Building resilience through psychological intervention

Michelle Escobar Carías, Victoria Baranov, Sonia Bhalotra,
Pietro Biroli and Joanna Maselko *

August 07, 2026

Abstract

Does past treatment for depression create resilience to future shocks? We study a randomized CBT-based intervention for prenatally depressed mothers in Pakistan and exploit variation in COVID-19 restriction stringency six years after childbirth. Treated women were protected against restriction-induced anxiety, depression, comorbidity, and functional impairment. Protection extended to spousal relationships, offsetting restriction-driven declines in relationship quality and increases in marital conflict. This resilience operates through psychological channels: the intervention builds locus of control and self-efficacy, strengthens family and partner support, and reduces how overwhelmed women feel when shocks arrive. The resulting psychological capital has returns well beyond the treated condition.

Keywords: Resilience, mental health, functioning, spousal conflict, restrictions
JEL: G18, I12, J12, J16

*Escobar Carías: Business School, Newcastle University (email: michelle.escobar@newcastle.ac.uk); Baranov: Department of Economics, University of Melbourne (email: victoria.baranov@unimelb.edu.au); Bhalotra: Department of Economics, University of Warwick (e-mail: sonia.bhalotra@warwick.ac.uk); Biroli: Department of Economics, University of Bologna (e-mail: pietro.biroli@unibo.it); Maselko: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina (email: jmaselko@unc.edu). We thank Abhirupa Das for outstanding research assistance. We gratefully acknowledge funding from the National Institutes of Health and the CAGE Centre at Warwick under ESRC Centre grant ES/Z504701/1. We thank attendees of various seminars and conferences for very helpful comments and feedback. Views expressed in this study are our own. The trial mentioned in this study was registered at ClinicalTrials.gov as NCT02111915 (THPP) and NCT02658994 (THPP+).

1 Introduction

A large body of evidence in economics examines how material resources such as cash transfers, insurance and credit access can buffer households against adverse shocks (Christian, Hensel and Roth, 2019; Lane, 2024; Premand and Stoeffler, 2022). The underlying logic is one of *financial resilience*: ensuring households have sufficient resources to smooth consumption when adversity strikes. But financial resilience may not be the only kind that matters. How individuals cope with shocks also depends on psychological capacities — stress regulation, self-efficacy, the ability to mobilize social support — that constitute *psychological resilience* (Alloy and Ahrens, 1987; Elmer and Stadtfeld, 2020). Depression, which affects over 330 million people worldwide and remains widely undertreated (Global Burden of Disease Collaborative Network, 2024; Ridley et al., 2020), erodes precisely these capacities. Thus, if left untreated, depression may amplify the damage of future shocks, turning transitory hardship into lasting harm.

In this paper we ask whether treating depression can build lasting psychological resilience. We study a Cognitive Behavioral Therapy (CBT)-based intervention delivered to prenatally depressed women in rural Pakistan, and exploit an aggregate shock, COVID-19, which struck five years after the main therapeutic component of the treatment had concluded. The interaction of randomized treatment assignment and plausibly exogenous variation in restriction stringency allows us to test whether treatment causally moderated the impact of a future unanticipated shock.

The intervention, the Thinking Healthy + Bachpan Care Program (TH-BCP), was a cluster-randomized trial that enrolled pregnant women screening positive for depression in 2014 (Maselko et al., 2020; Sikander et al., 2019*a*). Treatment was delivered by trained community volunteers using the core techniques of CBT-based interventions: identifying and reappraising negative thought patterns, behavioral activation, and structured problem-solving. Delivery by trained lay providers is a defining feature of this class of interventions and a key reason for their policy relevance: non-specialist delivery makes scaled implementation feasible where professional mental health care is scarce.

The study's sixth wave of longitudinal data collection was underway when the pandemic struck. Because data collection was fielded over 20 months, households were surveyed under different restriction regimes depending on their interview date, which had been scheduled in prior waves, before the pandemic began. We use this variation in restriction stringency, which we show is orthogonal to treatment assignment and other pre-determined household characteristics, as our source of adverse shocks. Using the Oxford COVID-19 Government Response Tracker (Hale et al., 2021), we measure stringency over the fourteen days preceding each household's survey date, aligning it with the two-week recall window of our depression measure.

We find that prior assignment to treatment provided significant protection against the psychological consequences of pandemic restrictions. Among control women, restrictions triggered a clinically meaningful escalation of symptoms: mild anxiety, moderate depression and comorbid anxiety and depression. Restrictions also increased the number of days women had difficulties carrying out daily activities, evidence that the psychological deterioration was severe enough to impair everyday functioning.[1] The intervention fully protected treated women against the negative psychological effects of pandemic restrictions. These results suggest that the coping skills acquired through treatment remained functional years after the conclusion of the psychoactive component.

We next explore the channels through which this protection operated. A broad consensus in psychology and neuroscience has converged on three core capacities underlying resilience to adverse shocks: the ability to appraise stressful situations as manageable, locus of control and self-efficacy, and access to social support (Kalisch, Müller and Tüscher, 2015; Masten, 2001; Southwick and Charney, 2012). We test whether the intervention built each of these, and whether it also operated through biological stress-response pathways. First, we examine whether the intervention changed how women experience stress, distinguishing between perceiving a situation as threatening and feeling overwhelmed by it. Restrictions increased both among control women, but the intervention buffered only the latter: treated women recognized the threat just as acutely as control women but were significantly less likely to feel overwhelmed as restrictions intensified. Second, we test whether the intervention increased internal locus of control and perceived self-efficacy. We find large treatment effects on both that persisted independently of restriction levels, suggesting the intervention equipped women with a stock of psychological capital they carried into the future. Third, we examine perceived social support and find that treated women reported significantly greater perceived support from family and partner, but not from friends, consistent with the intervention having had a focus on training in interpersonal skills within the household (and consistent with contact with friends being more limited under lockdown restrictions).

We further examine whether the intervention affected the body's biological stress-response system. Prolonged stress keeps this system, known as the hypothalamic-pituitary-adrenal (HPA) axis, chronically activated, and sustained activation of this kind is a well-established pathway linking hardship to worse mental and physical health over the life course (McEwen, 1998; Pariante and Lightman, 2008). Hair samples provide a cumulative record of this activation through concentrations of cortisol and dehydroepiandrosterone (DHEA), two hormones central to the stress response. In earlier work, we found that treated mothers had healthier levels of these hormones one year postpartum, closely resembling women who were never depressed (Baranov et al., 2022*a*). Six

[1] A number of studies have established that lockdown restrictions during the pandemic led to a deterioration of mental health, particularly for women (Adams-Prassl et al., 2022; Etheridge and Spantig, 2022).

years after the intervention, however, these differences have faded: we find no distinction in either hormone between treated and control women. This is consistent with the resilience we document operating primarily through psychological rather than physiological channels.

Importantly, the intervention did not shield women from adverse events: treated and control women were equally likely to know someone who contracted or died from COVID-19, and equally likely to experience permanent job loss in the family. Like insurance, the intervention did not prevent the shock from occurring; it protected women's psychological wellbeing in the face of it.

Lastly, we study whether the protective effects of the intervention extend beyond the treated individual. Previous research has documented that pandemic restrictions strained spousal relationships (Bhalotra et al., 2026) and harmed children's mental health (Racine et al., 2021) and learning (Betthäuser, Bach-Mortensen and Engzell, 2023). We find evidence consistent with those findings: restrictions significantly reduced spousal relationship quality and increased marital conflict among control women. The intervention fully offset both effects. We also investigated spillovers to child outcomes, finding suggestive evidence that children of treated mothers displayed better socio-emotional wellbeing, though these estimates are imprecise.

Taken together, our findings are consistent with a broader literature that characterizes resilience not as a fixed trait but as a dynamic, learnable process (McEwen, 1998; Southwick et al., 2014). The core protective mechanisms that this literature identifies are stress appraisal, locus of control, and the ability to mobilize social support. Our results demonstrate that these can be built through a structured intervention and remain functional when confronted with a shock the intervention was never designed to address.

Our paper contributes primarily to two strands of the literature. First, we advance the growing body of work on psychological interventions in low- and middle-income countries (Angelucci and Bennett, 2024; Baranov et al., 2020; Bhat et al., 2022; Barker et al., 2022). A central question in this literature is whether the benefits of such interventions last. Recent work has shown that they can: treatment effects on depression and economic outcomes have been documented years after therapy concludes (Baranov et al., 2020; Blattman et al., 2023). We show that the benefits are not only durable but *transferable*: they protect against a class of stressor entirely different from the one that motivated treatment. Durability within the treated condition is one thing; protection against a novel, unanticipated shock is another. The latter implies that these interventions build a form of psychological capital with returns that extend beyond the treated condition.

Second, we contribute to the literature on resilience to shocks. Earlier work in economics has focused on material buffers (cash transfers, insurance, credit, savings) rather than psychological ones (Aizer et al., 2024; Christian, Hensel and Roth, 2019; Hidrobo and Fernald, 2013). Meanwhile, a well-established framework in psychology and neuroscience, built on decades of observational and laboratory research, has converged on the core capacities underlying adaptive functioning (Kalisch,

Müller and Tüscher, 2015; Kalisch et al., 2017; Masten, 2001; McEwen, 1998; Southwick et al., 2014). Yet, this literature has lacked causal evidence that these capacities can be built by a scalable intervention, and a real-world test of whether they withstand adverse events that are exogenous to the individual. Our setting provides both, leveraging the fact that we have long follow-up of a randomized controlled trial, with data collection that included multiple indicators of resilience.[2]

The rest of the paper is organized as follows. Section 2 describes the intervention and how the COVID-19 pandemic unfolded in Pakistan. Section 3 describes the sample, the longitudinal dataset and our outcomes of interest. Section 4 outlines the empirical strategy. Section 5 presents the main results. Section 6 discusses mechanisms. Section 7 examines downstream effects on household members, and Section 8 concludes.

2 Background

2.1 The Intervention

The Thinking Healthy + Bachpan Care Program (TH-BCP) is a cluster-randomized controlled trial addressing prenatal depression in Kallar Syedan, a rural subdistrict of Rawalpindi, Punjab, Pakistan (Maselko et al., 2020). The area has a poverty rate of approximately 50%, a high prevalence of maternal depression, and little access to clinical mental health care (NIPS and ICF, 2013). We worked in forty village clusters, randomizing 20 to treatment and 20 to control, with clusters geographically separated to minimize contamination. Eligible participants were married women aged 18 or older, in their third trimester of pregnancy, and registered with a village-based Lady Health Worker (Sikander et al., 2019*b*).

Between October 2014 and February 2016, eligible women were screened for depression using the Urdu version of the Patient Health Questionnaire (PHQ-9), a nine-item instrument assessing the frequency of depressive symptoms over the preceding two weeks. Women scoring 10 or above screened positive for depression and were invited to participate (Sikander et al., 2019*a*). Of the 1,731 women screened, 572 (33%) were identified as depressed; after two declined, 570 were enrolled: 283 in treatment clusters and 287 in control clusters. An additional 584 non-depressed women who were pregnant at the same time were randomly selected as a reference group to track the evolution of maternal and child outcomes.

[2]Our paper also relates to the literature documenting the effects of the COVID-19 pandemic on mental health and domestic violence (Adams-Prassl et al., 2022; Brodeur et al., 2021; Etheridge and Spantig, 2022; Giuntella et al., 2021; Aksunger et al., 2023; Altindag, Erten and Keskin, 2022; Baranov et al., 2022*b*; Nichols et al., 2024; Agüero, 2021; Bhalotra et al., 2026; McNeil et al., 2023; Piquero et al., 2021).

Curriculum. TH-BCP is a low-intensity psychological intervention that adapts the WHO's CBT-based Thinking Healthy Programme for delivery by non-specialist, volunteer women from the participant's own community, rather than by trained health workers. Peers completed a five-day training workshop and a two-month supervised internship before beginning delivery, and received monthly group supervision throughout the trial (Turner et al., 2016).

The intervention consisted of three components. The core CBT-based component comprised ten individual and four group sessions from the third trimester through six months postpartum, focused on identifying negative thought patterns, behavioral activation, and structured problem-solving. From six months to three years after birth, booster sessions gradually shifted from reinforcing CBT content toward encouraging mother-infant interaction and play. From three to five years, sessions focused primarily on coaching mothers to engage in age-appropriate play activities to support their children's development. Throughout, group sessions provided an environment for women to share experiences and support one another. Both treatment and control groups received enhanced usual care, which consisted of informing participants about their depression status and ways to seek help during and after pregnancy. Online Appendix Figure 5 shows the distribution of session attendance among treated women.

2.2 COVID-19 Restrictions in Pakistan

Pakistan recorded its first COVID-19 cases in February 2020 and experienced five waves of infections and government-imposed restrictions between March 2020 and March 2022 (Ahmad et al., 2023). The first two waves preceded our survey period (Online Appendix Table 3). The year-six follow-up began in early 2021 and continued through mid-2022, coinciding with the third, fourth, and fifth pandemic waves. The third wave, beginning in February 2021, was the deadliest and prompted the most stringent combination of workplace closures, gathering restrictions, and stay-at-home orders observed during the study period. As it receded, restrictions were eased, then reimposed during the fourth wave (mid-2021) and again during the fifth (early 2022) wave.

Our data on restrictions comes from the Oxford COVID-19 Government Response Tracker (Ox-CGRT), which records daily, national-level data on eight categories of containment and closure policy for over 180 countries (Hale et al., 2021). Because national policies were not always representative of conditions in rural villages, we focus on the three restrictions most directly relevant to women's daily lives in this setting: workplace closures, stay-at-home requirements, and restrictions on gathering size. Online Appendix Figure 2 shows the variation in each.

From these three policies we construct two composite measures of restriction exposure. Our primary measure is the average number of mandatory restrictions in place during the 14 days preceding each woman's survey date. Each policy is dichotomized (1 if mandatory, 0 otherwise)

and the three are summed, yielding a daily composite ranging from 0 to 3, which we then average over the 14-day window. As a robustness check, we also construct an ordinal stringency measure that preserves the original OxCGRT scales, ranging from 0 to 10. Both measures vary widely across the rollout (Figure 2). Exposure was highest for women surveyed in early 2021 and during the resurgence of restrictions between September 2021 and March 2022, and fell to zero for those surveyed after April 2022, when all mandatory restrictions were lifted. Women faced an average of 2.02 mandatory restrictions in the two weeks before their survey. About 46 percent faced two mandatory restrictions, 30 percent faced all three, and 11 percent faced none (Figure 3).

3 Data

Study sample and follow-up. Our sample consists of depressed mothers in the treatment and control arms. Data collection on the mother-child dyads has taken place in eleven waves as of 2025 (3rd trimester of pregnancy, 3, 6, 12, 24, 36, 48, 72, 73, 84, 96 months postnatal), but our study focuses on year six (72 months postnatal) when Pakistan's third through fifth pandemic waves occurred, yielding an analytical sample of 431 mothers. The effectiveness of the intervention on maternal depression and child development at three years postnatal has been evaluated in previous studies (Maselko et al., 2020; Sevim et al., 2024). Detailed variable definitions are provided in the Data Appendix, and Online Appendix Table 1 reports summary statistics for the analytic sample.

Balance and Attrition. The experimental sample was slightly imbalanced at baseline (Table 1): women in the treatment arm were approximately 1.1 cm taller, had smaller waist circumference, and lived in larger households. A joint F-test shows balance of baseline characteristics ($p = 0.10$) at the start of treatment, albeit marginally so. By year six the sample is no longer balanced ($p = 0.06$).

Lost to follow-up at year six was 22.2% (Figure 1), primarily due to loss of the index child, which was balanced across arms. Attritors differed in nuclear family structure, grandmother presence, and whether the index child was a firstborn, and had somewhat higher depression and impaired-functioning scores, though none of these differences were statistically significant (Online Appendix Table 2). A joint test of attritor characteristics by treatment arm yields balance ($p = 0.60$). Our analyses account for imbalance by interacting all covariates with treatment.

3.1 Primary Outcomes

Mental health. We measure maternal mental health using the Patient Health Questionnaire (PHQ-9) for depression severity (range 0–27) (Kroenke, Spitzer and Williams, 2001) and the Generalized

Anxiety Disorder scale (GAD-7) for anxiety severity (range 0–21) (Spitzer et al., 2006). Higher scores indicate greater symptom burden. Online Appendix Figure 1 plots the distribution of both scores.

Daily functioning. We measure functional impairment using the WHO Disability Assessment Schedule (WHODAS 2.0), which asks respondents how many days in the past 30 they had difficulty carrying out their usual activities. This captures whether psychological deterioration translates into reduced capacity to carry daily activities.

3.2 Secondary Outcomes

Spousal Conflict. We measure relationship quality using the six-item Relationship with Father/Husband (RFH) questionnaire, which asks respondents about their husband's attentiveness to basic needs, provision of pocket money, emotional understanding, support during difficulties, behavior, and overall relationship quality. These items are aggregated into an RFH score. We separately measure marital conflict by asking women whether their marital relations have had problems in the last year.

Child wellbeing. We assess socio-emotional wellbeing using the Spence Children's Anxiety Scale (SCAS), which measures anxiety symptoms across domains including separation anxiety, social phobia, and generalized anxiety (Spence, 1998), and the Strengths and Difficulties Questionnaire (SDQ), which captures emotional symptoms, conduct problems, and peer difficulties (Goodman, 1997). Higher scores indicate greater difficulties on both measures. Academic outcomes are assessed using the Annual Status of Education Report (ASER) tool, a widely used direct assessment in South Asia that tests literacy (Urdu) and numeracy through one-on-one evaluation (ASER Pakistan, 2019).

3.3 Mechanisms

To assess the channels through which the intervention moderated the effects of restrictions, we measure three families of psychological resilience factors and one biological pathway (Kalisch, Müller and Tüscher, 2015; Masten, 2001).

Stress appraisal. We measure stress appraisal through the Stress Overload Scale (SOS), a ten-item questionnaire in which respondents rate subjective feelings and thoughts experienced over the prior week on a five-point Likert scale (Amirkhan, 2018). Using principal component analysis, we derive two subscale indices: threat appraisal, based on the Personal Vulnerability subscale (e.g., feeling like nothing was going right, like there was no escape), and stress overload, based on the Event

Load subscale (e.g., feeling swamped by responsibilities, like things kept piling up).

Self-efficacy and control. We construct an index of perceived control and self-efficacy as the first principal component of locus of control, items from the Generalized Self-Efficacy scale (GSE), and perceived control items from the Perceived Stress Scale (PSS).

Biological stress response. We measure cumulative hypothalamic-pituitary-adrenal (HPA) axis activity using cortisol and dehydroepiandrosterone (DHEA) concentrations in hair samples. Hair absorbs these hormones as it grows, providing a biological record of the body's stress response over the preceding months and allowing us to test whether the intervention conferred protection through physiological pathways.

Perceived social support. We use the Multidimensional Scale of Perceived Social Support (MSPSS), a 12-item instrument measuring perceived adequacy of support from three sources: family, friends, and significant other (Zimet et al., 1990). We report each subscale separately, which allows us to examine whether the intervention differentially strengthened support from sources within versus outside the household.

COVID-related Adverse Events. Lastly, we asked respondents about their experience of health and economic shocks due to the pandemic and about their family's economic prospects for the following year. We use these to estimate any protective effects of the intervention against these shocks.

4 Empirical Strategy

4.1 Empirical Specification

Our identification strategy exploits two sources of variation: randomized assignment to CBT-based treatment at baseline and quasi-random variation in the stringency of COVID-19 restrictions at the time of each woman's year-six survey. The interaction of these two variables is our parameter of interest, capturing the extent to which prior treatment moderated the effect of restrictions on outcomes. The empirical specification is as follows:

$$Y_{ict} = \beta_0 + \beta_1 Treat_c + \beta_2 Restrictions_t + \beta_3 (Treat_c * Restrictions_t) + \Gamma' X_{ict} + \varepsilon_{ict} \quad (1)$$

where Y_{ict} is the outcome for woman i in cluster c surveyed at time t. $Treat_c$ is an indicator equal to one if cluster c was assigned to the treatment arm and zero if assigned to control. $Restrictions_t$

measures the demeaned average number of mandatory restrictions in place during the 14 days preceding each woman's survey date, as described in Section 2.2. Since the three policies used to construct this measure were set at the national level, $Restrictions_t$ varies only over time; cross-sectional variation in exposure arises solely from differences in survey dates across participants.

Because $Restrictions_t$ is demeaned, the coefficients have the following interpretation: β_1 captures the treatment-control difference at the mean level of restrictions observed during the survey period. β_2 captures the effect of one additional mandatory restriction on outcomes for women in the control arm. Our parameter of interest, β_3, captures the differential effect of one additional restriction for treated relative to control women. A negative β_3 on adverse outcomes such as depression or anxiety indicates that prior treatment buffered the psychological impact of restrictions: treated women's outcomes deteriorated less per additional restriction than those of their control counterparts.

The vector X_{ict} includes baseline and contemporaneous controls, all demeaned and interacted with the treatment indicator following Baranov et al. (2020). This approach accounts for imperfect balance between the baseline and year-six samples and allows covariates to have differential associations with outcomes across arms. Baseline controls include household structure (an indicator for joint or extended families, number of adults, number of people per room), socioeconomic status (a standardized wealth index, maternal and paternal education), maternal anthropometrics (height and waist circumference), and pre-treatment mental health (PHQ-9, WHODAS, and PSS scores). Contemporaneous controls include enumerator fixed effects, month of survey, and its square. Because the last two are interacted with treatment like all other controls, they absorb not only time trends in outcomes over the survey rollout but also differential trends across arms, ensuring that β_3 is identified from wave-driven variation in restrictions rather than from differences in outcome trajectories. We also include union council and baseline enumerator fixed effects throughout. Standard errors are clustered at the village cluster, the unit of randomization. Because we test interaction effects for multiple outcome families, we report Romano-Wolf stepdown-adjusted p-values below the standard errors.

4.2 Identifying Assumptions

The causal interpretation of our three coefficients rests on the following assumptions: for β_1, identification requires that treatment was randomly assigned, which holds by design. For β_2, identification requires that the timing and intensity of restrictions are uncorrelated with other determinants of outcomes. For β_3, the assumption is more specific but also weaker: that no time-varying factor coincided with periods of high restrictions and differentially affected treated and control women. We discuss each in turn.

Exogeneity of restriction exposure. The key design feature supporting identification of β_2 is that each participant's year-six survey was scheduled within a 45-day window defined by the sixth birthday of the index child, beginning 15 days before and ending 30 days after. A woman whose child was born in March 2015 was surveyed during Pakistan's third pandemic wave; a woman whose child was born in September 2015 was surveyed under different restriction conditions. Neither woman chose her survey window. Because birth dates were determined years before the pandemic, variation in restriction exposure across participants is driven by the research protocol rather than by selection. In practice, 75% of women were surveyed within this window.[3] We further verify that restriction exposure is independent of maternal characteristics: regressing $Restrictions_t$ on the full set of baseline controls, only family structure and maternal education are significantly correlated with exposure, and both are included as controls in our primary specification.[4]

Identification of β_3. Having established that restriction exposure is plausibly unrelated to maternal characteristics, the remaining concern for β_3 is that some time-varying factor coincided with periods of high restrictions and differentially affected treated and control women. If, for instance, high-restriction periods coincided with seasonal economic hardship that hit some localities harder than others, and treated and control villages differ in their exposure to such local shocks, the interaction could reflect this confound rather than the moderating effect of treatment. We address this concern in two ways. First, because our specification includes the full set of baseline characteristics interacted with treatment, identification does not require that restriction timing is independent of time-invariant observable maternal characteristics. It requires only the weaker condition that restriction timing is independent of *trends* in these characteristics across arms. We verify this in Online Appendix Table 8: baseline predictors are balanced across arms at each level of restrictions. Second, in Section 5.2, we provide further evidence supporting this assumption, including placebo tests and controls for time-varying weather conditions.

[3] Online Appendix Tables 4 and 5 report compliance rates. Online Appendix Table 6 shows that deviations from the scheduled window are unrelated to baseline mental health; the only significant predictor of an out-of-window survey is the number of restrictions in effect, and in the direction opposite to selection: restrictions positively predict compliance, likely because participants were more often at home when enumerators arrived. Even among the 25% surveyed outside the window, restriction exposure is unrelated to observable baseline traits.

[4] Online Appendix Table 7 reports these results.

5 Results

5.1 Main Results: Psychological Resilience

If the intervention builds a durable stock of psychological capital that protects against future shocks, we would expect the treatment-control difference to widen as restrictions intensified. Table 2 presents estimates of the three coefficients that allow us to test this: the treatment effect at mean restrictions (β_1), the cost of each additional mandatory restriction on untreated women (β_2), and the extent to which prior treatment offsets that cost (β_3).

Among control women, each additional mandatory restriction increased anxiety scores by 1.197 points ($p < 0.01$), equivalent to 22% of the control mean, and increased the number of days with functional difficulties by 1.275 ($p < 0.10$), or 14% of the mean. The effect on depressive symptoms is positive but imprecisely estimated. Prior assignment to the intervention substantially offset these costs. For each additional mandatory restriction, treated women reported anxiety scores 1.209 points lower ($p < 0.05$), depressive symptoms 1.448 points lower ($p < 0.10$), and 2.979 fewer days with functional difficulties ($p < 0.05$) relative to their control counterparts, representing reductions of 22.4%, 18.7%, and 31.9% relative to the control means.[5]

Table 3 examines these effects at clinical severity thresholds. Restrictions significantly increased the prevalence of mild anxiety by 9.5 percentage points ($p < 0.05$), moderate depression by 10.2 percentage points ($p < 0.05$), and co-occurring anxiety and depression by 12.4 percentage points ($p < 0.01$). This pattern is consistent with anxiety serving as a gateway that, under sustained stress, cascaded into depression and comorbidity (Wittchen et al., 2000; Jacobson and Newman, 2017).

The intervention interrupted this cascade at every stage, significantly reducing the prevalence of mild anxiety ($\beta_3 = -0.152$, $p < 0.01$), moderate depression ($\beta_3 = -0.140$, $p < 0.05$), and comorbid anxiety and depression ($\beta_3 = -0.183$, $p < 0.01$).[6]

The durable treatment effect (β_1) is significant on the PHQ-9 (-0.958, $p < 0.05$) and on days with functional difficulties (-1.605, $p < 0.05$), consistent with an intervention that directly targeted prenatal depression. The interaction effects are somewhat larger, though not statistically different in magnitude. This pattern is consistent with the nature of the shock: at moderate levels, restrictions may not have been uniformly harmful, but at higher levels the consequences — job loss, disrupted

[5] Online Appendix Figures 3 and 4 plot the interaction coefficients for each GAD-7 and PHQ-9 item, demonstrating that the protective effects are not driven by any single symptom but reflect consistent reductions across the full range of anxiety and depression symptoms.

[6] In Online Appendix Table 11, we decompose the results by type of restriction. The protective effects on mental health are strongest for workplace closures, suggesting that the economic and emotional strain of restrictions on household members' ability to work was the primary channel through which restrictions harmed mental health.

social networks, prolonged confinement — dominate, and the coping skills that treatment built become essential. Resilience, in this view, is not a trait observable under normal conditions but a capacity that becomes visible when tested (Kalisch, Müller and Tüscher, 2015; Masten, 2001).

5.2 Robustness Checks and Placebo Tests

We next assess the sensitivity of our main findings to alternative specifications and potential confounds.

Alternative restriction measures. Our primary measure captures the average number of mandatory restrictions over a 14-day window constructed from three locally relevant policies. A natural concern is that our results depend on this particular operationalization. In Online Appendix Table 12, we replace it with four alternatives: the ordinal stringency of the same three restrictions, the Oxford stringency index based on all nine OxCGRT policies, the number of confirmed COVID-19 cases, and the number of confirmed deaths. All measures are standardized. Results using the ordinal stringency of the three locally relevant restrictions are highly consistent with our primary estimates. The broader Oxford index, which includes policies less relevant to rural settings such as school closures and international travel restrictions, yields estimates of the same sign but roughly half the magnitude, underscoring the importance of adapting restriction measures to the local context. Estimates using case counts and deaths are weaker, suggesting that the protective effects operate through the disruption caused by restrictions themselves rather than through fear of illness or bereavement. Moreover, actual infection and death rates are outcomes that depend upon potentially endogenous individual behaviours and environmental factors like crowding and poverty. As such, we rely upon measures of restrictions.

Additional controls and multiple hypothesis testing. Because Pakistan's pandemic waves coincided with seasonal changes, a remaining concern is that time-varying factors such as weather could co-vary with restriction intensity and differentially affect the two arms. In Online Appendix Table 13, we show that our results are robust to controlling for weather conditions (temperature and rainfall) and their interaction with treatment. While precision is reduced, our conclusions are qualitatively unchanged. Results also survive the Romano-Wolf stepdown procedure: all three interaction coefficients in Table 2 remain significant at the 10% level, and the interaction effect on comorbid anxiety and depression in Table 3 remains robust.

Placebo restrictions. Finally, we test whether the survey rollout itself, independent of restrictions, could generate the patterns we observe. We transpose the observed restriction series onto pre-pandemic survey waves (years three and four post-birth): each woman's actual survey date in these earlier waves is retained, but the restriction value she receives is drawn from the pandemic period

based on the corresponding calendar date. This preserves the rollout structure while breaking the link to the actual pandemic. In Online Appendix Table 14, neither the restriction coefficient nor its interaction with treatment is significant in any pre-pandemic wave, individually or when pooled, confirming that our results reflect actual pandemic restrictions rather than artifacts of survey timing.

6 Mechanisms

The results in Section 5 establish that prior assignment to treatment protected women against the mental health and functioning consequences of COVID-19 restrictions. We now ask: through what channels did this protection operate?

A growing body of research in psychology and neuroscience has sought to identify the mechanisms through which individuals maintain mental health in the face of shocks. This literature, reviewed in Southwick and Charney (2012) and Masten (2001), converges on three core protective factors: (i) the ability to reappraise stressful situations in ways that reduce their emotional impact, (ii) a sense of personal control over one's circumstances, and (iii) access to social support. While these capacities have innate components, evidence suggests they are also malleable and can be strengthened through intervention (Kalisch et al., 2017; Roberts, Walton and Viechtbauer, 2006; Bleidorn et al., 2022). In what follows, we investigate whether treatment operated through each of these channels, whether it conferred protection through biological stress-response pathways, and whether it reduced women's exposure to shocks or whether it only buffered the psychological consequences of shocks.

6.1 Stress Appraisal

A central finding in the resilience literature is that how individuals interpret a stressful situation matters as much as the stressor itself. Psychologists refer to this interpretation as *cognitive appraisal*: the process by which a person evaluates whether a situation is threatening, manageable, or overwhelming. A substantial body of evidence identifies the tendency to appraise difficult situations as manageable rather than catastrophic as a key mechanism for maintaining mental health under stress, and shows that successfully managing one stressor builds the capacity to manage the next (Kalisch, Müller and Tüscher, 2015; Kalisch et al., 2017; Zeier, Meine and Wessa, 2022).

CBT-based interventions directly target this mechanism. The TH-BCP curriculum trains women to recognize patterns of negative thinking, evaluate whether those thoughts are accurate, and generate alternative interpretations of their circumstances. If this training produced lasting changes in how women interpret stressful situations, we would expect treated women to show a less escalating stress response as restrictions intensified, even though the shock was entirely different from the condition

for which they were treated.

Columns (1) and (2) of Table 4 report results for two dimensions of the stress response. Threat appraisal measures how negatively individuals evaluate their situation (e.g., feeling that the odds are against them, that there is no escape). Stress overload captures the sense of being overwhelmed by demands (e.g., feeling swamped by responsibilities, that things keep piling up). Restrictions significantly worsened both for control women, increasing threat appraisal by 0.181 standard deviations ($p < 0.05$) and stress overload by 0.125 standard deviations ($p < 0.10$) per additional mandatory restriction. Treatment actively buffered the latter: for each additional restriction, the increase in stress overload was 0.259 standard deviations smaller for treated than for control women ($p < 0.05$). The interaction coefficient on threat appraisal is negative but imprecisely estimated.

This distinction is informative. Treated women recognized the threat posed by restrictions just as control women did, but were significantly less likely to feel overwhelmed as restrictions became more stringent. The intervention changed how women processed shocks, not whether they perceived them, consistent with the framework proposed by Kalisch, Müller and Tüscher (2015). Online Appendix Table 15 reports item-level results for the individual SOS items underlying these indices.

6.2 Locus of Control and Self-efficacy

A second mechanism identified in the resilience literature is *perceived control*: the belief that one can influence one's own outcomes, even when external circumstances are difficult. The classic finding is *learned helplessness*: when individuals repeatedly face situations they cannot control, they eventually stop trying, even when change becomes possible (Seligman, 1972). The converse is well established: perceived control reduces the body's stress response and prevents emotional deterioration, and it predicts coping independently of how much actual control individuals have over the situation (Meier et al., 2026, 2025).

CBT-based interventions address this mechanism by teaching structured problem-solving and by encouraging women to re-engage with activities they had abandoned during their depression. These techniques rebuild a sense of agency: the feeling that one's actions matter and that difficult situations can be managed. If this component of treatment produced lasting effects, we would expect treated women to report greater personal control at the time of the year-six survey.

Column (3) of Table 4 reports results for an index of perceived control and self-efficacy. The treatment effect is large and significant: treated women report 0.177 standard deviations greater perceived control and self-efficacy ($p < 0.05$). This effect is constant across restriction levels: the interaction term is small and insignificant. That the intervention shifted locus of control and self-efficacy six years after baseline is noteworthy given that these constructs have traditionally been considered relatively stable personality traits, though recent evidence suggests they are mutable

over the life course and can be shifted by important life events (Roberts, Walton and Viechtbauer, 2006; Bleidorn et al., 2022; Cobb-Clark and Schurer, 2013; Stillman and Velamuri, 2025).

Unlike stress appraisal, this effect did not depend on restriction severity. The intervention built a sense of control and self-efficacy that persisted regardless of how stringent restrictions were. This standing psychological resource may help explain why treated women were buffered against stress overload: when restrictions tightened, women who already believed they could manage difficulties were less likely to feel overwhelmed by them. Online Appendix Table 16 reports item-level results for the items underlying this index.

6.3 Biological Stress Response

The mechanisms examined so far are psychological. The intervention may also have altered the body's physiological response to stress. The hypothalamic-pituitary-adrenal (HPA) axis is the primary stress-response system, releasing cortisol, which mobilizes energy when a person perceives a threat, and DHEA, which counterbalances the effects of prolonged cortisol exposure. Dysregulation of this system is among the most consistent biological correlates of major depression (Pariante and Lightman, 2008), and it can run in either direction. Cortisol may remain elevated when the stress response fails to switch off, or decline when the system has been chronically overtaxed (McEwen, 1998; Miller, Chen and Zhou, 2007).

Baranov et al. (2022*a*) find that our intervention affected HPA-axis function in the short run, with treated women showing lower cortisol and higher DHEA one year after it began. If those changes persisted, they could plausibly explain differences in the psychological response to pandemic restrictions. We therefore estimate treatment and interaction effects on cortisol and DHEA in hair samples at the six-year wave. [7]

Columns (4) and (5) of Table 4 report the results. Six years after the intervention, cortisol and DHEA are indistinguishable between treated and control women, both in the mean and the distribution (Online Appendix Figure 6). Restrictions themselves move hair cortisol ($\beta_2 = -0.263$, $p < 0.01$), thus showing that the hair measures capture real hormonal variation in response to the stressor. The interaction effects are null, which suggests that the body's hormonal stress-response is largely automatic, but whether that response becomes overwhelming anxiety or manageable stress depends on the cognitive processing employed by the individual. CBT-based interventions target precisely this layer. While we find no evidence that the intervention's protective effects operated

[7] Unlike blood or saliva, which capture hormone levels at a single point in time, hair absorbs these hormones as it grows, providing a biological record of cumulative HPA axis activity over the preceding months (Stalder et al., 2017). This makes hair-based measures well suited to capturing the sustained hormonal changes associated with months of pandemic restrictions.

through the hormonal stress-response channels we measure, we cannot rule out all physiological pathways. These results are however, consistent with the resilience we document operating primarily through psychological rather than physiological channels.

6.4 Perceived Social Support

The final core resilience factor identified in the psychology literature is social support. Emotional and practical support from family and community is among the strongest and most consistently documented predictors of positive outcomes following trauma or chronic stress, across populations, stressor types, and income levels (Southwick and Charney, 2012; Masten, 2001; Lund et al., 2010).

The TH-BCP intervention targeted this mechanism through individual sessions focused on communication with family members and the spousal relationship, group sessions that provided space for women to share experiences and support one another, and a peer delivery model that embedded the therapeutic relationship within the woman's own community. If these components produced lasting effects on women's capacity to maintain social bonds, we would expect treated women to report higher perceived support, particularly from family and partner.

Columns (1) through (3) of Table 5 report results for the three subscales of the Multidimensional Scale of Perceived Social Support (MSPSS). Treated women report perceived family support that is 0.142 standard deviations higher ($p < 0.10$) and partner support that is 0.146 standard deviations higher ($p < 0.05$) than control women, but there is no significant difference in support from friends. While restrictions reduced friend support among control women by 0.269 standard deviations per additional mandatory restriction ($p < 0.10$), consistent with the disruption of social gatherings, the intervention did not differentially protect the friend channel. Instead, treated women maintained stronger support from the relationships they could access under confinement: spouse and family.[8]

6.5 Exposure to Adverse Events

A central feature of the interpretation advanced in this paper is that the intervention reduced the *damage* of restrictions conditional on exposure, rather than reducing exposure itself. If treated women simply experienced fewer adverse events during high-restriction periods, because their

[8]This result helps distinguish between two potential explanations for the social support effects. If the gains in perceived support were driven by the social contact inherent in the peer delivery model, the relationship with the peer counselor and with other women in group sessions, we would expect the friend subscale to show the largest treatment effects. Instead, the effects are concentrated in family and partner support, the channels most directly targeted by the therapeutic content. In Online Appendix Table 17, we further show that higher session attendance is positively and significantly correlated with perceived social support from family and significant other, but not from friends, reinforcing the interpretation that therapeutic content operating within close relationships, rather than social contact with peers, drives the support effects.

households were wealthier or better connected, for instance, the interaction effects in Table 2 could reflect reduced exposure rather than enhanced coping, and the policy implications would be fundamentally different.

We test this directly in Table 7. Using self-reported health and economic shocks linked to the pandemic as dependent variables, we estimate the same interaction specification. Assignment to treatment did not reduce women's likelihood of contracting the virus, experiencing bereavement, or suffering economic hardship during high-stringency periods. It also did not increase their likelihood of receiving financial assistance from the government. The interaction terms are small and statistically insignificant. Thus the intervention did not prevent exposure to shocks; it buffered their psychological consequences.

7 Household Spillovers

The previous sections establish that the intervention protected treated women's mental health and functioning, and that this protection operated through psychological mechanisms. A natural question is whether these benefits extended beyond the treated individual. If the intervention preserved a woman's mental health and daily functioning under stress, this may have shielded her closest relationships and, through them, her children's wellbeing.

7.1 Spousal Relationships

Columns (4) and (5) of Table 5 examine two dimensions of the spousal relationship: relationship quality and marital conflict. Restrictions significantly reduced relationship quality among control women ($\beta_2 = -1.278$, $p < 0.05$) and increased the probability of marital conflict ($\beta_2 = 0.095$, $p < 0.05$). The intervention fully offset both effects. For each additional mandatory restriction, treated women reported relationship quality 2.427 points higher than their control counterparts ($p < 0.01$), equivalent to 10.3% of the control mean, and a 19.5 percentage point reduction in the probability of marital conflict ($p < 0.01$). Online Appendix Table 9 decomposes the relationship quality score and shows that these protective effects were specific to partner behaviors and did not extend to other household members (Online Appendix Table 10), consistent with an intervention that trained communication and problem-solving skills within the marital relationship.[9]

[9]Online Appendix Table 11 decomposes the results by type of restriction and shows that the spousal effects, like the mental health effects, are strongest for workplace closures, suggesting that the economic and emotional strain of restrictions on household members' ability to work was an important channel through which restrictions harmed relationships.

7.2 Child Outcomes

Table 6 reports effects on children's socio-emotional and cognitive development, measured when children were approximately six years old. Restrictions were harmful for children's socio-emotional wellbeing: behavioral difficulties on the Strengths and Difficulties Questionnaire (Goodman, 1997) increased by 0.177 standard deviations ($p < 0.10$) per additional mandatory restriction, while the corresponding increase in anxiety on the Spence Children's Anxiety Scale (Spence, 1998) (0.078 SD) was imprecisely estimated. Maternal assignment to treatment was clearly beneficial at mean restriction levels: children of treated mothers scored 0.230 standard deviations lower on anxiety ($p < 0.01$) and 0.149 standard deviations lower on behavioral difficulties ($p < 0.10$). The interaction coefficients suggest the intervention may have further offset the harmful effects of restrictions on children's socio-emotional wellbeing, but these estimates are imprecise.

Literacy and numeracy, measured through the Annual Status of Education Report (ASER Pakistan, 2019), both deteriorated during periods of stringent restrictions ($\beta_2 = -0.218$ SD and -0.200 SD, $p < 0.05$), in line with broader evidence from other countries (Werner and Woessmann, 2023). However, we find neither main treatment effects nor protective interaction effects. This is unsurprising: the intervention taught mothers to manage their own mental health and relationships, not to deliver academic instruction.

8 Discussion

We study whether psychological resilience to unanticipated shocks can be built. To do so, we leverage a CBT-based intervention delivered to prenatally depressed women in rural Pakistan together with COVID-19 restrictions as a shock that arrived years after treatment ended. We find that it can. As restrictions intensified, treated women were significantly less likely to develop anxiety, depression, and comorbid mental health conditions. Without the intervention, restrictions triggered a clinical escalation: mild anxiety first, then moderate depression, then comorbidity. The intervention interrupted this cascade, preserving women's daily functioning in the process. These protective effects extended beyond the treated individual, shielding spousal relationships from the strain of restrictions and suggestively improving children's socio-emotional wellbeing. The main results are robust to alternative restriction measures, additional controls, and placebo tests using pre-pandemic survey waves.

Three families of mechanisms explain how this protection operated. The intervention changed how women process stress: treated women perceived the threat of restrictions just as acutely as control women but were significantly less likely to feel overwhelmed as restrictions became more

stringent. This active buffering rested on a stock of psychological capital. Treatment had built a durable sense of personal control and self-efficacy that was present regardless of restriction severity. It had also strengthened perceived social support from family and partner, though not from friends, consistent with an intervention that trained interpersonal skills within the household rather than building external social networks. Hair cortisol and DHEA concentrations confirm that restrictions altered biological stress markers, but we find no significant treatment effects on either hormone, consistent with the resilience we document operating primarily through psychological rather than physiological channels. Nor did treatment reduce women's exposure to adverse events. Instead, it changed how they weathered them.

The skills that conferred this protection are not condition-specific. They were built to treat prenatal depression but transferred to a class of stressor that was entirely unrelated to the condition for which women were originally treated. If these skills are general-purpose, they represent a form of psychological capital with returns well beyond the treated condition. Given the persistent exposure of populations in low- and middle-income countries to climatic, economic, and health shocks, and the low cost of peer-delivered interventions, building psychological resilience may be among the more cost-effective forms of disaster preparedness available to policymakers.

References

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh. 2022. "The impact of the coronavirus lockdown on mental health: evidence from the United States." *Economic Policy*, 37(109): 139–155.

Agüero, Jorge M. 2021. "COVID-19 and the rise of intimate partner violence." *World Development*, 137.

Ahmad, Taimoor, Mujahid Abdullah, Abdul Mueed, Faisal Sultan, Ayesha Khan, and Adnan Ahmad Khan. 2023. "COVID-19 in Pakistan: A national analysis of five pandemic waves." *PLOS ONE*, 18(12): e0281326.

Aizer, Anna, Sungwoo Cho, Shari Eli, and Adriana Lleras-Muney. 2024. "The Impact of Cash Transfers to Poor Mothers on Family Structure and Maternal Well-Being." *American Economic Journal: Applied Economics*, 16(2): 492–529.

Aksunger, Nursena, Corey Vernot, Rebecca Littman, Maarten Voors, Niccolò F. Meriggi, Amanuel Abajobir, Bernd Beber, et al. 2023. "COVID-19 and mental health in 8 low- and middle-income countries: A prospective cohort study." *PLoS Medicine*, 20(4): 1–22.

Alloy, Lauren B., and Anthony H. Ahrens. 1987. "Depression and pessimism for the future: Biased use of statistically relevant information in predictions for self versus others." *Journal of Personality and Social Psychology*, 52(2): 366–378.

Altindag, Onur, Bilge Erten, and Pinar Keskin. 2022. "Mental Health Costs of Lockdowns: Evidence from Age-Specific Curfews in Turkey." *American Economic Journal: Applied Economics*,

14(2): 320–343.

Amirkhan, James H. 2018. "A Brief Stress Diagnostic Tool: The Short Stress Overload Scale." *Assessment*, 25(8): 1001–1013.

Angelucci, Manuela, and Daniel Bennett. 2024. "The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy." *American Economic Review*, 114(1): 169–198.

ASER Pakistan. 2019. "Annual Status of Education Report: ASER-Pakistan 2019." Idara-e-Taleem-o-Aagahi, Lahore, Pakistan.

Baranov, Victoria, Allison Frost, Ashley Hagaman, Julian G. Simmons, Muhammad S. Manzoor, Pietro Biroli, Sonia Bhalotra, Atif Rahman, Siham Sikander, and Joanna Maselko. 2022*a*. "Effects of a maternal psychosocial intervention on hair derived biomarkers of HPA axis function in mothers and children in rural Pakistan." *SSM - Mental Health*, 2(November 2021): 100082.

Baranov, Victoria, Pauline Grosjean, Fatima Jamal Khan, and Sarah Walker. 2022*b*. "The impact of COVID-related economic shocks on household mental health in Pakistan." *Health Economics*, 31(10): 2208–2228.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial." *American Economic Review*, 110(3): 824–859.

Barker, Nathan, Gharad Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher Udry. 2022. "Cognitive Behavioral Therapy among Ghana's Rural Poor Is Effective Regardless of Baseline Mental Distress." *American Economic Review: Insights*, 4(4): 527–545.

Betthäuser, Bastian A., Anders M. Bach-Mortensen, and Per Engzell. 2023. "A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic." *Nature Human Behaviour*, 7: 375–385.

Bhalotra, Sonia, Emilia Brito, Damian Clarke, Pilar Larroulet, and Francisco Pino. 2026. "Dynamic Impacts of Lockdown on Domestic Violence: Evidence from Multiple Policy Shifts in Chile." *Review of Economics and Statistics*, 108(2): 344–354.

Bhat, Bhargav, Jonathan de Quidt, Johannes Haushofer, Vikram Patel, Gautam Rao, Frank Schilbach, and Pierre-Luc Vautrey. 2022. "The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Economic Outcomes." *NBER Working Paper Series*, , (30011).

Blattman, Christopher, Sebastian Chaskel, Julian C. Jamison, and Margaret Sheridan. 2023. "Cognitive Behavioral Therapy Reduces Crime and Violence over Ten Years: Experimental Evidence." *American Economic Review: Insights*, 5(4): 527–545.

Bleidorn, Wiebke, Ted Schwaba, Anqing Zheng, Christopher J. Hopwood, Susana S. Sosa, Brent W. Roberts, and D. A. Briley. 2022. "Personality stability and change: A meta-analysis of longitudinal studies." *Psychological Bulletin*, 148(7-8): 588–619.

Brodeur, Abel, Andrew E Clark, Sarah Fleche, and Nattavudh Powdthavee. 2021. "COVID-19, lockdowns and well-being: Evidence from Google Trends." *Journal of Public Economics*, 193: 104346.

Christian, Cornelius, Lukas Hensel, and Christopher Roth. 2019. "Income Shocks and Suicides: Causal Evidence From Indonesia." *The Review of Economics and Statistics*, 101(5): 905–920.

Cobb-Clark, Deborah A, and Stefanie Schurer. 2013. "Two economists' musings on the stability of locus of control." *The Economic Journal*, 123(570): F358–F400.

Elmer, Timon, and Christoph Stadtfeld. 2020. "Depressive symptoms are associated with social isolation in face-to-face interaction networks." *Scientific Reports*, 10(1): 1444.

Etheridge, Ben, and Lisa Spantig. 2022. "The gender gap in mental well-being at the onset of the Covid-19 pandemic: Evidence from the UK." *European Economic Review*, 145: 104114.

Giuntella, Osea, Kelly Hyde, Silvia Saccardo, and Sally Sadoff. 2021. "Lifestyle and mental health disruptions during COVID-19." *Proceedings of the National Academy of Sciences*, 118(9).

Global Burden of Disease Collaborative Network. 2024. "Global Burden of Disease Study 2021 (GBD 2021) Results." Accessed 25 July 2026.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry*, 38(5): 581–586.

Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, et al. 2021. "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)." *Nature Human Behaviour*, 5(4): 529–538.

Hidrobo, Melissa, and Lia Fernald. 2013. "Cash transfers and domestic violence." *Journal of Health Economics*, 32(1): 304–319.

Jacobson, Nicholas C., and Michelle G. Newman. 2017. "Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies." *Psychological Bulletin*, 143(11): 1155–1200.

Kalisch, Raffael, Dewleen G. Baker, Ulrike Basten, Marco P. Boks, George A. Bonanno, Eddie Brummelman, Andrea Chmitorz, et al. 2017. "The resilience framework as a strategy to combat stress-related disorders." *Nature Human Behaviour*, 1(11): 784–790.

Kalisch, Raffael, Marianne B. Müller, and Oliver Tüscher. 2015. "A conceptual framework for the neurobiological study of resilience." *Behavioral and Brain Sciences*, 38: e92.

Kroenke, Kurt, Robert L Spitzer, and Janet B W Williams. 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine*, 16(9): 606–613.

Lane, Gregory. 2024. "Adapting to Climate Risk With Guaranteed Credit: Evidence From Bangladesh." *Econometrica*, 92(2): 355–386.

Lund, Crick, Alison Breen, Alan J. Flisher, Ritsuko Kakuma, Joanne Corrigall, John A. Joska, Leslie Swartz, and Vikram Patel. 2010. "Poverty and common mental disorders in low and middle income countries: A systematic review." *Social Science and Medicine*, 71(3): 517–528.

Maselko, Joanna, Siham Sikander, Elizabeth L. Turner, Lisa M. Bates, Ikhlaq Ahmad, Najia Atif, Victoria Baranov, et al. 2020. "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*, 7(9): 775–787.

Masten, Ann S. 2001. "Ordinary magic: Resilience processes in development." *American Psychologist*, 56(3): 227–238.

McEwen, Bruce S. 1998. "Protective and Damaging Effects of Stress Mediators." *New England Journal of Medicine*, 338(3): 171–179.

McNeil, Aliya, Lydia Hicks, Busra Yalcinoz-Ucan, and Dillon T. Browne. 2023. "Prevalence & Correlates of Intimate Partner Violence During COVID-19: A Rapid Review." *Journal of Family Violence*, 38(2): 241–261.

Meier, Jana, Bianca Kollmann, Laura E. Meine, Benjamin Meyer, Kenneth Yuen, Magdalena Stork, Oliver Tüscher, and Michèle Wessa. 2026. "Perceived control as a resilience factor: associations with neural, physiological and affective stress responses and mental health." *Translational Psychiatry*, 16(1): 39.

Meier, Jana, Laura E. Meine, Katja Schüler, and Michèle Wessa. 2025. "Distinct trajectories of perceived control over aversive stimulation predict affective reactions to stressors over and above objective control." *Scientific Reports*, 15(1): 35009.

Miller, Gregory E., Edith Chen, and Eric S. Zhou. 2007. "If It Goes Up, Must It Come Down? Chronic Stress and the Hypothalamic-Pituitary-Adrenocortical Axis in Humans." *Psychological Bulletin*, 133(1): 25–45.

Nichols, Emma, Sarah Petrosyan, Pranali Khobragade, Joyita Banerjee, Marco Angrisani, Sharmistha Dey, David E. Bloom, Simone Schaner, Aparajit B. Dey, and Jinkook Lee. 2024. "Trajectories and correlates of poor mental health in India over the course of the COVID-19 pandemic: a nationwide survey." *BMJ Global Health*, 9(1).

NIPS, and ICF. 2013. "Pakistan Demographic and Health Survey 2012-2013."

Pariante, Carmine M., and Stafford L. Lightman. 2008. "The HPA axis in major depression: classical theories and new developments." *Trends in Neurosciences*, 31(9): 464–468.

Piquero, Alex R, Wesley G Jennings, Erin Jemison, Catherine Kaukinen, and Felicia Marie Knaul. 2021. "Domestic violence during the COVID-19 pandemic - Evidence from a systematic review and meta-analysis." *Journal of Criminal Justice*, 74(March): 101806.

Premand, Patrick, and Quentin Stoeffler. 2022. "Cash transfers, climatic shocks and resilience in the Sahel." *Journal of Environmental Economics and Management*, 116: 102744.

Racine, Nicole, Brae Anne McArthur, Jessica E. Cooke, Rachel Eirich, Jenney Zhu, and Sheri Madigan. 2021. "Global Prevalence of Depressive and Anxiety Symptoms in Children and Adolescents During COVID-19: A Meta-analysis." *JAMA Pediatrics*, 175(11): 1142–1150.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. "Poverty, depression, and anxiety: Causal evidence and mechanisms." *Science*, 370(6522).

Roberts, Brent W., Kate E. Walton, and Wolfgang Viechtbauer. 2006. "Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies." *Psychological Bulletin*, 132(1): 1–25.

Seligman, M E P. 1972. "Learned Helplessness." *Annual Review of Medicine*, 23(1): 407–412.

Sevim, Dilek, Victoria Baranov, Sonia Bhalotra, Joanna Maselko, and Pietro Biroli. 2024. "Trajectories of Early Childhood Skill Development and Maternal Mental Health." *Journal of Human Resources*, 59(S): S365–S401.

Sikander, Siham, Ikhlaq Ahmad, Lisa M Bates, John Gallis, Ashley Hagaman, Karen O'Donnell, Elizabeth Louise Turner, Ahmed Zaidi, Atif Rahman, and Joanna Maselko. 2019*a*. "Cohort profile: Perinatal depression and child socioemotional development; The bachpan cohort study from rural Pakistan." *BMJ Open*, 9(5): 1–11.

Sikander, Siham, Ikhlaq Ahmad, Najia Atif, Ahmed Zaidi, Fiona Vanobberghen, Helen A Weiss, Anum Nisar, et al. 2019*b*. "Delivering the Thinking Healthy Programme for perinatal depression through volunteer peers: a cluster randomised controlled trial in Pakistan." *The Lancet Psychiatry*, 6(2): 128–139.

Southwick, Steven M., and Dennis S. Charney. 2012. "The Science of Resilience: Implications for the Prevention and Treatment of Depression." *Science*, 338(6103): 79–82.

Southwick, Steven M., George A. Bonanno, Ann S. Masten, Catherine Panter-Brick, and Rachel Yehuda. 2014. "Resilience definitions, theory, and challenges: interdisciplinary perspectives." *European Journal of Psychotraumatology*, 5(1).

Spence, Susan H. 1998. "A measure of anxiety symptoms among children." *Behaviour Research and Therapy*, 36(5): 545–566.

Spitzer, Robert L, Kurt Kroenke, Janet B W Williams, and Bernd Löwe. 2006. "A Brief Measure for Assessing Generalized Anxiety Disorder." *Archives of Internal Medicine*, 166(10): 1092.

Stalder, Tobias, Susann Steudte-Schmiedgen, Nina Alexander, Tim Klucken, Aline Vater, Stephan Wichmann, Clemens Kirschbaum, and Robert Miller. 2017. "Stress-related and basic determinants of hair cortisol in humans: A meta-analysis." *Psychoneuroendocrinology*, 77: 261–274.

Stillman, Steven, and Malathi Velamuri. 2025. "The sensitivity of locus of control to life events and implications for its moderating influence on wellbeing." *Scientific Reports*, 15(1): 29067.

Turner, Elizabeth L., Siham Sikander, Omer Bangash, Ahmed Zaidi, Lisa Bates, John Gallis, Nima Ganga, Karen O'Donnell, Atif Rahman, and Joanna Maselko. 2016. "The effectiveness of the peer delivered Thinking Healthy Plus (THPP+) Programme for maternal depression and child socio-emotional development in Pakistan: study protocol for a three-year cluster randomized controlled trial." *Trials*, 17(1): 442.

Werner, Katharina, and Ludger Woessmann. 2023. "The legacy of COVID-19 in education." *Economic Policy*, 38(115): 609–668.

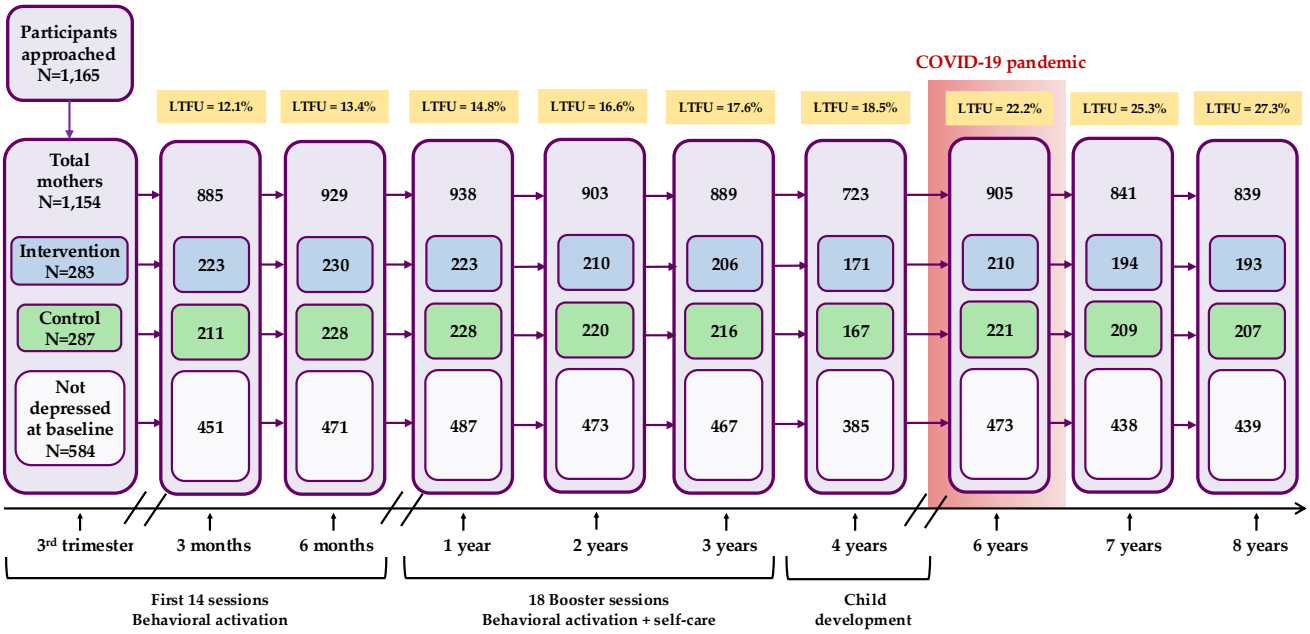
Wittchen, Hans-Ulrich, Ronald C. Kessler, Hildegard Pfister, Michael Höfler, and Roselind Lieb. 2000. "Why do people with anxiety disorders become depressed? A prospective-longitudinal community study." *Acta Psychiatrica Scandinavica*, 102(s406): 14–23.

Zeier, Peter, Laura E. Meine, and Michèle Wessa. 2022. "It's worth the trouble: Stressor exposure is related to increased cognitive reappraisal ability." *Stress and Health*, 38(3): 602–609.

Zimet, Gregory D, Suzanne S Powell, Gordon K Farley, Sidney Werkman, and Karen A Berkoff. 1990. "Psychometric Characteristics of the Multidimensional Scale of Perceived Social Support." *Journal of Personality Assessment*, 55(3-4): 610–617.

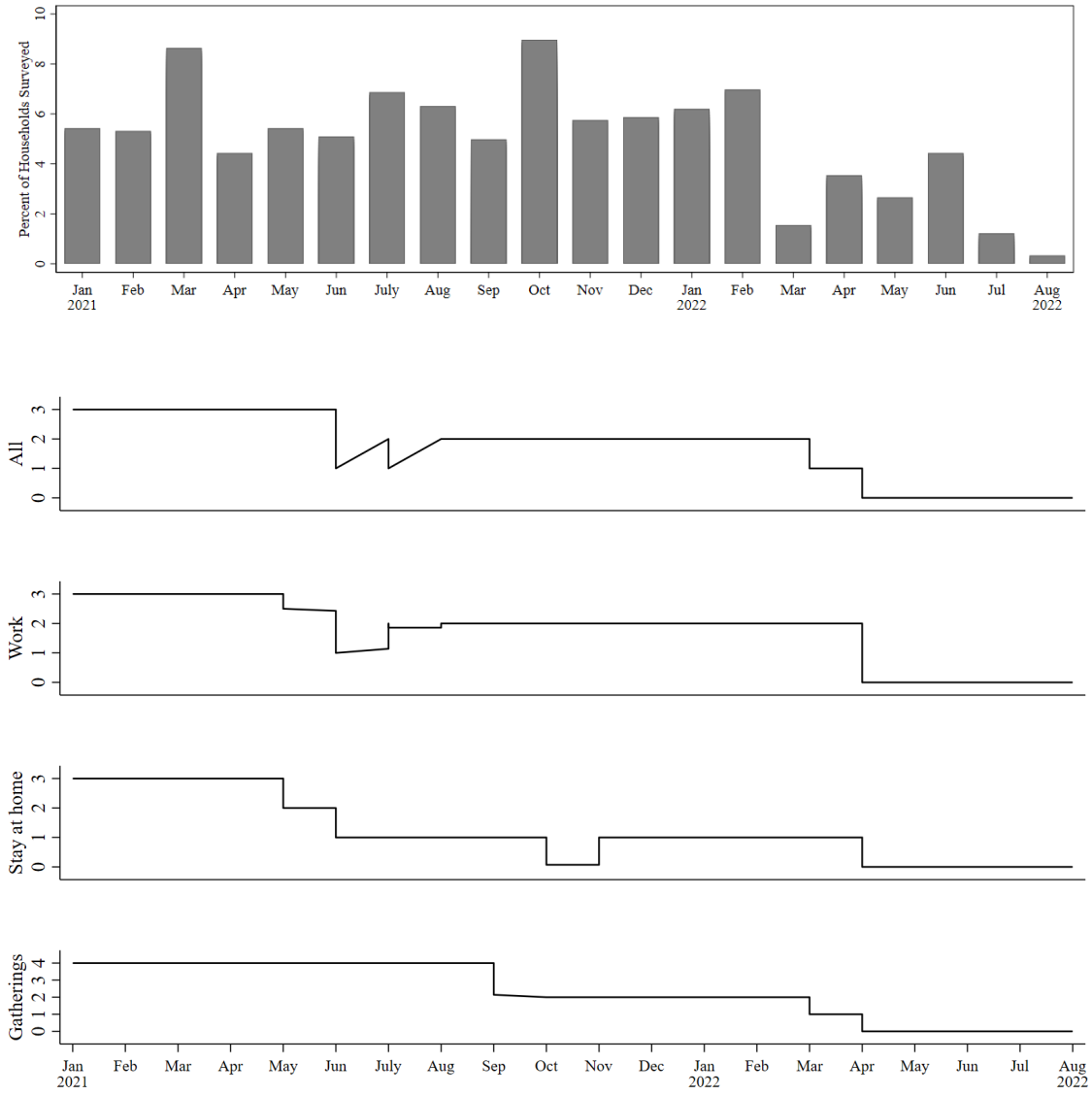
9 Figures

Figure 1: Timeline of TH-BCP Intervention and Follow-ups



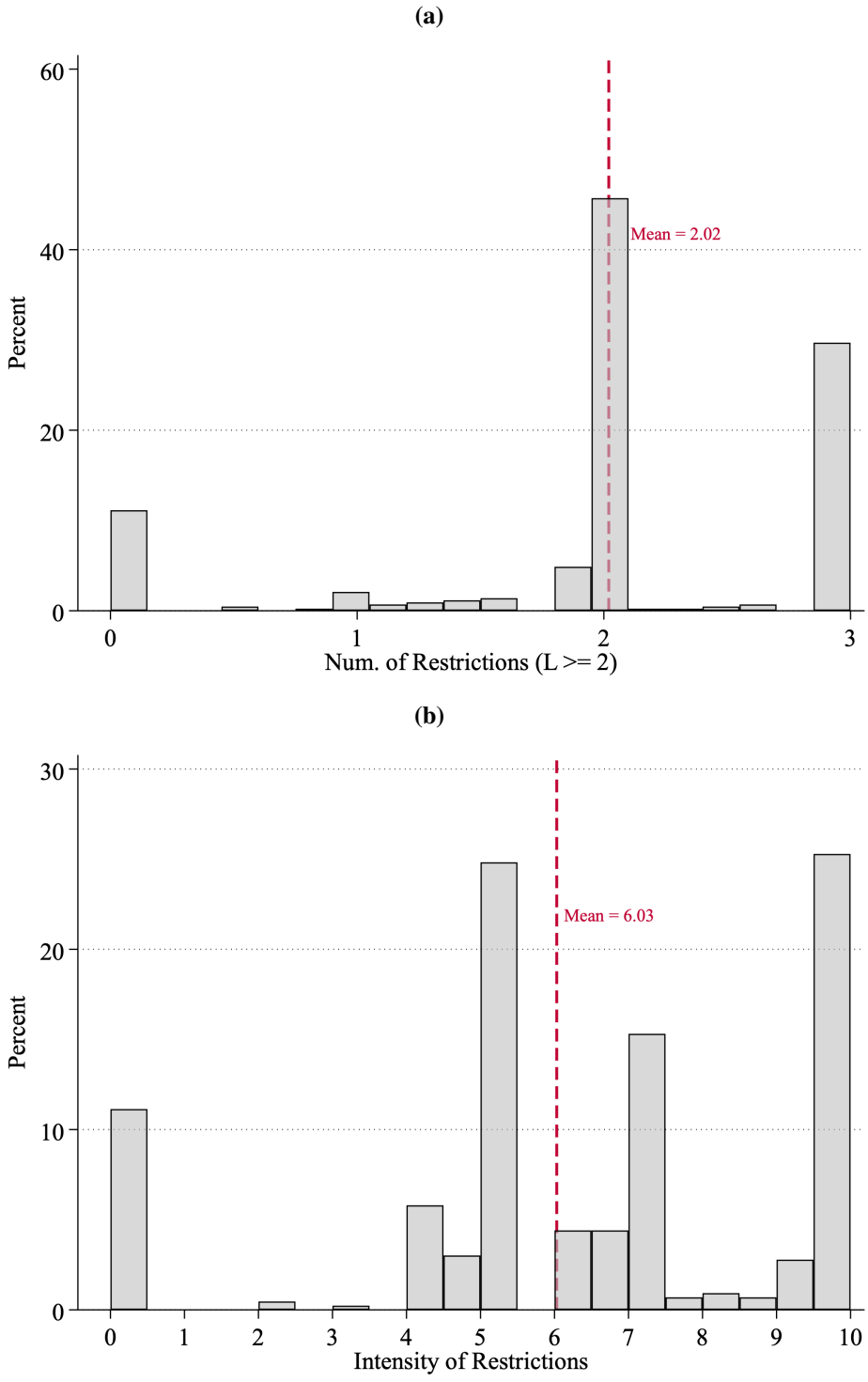
Note: In year 4, surveys were collected by mobile phone, which explains the lower recontact rate that year.

Figure 2: Timeline of the Year-6 survey rollout and national restrictions



Note: Figure 2a plots the percentage of households surveyed in each month of the year six rollout between January of 2021 and August of 2022. Figure 2b plots i) the average number of mandatory restrictions that were in place during the two weeks prior to the survey, ii) the intensity of workplace closures (0 – no restrictions, 1 – suggested restrictions, 2 – mandated from some segments of the population and 3 – mandated for everyone), iii) the intensity of stay-at-home orders (0 – no restrictions, 1 – suggested restrictions, 2 – mandated from some segments of the population and 3 – mandated for everyone), and iv) the stringency of restrictions on gatherings in terms of number of people allowed in them (0 – no restrictions, 1 – 1000 people at most, 2 – 101 to 1000 people, 3 – 10 to 100 people at most, 4 – 4 to 10 or less people allowed).

Figure 3: Aggregate restriction measures during year six of the study



Notes: Figure 3a plots the proportion of women surveyed by the average number of mandatory restrictions over the two weeks prior to (and including) the survey. Figure 3b plots the corresponding distribution by intensity of restrictions, which is the sum of the ordinal measures of workplace closures, stay-at-home orders and restrictions on gatherings, averaged over the two weeks preceding the survey date.

10 Tables

Table 1: Balance in baseline characteristics across waves

	Baseline Sample N=570					Year 6 N=431	
	(1) Control Mean	(2) SD	(3) Treatment Mean	(4) Diff T-C	(5) *p*-val	(6) Diff T-C	(7) *p*-val
Mother's age	27.29	(5.0)	26.80	-0.49	0.23	-0.49	0.28
Mother's height (cm)	156.33	(6.1)	157.43	1.10	0.03	1.17	0.05
Mother's weight (kg)	61.24	(12.9)	60.17	-1.07	0.31	-1.43	0.25
Mother's waist circ. (cm)	95.39	(10.4)	93.60	-1.79	0.05	-2.10	0.04
Mother's blood pressure	71.82	(14.1)	70.66	-1.16	0.29	-0.61	0.62
PHQ-9 Score	14.48	(3.6)	14.89	0.41	0.18	0.27	0.44
WHODAS Score	16.11	(9.1)	16.71	0.60	0.42	0.63	0.46
PSS Score	22.90	(7.5)	23.84	0.94	0.14	1.05	0.15
Current major dep. episode	0.73	(0.4)	0.78	0.05	0.21	0.05	0.28
Nuclear family	0.25	(0.4)	0.25	0.00	0.92	-0.02	0.62
Joint/extended family	0.63	(0.5)	0.58	-0.05	0.18	-0.04	0.40
Grandmother present	0.67	(0.5)	0.63	-0.04	0.36	-0.01	0.84
Total adults in the household	5.70	(3.0)	5.33	-0.37	0.13	-0.08	0.77
First child	0.25	(0.4)	0.23	-0.02	0.55	-0.00	0.97
Number of boys	0.79	(1.0)	0.86	0.07	0.40	-0.05	0.57
Number of girls	0.85	(1.1)	0.96	0.10	0.26	0.05	0.63
People per room	2.47	(1.9)	2.79	0.32	0.05	0.35	0.04
SES asset index	-0.32	(1.7)	-0.56	-0.24	0.09	-0.08	0.61
Mother's education	6.80	(4.5)	6.83	0.03	0.95	0.24	0.56
Father's education	8.33	(3.3)	7.85	-0.48	0.10	-0.61	0.06
Joint test (*p*-value)					0.10		0.06

Note: This table tests for balance in baseline characteristics. Columns 1 and 3 show the mean in the control and treatment at baseline, column 4 their difference, and column 5 the p-value of a test for whether this is statistically different from zero. Columns 6 and 7 report the same but for the sample that did not attrit until the survey in Year 6.

Table 2: Mental health and daily functioning

	Anxiety Score	Depression Score	Days with Functional Difficulties
	(1)	(2)	(3)
Intervention	-0.107	-0.958**	-1.605**
	(0.372)	(0.381)	(0.657)
Average # of restrictions	1.197***	0.763	1.275*
	(0.404)	(0.475)	(0.725)
Intervention × Restrictions	-1.209**	-1.448*	-2.979**
	(0.562)	(0.754)	(1.175)
Romano-Wolf *p*-value	0.088	0.088	0.055
R^2	0.316	0.302	0.297
Outcome Mean	5.383	7.735	9.325
Observations	431	431	431

Note: *Average # of restrictions* sums three dichotomized indicators (1 if mandatory, 0 otherwise) for workplace closures, stay-at-home requirements, and restrictions on social gatherings (range: 0–3). Anxiety is measured using the Generalized Anxiety Disorder Scale (GAD-7). Depression is measured using the 9-item Patient Health Questionnaire (PHQ-9). Days with functional difficulties comes from the WHO Disability Assessment Schedule (WHO-DAS) item: Overall, in the past 30 days how many days have these difficulties (battery of 12 functional difficulties) interfere with your life? The main controls include: At baseline (PHQ-9, WHODAS, PSS, woman's height, waist circumference, joint household, adults in the house, people in the room, asset score, mother's education, father's education, sexual IPV, Union Council FE, assessor FE), at year 6 (assessor FE, month and month squared). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Clinical severity and comorbidity

	Anxiety		Depression		Comorbidity	
	Mild	Moderate	Mild	Moderate	Mild	Moderate
	(1)	(2)	(3)	(4)	(5)	(6)
Intervention	-0.007	0.020	-0.109***	-0.069**	-0.030	-0.001
	(0.034)	(0.026)	(0.035)	(0.032)	(0.035)	(0.025)
Average # of restrictions	0.095**	0.073*	0.031	0.102**	0.124***	0.088**
	(0.040)	(0.040)	(0.055)	(0.040)	(0.041)	(0.033)
Intervention × Restrictions	-0.152***	-0.067	-0.062	-0.140**	-0.183***	-0.097*
	(0.054)	(0.059)	(0.070)	(0.064)	(0.055)	(0.055)
R^2	0.308	0.234	0.259	0.284	0.302	0.229
Outcome Mean	0.455	0.200	0.645	0.334	0.418	0.165
Observations	431	431	431	431	431	431

Note: *Average # of restrictions* sums three dichotomized indicators (1 if mandatory, 0 otherwise) for workplace closures, stay-at-home requirements, and restrictions on social gatherings (range: 0–3). Women with mild and moderate anxiety have GAD-7 scores greater or equal to 5 and 10, respectively. Women with mild and moderate depression have PHQ-9 scores greater or equal to 5 and 10, respectively. Mild comorbidity indicates a person who has both mild anxiety and mild depression. Moderate comorbidity indicates a person who has both moderate anxiety and moderate depression. The main controls include: At baseline (PHQ-9, WHODAS, PSS, woman's height, waist circumference, joint household, adults in the house, people in the room, asset score, mother's education, father's education, sexual IPV, Union Council FE, assessor FE), at year 6 (assessor FE, month and month squared). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Psychological and Biological Mechanisms

	Stress Appraisal and Locus of Control			HPA Axis	
	Threat Appraisal (1)	Stress Overload (2)	Control and Self-Efficacy (3)	Log Cortisol (4)	Log DHEA (5)
Intervention	-0.061	-0.031	0.177**	0.055	-0.020
	(0.084)	(0.090)	(0.074)	(0.057)	(0.135)
Average # of restrictions	0.181**	0.125*	-0.107	-0.263***	0.202
	(0.077)	(0.065)	(0.096)	(0.085)	(0.162)
Intervention × Restrictions	-0.094	-0.259**	0.048	0.101	-0.078
	(0.125)	(0.109)	(0.137)	(0.129)	(0.230)
Romano-Wolf p-value	0.881	0.106	0.934	0.881	0.934
R^2	0.278	0.275	0.282	0.456	0.225
Outcome Mean	-0.000	0.000	0.000	1.366	0.522
Observations	431	431	431	391	391

Note: Outcomes in columns (1)–(3) are standardized to have mean zero and standard deviation one. Column (1) is the first principal component of the Stress Overload Scale (SOS) Personal Vulnerability subscale, capturing threat appraisal and helplessness. Column (2) is the first principal component of the SOS Event Load subscale, capturing perceived stress burden. Column (3) is the first principal component of locus of control, items from the Generalized Self-Efficacy scale (GSE), and items from the Perceived Stress Scale (PSS). Columns (4)–(5) are the natural logarithm of cortisol and DHEA concentrations in hair samples, capturing hypothalamic–pituitary–adrenal (HPA) axis activity. All regressions include the same controls as in Tables 2 and 3. Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Social support and spousal relationships

	Perceived Social Support			Spousal Relationships	
	Family	Friends	Significant Other	Relationship score	Marital conflict
	(1)	(2)	(3)	(4)	(5)
Intervention	0.142*	0.076	0.146**	-0.175	0.010
	(0.075)	(0.086)	(0.070)	(0.430)	(0.028)
Average # of restrictions	-0.091	-0.269*	-0.118	-1.278**	0.095**
	(0.105)	(0.139)	(0.104)	(0.555)	(0.046)
Intervention × Restrictions	0.055	-0.124	0.097	2.427***	-0.195***
	(0.131)	(0.163)	(0.133)	(0.876)	(0.062)
Romano-Wolf p-value	0.826	0.826	0.826	0.070	0.034
R^2	0.259	0.243	0.287	0.300	0.206
Outcome Mean	0.000	0.000	-0.000	23.574	0.230
Observations	431	431	431	427	427

Note: *Average # of restrictions* sums three dichotomized indicators (1 if mandatory, 0 otherwise) for workplace closures, stay-at-home requirements, and restrictions on social gatherings (range: 0–3). The main controls include: At baseline (PHQ-9, WHODAS, PSS, woman's height, waist circumference, joint household, adults in the house, people in the room, asset score, mother's education, father's education, sexual IPV, Union Council FE, assessor FE), at year 6 (assessor FE, month and month squared). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Child wellbeing

	Socio-emotional wellbeing		Academic outcomes	
	Anxiety score	Emotional and behavioral problems	Literacy score	Numeracy score
	(1)	(2)	(3)	(4)
Intervention	-0.230***	-0.149*	-0.052	-0.023
	(0.078)	(0.083)	(0.066)	(0.070)
Average # of restrictions	0.078	0.177*	-0.218**	-0.200**
	(0.100)	(0.101)	(0.084)	(0.092)
Intervention × Restrictions	-0.323	-0.229	0.057	0.099
	(0.197)	(0.152)	(0.130)	(0.161)
Romano-Wolf p-value	0.896	0.357	0.483	0.484
R^2	0.202	0.260	0.374	0.317
Outcome Mean	0.000	0.000	-0.000	-0.000
Observations	430	430	425	425

Note: *Average # of restrictions* sums three dichotomized indicators (1 if mandatory, 0 otherwise) for workplace closures, stay-at-home requirements, and restrictions on social gatherings (range: 0–3). The main controls include: At baseline (PHQ-9, WHODAS, PSS, woman's height, waist circumference, joint household, adults in the house, people in the room, asset score, mother's education, father's education, sexual IPV, Union Council FE, assessor FE), at year 6 (assessor FE, month and month squared). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Covid-related Adverse Events

	Health Shocks		Economic Shocks	
	Know someone who contracted Covid-19	Know someone who died from Covid-19	Received financial assistance	Permanent job loss in family
	(1)	(2)	(3)	(4)
Intervention	0.017	-0.061	-0.001	0.045
	(0.059)	(0.052)	(0.026)	(0.044)
Average # of restrictions	-0.038	0.024	0.050*	0.024
	(0.074)	(0.070)	(0.028)	(0.047)
Intervention × Restrictions	0.022	0.000	-0.057	-0.016
	(0.112)	(0.091)	(0.050)	(0.075)
R^2	0.314	0.267	0.221	0.220
Outcome Mean	0.752	0.413	0.123	0.167
Observations	431	429	430	431

Note: The response options for each of the outcomes are coded as follows: column (1) is coded as 0 if no one contracted Covid, 1 if someone in the community contracted Covid, and 2 if friends, relatives, a household member or the respondent herself contracted Covid. Column (2) takes a value of 0 if the respondent knows no one who died from Covid, 1 if someone in the community died, and 2 if a friend or relative died. Column (3) on financial assistance is a binary variable. Column (4) ranges from 0 if not true for the respondent, 1 if somewhat true, and 2 if very true that there was permanent job loss in the family. *Average # of restrictions* sums three dichotomized indicators (1 if mandatory, 0 otherwise) for workplace closures, stay-at-home requirements, and restrictions on social gatherings (range: 0–3). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix for "Building Resilience through Psychological Intervention"

Michelle Escobar Carías, Victoria Baranov, Sonia Bhalotra,
Pietro Biroli and Joanna Maselko

August 07, 2026

[For Online Publication Only]

1 Data Appendix

Variable Definitions

Primary Outcome Variables:

- **Depression (PHQ-9):** The Patient Health Questionnaire is a nine-item instrument assessing the frequency of depressive symptoms over the preceding two weeks (Kroenke, Spitzer and Williams, 2001). Respondents rate each item on a four-point scale (0 = not at all, 1 = several days, 2 = more than half the days, 3 = nearly every day), yielding a total score of 0–27. The nine items ask how often the respondent has been bothered by: (1) little interest or pleasure in doing things; (2) feeling down, depressed, or hopeless; (3) trouble falling or staying asleep, or sleeping too much; (4) feeling tired or having little energy; (5) poor appetite or overeating; (6) feeling bad about yourself, or that you are a failure or have let yourself or your family down; (7) trouble concentrating on things; (8) moving or speaking so slowly that other people could have noticed, or the opposite; and (9) thoughts that you would be better off dead, or of hurting yourself. Higher scores indicate greater symptom burden.

- **Anxiety (GAD-7):** The Generalized Anxiety Disorder scale is a seven-item instrument measuring anxiety severity over the preceding two weeks (Spitzer et al., 2006). Items are scored on the same four-point scale as the PHQ-9, yielding a total score of 0–21. The seven items ask how often the respondent has been bothered by: (1) feeling nervous, anxious, or on edge; (2) not being able to stop or control worrying; (3) worrying too much about different things; (4) trouble relaxing; (5) being so restless that it is hard to sit still; (6) becoming easily annoyed or irritable; and (7) feeling afraid as if something awful might happen. Higher scores indicate greater symptom burden.

- **Mild and moderate severity:** We define clinical severity thresholds following standard cutoffs. Women with mild and moderate anxiety have GAD-7 scores greater or equal to 5 and 10, respectively. Women with mild and moderate depression have PHQ-9 scores greater or equal to 5 and 10, respectively.

- **Comorbid anxiety and depression:** Mild comorbidity indicates a person who has both mild anxiety and mild depression. Moderate comorbidity indicates a person who has both moderate anxiety and moderate depression.

- **Functional impairment (WHODAS 2.0):** The WHO Disability Assessment Schedule asks respondents how many days in the past 30 they had difficulty carrying out their usual activities.

This captures whether psychological deterioration translates into reduced capacity to manage daily responsibilities.

Secondary Outcome Variables:

- **Relationship quality (RFH):** The six-item Relationship with Father/Husband questionnaire asks respondents to rate: (1) their husband's attentiveness to basic needs; (2) provision of pocket money; (3) emotional understanding; (4) support during difficulties; (5) behavior toward the respondent; and (6) overall relationship quality. Items are aggregated into an RFH score.
- **Marital conflict:** A binary indicator equal to one if the respondent reports that her marital relations have had problems in the last year.
- **Child anxiety (SCAS):** The Spence Children's Anxiety Scale is a six-item parent-report instrument measuring anxiety symptoms in children (Spence, 1998). The six items are: (1) my child worries about things; (2) when my child has a problem, s/he complains of having a funny feeling in his/her stomach; (3) my child complains of feeling afraid; (4) when my child has a problem, s/he complains of his/her heart beating really fast; (5) my child worries that something bad will happen to him/her; and (6) when my child has a problem, s/he feels shaky. Items are summed into a total score; higher scores indicate greater anxiety.
- **Child difficulties (SDQ):** The Strengths and Difficulties Questionnaire is a 25-item parent-report behavioral screening instrument (Goodman, 1997). Each item is scored 0–2 (not true, somewhat true, certainly true). The 25 items are: (1) considerate of other people's feelings; (2) restless, overactive, cannot stay still for long; (3) often complains of headaches, stomachaches, or sickness; (4) shares readily with other children; (5) often loses temper; (6) rather solitary; (7) generally well-behaved; (8) many worries or often seems worried; (9) helpful if someone is hurt, upset, or feeling ill; (10) constantly fidgeting or squirming; (11) has at least one good friend; (12) often fights with other children or bullies them; (13) often unhappy, depressed, or tearful; (14) generally liked by other children; (15) easily distracted, concentration wanders; (16) nervous or clingy in new situations, easily loses confidence; (17) kind to younger children; (18) often argumentative with adults; (18A) often lies or cheats; (19) picked on or bullied by other children; (20) often offers to help others; (21) can stop and think things out before acting; (22) can be spiteful to others; (22A) steals from home, school or elsewhere; (23) gets along better with adults than with other children; (24) many fears, easily scared; and (25) good attention span, sees work through to the end. Higher scores indicate greater difficulties.
- **Academic outcomes (ASER):** The Annual Status of Education Report tool is a widely used direct assessment in South Asia that tests literacy (Urdu) and numeracy through one-on-one

evaluation (ASER Pakistan, 2019). Urdu literacy is scored on a five-level scale: beginner (0), alphabet (1), word (2), sentence (3), and story (4). Math numeracy is scored on a six-level scale: beginner (0), single digit (1), double digit (2), three digit (3), subtraction (4), and division (5).

Mechanism Variables:

- **Threat appraisal (SOS – Personal Vulnerability):** Derived from the Personal Vulnerability subscale of the Stress Overload Scale (Amirkhan, 2018). Respondents rate how often they experienced each feeling over the prior week on a five-point Likert scale (1 = not at all, 5 = a lot). The four items are: (1) feel inadequate; (2) odds are against me; (3) nothing going right; and (4) there is no escape. The index is constructed using principal component analysis.

- **Stress overload (SOS – Event Load):** Derived from the Event Load subscale of the SOS, scored on the same five-point Likert scale. The five items are: (1) swamped by responsibility; (2) no time to get to everything; (3) felt rushed; (4) things keep piling up; and (5) carrying heavy load. The index is constructed using principal component analysis.

- **Self-efficacy and control:** An index constructed as the first principal component of six items drawn from three instruments. From the locus of control scale: (1) locus of control. From the Generalized Self-Efficacy scale (GSE): (2) I can manage to solve difficult problems if I try; (3) I can find several solutions to problems; (4) if I am in trouble, I can think of a solution. From the Perceived Stress Scale (PSS): (5) unable to control important things in life; (6) things not going my way.

- **Hair cortisol:** Cortisol concentration measured in hair samples collected at year six. Hair absorbs cortisol as it grows (approximately 1 cm per month), providing a biological record of cumulative hypothalamic-pituitary-adrenal (HPA) axis activity over the preceding months.

- **Hair DHEA:** Dehydroepiandrosterone concentration measured in the same hair samples. DHEA is an antagonist of cortisol and serves as an additional indicator of HPA axis regulation.

- **Family support (MSPSS):** The family subscale of the Multidimensional Scale of Perceived Social Support (Zimet et al., 1990). Each item is rated on a seven-point Likert scale (1 = very strongly disagree, 7 = very strongly agree). The four items are: (1) my family really tries to help me; (2) I get the emotional help and support I need from my family; (3) I can talk about my problems with my family; and (4) my family is willing to help me make decisions.

- **Friends support (MSPSS):** The friends subscale. The four items are: (1) my friends really try to help me; (2) I can count on my friends when things go wrong; (3) I have friends with whom I can share my joys and sorrows; and (4) I can talk about my problems with my friends.

- **Partner support (MSPSS):** The significant other subscale. The four items are: (1) there is a special person who is around when I am in need; (2) there is a special person with whom I can share my joys and sorrows; (3) I have a special person who is a real source of comfort to me; and (4) there is a special person in my life who cares about my feelings.

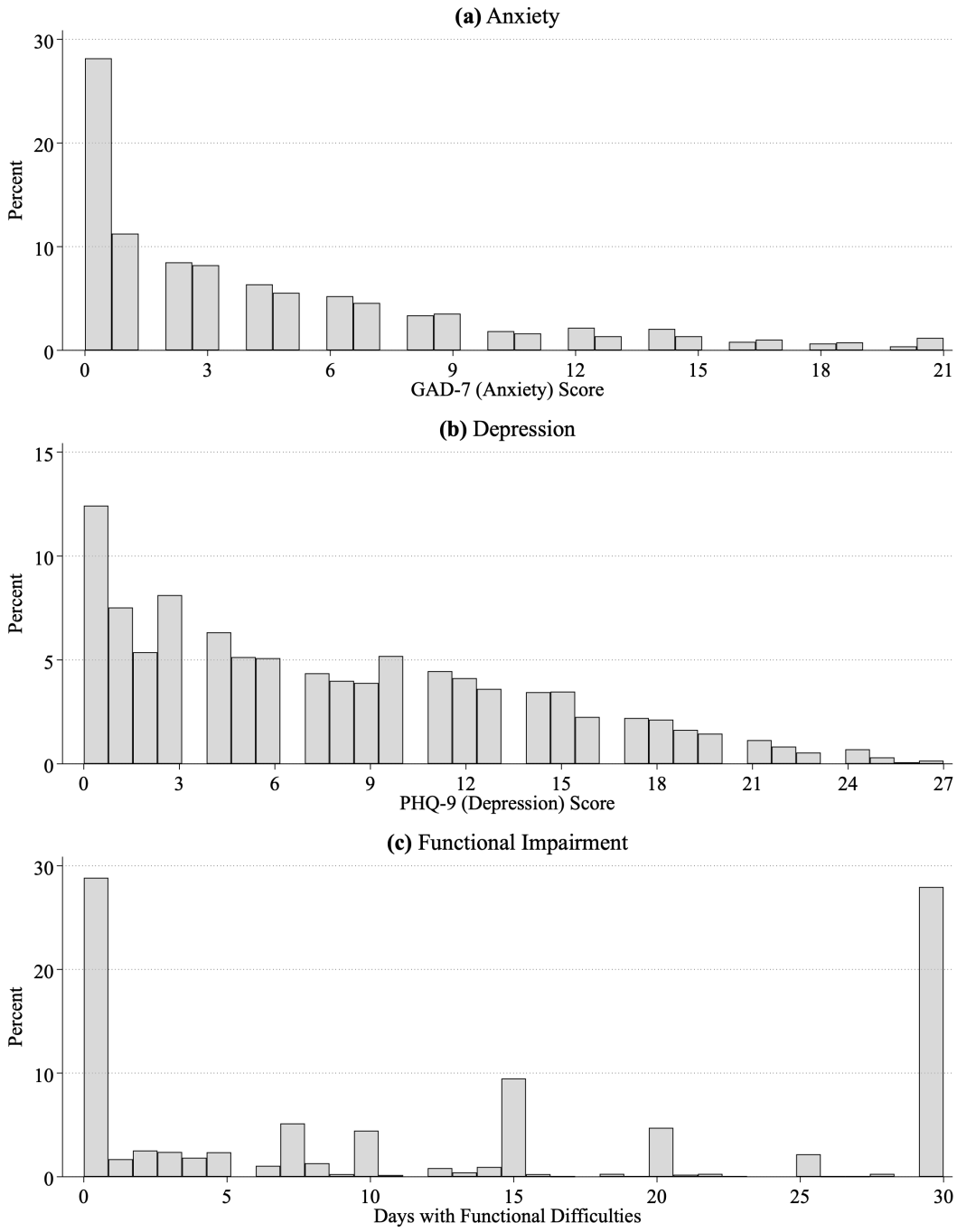
- **Self-reported COVID-related shocks:** Respondents were asked about health and economic shocks experienced during the pandemic. Health shocks include: (1) whether the respondent knows someone who contracted COVID-19 (coded 0 if no one contracted COVID, 1 if someone in the community, 2 if friends, relatives, a household member or the respondent herself); and (2) whether the respondent knows someone who died from COVID-19 (coded 0 if no one, 1 if someone in the community, 2 if a friend or relative). Economic shocks include: (3) whether the family received financial assistance (binary); and (4) whether there was permanent job loss in the family (coded 0 if not true, 1 if somewhat true, 2 if very true).

Treatment and Restriction Variables:

- **Treatment:** A binary indicator equal to one if the respondent resides in a village cluster randomized to receive the Thinking Healthy + Bachpan Care Program (TH-BCP).

- **COVID-19 restrictions:** The average number of mandatory restrictions in place during the 14 days preceding each woman's survey date. Each of three locally relevant policies (workplace closures, stay-at-home requirements, and restrictions on gathering size) is dichotomized (1 if mandatory, 0 otherwise) and summed, yielding a daily composite ranging from 0 to 3, which is then averaged over the 14-day window.

Appendix Figure 1: Distribution of primary mental health and functioning outcomes



Note: The above figures show the distribution of PHQ-9 and GAD-7 Scores, and days with functional impairment.

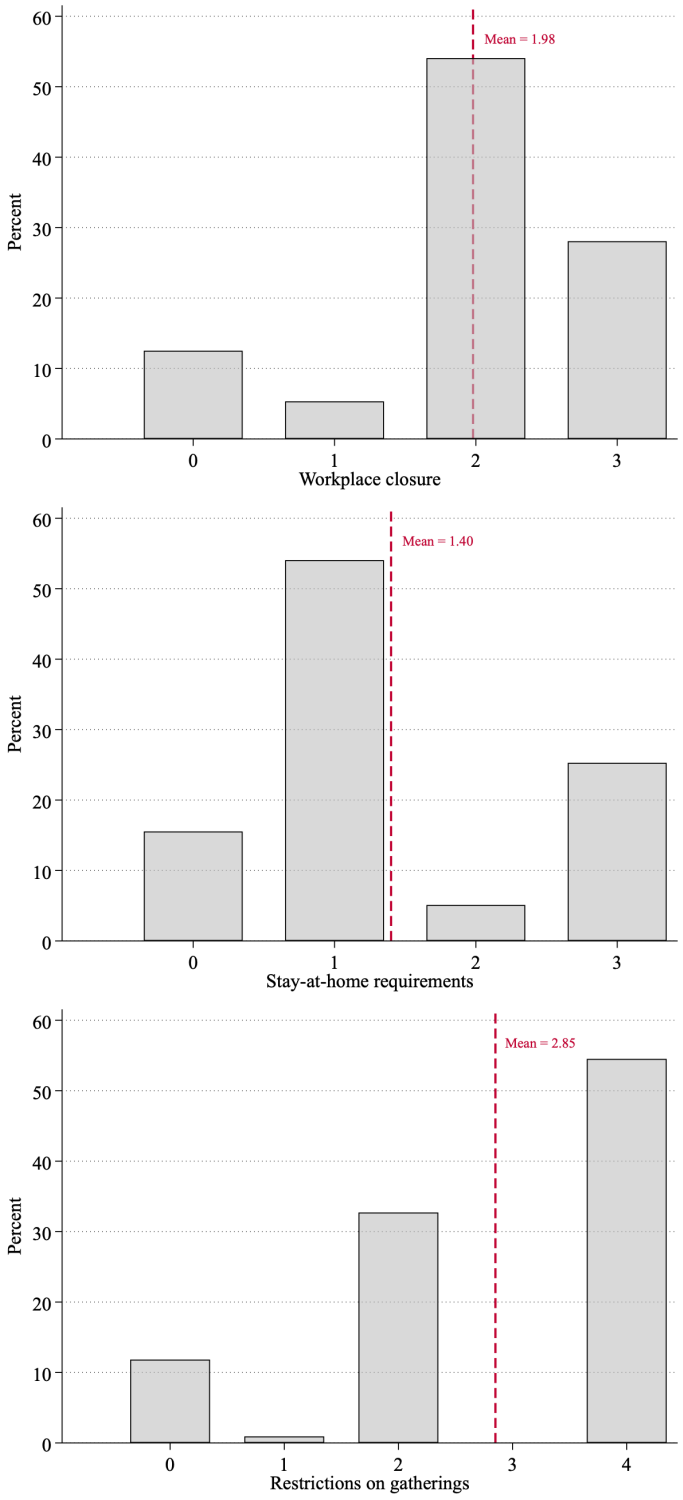
Appendix Table 1: Summary Statistics

	N	Mean	SD	Min	Max
Anxiety (GAD-7)	431	5.383	5.524	0.000	21.000
Depression (PHQ-9)	431	7.735	6.070	0.000	25.000
Days of functional impairment (WHODAS)	431	9.325	9.255	0.000	30.000
Mild depression	431	0.645	0.479	0.000	1.000
Moderate depression	431	0.334	0.472	0.000	1.000
Mild anxiety	431	0.455	0.499	0.000	1.000
Moderate anxiety	431	0.200	0.400	0.000	1.000
Mild comorbidity	431	0.418	0.494	0.000	1.000
Moderate comorbidity	431	0.165	0.371	0.000	1.000
Threat appraisal (PCA)	431	-0.000	1.000	-0.868	2.338
Stress overload (PCA)	431	0.000	1.000	-1.134	1.972
Self-efficacy and control (PCA)	431	0.000	1.000	-2.661	1.799
Hair cortisol (log)	392	1.364	0.892	-0.515	3.911
Hair DHEA (log)	392	0.522	1.549	-3.058	3.438
Family support (MSPSS)	431	0.000	1.000	-2.886	0.627
Friends support (MSPSS)	431	0.000	1.000	-1.030	1.269
Partner support (MSPSS)	431	-0.000	1.000	-3.100	0.552
Relationship quality (RFH)	427	23.574	6.380	6.000	30.000
Marital conflict	427	0.230	0.421	0.000	1.000
Child anxiety (SCAS)	430	0.000	1.000	-1.453	4.561
Child difficulties (SDQ)	430	0.000	1.000	-2.262	2.809
Urdu literacy (ASER)	425	-0.000	1.000	-1.569	3.707
Math numeracy (ASER)	425	-0.000	1.000	-1.631	2.813
Know someone who contracted COVID-19	431	0.752	0.827	0.000	2.000
Know someone who died from COVID-19	429	0.413	0.641	0.000	2.000
Received financial assistance	430	0.123	0.329	0.000	1.000
Permanent job loss in family	431	0.167	0.528	0.000	2.000

Sample restricted to depressed mothers in treatment and control arms at year-six follow-up (timepoint 7).

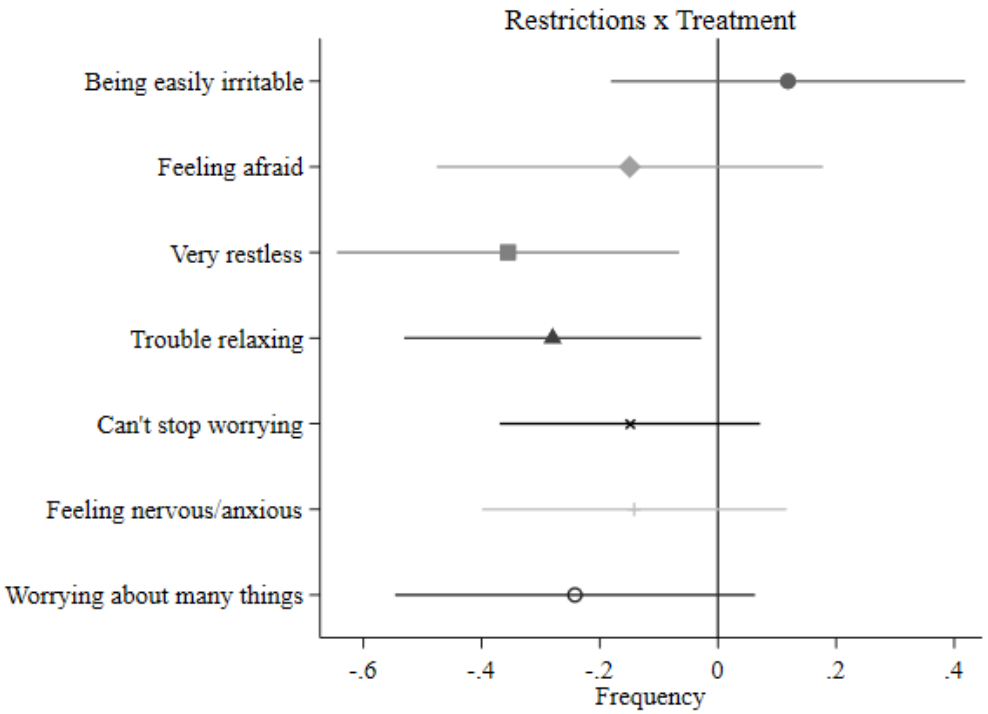
2 Additional Figures

Appendix Figure 2: Distribution of individual Covid-19 restrictions



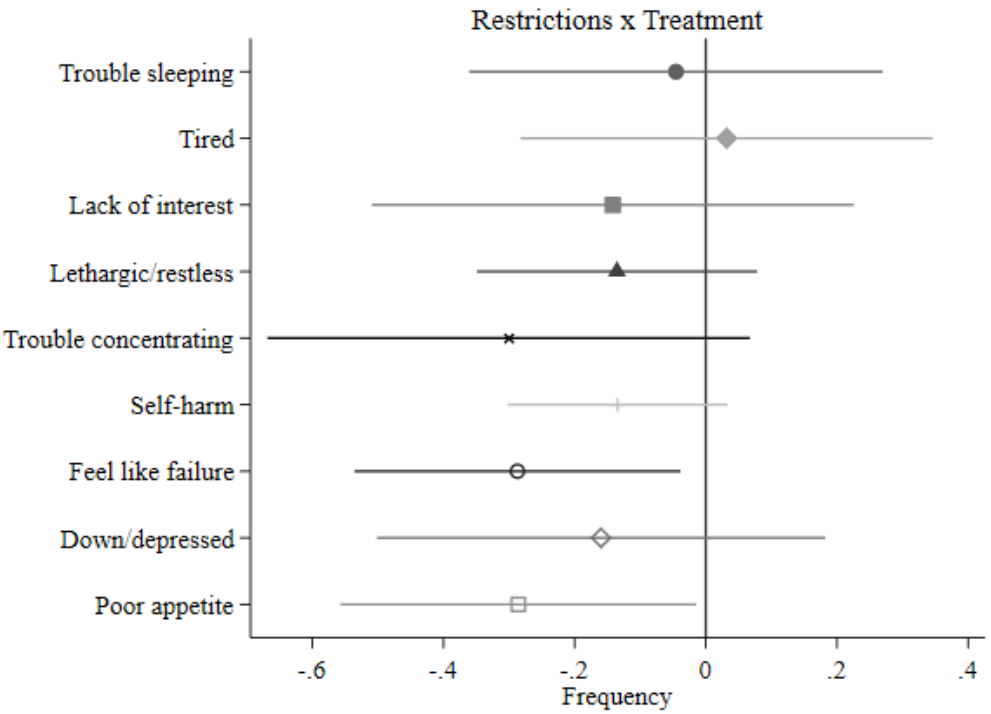
Note: The above figures show the distribution of the Covid restrictions that make up our main summary measure.

Appendix Figure 3: Interaction effects on anxiety symptoms



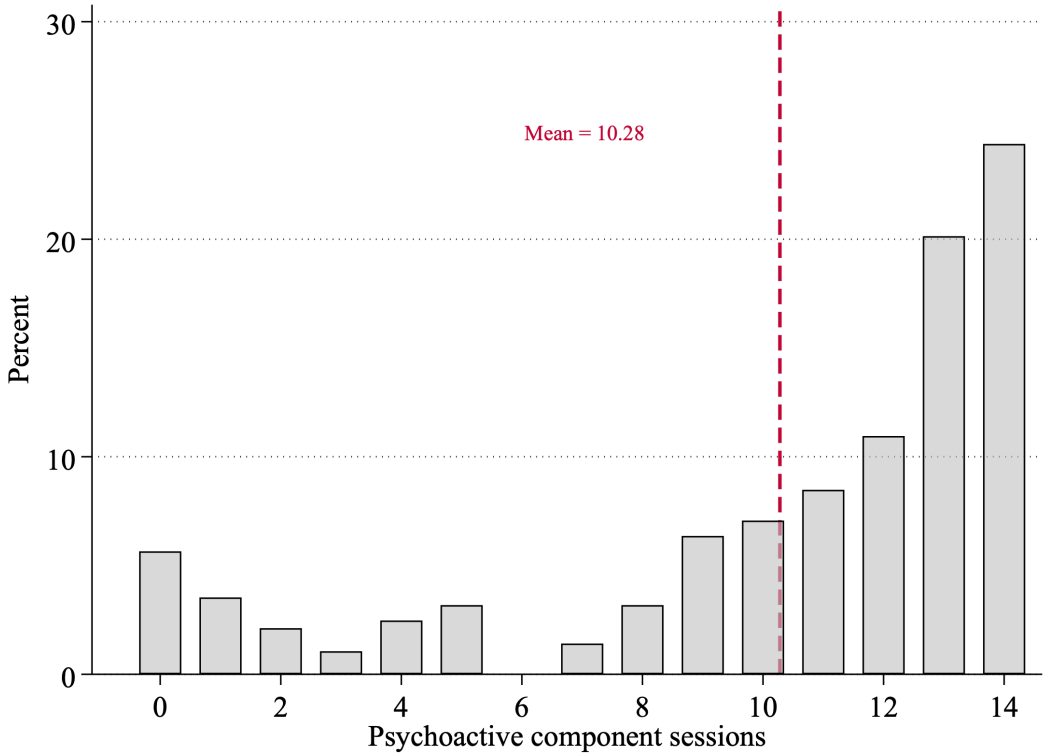
Note: The above figure plots the coefficient for the interaction of CBT treatment assignment and the average number of mandatory restrictions in the past 2 weeks on GAD-7 (Anxiety) symptoms.

Appendix Figure 4: Interaction effects on depression symptoms



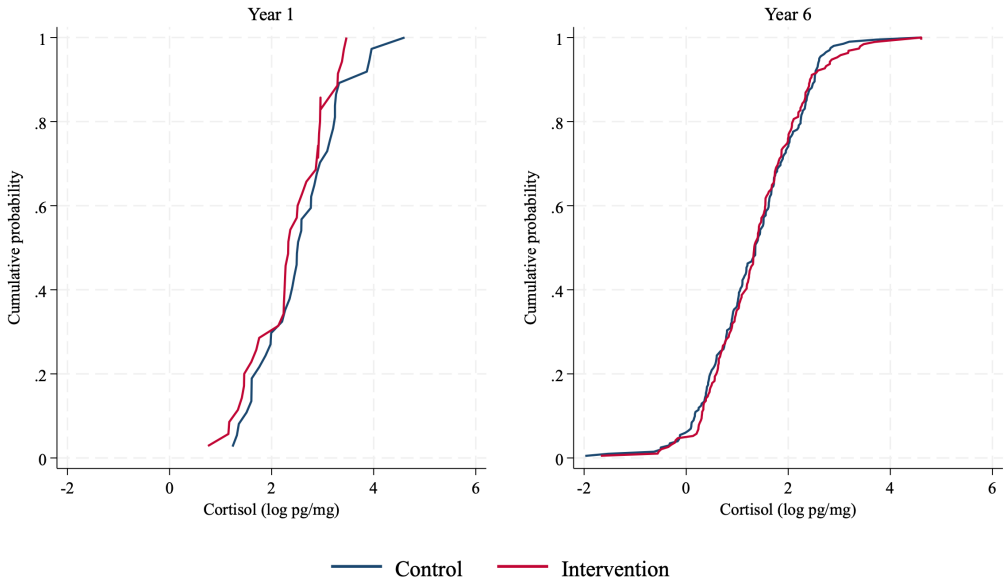
Note: The above figure plots the coefficient for the interaction of CBT treatment assignment and the average number of mandatory restrictions in the past 2 weeks on PHQ-9 (Depression) symptoms.

Appendix Figure 5: Compliance with treatment: Distribution of session attendance



Note: The behavioral component lasted from the third trimester of pregnancy until year 1 postnatal.

Appendix Figure 6: Distribution of hair cortisol by treatment arm



Note: Empirical cumulative distribution functions of log hair cortisol (pg/mg) for the intervention and control arms, one year and six years after the intervention began. Hair samples were collected from a subsample at the one-year wave (N = 72) and from the full sample at the six-year wave (N = 392). Cortisol is winsorized at 100 pg/mg. Distributions are unconditional and include no covariates.

3 Additional Tables

Appendix Table 2: Balance in attritor characteristics at Year 6

	Year 6			
	(1) Control Mean	(2) Treatment Mean	(3) Diff T-C	(4) *p*-val
Mother's age	26.94	26.53	-0.41	0.63
Mother's height (cm)	156.11	157.06	0.94	0.35
Mother's weight (kg)	59.44	59.66	0.22	0.92
Mother's waist circ. (cm)	94.13	93.42	-0.71	0.71
Mother's blood pressure	73.92	70.89	-3.03	0.22
PHQ-9 Score	14.53	15.32	0.79	0.20
WHODAS Score	16.53	17.00	0.47	0.76
PSS Score	22.97	23.59	0.62	0.63
Current major dep. episode	0.77	0.81	0.04	0.58
Nuclear family	0.29	0.35	0.06	0.43
Joint/extended family	0.61	0.51	-0.09	0.27
Grandmother present	0.62	0.51	-0.11	0.20
Total adults in the household	6.08	4.84	-1.24	0.02
First child	0.35	0.26	-0.09	0.24
Number of boys	0.61	1.04	0.43	0.01
Number of girls	0.86	1.12	0.26	0.21
People per room	2.71	2.92	0.21	0.62
SES asset index	-0.22	-0.92	-0.69	0.02
Mother's education	7.64	6.92	-0.72	0.38
Father's education	7.77	7.74	-0.03	0.96
Any psychological IPV past 12m	0.41	0.55	0.14	0.10
Any physical IPV past 12m	0.26	0.25	-0.01	0.88
Any sex-related IPV past 12m	-0.01	-0.02	-0.01	0.87
Frequency of psychological IPV	1.17	1.60	0.44	0.33
Frequency of physical IPV	0.54	0.41	-0.13	0.56
Frequency of sex-related IPV	1.93	1.62	-0.30	0.60
Joint test (*p*-value)				0.60

Note: This table tests for balance in baseline characteristics between attritors in the control and treatment arms.

Appendix Table 3: Characteristics of COVID-19 waves in Pakistan

COVID-19 Wave	Dates	Confirmed Cases	Deaths	Vaccinations
Wave 1	March 2020 – October 2020	321,877	6,621	–
Wave 2	October 2020 – February 2021	242,200	5,712	–
Wave 3	February 2021 – June 2021	379,988	9,448	3,136,386
Wave 4	June 2021 – September 2021	309,803	6,175	45,291,948
Wave 5	December 2021 – February 2022	221,825	1,401	45,667,572

Note: Data come from Ahmad et al. (2023) and Emmanuel et al. (2023). COVID-19 vaccines were not being administered during waves 1 and 2. The peaks of each wave were usually reached around the middle of each timeframe: wave 1 – June 2020, wave 2 – December 2020, wave 3 – May 2021, wave 4 – August 2021 and wave 5 – January 2022.

Appendix Table 4: Success in assessment of respondents within scheduled window by arm

	Treatment		Control		Total	
	N	%	N	%	N	%
	(1)	(2)	(3)	(4)	(5)	(6)
Assessed and DOI in window	150	53.00	171	59.58	321	56.32
Assessed outside window	59	20.85	50	17.42	109	19.12
Reached within window but not surveyed	12	4.24	7	2.44	19	3.33
Reached outside window and not surveyed	5	1.77	2	0.70	7	1.23
Complete earlier LTFU	57	20.14	57	19.86	114	20.00

Note: Table shows total number of people and % within categories of success of assessment within scheduled window by arm. A successful assessment within window would take place if the woman was surveyed within a 45-day period that starts 15 days before and 30 days after the index child's 6th birthday.

Appendix Table 5: Reasons for differences in schedule

	Assessed and DOI in window (1)	Assessed outside window (2)	Reached within window but not surveyed (3)	Reached outside window and not surveyed (4)	Complete earlier LTFU (5)	Total (6)
Illness or death in family	3	1	2	0	0	6
Harvest or Eid	0	3	0	0	0	3
Husband didn't give permission	1	0	1	1	0	3
Health worker unavailable	0	4	0	0	0	4
Mother only	0	0	1	0	0	1
Moved	3	0	12	6	0	21
Participant unavailable	0	1	0	0	0	1
Harvest (but still in window)	1	0	0	0	0	1
Sensitive info	0	0	0	0	1	1
No notes	313	100	3	0	113	529
Total	321	109	19	7	114	570

Note: In some instances, fieldworkers took notes for the reasons why people were reached and/or successfully surveyed outside the scheduled window. This table is a cross tabulation of their notes against categories of success in reaching and surveying participants.

Appendix Table 6: Baseline characteristics that predict DOI outside range

	DOI outside range		Due to Illness	
	Beta	SE	Beta	SE
	(1)	(2)	(3)	(4)
Intervention	0.066	(0.041)	0.002	(0.013)
Num. Restrictions over past 2 weeks	-0.105**	(0.047)	-0.008	(0.006)
Intervention x Num. Restrictions	-0.018	(0.062)	0.005	(0.007)
Mother's age	0.005	(0.006)	0.001	(0.001)
Mother's height (cm)	-0.001	(0.003)	-0.000	(0.001)
Mother's weight (kg)	0.004*	(0.002)	0.000	(0.000)
Mother's waist circ. (in)	-0.006	(0.007)	0.002	(0.001)
Mother's blood pressure	0.002	(0.002)	0.000	(0.000)
PHQ-9 Score	0.003	(0.006)	-0.001	(0.002)
WHODAS Score	-0.005*	(0.003)	-0.001	(0.000)
PSS Score	-0.000	(0.003)	0.001	(0.001)
Current major dep. episode	-0.031	(0.053)	0.011	(0.013)
Nuclear family	-0.002	(0.072)	-0.012	(0.019)
Joint/extended family	0.053	(0.057)	0.010	(0.016)
Grandmother present	0.044	(0.054)	0.005	(0.009)
Total adults in the household	-0.010	(0.009)	-0.001	(0.002)
First child	0.007	(0.064)	0.007	(0.012)
Number of boys	-0.007	(0.025)	0.008	(0.007)
Number of girls	-0.036	(0.024)	-0.000	(0.006)
People per room	-0.010	(0.013)	0.003	(0.004)
SES asset index	0.002	(0.016)	0.001	(0.006)
Mother's education	-0.009*	(0.006)	0.002	(0.002)
Father's education	-0.007	(0.007)	-0.000	(0.002)
Any psychological IPV past 12m	0.041	(0.051)	0.022	(0.015)
Any physical IPV past 12m	-0.061	(0.058)	0.000	(0.017)
Any sex-related IPV past 12m	0.062	(0.049)	-0.011	(0.010)

Note: (DOI = Date of interview). A DOI outside range means that the woman was assessed either before or after the 45-day window (15 days before and 30 days after the index child's 6th birthday). Columns (3)-(4) show the characteristics that predict DOI outside range due to illness as identified by the fieldworkers.

Appendix Table 7: Balance in COVID restrictions by baseline characteristics

	Aggregate measures over past 2W		Individual restrictions		
	Average # of restrictions (1)	Intensity of restrictions (2)	Work Restrictions (3)	Stay home restrictions (4)	Restrictions on Gatherings (5)
PHQ-9 Score	-0.004	-0.026	-0.007	-0.001	-0.017
	(0.011)	(0.038)	(0.011)	(0.015)	(0.016)
WHODAS Score	0.004	-0.003	0.004	-0.011	0.003
	(0.007)	(0.022)	(0.006)	(0.007)	(0.011)
PSS Score	0.008	0.038*	0.009	0.019***	0.009
	(0.006)	(0.020)	(0.006)	(0.007)	(0.010)
Mother's age	0.010	0.026	0.011	0.012	0.004
	(0.011)	(0.037)	(0.011)	(0.012)	(0.018)
Mother's height (cm)	0.009	0.026	0.009	0.007	0.010
	(0.007)	(0.023)	(0.007)	(0.008)	(0.011)
Mother's weight (kg)	-0.001	-0.005**	-0.001	-0.001***	-0.002**
	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Mother's waist circ. (in)	-0.016	-0.027	-0.016	-0.005	-0.007
	(0.014)	(0.048)	(0.014)	(0.016)	(0.021)
Mother's blood pressure	-0.004	-0.017	-0.003	-0.008**	-0.005
	(0.003)	(0.012)	(0.004)	(0.004)	(0.005)
Nuclear family	-0.274***	-0.781**	-0.235**	-0.110	-0.444***
	(0.098)	(0.325)	(0.098)	(0.094)	(0.158)
Joint/extended family	-0.220**	-0.567*	-0.223**	-0.166*	-0.178
	(0.089)	(0.295)	(0.093)	(0.092)	(0.137)
Total adults in the household	0.016	0.045	0.019	0.022	0.003
	(0.015)	(0.056)	(0.015)	(0.020)	(0.026)
First child	-0.167	-0.347	-0.170	-0.074	-0.107
	(0.165)	(0.505)	(0.162)	(0.161)	(0.213)
Number of boys	-0.073	-0.175	-0.073	-0.047	-0.058
	(0.058)	(0.186)	(0.056)	(0.076)	(0.075)
Number of girls	0.011	0.196	0.004	0.054	0.135*
	(0.059)	(0.187)	(0.059)	(0.065)	(0.078)
People per room	0.011	0.015	0.009	-0.008	0.012
	(0.026)	(0.088)	(0.025)	(0.031)	(0.040)
SES asset index	-0.040	-0.177	-0.028	-0.062	-0.089
	(0.033)	(0.122)	(0.030)	(0.039)	(0.063)
Mother's education	-0.023**	-0.089**	-0.028***	-0.026**	-0.035*
	(0.009)	(0.035)	(0.009)	(0.011)	(0.018)
Father's education	0.012	0.040	0.007	0.010	0.023
	(0.013)	(0.040)	(0.013)	(0.012)	(0.020)
Any psychological IPV past 12m	-0.127	-0.575	-0.094	-0.149	-0.330*
	(0.115)	(0.375)	(0.108)	(0.108)	(0.186)
Any physical IPV past 12m	0.142	0.615	0.107	0.082	0.424**
	(0.109)	(0.383)	(0.106)	(0.126)	(0.185)
Any sex-related IPV past 12m	-0.136	-0.242	-0.105	0.014	-0.156
	(0.201)	(0.710)	(0.196)	(0.220)	(0.340)
Joint test (p-value)	0.0043	0.0003	0.0021	0.0005	0.0051
Observations	457	457	457	457	457

Note: Table shows the association between baseline characteristics and restrictions during the year-6 survey rollout. The p-values at the bottom come from F-tests that jointly test all coefficients with the null hypothesis of balance among respondents surveyed in times with more restrictions. Robust and clustered standard errors at the cluster level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Appendix Table 8: Differences in baseline characteristics between arms by COVID restrictions

	No restrictions		1 restriction		2 restrictions		3 restrictions	
	T-C (1)	p-val (2)	T-C (3)	p-val (4)	T-C (5)	p-val (6)	T-C (7)	p-val (8)
Mother's age	-3.57	0.111	2.40	0.436	-0.36	0.561	-0.33	0.668
Mother's height (cm)	0.83	0.794	0.57	0.921	0.81	0.328	1.46	0.126
Mother's weight (kg)	-2.49	0.602	0.96	0.942	-0.95	0.565	-0.98	0.656
Mother's waist circ. (in)	-0.94	0.600	0.78	0.856	-0.70	0.189	-0.80	0.374
Mother's blood pressure	0.64	0.809	3.80	0.423	-0.40	0.810	-2.35	0.245
PHQ-9 Score	-2.41	0.111	-0.10	0.978	0.44	0.387	0.52	0.280
WHODAS Score	-9.07	0.027*	3.60	0.634	1.05	0.373	1.99	0.179
PSS Score	-4.80	0.145	1.80	0.761	0.95	0.280	1.94	0.022*
Current major dep. episode	0.04	0.842	0.15	0.679	0.04	0.637	0.06	0.395
Nuclear family	-0.26	0.186	0.00	.	0.06	0.298	-0.03	0.686
Joint/extended family	0.26	0.186	-0.10	0.797	-0.08	0.191	-0.08	0.251
Grandmother present	0.01	0.943	0.15	0.679	-0.04	0.526	-0.02	0.793
Total adults in the household	0.27	0.807	0.05	0.972	-0.70	0.059	0.27	0.607
First child	0.33	0.074	0.25	0.338	-0.01	0.853	-0.08	0.268
Number of boys	0.00	1.000	-0.45	0.623	0.05	0.646	-0.05	0.819
Number of girls	-0.07	0.876	1.10	0.250	-0.02	0.881	0.21	0.282
People per room	1.50	0.074	0.31	0.293	0.18	0.456	0.47	0.074
SES asset index	-2.30	0.001*	-0.24	0.828	0.14	0.572	-0.37	0.172
Mother's education	-2.64	0.183	1.00	0.768	0.88	0.184	-0.60	0.347
Father's education	-2.10	0.167	-1.20	0.666	-0.42	0.342	-0.96	0.035*
Any psychological IPV past 12m	-0.03	0.915	0.20	0.353	-0.02	0.792	0.02	0.813
Any physical IPV past 12m	-0.09	0.694	0.50	0.116	0.01	0.780	0.03	0.583
Any sex-related IPV past 12m	0.13	0.636	0.35	0.346	0.06	0.379	0.07	0.358
Observations	24		9		263		165	

Note: This table shows the difference in baseline characteristics between treatment and control groups by restrictions. Imbalance at 5% level or lower indicated with a *.

Appendix Table 9: Components of relationship with child's father/husband score

	Husband looks after my needs (1)	I get pocket money (2)	Husband understands my feelings (3)	Husband supports in difficult sit. (4)	Happy with husband's behavior (5)	Relationship with husband 1-bad, 5-VG (6)
Intervention	-0.186***	-0.043	-0.040	0.029	0.017	0.047
	(0.057)	(0.102)	(0.109)	(0.099)	(0.096)	(0.060)
Average # of restrictions	-0.164	-0.078	-0.257*	-0.266**	-0.275**	-0.238**
	(0.106)	(0.128)	(0.135)	(0.118)	(0.128)	(0.094)
Intervention × Restrictions	0.293*	0.354	0.491**	0.532***	0.459**	0.298**
	(0.170)	(0.229)	(0.207)	(0.177)	(0.175)	(0.127)
R^2	0.315	0.258	0.255	0.268	0.273	0.299
Outcome Mean	4.009	3.300	3.993	4.112	4.044	4.115
Observations	427	427	427	427	427	427

Note: This table displays the estimates for the average number of mandated restrictions, assignment to treatment, and their interactions. The response options for outcomes in columns (1)-(5) are 1-never, 2-almost never, 3-sometimes, 4-fairly often, 5-very often. Response options for the outcome in column (6) include: 1-very bad, 2-bad, 3-moderate, 4-good, 5-very good. Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 10: Other forms of conflict in the house

	Quarrels with family	Worse relations with children	Increase in family arguments
	(1)	(2)	(3)
Intervention	-0.092**	-0.018	-0.067**
	(0.039)	(0.033)	(0.033)
Average # of restrictions	0.035	-0.010	-0.048
	(0.053)	(0.038)	(0.049)
Intervention × Restrictions	-0.037	0.060	0.007
	(0.066)	(0.059)	(0.065)
R^2	0.170	0.218	0.232
Outcome Mean	0.292	0.223	0.316
Observations	431	431	431

Note: This table displays the estimates for the average number of mandated restrictions, assignment to treatment, and their interactions. The response options for outcomes in columns (1)-(3) are all binary. Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 11: Individual COVID-19 restrictions

	Mental Health			Spousal Conflict	
	GAD-7 Score	PHQ-9 Score	Days with Functional Difficulties	Relationship w. husband Score	Marital problems (0/1)
	(1)	(2)	(3)	(4)	(5)
Panel A: Work					
Intervention	-0.117	-0.970**	-1.630**	-0.158	0.008
	(0.376)	(0.384)	(0.669)	(0.433)	(0.028)
Restriction stringency	1.214**	0.715	1.131	-1.125*	0.097**
	(0.499)	(0.494)	(0.745)	(0.583)	(0.047)
Intervention x Restriction stringency	-1.361*	-1.289	-3.165**	2.429**	-0.190***
	(0.719)	(0.776)	(1.270)	(0.900)	(0.064)
R^2	0.316	0.301	0.298	0.299	0.205
Panel B: Stay-at-home					
Intervention	-0.022	-0.928**	-1.537**	-0.247	0.015
	(0.379)	(0.399)	(0.640)	(0.435)	(0.027)
Restriction stringency	1.254**	0.411	1.200*	-1.139**	0.091**
	(0.473)	(0.591)	(0.678)	(0.496)	(0.038)
Intervention x Restriction stringency	-1.049	-0.547	-2.351**	2.440***	-0.175***
	(0.685)	(0.888)	(1.053)	(0.811)	(0.053)
R^2	0.319	0.297	0.294	0.302	0.205
Panel C: Gatherings					
Intervention	-0.025	-0.880**	-1.444**	-0.340	0.021
	(0.370)	(0.384)	(0.674)	(0.399)	(0.026)
Restriction stringency	0.621**	0.323	0.621	-0.706*	0.051*
	(0.244)	(0.350)	(0.528)	(0.419)	(0.029)
Intervention x Restriction stringency	-0.677**	-0.832	-1.833**	1.656***	-0.114***
	(0.330)	(0.520)	(0.709)	(0.597)	(0.039)
R^2	0.312	0.301	0.297	0.302	0.203
Outcome Mean	5.383	7.735	9.325	23.574	0.230
Observations	431	431	431	427	427

Note: The restrictions are continuous variables between 0 and 3 (4 for gatherings), denoting whether there were no restrictions (0), whether restrictions were recommended but no mandated (1), whether they were required but only for segments of the population (2), and whether they were required for all except essential workers (3). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 12: Alternative Covid-19 pandemic markers

	Mental Health			Spousal Conflict	
	GAD-7 Score	PHQ-9 Score	Days with Functional Difficulties	Relationship w. husband Score	Marital problems (0/1)
	(1)	(2)	(3)	(4)	(5)
Panel A					
Intervention × Number of restrictions	-1.079**	-1.293*	-2.660**	2.167***	-0.174***
	(0.502)	(0.673)	(1.049)	(0.782)	(0.055)
R^2	0.316	0.302	0.297	0.300	0.206
Panel B					
Intervention × Intensity of restrictions	-1.165**	-1.094	-2.820**	2.532***	-0.184***
	(0.532)	(0.739)	(1.045)	(0.848)	(0.055)
R^2	0.317	0.300	0.298	0.304	0.208
Panel C					
Intervention × Stringency Index	-0.089	-0.475	-2.193**	1.244	-0.091*
	(0.611)	(0.688)	(0.938)	(0.802)	(0.048)
R^2	0.309	0.298	0.295	0.291	0.192
Panel D					
Intervention × New cases	-0.558	-0.428	-1.704*	0.762	-0.094**
	(0.573)	(0.606)	(0.920)	(0.801)	(0.046)
R^2	0.310	0.297	0.294	0.288	0.195
Panel E					
Intervention × New deaths	-0.510	-0.176	-2.030**	0.635	-0.111**
	(0.571)	(0.697)	(0.917)	(0.847)	(0.051)
R^2	0.313	0.297	0.295	0.286	0.196
Outcome Mean	5.383	7.735	9.325	23.574	0.230
Observations	431	431	431	427	427

Note: All 5 pandemic measures have been standardized (with a mean of zero and s.d. of one) to facilitate comparisons across regressions for each outcome. All markers are averaged over the 14 days prior to (but including) the date of the survey. Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 13: Controlling for weather

	Mental Health			Spousal Conflict	
	GAD-7 Score	PHQ-9 Score	Days with Functional Difficulties	Relationship w. husband Score	Marital problems (0/1)
	(1)	(2)	(3)	(4)	(5)
Intervention	-0.119	-1.022**	-1.647**	-0.202	0.011
	(0.389)	(0.388)	(0.670)	(0.421)	(0.027)
Average # of restrictions	1.225**	0.493	0.676	-1.269*	0.099**
	(0.456)	(0.543)	(0.990)	(0.650)	(0.047)
Intervention × Restrictions	-1.301**	-1.039	-2.370*	2.598**	-0.215***
	(0.636)	(0.810)	(1.316)	(0.968)	(0.062)
R^2	0.318	0.315	0.308	0.302	0.213
Outcome Mean	5.383	7.735	9.325	23.574	0.230
Observations	431	431	431	427	427
Main Controls	Y	Y	Y	Y	Y
Maximum temperature and total rain on DOI	Y	Y	Y	Y	Y

Note: DOI (day of the interview). The main controls include: At baseline (PHQ-9, WHODAS, PSS, woman's height, waist circumference, joint household, adults in the house, people in the room, asset score, mother's education, father's education, sexual IPV, Union Council FE, assessor FE), at year 6 (assessor FE, month and month squared). Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 14: Placebo restrictions and mental health

	GAD-7 Score			PHQ-9 Score			Functional Difficulties
	Combined	Year 3	Year 4	Combined	Year 3	Year 4	Year 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intervention	-0.227	-0.694	0.838	-0.167	-0.844	0.668	3.615
	(1.473)	(2.495)	(0.730)	(0.863)	(1.288)	(0.685)	(3.387)
Average # of restrictions	-0.201	-1.016	0.023	0.251	0.238	0.130	0.885
	(0.359)	(1.300)	(1.071)	(0.359)	(0.645)	(0.941)	(1.111)
Intervention × Restrictions	-0.227	0.468	-1.239	-0.455	-0.190	0.063	0.148
	(0.465)	(1.551)	(1.324)	(0.489)	(1.048)	(1.277)	(2.004)
R^2	0.272	0.389	0.324	0.263	0.348	0.321	0.306
Observations	637	299	338	760	422	338	422

Note: Placebo restrictions are constructed by taking the of average number of restrictions in the 2 weeks prior to the survey from year 6 and imposing them on all observations in years 3 and 4 based on the order of the rollout in each year. In each wave, the rollout started in the month of January and lasted for approximately 20 months. Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 15: Stress appraisal items

	Threat Appraisal				Stress Overload				
	Feel inadequate	Odds are against me	Nothing going right	There is no escape	Swamped by responsibility	No time to get to everything	Felt rushed	Things keep piling up	Carrying heavy load
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intervention	-0.055	-0.177	-0.060	-0.019	-0.036	-0.071	0.006	0.013	-0.104
	(0.136)	(0.118)	(0.127)	(0.103)	(0.155)	(0.124)	(0.136)	(0.139)	(0.134)
Average # of restrictions	0.135	0.334**	0.253**	0.180*	0.229**	0.136	-0.037	0.131	0.340***
	(0.124)	(0.124)	(0.125)	(0.101)	(0.090)	(0.120)	(0.147)	(0.122)	(0.122)
Intervention x Restrictions	-0.033	-0.122	-0.208	-0.098	-0.529***	-0.225	-0.085	-0.305	-0.532***
	(0.213)	(0.189)	(0.200)	(0.185)	(0.175)	(0.177)	(0.213)	(0.182)	(0.184)
R^2	0.222	0.264	0.283	0.250	0.203	0.237	0.281	0.245	0.260
Outcome Mean	2.202	2.044	2.181	1.923	2.538	2.153	2.413	2.905	2.353
Observations	431	431	431	431	431	431	431	431	431

Note: Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 16: Perceived control and self-efficacy items

	Perceived Control and Self-Efficacy					
	Locus of control	Manage to solve difficult problems if try	I can find several solutions to problems	If in trouble can think of a solution	Unable to control important things in life	Things not going my way
	(1)	(2)	(3)	(4)	(5)	(6)
Intervention	0.200**	0.157*	0.108*	0.176**	-0.091	0.032
	(0.078)	(0.091)	(0.062)	(0.067)	(0.082)	(0.075)
Average # of restrictions	-0.056	0.007	0.071	-0.093	0.204**	-0.212**
	(0.112)	(0.101)	(0.100)	(0.101)	(0.088)	(0.096)
Intervention x Restrictions	0.136	-0.057	-0.248*	-0.038	-0.301**	0.270*
	(0.162)	(0.145)	(0.131)	(0.139)	(0.135)	(0.145)
R^2	0.245	0.177	0.283	0.243	0.238	0.233
Outcome Mean	-0.000	-0.000	-0.000	-0.000	-0.000	0.000
Observations	431	431	431	431	431	431

Note: Clustered-robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix Table 17: Session attendance and perceptions of social support

	Family			Friends			Significant Other		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sum of all sessions	0.010*			0.011			0.010**		
	(0.006)			(0.011)			(0.005)		
# Sessions Psych Component		0.018			0.026			0.022**	
		(0.012)			(0.021)			(0.009)	
# Booster			0.020*			0.016			0.016
			(0.011)			(0.023)			(0.010)
Outcome Mean	5.93	5.93	5.93	3.69	3.69	3.69	6.09	6.09	6.09
Observations	431	431	431	431	431	431	431	431	431

Note: Robust and clustered standard errors at the cluster level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Ahmad, Taimoor, Mujahid Abdullah, Abdul Mueed, Faisal Sultan, Ayesha Khan, and Adnan Ahmad Khan. 2023. "COVID-19 in Pakistan: A national analysis of five pandemic waves." *PLOS ONE*, 18(12): e0281326.

Amirkhan, James H. 2018. "A Brief Stress Diagnostic Tool: The Short Stress Overload Scale." *Assessment*, 25(8): 1001–1013.

ASER Pakistan. 2019. "Annual Status of Education Report: ASER-Pakistan 2019." Idara-e-Taleem-o-Aagahi, Lahore, Pakistan.

Emmanuel, Faran, Anusheh Hassan, Ahsan Ahmad, and Tahira E Reza. 2023. "Pakistan's COVID-19 Prevention and Control Response Using the World Health Organization's Guidelines for Epidemic Response Interventions." *Cureus*.

Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: A Research Note." *Journal of Child Psychology and Psychiatry*, 38(5): 581–586.

Kroenke, Kurt, Robert L Spitzer, and Janet B W Williams. 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine*, 16(9): 606–613.

Spence, Susan H. 1998. "A measure of anxiety symptoms among children." *Behaviour Research and Therapy*, 36(5): 545–566.

Spitzer, Robert L, Kurt Kroenke, Janet B W Williams, and Bernd Löwe. 2006. "A Brief Measure for Assessing Generalized Anxiety Disorder." *Archives of Internal Medicine*, 166(10): 1092.

Zimet, Gregory D, Suzanne S Powell, Gordon K Farley, Sidney Werkman, and Karen A Berkoff. 1990. "Psychometric Characteristics of the Multidimensional Scale of Perceived Social Support." *Journal of Personality Assessment*, 55(3-4): 610–617.