (Figure A2) to explore further heterogeneities.
- Hypothesis 4 pre-registers the spiritual intervention as a moderator of treatment effects on mental health and social cohesion. We report results for mental health (Table A15), but omit the social cohesion results because the cross-randomization produced no differential impact on those outcomes.
- Hypothesis 5 states that we explore impacts on economic well-being. We do not report these (null) results as the endline data collection happened shortly after the intervention (and because we find null effects on mental health).
- The PAP additionally mentions several further outcomes intended to help address concerns related to stigma (above and beyond those included in the main paper), including a list experiment, alcohol, tobacco, and drug consumption, self-efficacy (optimistic self-belief), and life satisfaction. Unfortunately, the list experiment was implemented incorrectly and therefore cannot be reliably analyzed. Questions on smoking and drug consumption were removed from the endline survey because of very low response rates at baseline. Finally, we find no treatment effects on alcohol consumption, self-efficacy, or life satisfaction; these results are available upon request.

- We do not report the 2SLS estimation strategy. This is because attendance (see Figure 1) is very high, and thus our instruments are extremely strong predictors of compliance (to the point where ITT and IV estimates are nearly identical).

**AEARCTR-0015406.** This PAP covers the results on local responsibility. We had obtained additional funding from J-PAL to collect additional data on the impact of public service provision on preferences about governance in conflict-affected areas, and the PAP was written with this objective in mind. Unfortunately, due to the deteriorating security situation in Myanmar, we were unable to carry out this data collection. From this PAP, we can therefore only test pre-registered Hypothesis 1. Hypotheses 2-4 were contingent on the additional data collection.