Quantitative Political Economy Research Group
Department of Political Economy
King's College London

QPE

Forced Displacement, Mental Health, and Child Development: Evidence from the Rohingya Refugees

QPE Working Paper 2022-45

Asad Islam
Tanvir Ahmed Mozumder
Tabassum Rahman
Tanvir Shatil
Abu Siddique

March 14, 2022

Forced Displacement, Mental Health, and Child Development: Evidence from the Rohingya Refugees*

Asad Islam† Tanvir Ahmed Mozumder‡ Tabassum Rahman§ Tanvir Shatil¶
Abu Siddique‖

March 14, 2022

Abstract

Forced displacement is a major driver of mental disorders among refugees worldwide. Poor mental health of adult refugees, particularly mothers, is also considered a risk factor for the psychological well-being and development of their children. In this study, we experimentally examine the extent to which a multifaceted psychosocial program improves the mental well-being of refugee mothers, and facilitates growth and development among children under the age of two. In partnership with BRAC, we ran a cluster randomized controlled trial on 3,500 Rohingya mother-child dyads in refugee camps in Bangladesh. Participants were given weekly psychosocial support for a year that includes psychoeducation and parenting support for mothers and play activities for both mothers and children. The intervention was largely successful and led to: (i) reductions in the psychological trauma and depression severity of mothers and children, (ii) improvements in communication, gross-motor, problem-solving, and social skills of children, and (iii) reductions in stunting, underweight, and wasting among children in the treatment group. The intervention also caused the mental health of children to be more aligned with the mental health of their mothers, implying policies targeting the mental well-being of displaced mothers can be an important stepping stone to developing psychological resilience among their children, which can help them grow into well-rounded, healthy adults.

JEL Classification: I15, J15, O12, O15

Keywords: Mental health, forced displacement, early childhood development, randomized experiment, psychoeducation, Rohingya refugees.

*We thank Sonia Bhalotra, Victoria Baranov, Narayan C Das, Fabian Dehos, Adeline Delavande, Amrita Dhillon, Steven Dieterle, Eliana La Ferrara, Soledad Giardili, Sebastian Goerg, Patricia Justino, Tatiana Kornienko, Kostas Matakos, Imran Matin, Mushfiq Mobarak, Ana Nuevo-Chiquero, Vikram Patel, Catherine Porter, Smriti Sharma, and Saurabh Singhal for many helpful comments and suggestions. We also thank seminar participants at Australian National University, Bath, Duisburg-Essen, Edinburgh, Indian Statistical Institute-Delhi, King's College London, Lancaster, Monash, Newcastle, Technical University of Munich, and conference participants at AEA Meeting 2022, CEPR/LEAP Workshop in Development Economics, Bavarian Micro Day Workshop, Maastricht-BEPS, Measuring Development at Berkeley, Mental Health Workshop at Max Planck Institute, MWIEDC at Northwestern, NEUDC at Boston, Workshop on Conflict, Migration, and Displacement, Young Economists Meeting at Masaryk, and Field Days Workshop at Cologne for many helpful comments. We also thank Erum Marium and Sakila Yesmin from the BRAC Institute of Educational Development and BRAC fieldworkers in Rohingya refugee camps for extensive support and collaboration for the entire project. This project has received ethics approval from BRAC University (ref no. 2019-028-ER) and is registered at the AEA RCT Registry (AEARCTR-0004516).

†Centre for Development Economics and Sustainability (CDES) and Department of Economics, Monash University; J-PAL. Email: asadul.islam@monash.edu

‡BRAC Institute of Governance & Development, BRAC University. Email: tanvir.mozumder@bracu.ac.bd

§School of Medicine & Public Health, University of Newcastle. Email: tabassum.rahman@uon.edu.au

¶BRAC Institute of Governance & Development, BRAC University. Email: tanvir.shatil@bracu.ac.bd

‖Department of Political Economy, King's College London; CDES; TUM. Email: abu.b.siddique@kcl.ac.uk

1 Introduction

Evergrowing conflict, persecution, natural disasters, and famine have forcibly displaced over 84 million people worldwide (UNHCR, 2021b). Among the forcibly displaced, one-third are refugees and about 85% of refugees are hosted by developing countries (UNHCR, 2021b). The cost of hosting refugees is also substantial for low- and middle-income countries (Chambers, 1986; Jacobsen, 2005; Taylor et al., 2016). As a result, governments in developing countries primarily rely on foreign emergency assistance and donations to support refugees (Taylor et al., 2016). However, the budget for basic needs such as food, healthcare, and education for refugees often falls short due to inadequate donations (Jacobsen, 2005).[1] Such shortfalls have several negative repercussions in poor settings, including deteriorated living conditions, increased chance of mortality and morbidity, and impaired human capital formation during early childhood (UNHCR, 2021a).

A key factor that further aggravates this problem is the declining mental health of refugees. Because of adverse life experiences and exposure to various stress factors—such as violence and persecution, separation, financial strain, and uncertainty—common mental disorders such as depression, trauma, and anxiety among refugees are widespread (Porter & Haslam, 2005; Steel et al., 2009). The most vulnerable among the forcibly displaced refugees are children and women—they are at a higher risk of dying of preventable causes and their poor mental health may exacerbate this risk (UNHCR, 2021a,b). The economic costs of poor mental health and malnutrition during childhood are also substantial, as it can hinder human capital accumulation, cause poor mental health during adulthood, and can prolong the cycle of their poverty (Currie & Stabile, 2006; Heckman et al., 2006; Currie, 2009; Adhvaryu et al., 2019; Ridley et al., 2020). On the other hand, women easily develop complex traumatic syndrome from trauma caused by terror and violence (Herman, 2015), and have a higher depression burden than men (Baranov et al., 2020). Therefore, poor mental health coupled with resource and healthcare shortages in the camps can harm the physical health, productivity, and network formation of adult refugees with long-term detrimental economic consequences. In addition, poor mental health of primary caregivers of children, such as mothers, can increase their mental cost of childcare and, hence, their investment and practices during early childhood (Patel et al., 2004; Rahman et al., 2008). Given the high prevalence of malnutrition and diarrhoeal diseases among child refugees, neglected childcare can also adversely affect children's cognitive, non-cognitive, and anthropometric development (Mensah & Kiernan, 2010; UNHCR, 2021a). Therefore, governments and humanitarian agencies in poor settings are in need of a low-cost and scalable program that can concurrently address the concerns of mental health and early childhood development of refugees.

In this study, we present evidence from a cluster randomized controlled trial to evaluate the impact of a multifaceted psychosocial support program that was designed to improve the psychological well-being of refugee women (mothers, henceforth) and mental, socioemotional, physical, cognitive, and anthropometric development of their children under the age of two. The intervention, known as the home-based Humanitarian Play Lab (HPL), was pioneered and implemented by BRAC Bangladesh on a sample of 3,500 Rohingya refugee mother-child dyads

[1] For instance, Bangladesh requires about USD 1.2 billion every year to support the one million Rohingyas residing in refugee camps in Bangladesh; however, only 40% of the total could be managed in 2021 due to a sudden decline in foreign donations for protracted crises (Tayeb, 2021a).

located in refugee camps in Bangladesh. The focus was on the Rohingya people—a severely persecuted ethnic and religious minority from Myanmar. The 2017 incidence of mass genocide and community violence in Myanmar caused a mass displacement of about 750,000 Rohingya people to Bangladesh, where currently a million Rohingyas live as refugees in confined camps. In the camps, the mental health of Rohingya women and children is alarmingly poor, and acute malnutrition, anemia, and stunting are rampant among children (The Lancet, 2019; Hossain et al., 2019). Moreover, the legal status of Rohingya refugees in Bangladesh does not allow their social inclusion, participation in employment, or mobility out of camps, which imposes an additional mental toll on the refugees. Besides, insufficient access to health services, scarce food, and basic shelters in the camps make the refugees vulnerable to contracting potentially fatal diseases (Ahmed et al., 2018). Against this background, the home-based HPL intervention was carefully designed by the BRAC Institute of Education and Development with support from psychologists and early childhood experts. The primary aim is to address the health concerns of refugee mothers and children in poor settings using a low-cost and scalable solution. Therefore, the intervention involved psychoeducation and parenting support for mothers—carefully delivered by community peers who worked as volunteers—and play activities for mothers and children.[2] It was provided on a weekly basis for a year, from October 2019 until September 2020. The intervention was largely successful and very cost-effective—costing about USD 45 per mother-child dyad—and is currently being scaled-up by BRAC in the Rohingya refugee camps.

We find that the intervention significantly improved the psychological well-being of treated mothers. Specifically, mothers that received psychosocial support experienced a 0.23 standard deviation (SD) reduction in trauma and a 0.14 SD reduction in depression relative to mothers in the control group that did not receive the psychosocial intervention. Among the mothers that were identified to have trauma and depression at baseline, we observe a sizeable improvement in their mental well-being—effects that are comparable to the short-run impacts of recent interventions targeting the mental health in poor settings (Hussam et al., 2021; Barker et al., 2021). In addition, treated mothers also experienced an improvement in their self-reported level of happiness (0.12 SD) and sense of belongingness in the host community (0.18 SD) following the intervention, but we do not find any noticeable impact on their aspirations for the future. Furthermore, mentally unhealthy mothers that received the treatment caught up to, and often surpassed, the mental health of the 'mentally-healthy' mothers in the control group following the intervention. Thus, the intervention was largely successful at lifting refugee mothers out of psychological distress and potentially prevented long-term mental illness, such as persistent depressive disorder.

With respect to child outcomes, children in the treatment group also experienced statistically significant reductions in trauma (0.10 SD) and depression (0.12 SD) relative to children in the control group. We also observe that the intervention improved treated children's communication skills (i.e., speech and language development) by 0.23 SD, gross-motor skills (i.e., physical activities and whole-body movements) by 0.18 SD, problem-solving skills (i.e., learn to

[2]Psychoeducation is an established psychosocial support method that integrates light-touch psychotherapeutic and educational interventions to help people cope with mental health problems (American Psychological Association, 1995). It educates people in mental hardship about the possible reasons of their distress and simple ways of addressing it. It also facilitates discussion and sharing of various positive and negative feelings with others, which helps people identify the challenges they are facing and their personal coping abilities (Lukens & McFarlane, 2004). Given its simplicity, psychoeducation can be easily delivered by non-experts from poor settings with limited educational background.

play with toys and solve puzzles) by 0.18 SD, and personal-social skills (i.e., care for themselves and interactions with others) by 0.13 SD relative to the control-group children. In addition, we also estimate the impact on the prevalence of malnutrition among children, such as being underweight (weigh too little or too thin for age), stunting (skeletal growth retardation), and wasting (acute undernutrition). We find that children in the treatment group experienced: (i) a large increase in weight-for-age z-score by 0.64 SD (20% or 635 grams higher weight) than children in the control group, which also translates to 7 percentage points (or 10%) reduction in being underweight and 16 percentage points (or 26%) reduction in being severe underweight; and, (ii) a large increase in height-for-age z-score by 0.52 SD (19% or 1.58 centimeters taller) than children in the control group, which also translates to 7 percentage points (or 10%) reduction in stunting and 13 percentage points (or 22%) reduction in severe stunting. Both improvements in weight and growth are also reflected in the incidence of wasting among children, where treated children experienced a large increase in weight-for-height z-score by 0.51 SD (21% higher) than children in the control group, which also translates to 8 percentage points (or 14%) reduction in wasting and 13 percentage points (or 30%) reduction in severe wasting. Thus, BRAC's home-based HPL intervention was largely successful at improving the nutritional, mental, socioemotional, physical, and cognitive development of the Rohingya children. These impacts have strong policy implications, particularly for low- and middle-income countries hosting refugees where resources for refugees are often limited and more-costly child development approaches, such as cash transfers, pediatric support, nutritional supplements, special healthcare needs, etc., are often unavailable and not prioritized by host governments.

Our unique design with mother-child dyads also allows us to examine the intergenerational correlation of mental health between mothers and children, and the causal impact of the intervention on the mother-child mental health gap. We find that the mental health of mothers is strongly and positively correlated with the mental health of their children at baseline and this pairwise correlation strengthened following the intervention, implying more aligned mental health between mothers and their children at endline. In fact, the intervention was successful at reducing the mental health gap (both in terms of trauma and depression) between mothers and their children, possibly because it allowed children to absorb and integrate the mental health of their mothers due to increased interactions. This further suggests that interventions targeting the mental well-being of mothers can be an important stepping stone to developing psychological resilience among their children. This is very important in contexts where psychosocial support facilities for infants and small children are scarce and unavailable.

To understand the mechanisms at play, we first estimate the effect of the intervention on various intermediate outcomes. We do not find any statistically significant evidence that the intervention led to improvements in mothers' physical health, relationships with spouses, or other potential channels. Therefore, either directly or via unobserved channels, the intervention influenced the mental well-being of mothers. With respect to children's mental health and development, we find strong evidence that the intervention increased mothers' daily involvement with their children by about 1.5 hours, but muted impacts on fathers' time input, implying maternal time investment on children is a potential channel. We also find that treated mothers are less likely to allow their children to play or walk barefoot (which prevents hookworm infections and exposure to various bacterial and fungal organisms) and less likely to engage in negative

parenting, suggesting improvements in mothers' health behaviors toward their children are other potential mechanisms for their children's development. However, other maternal behaviors and practices, such as breastfeeding time, feeding frequency, seeking others' help to babysit, and discouraging fathers to smoke indoors, are unlikely to be the channels of impact. We also consider a formal mediation analysis following Heckman et al. (2013); Heckman & Pinto (2015) to quantify the mediated impacts: (i) how much of mother's mental well-being and children's development contributed to children's improvements in mental well-being; and, (ii) how much of mothers' and children's mental health improvements contributed to children's development outcomes. We find that 55% of depression reductions and 83% of trauma reductions among children were due to improvements in their mothers' mental health and their own socioemotional, physical, and cognitive development. We also find that 20% of improvements in child development outcomes can be jointly explained by mothers' and children's mental health improvements. The remaining effects were either due to the direct effect of the intervention or other unobserved mediators.

Finally, we examine heterogeneity in treatment effects using machine learning (Chernozhukov et al., 2020). With respect to the mental well-being of mothers, those that had poor mental health at baseline, high exposure to violent conflict in Myanmar, and experienced relatively more abuse in refugee camps benefited the most from the intervention. Furthermore, there are indications, although weaker, that older and illiterate mothers benefited the most. In terms of children's skills development and anthropometric outcomes, older children benefited the most throughout. Moreover, boys on average experienced higher nutritional benefits from the program than girls. However, results on the heterogeneity by exposure to violence in Myanmar and abuse in camps are rather mixed for child development. We also do not find any heterogeneity by baseline mental health on child development and anthropometric outcomes.

To summarize, using a randomized experiment on the Rohingya refugees, we show that a multifaceted psychoeducation, parenting support, and play-activity program can be successfully set up in resource-poor settings, such as in refugee camps in developing countries, where both mothers and children between 0-2 years can be simultaneously targeted and provided weekly support through peers. We show that this multifaceted psychosocial program was largely successful in many ways and very cost-effective. To the best of our knowledge, our study provides the first experimental evidence on improving the mental health of vulnerable Rohingya refugees through psychosocial support. We also provide the first experimental evidence on the impact of a psychosocial intervention on the intergenerational transmission of mental health and early childhood development in a post-conflict setting. In all, we contribute to several strands of literature. First, the emerging literature on the impacts of psycho- and non-psycho-therapeutic interventions on mental health, behavioral change, and economic decision-making and outcomes (Bolton et al., 2003; Rahman et al., 2008; Blattman et al., 2017; Baranov et al., 2020; Haushofer et al., 2020; Angelucci & Bennett, 2021; Hussam et al., 2021; Vlassopoulos et al., 2021; Barker et al., 2021),.[3] Second, our paper is also related to the literature on the effect of health, parenting, and maternal mental health on early childhood development (Patel et al., 2004; Rahman et al., 2008; Bhalotra et al., 2020; Attanasio et al., 2020; Baranov et al., 2020; Cappelen et al., 2020; Doyle, 2020), as well as the literature on the intergenerational transmission of mental health, such as depression

[3]Similar to our context, Hussam et al. (2021) offered eight-weeks long employment opportunities to Rohingyas living in refugee camps in Bangladesh and finds that the mental benefits from being employed surpasses the mental benefits of receiving cash transfers among the refugees.

and trauma (Ahlburg, 1998; Ackard et al., 2006; Johnston et al., 2013; Gonçalves et al., 2016; Eyal & Burns, 2019). Third, we also contribute towards the understanding of the detrimental consequences of violent conflicts, adverse migration experiences, intimate partner violence, and poverty on mental health (Fischbach & Herbert, 1997; Summerfield, 2000; Schilbach et al., 2016; Ridley et al., 2020). Finally, more broadly, our study also relates to the literature on the negative consequences of adverse life experiences during childhood (Almond et al., 2009; Minoiu & Shemyakina, 2012; Akresh et al., 2012; Hanna & Oliva, 2016; Singhal, 2019), and the importance of early-life interventions on child development and human capital accumulation (Miguel & Kremer, 2004; Breierova & Duflo, 2004; Bleakley, 2007; Bharadwaj et al., 2013; Amadu et al., 2019; Alan et al., 2021; Carneiro et al., 2021; Attanasio et al., 2022).

We have organized the rest of our paper as follows. Section 2 provides the context and section 3 provides the experimental design and details of the psychoeducation program. Section 4 reports the main results, while section 5 explores some possible mechanisms. Section 6 reports the results on the heterogeneous treatment effects using machine learning. In section 7, we discuss some fieldwork challenges and the program's cost-effectiveness. We briefly conclude in section 8.

2 The context

> "When I try to sleep, I imagine what the military has done to me. I feel like they are coming, chasing, and shooting me... When I am in bed, the imagination of the torture appears in my mind."
>
> Rashida Begum, a Rohingya woman (Fortify Rights, 2020).

> "At that moment I felt like I was already dead. I think I am only alive to tell the world about what I saw."
>
> Rajuma, a Rohingya woman (Motlagh, 2018).

> "As Rahima ran through a forest on the way to neighboring Bangladesh, three uniformed Myanmar soldiers grabbed her. For two nights, they kept her in a jungle clearing and gang-raped her, smoking methamphetamine to sustain the torture."
>
> Beech (2017) on Rahima's experience during the genocide.

The Rohingya people of Myanmar (previously Burma) are an ethnic, linguistic, and religious minority who are usually referred to as 'foreigners' by other ethnic groups in Myanmar. They have lived in the Rakhine State for centuries until they were stripped of their citizenship in 1982 (Cheung, 2011; Mahmood et al., 2017). Rohingyas have been subject to repeated waves of persecution and forced displacement since Myanmar's independence in 1948. Around 200,000 Rohingyas fled to Bangladesh in 1978 when the Burmese military started a violent operation to screen out 'foreigners' from citizens (Cheung, 2011). Similar operations and displacement also took place after the 1991-92 elections. Likewise, in late 2012, due to increased communal violence and conflict against the Rohingya by the Rakhine residents, and later by the Burmese military, Rohingyas once again began fleeing Myanmar to their neighboring countries. A new wave of violence against the Rohingya people spurred in 2017, also known as 'ethnic cleansing' by the Burmese military, forced the majority of Rohingyas to seek refuge in neighboring Bangladesh

(Beyrer & Kamarulzaman, 2017). During this incident, about 24 thousand Rohingyas were killed, 18 thousand women and girls were raped, 34 thousand were thrown in the fire, 114 thousand were severely beaten, and over 100 thousand households were burned down or vandalized (Habib et al., 2018). Since 2017, almost 1 million Rohingya people have been residing in crowded settlements in southern Bangladesh, among which 81% arrived after the 2017 incident (UNHCR Population Factsheet, 2019). This makes them one of the largest groups of stateless people in the world.

Compared to previous attempts of ethnic cleansing by the Myanmar government, the most recent attempt had the largest impact by displacing more than two-thirds of the entire 1.5 million Rohingya population from Myanmar (Mahmood et al., 2017). According to UNHCR Population Factsheet (2019) and UNHCR Camp Profiles (2019), among the 1 million refugees currently residing in Cox's Bazar, Bangladesh, 55% are children with 41% being below the age of 11 and 18% below the age of 4. Also, 52% of the overall refugees are female. Besides, 16% of 211,383 Rohingya families residing in refugee camps are run by single mothers and 63% of families have over 4 family members. Moreover, these camps consist of 31% vulnerable families, such as separated children and families with single mothers, with at least one protection vulnerability.[4]

According to WHO Situation Report (2017), around half of the Rohingya children in refugee camps in Bangladesh are malnourished, underweight, and suffering from anemia, and 25% of the children under 5 have acute malnutrition. Moreover, 38% of children have stunted growth—very close to the WHO critical health emergency threshold of 40% (Save the Children, 2018)— and over 80,000 children have severe mental distress, which is one in every five children in the camps (Save the Children, 2019). Malnutrition is also rampant among adult men and women in the camps, with many experiencing chronic starvation (Hodal, 2017).

Immediately after fleeing Myanmar, over 80% of Rohingya women reported having depressive and emotional distress symptoms, and 60% had post-traumatic stress disorders (Fortify Rights, 2020). In the refugee camps, gender-based violence is very common, where most violence is initiated by intimate partners, relatives, or other camp members (Beech, 2017). With respect to healthcare access, two-thirds of pregnant women have minimal to no access to gynecological and obstetrical care services (WHO Situation Report, 2017). This is because, camps are grossly underfunded with few healthcare centers operating within camps (Hodal, 2017). Seeking healthcare outside camps is also not possible due to mobility restrictions, as mobility is confined within their designated camps. Moreover, they cannot be employed, start new income-generating activities, or send their children to schools outside the camps due to legal restrictions. Therefore, they rely entirely on government support, foreign donors, and humanitarian agencies for food, healthcare, and shelter. For food, a senior woman in each household receives a food voucher every month from camp authorities. Camps are also dense, with about 90 thousand people living in one square kilometer. Due to the lack of privacy, many women "...avoid eating because they feel they do not have proper access to toilets (within camps)" (Beech, 2017).

The Bangladeshi government needs at least USD 1.2 billion every year to support Rohingya people, where 10% is required for basic healthcare support (Tayeb, 2021a). Moreover, over 30 thousand infants are born every year in camps that require frequent and quality health support from mothers and health experts (Tayeb, 2021a). Recent estimates also show that it will take 12

[4]More details on demographics and camp profiles can be found in UNHCR Population Factsheet (2019) and UNHCR Camp Profiles (2019).

years if the Bangladeshi government repatriates 300 Rohingyas every day but, without repatriation, the government needs about USD 7 billion to provide basic support to Rohingya people for the next 5 years (Tayeb, 2021a).

3 Experimental design

3.1 The home-based Humanitarian Play Lab program

The program. The international development organization BRAC Bangladesh has developed a multifaceted psychosocial program—part of their umbrella program 'Humanitarian Play Lab (HPL)'—to foster the psychosocial well-being of Rohingya women that are mothers of children below 2 as well as mental, nutritional, socioemotional, physical, and cognitive development of their children. Specifically, experts from the BRAC Institute of Education and Development (BIED) with support from psychologists and early-childhood experts developed this low-cost, easy-to-implement one year program for the Rohingya refugee mothers and children. It includes 44 different session modules to be delivered on a weekly basis, in a home setting. This program was developed as an urgent measure for the severely persecuted and forcibly displaced Rohingya mothers and children, with a view to implementing it at scale shortly after evaluating its impact through a cluster randomized controlled trial.

The multifaceted home-based HPL program has three important components: (i) *Psychoeducation*, which is an established psychosocial support method that integrates light-touch psychotherapeutic and educational interventions to help people cope with mental health problems (American Psychological Association, 1995; Lukens & McFarlane, 2004). Therefore, this feature helps the Rohingya refugee mothers cope with various mental distresses and psychological trauma that they have endured during the 'ethnic cleansing' attempt in Myanmar by educating them about the possible reasons of their distress and simple ways of addressing it. As a result, mothers get a thorough understanding of the challenges they are facing, their (own) coping ability, and areas of strengths and weaknesses. This allows mothers to better address their day-to-day problems, feel more in control of their feelings and actions, and have a higher internal capacity to achieve psychological well-being. (ii) Another feature of this multifaceted program is that it provided *parenting support* to mothers, highlighting the importance of childcare and playing with children for their healthy development. (iii) Finally, mothers also engaged in various *play activities* with their children and other participating mothers. Children also engaged in free-play activities with age-appropriate toys, such as balls, dolls, blocks, legos, and so on, as a play-based early stimulation.

The program was delivered by trained Rohingya women who volunteered for the task (known as mother volunteers or MV) and are from the same neighborhood as the participants.[5] Thus, it was a peer-delivered intervention, provided to small groups of participating mothers at MV's home on a weekly basis, with each session lasting for 60 minutes. MVs were trained by mental health and early-childhood experts and received support (when required) by psychosocial expert counselors during the intervention period. More details on MVs are provided in Appendix C.1.

[5] MVs were hired by BRAC program managers and camp-in-charges based on their level of education, fluency in Bangla and Rohingya languages, and field management skills. Priority was given to women who knew how to read and write, and willing to set-up sessions at their homes.

Session procedure. Enrolled mothers in the treatment arm were required to attend a session (which is the home of an MV) along with their children on a weekly basis. Each session was divided into four steps: (i) greetings, (ii) my well-being, (iii) the baby grows up, and (iv) homework. The order and description of the four steps in each session are provided below:

- **Step 1:** The first step is called ***Greetings***, which involved greeting one another and doing breathing exercises, primarily to induce relaxation among attendees. MVs also asked participants about their homework from the previous week and participants shared their experiences. This step was identical in every session and ran for 15 minutes. Pictorial directions of this step are given in Figure A1.
- **Step 2:** The second step is called *"Amar Bhalo Thaka"* or ***My well-being***, which focused on providing mental health support to mothers. The topic varied every week, including sharing positive and negative feelings with other participating mothers and MV, sharing happy memories, advice on self-care (such as healthy diet, the importance of sleep, nurturing hobbies, etc.), the importance of communication, positive thinking and how it can help coping, acceptance and tolerance, emotional development, and so on. Mothers also played various games, such as hole tarp, bank-a-ball, etc., and participated in various art activities. During this step, children also engaged in free-play activities with age-appropriate toys. This step differed across sessions, where mothers received different self-care and relaxing advice and engaged in different play and art activities across sessions. This step ran for 20 minutes. Some pictorial directions of this step are given in Figure A2.
- **Step 3:** The third step is called *"Shishur Bere Otha"* or ***'Baby's growing up***, which focused on parenting support. Participating mothers received advice from MV on childcare, such as cleanliness, feeding, nutrition, ways to massage a baby for better sleep, etc., and the importance of playing with children and its influence on psychological and physical well-being. Mothers were taught how they can play with their toddlers and children with various household items, such as using a pillow, handkerchief, etc. Mothers also engaged in play activities with their children during this step (e.g., peekaboo, toy hunt, counting fingers, etc.). Analogous to Step 2, topics in this step also differed across sessions, where mothers received different childcare advice and engaged in different play activities with their children across sessions. This step ran for 20 minutes. Some pictorial directions of this step are given in Figure A3.
- **Step 4:** The final step is called ***Homework***, where participating mothers were assigned homework based on the weekly topic. This step ran for 5 minutes.

The full curriculum (translated into English) is available here. Randomly selected mother-child dyads in the treatment group received the psychosocial program. In contrast, mothers in the control arm did not receive the psychosocial program but attended an *unstructured* social gathering with other mothers (thus, there was no curriculum, structured discussions, or MV to facilitate psychoeducation) on a weekly basis. This allows us to disentangle the effect of the program from the effect of attending social gatherings. All sessions were conducted in the local Rohingya language. We describe our sampling and randomization in detail in section 3.2.

COVID-19 and mobile phone sessions. Due to the COVID-19 pandemic, Bangladesh went into a nationwide lockdown on March 26, 2020. Thus, after delivering 24 in-person sessions, the remaining 20 sessions were conducted over mobile phones (via basic feature phones) due

to strict social distancing rules. This intervention was not stopped after 24 in-person sessions for two reasons: first, experts from BIED recommended completing the entire curriculum of 44 sessions; and, second, due to humanitarian reasons, as the COVID-19 lockdown and uncertainty were likely to impose a further mental toll on these vulnerable refugees.

To accommodate over-the-phone sessions, session duration and structure were carefully revised by experts from BIED. Thus, various play activities, group activities, and group discussions could not be conducted over mobile phones. Otherwise, session steps, topics, and the order remained unchanged. These sessions were conducted by the same MVs as in the group sessions and were implemented on a one-to-one basis. Each over-the-phone session lasted for 20 minutes. In total, 87% of enrolled women from baseline had access to mobile phones (statistically similar across treatment arms); however, camp managers and block-*Majhis* (leaders of each block) managed to lend their mobile phones to the remaining 13% of women.[6] During this period, control group women did not receive any (placebo) calls or did not engage in any unstructured social gatherings. There were 20 weekly over-the-phone sessions. More details on over-the-phone sessions are provided in Appendix C. In section 3.6, we also show that the characteristics of women that had mobile phones are very similar to the characteristics of those without mobile phones.

Program Timeline. Figure 1 shows the intervention timeline. The program began in early-October 2019 and ended in late-September 2020. However, it temporarily ceased its operation in late March 2020 due to the COVID-19 lockdown. Later, over-the-phone support replaced face-to-face sessions, which began in mid-May 2020 and ended in late September 2020. Our baseline data collection began in late July and ended in late September 2019, whereas the endline data collection began in early October and ended in late October 2020. The endline survey was conducted over the phone due to COVID-19 restrictions. No midline data collection was carried out prior to starting over-the-phone sessions. We describe data collection in detail in section 3.3.1.

3.2 Research design

We evaluate the home-based HPL program using a cluster randomized control trial. Following arrival in Bangladesh, Rohingya people were provided with shelter in confined refugee camps in Cox's Bazar district. Each of these camps consists of many *Majhee* blocks (blocks hereinafter), which are clusters of many households and can be considered "neighborhoods".[7] We use this geographic-level information, which is blocks within the camps, for randomization. At the time of randomization, there were over 2,000 blocks distributed across 17 refugee camps where BRAC operates (out of 34 camps in total). We randomly selected 251 blocks from the universe of over 2,000 blocks, of which 137 were assigned to the treatment (55%) and 114 were assigned to the control group (45%). Figure A4 in Appendix A shows a camp map and blocks therein, highlighting the treatment and control blocks.

Within each block, we randomly created two groups, where each group attended an MV's

[6] Note that participants were not forced to borrow mobile phones. Every week, prior to a scheduled session, *Majhis* went to participants' doors and offered them their mobile phone for the session. After about an hour, *Majhis* went back to collect the mobile phone at the door. Qualitative feedback from *Majihs* suggest that mothers were not reluctant to borrow mobiles, rather were very enthusiastic. Note also that social distancing measures were strictly followed and disinfectants provided by BRAC were applied on mobiles after each use.

[7] *Majhees* are the (unelected) leaders of blocks. Each block has its own *Majhee*.

home (or pocket) throughout the year for the sessions. We had a total of 226 pockets in treatment and 191 in control blocks. For each session, we randomly invited roughly 7 mother-child dyads on average. From BRAC's list of Rohingya households, project assistants and MVs randomly visited households that met the selection criteria—mothers with at least one child between the age of 46 days and 24 months—and invited the mothers to participate in the home-based HPL program. In case a mother had multiple children within this age category, we randomly selected one child for the intervention. A total of 3,499 mother-child dyads were enrolled to participate in this program. Only mothers in the treatment arm received our weekly treatment, while mothers in the control arm participated in unstructured (or unsupervised by an MV) social gatherings that did not involve psychoeducation, parenting support, or play activities.

3.3 Data

3.3.1 Surveys

We collected survey data both at the baseline and endline of this program (see Figure 1 for the project timeline). The baseline was conducted by enumerators from BRAC on a one-to-one basis (face-to-face) with participating mothers. The endline, on the other hand, was conducted over mobile phones due to COVID-19 restrictions. Enumerators—both females and males—are Bangladeshi from the Ukhiya region in the Cox's Bazar district and are fluent in the *Rohingya* language (in fact, the *Chittagonian* language is very similar to the *Rohingya* language). They are highly trained with several years of survey experience. Our baseline questionnaires were divided into three broad parts that cover: (i) socioeconomic background; (ii) mother's characteristics—demographics, migration and camp experiences, exposure to violence both during the conflict in Myanmar and at the refugee camp, mental health conditions, the current state of life, etc.; and, (iii) adverse life experience of children—demographics and mental health conditions; and age-specific cognitive skills development of children, such as communication, motor, problem-solving, and personal-social skills. At endline, only outcome and mechanism variables were collected. All survey questions were answered by the mothers.

Dedicated and highly trained anthropometric enumerators also measured and collected children's height and weight using appropriate scales at baseline. At endline, due to COVID-19 restrictions, trained anthropometric enumerators instructed mothers over the phone to measure children's height and weight. Height was measured by mothers themselves, using their right hand and index finger. Although 'hand' or 'palm' measures are obsolete anthropic units of length, it was the only way we could measure anthropometric outcomes during the pandemic without risking the health of mothers, children, enumerators, and other Rohingyas in the camp. 'Hand' length is the length between the mid-point of the wrist's distal transverse crease and the tip of the middle finger, and 'finger' width is the width of the index finger (see Figure A12 in Appendix A). Mothers reported lengths to enumerators in 'hand' and 'finger' units (also, 'half-hand' or 'half-finger' units were considered). Later, 'hand' and 'finger' units were converted to centimeters (cm) following Asadujjaman et al. (2019), where one 'hand' equals 16 cm and one 'finger' equals 2 cm. Weight, on the other hand, was measured by asking mothers to first weigh their child by lifting her/him using both hands and then weighing one kilogram of rice sack (holding using both hands) and then reporting back to the enumerator (see Figure A13 in Appendix A). Mothers

reported the weight of the child in 'rice sack' units (also, half, quarter, fifth, and tenth of the rice sack units were considered). Later, the 'rice sack' unit was converted to grams and kilograms. In section 4.4, we discuss whether these measures are too noisy and also subject to experimenter demand effects.

At baseline, a total of 3,499 mothers were surveyed: 1,911 from the treatment and 1,588 from the control groups. At endline, 2,845 mothers were surveyed (using mobile phones), 1,679 from the treatment, and 1,166 from the control arm. Therefore, by the endline, roughly 19% of mothers could not be surveyed. We discuss attrition in detail in section 3.6.

3.3.2 Outcomes

We have five mother and ten child outcomes. Outcomes, other than child anthropometrics, were constructed by aggregating survey questions: (i) converting each response into a dummy, so if a question was answered on a 5-point Likert scale, then the maximum two response points were coded as 1 and the remaining three points as 0; (ii) then aggregating the dummy responses into a scale; (iii) from each scale, subtracting the control group mean and dividing this difference by the control group standard deviation. Eventually, we got indices such that the control group has mean zero and standard deviation one. All outcomes and the specific survey questions used to construct indices were pre-registered at the AEA RCT Registry (AEARCTR-0004516). We define our outcomes below (with specific survey questions listed in Appendix B):

Mental health outcomes

Psychological trauma. We broadly explore psychological trauma by combining post-traumatic stress disorder (PTSD) and acute stress disorder (ASD), as their symptoms overlap.[8] According to the American Psychiatric Association (2013), the primary etiological reason for developing PTSD and ASD is experiencing traumatic events that involve exposure to actual or threatened violence (sexual, physical, verbal, etc.), serious injury, or death. An adult or a child is said to suffer from trauma if s/he experiences any combination of the following four major symptoms: (i) intrusion (such as having recurrent distressing memories or flashbacks, nightmares, etc.), (ii) avoidance (such as avoiding memories or thoughts of the traumatic event or people, conversations, or activities that remind of such events), (iii) negative mood (such as unable to feel happy or satisfaction), (iv) marked alterations in arousal and reactivity (such as sleep disturbance, being easily startled, or emotional outbursts). To measure mothers' trauma through survey questions, we partially use the simplified Kessler Psychological Distress Scale (Andrews & Slade, 2001) and combine it with other survey questions that were developed following the diagnostic criteria for PTSD and ASD as laid out in American Psychiatric Association (2013). We measure children's psychological trauma through an adverse life experience survey by Dyregrov et al. (2000); Neugebauer et al. (2009).

Depression. People develop depression due to various reasons, where experiencing violence

[8]Trauma and stressor-related disorders emerge due to exposure to traumatic and stressful events and include post-traumatic stress disorder (PTSD), acute stress disorder (ASD), reactive attachment disorder, disinhibited social engagement disorder, and adjustment disorders (American Psychiatric Association, 2013). We only look at the impact on PTSD and ASD, the two most common disorders developed from exposure to injury, violence, or death, and leave out the latter three because reactive attachment disorder and disinhibited social engagement disorder occur mostly among children due to inadequate parenting and neglect. Moreover, we leave out adjustment disorders because the symptoms are very similar to that of depression and is also a primary driver of depressive disorder at a later stage.

(sexual, physical, verbal, etc.) or loss of family and identity are thought to be the major drivers of depression. According to American Psychiatric Association (2013), the diagnostic criteria for a major depressive episode, or depression, is experiencing at least five of the following nine symptoms: (i) feeling sad, empty, or hopeless almost every day, (ii) loss of interest in various activities, (iii) dramatic change in body weight or appetite, (iv) difficulty sleeping or insomnia, (v) feeling restless, (vi) feeling tired or fatigue, (vii) feeling worthless or guilt, (viii) unable to concentrate or make decisions, and (ix) recurrent thoughts of death, suicide, or self-harm. To measure depression, we use the Center for Epidemiologic Studies Depression Scale-20 that consists of self-reported measures of depressive symptoms (Radloff, 1977). We measure children's depressive symptoms using an adverse life experience survey by Dyregrov et al. (2000); Neugebauer et al. (2009).

Subjective well-being of mothers

We measure the subjective well-being of mothers that look into their levels of **happiness**, hope and **aspirations** about the future, and their sense of **belongingness**. As refugees go through the psychological stress of searching for identity (Kumsa, 2006), measures of belongingness inform us about well-being related to their general social identity. For more details see Appendix B.

Child development

We investigate development outcomes of children that explore different developmental progress associated with their socioemotional, cognitive, and physical development at an early age. These are **communication**, **gross-motor**, **fine-motor**, **problem-solving**, and **personal-social skills**. To measure this five-set of skills, we use the widely used Ages and Stages Questionnaire (ASQ-3) (Squires & Bricker, 2009), where questions are grouped into categories dedicated to assessing a specific set of skills and are also age-specific. That is, this questionnaire has different questions for 2, 4, 6, etc., months-old children.[9] All survey questions are answered by the mothers. A detailed description of these outcomes is given in Appendix B.

Child anthropometrics

We explore children's stunting (or shortness), underweight (or lightness), and wasting (acute undernutrition) by looking at their height-for-age (HAZ), weight-for-age (WAZ), and weight-for-height (WHZ) z-scores respectively following the World Health Organization (WHO) defined guidelines (WHO, 2009).[10] According to WHO, the criterion for stunting is when $HAZ < -2$ standard deviations (i.e. 2 standard deviations/SD below the median in reference population), severe stunting is when $HAZ < -3$ SD, underweight is when $WAZ < -2$ SD, severe underweight is when $WAZ < -3$, wasting is when $WHZ < -2$, and severe wasting is when $WHZ < -3$. We use both z-scores and dummy variables constructed using these cut-offs as our anthropometric outcomes.[11]

[9] Questionnaires for 2 months-old children applies to 0-2 months-olds, 4 months-old children applies to 3-4 months-olds, 6 months-old children applies to 5-6 months-olds, and so on.

[10] We compute anthropometric z-scores using the 2006 WHO child growth standards. The formula to compute z-score of child i is: $z - score = [(observed\ value) - (median\ reference\ value)] \div (standard\ deviation\ of\ the\ reference\ population)$, where the median reference value is the median height or weight of all girls/boys who are i's age and the reference population is (American) children under 2 years who are chosen by WHO and National Center for Health Statistics. See WHO (2006, 2009) for more details.

[11] We also pre-registered mother-child relationship as an outcome (also measured at baseline) using the Ages and Stages Questionnaires: Social-Emotional (ASQSE-2) (Squires et al., 2015). However, due to limitations on the

3.4 Sample characteristics and balance checks

We report the balance on observables at baseline between treatment and control groups in Table 1 and balance on baseline outcomes in Table A1. In both tables, we report means and standard deviations. To derive p-values on tests of equality of means across treatment and control groups, we regress the variable of interest on the binary treatment (=1 if treatment group and 0 otherwise) with camp fixed effects and standard errors clustered at the unit of randomization (which is blocks). Our mother and child samples are well balanced across individual and household characteristics and average differences in almost all observables are very small. In terms of outcomes measured at baseline (Table A1), again our samples are well balanced. Comparing the differences in distributions of mental health at baseline (shown in Figure A6 in Appendix A), we find that treatment group distributions are statistically similar to control group distributions using a Kolmogorov-Smirnov test (all $p > 0.10$).

Note that, we did 33 independent tests. Therefore, corrections for multiple-hypothesis testing substantially reduce the significance threshold, and, thus, the two significant differences that we observe disappear following such adjustments. In addition, following Imbens & Wooldridge (2009); Imbens & Rubin (2015) we also compute the normalized differences in means for all variables to show the scale-free differences.[12] The idea is that increasing the sample size can increase the t-statistic but it does not systematically affect the normalized difference. We find that, out of the 33 normalized differences, 32 differences are lower than $1/8^{th}$ of the combined sample variation and only one difference is below $1/3^{rd}$ (variable 'child victim of at least one camp abuse'). The general rule of thumb is that if a difference exceeds one quarter, then linear regression methods are likely to be sensitive to specification changes (Imbens & Wooldridge, 2009). In any case, we control for all characteristics that differ in terms of mean or normalized differences when estimating treatment effects (see section 3.7).

3.5 Program take-up and attendance

The initial program take-up rate (agreed to attend weekly sessions and, thus, participated in baseline survey) following invitation was about 95% (3,499 out of 3,700 invited). Less than 5% did not take it up after the initial invitation because they were either responsible to take care of the elderly or they needed permission from their spouses. For the 95% that initially enrolled for the program, we also kept a record of their weekly participation in sessions through MV. As there were no MVs involved in managing control group social gatherings, we only managed to keep the weekly participation record in the treatment group.

Out of 1,911 treatment participants, 11 (or 0.6%) participants never attended any sessions, while the remaining participants attended at least one out of forty-four sessions.[13] Therefore, the actual take-up among the enrolled is over 99%. On average, participants attended 20.4 sessions (median is 20 sessions), which can be seen in Figure A5 in Appendix A. We have a very high

questionnaire length and interview duration at endline (during COVID-19), we could not administer ASQSE-2.

[12] We compute the normalized difference for each variable by first taking the difference in means (treatment group mean minus control group mean) and then dividing this difference by the square root of the sum of the variances.

[13] Only five participants' attendance is missing, as we could not match their names in the attendance register to their names in the initial enrolment sheet. Thus, we have attendance record of 1,906 out of 1,911 in the treatment group. If we consider these 5 participants as 'never-attended', then the total number of participants that never attended any sessions is 16 (or 0.8% of 1,911).

program take-up and attendance, possibly because—as mentioned in section 2—refugees in these camps cannot leave their designated camps or work in income-generating activities outside. As a result, mothers mostly spend idle time in their home after finishing household chores and, thus, have ample discretionary time. Moreover, the program was provided by BRAC Bangladesh—an organization playing a pivotal role in refugee well-being within the camps—and sessions were conducted within neighborhoods at another (neighbor) Rohingya woman's home, whom participants possibly trust and are familiar with.

3.6 Attrition

We successfully followed up on 2,845 mother-child pairs (out of 3,499) at endline, 1,679 in the treatment group (out of 1,911) and 1,166 in the control group (out of 1,588). Given the follow-up survey was conducted during the COVID-19 pandemic, the overall attrition rate was surprisingly low at 19% (or 654 women). We also observe the attrition rate to be about 14 percentage points higher in the control group relative to the treatment group (T-test: $p < 0.01$), and this 14 pp difference translates to 190 women. To check if individual and household characteristics measured at baseline predict attrition at endline, we compare baseline characteristics of mothers/children that attrited to the baseline characteristics of mothers/children that remained at endline. This is reported in Table A2 in Appendix A. We find that characteristics are fairly similar, with the exception of mothers being the household head (significant at 5% level). We also observe that attrited mothers are 'marginally' newer in the camp and children are shorter in height (both significant at 10% level), but the magnitudes of these differences are very small. To further examine the differences in baseline observables, we regress attrition (equals to 1 if attrited at endline and 0 otherwise) on baseline characteristics, treatment dummy, and the interaction between baseline characteristics and the treatment dummy (see Table A3 in Appendix A). A joint F-test in columns 1 and 2 suggest that baseline characteristics can jointly explain attrition within the control group, but not in the treatment group. However, a joint F-test on the interactions yields a p-value=0.19, suggesting attrition was not differential by baseline characteristics.[14] Although we find that observable characteristics of those who attrited versus those who did not are fairly similar across treatment arms, the 14 percentage points gap in endline participation between treatment and control raises the concern that attrition may bias the treatment effects estimated later in section 4. We address this concern using four different approaches in section 4.5.

As mentioned in section 3.1, about 87% of our respondents (or another household member) had at least one mobile phone, and mobile phone ownership was similar across treatment arms (T-test: $p = 0.916$). The remaining participants were lent mobile phones owned by *Majhis* (leaders of each block) and camp managers. Therefore, out of 654 women that attrited at endline, 75 women did not own a phone (mobile was offered but they did not borrow it) and the remaining 579 attrited women had access to mobile phones but did not participate in the endline survey. Among those that participated in the endline survey (2,845 women), 381 participants did not own a mobile phone but chose to borrow the phone to participate in the endline survey. To check if baseline characteristics explain mobile phone ownership in the camps and whether it also differs

[14] Surprisingly, in the treatment arm, average session attendance of those who attritted is 22 compared to 20 among those that did not attrit, and this difference of 2 sessions is statistically significant at 1% level.

by treatment, we regress the mobile ownership dummy on the treatment, baseline characteristics, and interactions between these two. This result is reported in Table A4 in Appendix A. We find that baseline characteristics jointly explains (although marginal) phone ownership within treatment group (joint $p = 0.053$, column 1), but not control group (joint $p = 0.290$, column 2). However, a joint test suggests that these characteristics do not jointly differ across treatment groups ($p = 0.60$, column 3).

3.7 Empirical strategy

Impact on outcomes. To test the impact of the program on mothers and children outcomes, we postulate our main empirical model as follows:

$$Y_{1ijc} = \beta_0 + \beta_1 Treat_{jc} + \beta_2 Y_{0ijc} + \mathbf{\Gamma}' \boldsymbol{X}_{ijc} + \theta_c + \epsilon_{ijc} \tag{1}$$

where Y_{1ijc} denotes the outcome of mother/child i in block j located in camp c, measured at the endline. $Treat_{jc}$ is a binary variable that indicates the treatment status of block j in camp c; and $\boldsymbol{X}_{ijc}$ is a vector of controls, measured at the baseline, that include age, gender of the child, whether mother attended school, monthly household spending, household size, number of children (or siblings for child outcomes), months spent in the refugee camp, whether the mother receives food vouchers, whether the mother is the household head, whether the husband is alive, whether at least one family member remains stranded in Myanmar, household's victimization based on the conflict in Myanmar (an index between 0 and 1, where higher value corresponds to more exposure to violence), mother's and child's victimization based on refugee camp abuse (both indices, between 0 and 1). Y_{0ijc} is the baseline analogue of the outcome. θ_c is camp fixed effects, so that the comparisons are between blocks in the same refugee camp. Error terms are assumed to be independent across blocks, but we allow them to be correlated within the same block (i.e., the standard errors are clustered at the block level, which is our unit of randomization). As outcomes other than anthropometrics are indices with the control group having mean 0 and standard deviation 1, our estimated β_1 for such outcomes determines where the mean of the treatment group lies in the distribution of the control group in terms of standard deviations (Kling et al., 2007). Since trauma and depression indices are based on 'negative' feelings, negative $\hat{\beta}_1$ correspond to an improvement in mental health. For the remaining outcomes, positive coefficients correspond to more favorable outcomes. We estimate equation 1 using OLS, where β_1 is the intent-to-treat (ITT) effect. As session attendance is very high with over 99% of participants attending at least one session (i.e., taken-up the program) and less than 1% not attending any session, we can also interpret the ITT effects as the treatment-on-treated (TOT) effects. As less than 1% did not participate in any sessions, we decided not to carry out a separate TOT analysis using a two-stage least squares (2SLS) regression analysis.

Inference. First, we cluster standard errors at the unit of our randomization, which is by blocks. Second, even though the number of clusters per arm is somewhat large (more than 110 clusters in each arm), for robustness, we also compute p-values using randomization-based inference (RI) with randomization permuted at the cluster level (Young, 2019). For this, we use 1,000 replications.[15] In regression tables that report treatment effect estimates, we also report

[15] Young (2019) suggests that draws beyond 2,000 make little to no difference to p-values. Our conclusions do

the Young (2019) RI p-values. Results reported in the following section are largely robust to using this method.

Correction for multiple hypotheses testing. With many outcomes of interest (15 outcomes), it is crucial to correct p-values for each outcome that we test. To address this issue, we check the robustness of our results using the List-Shaikh-Xu procedure that uses bootstrapping (with 3,000 replications) to account for joint correlation across different tests and then controls the probability of making any type-I error (or the familywise error rate (FWER)) (List et al., 2019). In each regression table, where we report the treatment effects, we also report the FWER-adjusted p-values for each test. We also check the robustness of our results using the Westfall-Young adjustment (Westfall & Young, 1993). Though we do not report FWER p-values using Westfall & Young (1993) in the tables, our conclusions are largely consistent using both methods. Moreover, we aggregate the mental health, mothers' subjective well-being, and child development outcome measures into composite indices to reduce the number of tests (also reported in the main table). Our results are also robust to this adjustment.

4 Results

We divide this section into five parts. First, we present results on the impact of the intervention on mothers (section 4.1), and, second, on children (section 4.2). Third, we present results on whether the intervention affected the transmission of mental health from mothers to children (section 4.3). Fourth, we carry out a series of robustness checks (section 4.4). Finally, we explore the impact of differential attrition on our treatment effect estimates (section 4.5).

4.1 Impact of the intervention on mothers

Mental health. Figure 2 and Table 2 report the impact of the intervention on mothers' mental health (Panel A1) and subjective well-being outcomes, such as happiness, aspirations, and belongingness (Panel A2). Column 1 reports treatment effects without controlling for baseline controls and column 2 reports estimates with the full set of controls. Since results with and without controlling for baseline characteristics are similar, we focus our discussions below only based on estimates reported in column 2.

We find that the intervention has significantly improved the mental health of Rohingya mothers. Specifically, mothers that received mental health counseling experienced a 0.23 standard deviation (SD) reduction in trauma ($p < 0.01$) and 0.14 SD reduction in depression severity ($p < 0.01$) relative to mothers in the control group that did not receive the psychosocial program (Panel A1, Table 2). When we explore the impact of the intervention only among mothers that were identified as traumatized (1,557 out of 3,499 mothers) and depressed (645 out of 3,499 mothers) at baseline, we find that the reduction in trauma is 0.26 SD among the traumatized, which is slightly higher than the aggregate effect of 0.23 SD. However, the reduction in depression among the depressed is sizable at 0.29 SD (a twofold improvement) relative to depressed mothers in the control group. As over 99% of mothers attended one or more sessions, we believe these ITT effects $\approx$ TOT effects. In Figure A7 in Appendix A, we also explore the correlation between the number of sessions attended and the mental health of mothers in the treatment arm (note

not change if we use 2,000 replications.

attendance was only recorded in the treatment arm). The negative linear fits suggest that more attendance is correlated with the better mental health of mothers (pairwise correlation tests: $p < 0.01$ in Plot A and $p = 0.01$ in Plot B).

In comparison to the short-run impacts of other mental health interventions in developing countries (Baranov et al., 2020; Vlassopoulos et al., 2021), our estimated impacts are relatively smaller. One possible reason is that participants in the control group in our study also participated in social gatherings on a weekly basis (which is mostly muted in existing studies), and as social interaction has a positive impact on people's mental health (Nezlek et al., 1994), improvements in the mental well-being of our control group participants may have reduced the size of the impact. However, even with a placebo, our effect sizes are largely comparable to the effect sizes of a recent non-therapeutic study conducted on the Rohingya refugees in Bangladesh (Hussam et al., 2021).

Subjective well-being. We then consider outcomes related to mothers' subjective well-being in terms of happiness, aspirations for the future, and belongingness (Panel A2). We find that the happiness and belongingness of mothers in the treatment group increased by 0.12 SD ($p < 0.05$) and 0.18 SD ($p < 0.01$) respectively relative to mothers in the control group. However, in terms of aspirations, the treatment effect is muted. We also illustrate these treatment effect estimates in Figure 2, where we show where the mean of the treatment group lies in the distribution of the control group in terms of SD units. Under each 'pooled' result, we also present results by child's gender (graph A) and by exposure to violence during the conflict in Myanmar (graph B). We do not find the impacts to vary by these characteristics.

Are the mentally unhealthy catching up to the healthy? Next, in Panel A1, Table A5, we examine whether the mentally-unhealthy mothers in the treatment group are catching up to the mentally-healthy mothers in the control group in terms of mental health and subjective well-being. We find that our intervention was successful at aligning the depression severity of the treated mothers that were depressed at baseline with the depression severity of the control mothers that were healthy at baseline (suggested by the statistically insignificant coefficients). It implies that the mentally unhealthy in the treatment group successfully caught up to the mentally healthy in terms of depression. On the other hand, in terms of trauma, we find that the mentally unhealthy mothers in the treatment group, in fact, surpassed the mentally healthy mothers in the control group by 0.20 SD, which is sizable and statistically significant at 1% level.[16] Finally, in terms of subjective well-being (Panel A2), the mentally unhealthy mothers in the treatment group also surpassed the subjective well-being of the mentally healthy mothers in the control group. Thus, the treatment was very effective at not only lifting mothers from mental distress but also making their mental health better than those who were considered mentally healthy at baseline.

4.2 Impact of the intervention on children

Mental health. Figure 2 and Table 2 also report treatment effects on child outcomes (Panels B1 and B2). We find that children experienced an improvement in both trauma and

[16]Note that mental health is measured using scales to create depression/trauma scores (where higher score corresponds to poor mental well-being). Here, crossing a certain threshold in the score implies mentally unhealthy. In the literature, the threshold is $1/4^{th}$ of the aggregated score. For instance, CESD-20 is scored between 0-60, and exceeding 15 implies being depressed. Therefore, it is possible for the mentally unhealthy to surpass the scores of the mentally healthy.

depression severity (Panel B1)—a 0.10 SD reduction in trauma ($p < 0.10$) and a 0.12 SD reduction in depression ($p < 0.05$) relative to children in the control group. When we examine treatment effects only among the traumatized and depressed (at baseline), we find conclusions analogous to that among mothers. That is, the effect on trauma among the traumatized is 0.13 SD ($p < 0.10$), which is slightly larger than the aggregate impact of 0.1 SD. On the other hand, the effect on depression among the depressed is twofold larger (a 0.24 SD reduction) than the pooled effect ($p < 0.05$). In addition, we also find that children whose mothers in the treatment arm had trauma or depression at baseline, their mental well-being surpassed the mental well-being of children in the control arm whose mothers were mentally healthy at baseline (Panel B1, Table A5). In other words, children's mental health became better than the mental of children to healthy mothers, which is very similar to the results on the mothers that we observe in section 4.1. One plausible reason for such a parallel shift in mental health is that infants and young children spend most of their time, if not their entire time, with their mothers, which leads to strong transmissions of mental health from mothers to their children. Therefore, it is possible that improving the mental health of mothers led to improvements in children's mental health through this transmission channel. We discuss this mechanism in detail in sections 4.3 and 5.2. In Figure A7 in Appendix A (plots C and D), we also explore the correlation between the number of sessions attended by mothers and the mental health of children in the treatment arm. Analogous to the results of mothers, these correlations are also negative, implying that more attendance is correlated with the better mental health of children (pairwise correlation tests: $p = 0.09$ in Plot C and $p = 0.03$ in Plot D).

Socioemotional, physical, and cognitive development. Next, we examine the impact on children's various skills development (Panel B2 and Figure A8 in Appendix A). We find that the intervention significantly improved communications skills of children by 0.23 SD ($p < 0.01$), gross-motor skills by 0.18 SD ($p < 0.01$), problem-solving skills by 0.18 SD ($p < 0.01$), and social skills by 0.13 SD ($p < 0.10$) relative to children in the control group. While the impacts on the former three domains are statistically significant at 1% level, the impact on children's social skills is rather weak and only marginally significant at 10% level. In terms of fine-motor skills, we do not find any statistically significant treatment effect. In addition, analogous to results on children's mental health, we again find that children in the treatment arm whose mothers were mentally unhealthy at baseline, their socioemotional, physical, and cognitive development outcomes either surpass or become similar to that of children in the control arm whose mothers were mentally healthy at baseline (Panel B2, Table A5). This implies that development of children is closely associated with the mental health of their mothers.

Anthropometrics. In terms of child malnutrition (Table 3 and Figure A8 in Appendix A), we estimate treatment effects on children's underweight, stunting, and wasting. First, we find that the treatment improved children's weight-for-age z-score (WAZ) by 0.64 SD, which is equivalent to increasing the weight of treated children by 635 grams relative to children in the control group (column 3, Panel A). Also, we observe a significant reduction in underweight by 7pp (which is 10% relative to the control group) and a reduction in severe underweight by 16pp (which is 26% relative to the control group), implying children at the lower and extreme lower tails of the distribution benefited greatly by the intervention. Second, we consider height-for-age z-scores (HAZ) that also measure malnutrition but focus on skeletal growth retardation of children. We

find that our intervention was successful at increasing the HAZ of treated children by 0.52 SD, which is equivalent to 1.58 centimeters (column 3, Panel B). At the lower and extreme lower tails of the distribution, we observe stunting and extreme stunting fell by 7pp (or 10%) and 13pp (or 22%), respectively. Finally, we examine impacts on children's prevalence of wasting, measured by weight-for-height z-scores (WHZ). Both improvements in weight and height are reflected in the incidence of wasting, where we observe WHZ of treated children to increase by 0.51 SD. Moreover, treated children's prevalence of wasting reduced by 8pp, and severe wasting was reduced by 13pp relative to children in the control group.

These differences that we observe do not differ by children's gender (columns 4-6), suggesting both female and male children experienced reductions in nutritional deprivation relatively equally.

4.3 Intergenerational transmission of mental health

Our unique sample consists of mother-child dyads with measures of trauma and depression at both baseline and endline. Therefore, for the next set of results, we focus on the transmission of mental health from mothers to children both before and after the psychosocial support intervention, as it allows us to dig deeper into the mechanisms for children's mental and physical development.

Correlation. As children under 2 years spend most of their time at home with their mothers, we hypothesized that mother-child mental health would be positively correlated. Thus, we test the transmission of mental health from mothers to children using measures of trauma and depression by looking at correlations in the spirit of Dohmen et al. (2012). Thus, while discussing correlations we do not claim the relationship to be causal. To investigate this, both at baseline and endline, we estimate the following regression using OLS:

$$y_{ijc} = \phi_0 + \phi_1 Y_{ijc} + \mathbf{\Gamma}' \boldsymbol{X}_{ijc} + \theta_c + \sigma_{ijc} \tag{2}$$

where y_{ijc} denotes the outcome (trauma or depression) of a child of mother i in block j located in camp c. Similarly, Y_{ijc} is either trauma or depression of mothers i in block j located in camp c. Here, we test the null that $\phi_1 = 0$.

In Table A6, we report the correlation coefficients using data collected at both baseline (Panel A) and endline (Panel B). In terms of intergenerational correlation in trauma and depression, we find that the coefficients are positive and statistically significant (columns 1 and 4) (both $p < 0.01$), implying mental health of children tends to be similar to the mental health of their mothers. Figure A6 in Appendix A also illustrates similarities in mental health between mothers and children when graphs are vertically compared (A1 versus B1 and A2 versus B2). In fact, correlation coefficients became larger in Panel B (Table A6), suggesting our intervention must have made the mental health of children to be more aligned with that of their mothers. When we look at correlations by child's gender, we find that the mental health of mothers is positively correlated with both female and male children (columns 2-3 and 5-6, all $p < 0.01$). Also, coefficients on the interaction between a child's gender and mother's mental health are not statistically different than zero, suggesting mental health correlations do not differ by child's gender.

Causal impact. We then investigate how strongly mothers transmitted trauma and depression to their children, or vice versa, following the intervention. We claim this relationship

to be causal as we exploit the variation caused by randomly assigning blocks to either treatment or control arms. To check the impact of the program on the transmission of mental health from mothers to children, we estimate the following equation using OLS:

$$\triangle_{1ijc} = \kappa_0 + \kappa_1 Treat_{jc} + \triangle_{0ijc} + \mathbf{\Gamma}' \boldsymbol{X}_{ijc} + \theta_c + \psi_{ijc} \tag{3}$$

where $\triangle_{1ijc} = |Y_{ijc} - y_{ijc}|$ is the absolute difference in mental health (trauma or depression) between mothers (Y_{ijc}) and children (y_{ijc}) at the endline, and $\triangle_{0ijc}$ is the baseline analogue of the outcome.[17] In this specification, if κ_1 is negative and significant, then it means the program narrowed the mental health gap between mothers and children and, thus, will imply a strong transmission of mental health from mothers to children following the intervention. In other words, it would be evidence that children's mental health became more aligned with the mental health of their mothers due to the psychosocial support program.

Table 4 reports treatment effects on the mother-child mental health gap. We find that the intervention reduced the difference in trauma between mothers and children (the negative and statistically significant coefficient on the treatment indicator). While the reduction in the depression gap is also significant, it is only significant at 5% level (column 2). When we disaggregate the impact to check if the treatment effect is similar across mothers with female and male children, we do not find any evidence of such heterogeneity. That is, the intervention reduced the difference in both trauma and depression between mothers and their sons and daughters equally (Panels A2 and B2).[18]

This result implies that children absorb and integrate the mental health of their mothers at a very young age; thus, interventions targeting the mental well-being of mothers can also be an important stepping stone to developing psychological resilience among their children. This is very important in contexts where psychosocial facilities or expertise for young children are scarce and unavailable. While this is an important finding on its own, it does not entirely answer why transmission became stronger following the intervention or how much mothers' mental health contributes to children's mental well-being. In section 5, we discuss some possible mediators and carry out a mediation analysis to understand the channels.

4.4 Robustness checks

We examine the robustness of our main results in several ways. First, we examine whether mothers' tendency to give socially desirable responses biased our estimated treatment effects. Second, as anthropometrics at endline were measured remotely, we check whether these measures are too *noisy* and whether it is also subject to experimenter demand effects. Third, we check whether the mental health of mothers might be affecting their judgments on child development, which as a result might have biased survey responses on child development. Finally, given refugee camps are overcrowded, we empirically check for possible contamination in the control blocks.

Social desirability bias. One key concern with self-reported outcomes is that respondents

[17] Before taking the absolute differences, we normalized both mothers' and children's mental health outcomes so that both normalized outcomes are between 0 and 1, and have the same range. Then, we control-group standardized this absolute difference, such that the control group has mean 0 and SD 1.

[18] To check if coefficients reported under Panel A2 (in column 2) statistically differ, we interact child's gender with the treatment dummy and find that the coefficients on this interaction term fails to reach statistical significance at conventional levels ($p > 0.10$).

often have the tendency to provide responses to survey questions that might be deemed favorable by surveyors (social desirability bias), and receiving some 'treatment' from surveyors or their employers might trigger such behavior (experimenter demand effects). For instance, in our context, treated respondents that received psychosocial support for a year might feel more inclined to provide favorable responses to enumerators relative to control group respondents. However, in this study, control group participants also participated in social gatherings (pre-pandemic) organized and invited by BRAC Bangladesh. This provides some reassurance that experimenter demand effects might also be present among the control group participants (thus, possibly constant across treatment arms). However, our program objectives were very salient to women in the treatment group and, hence, experimenter demand effects might remain a concern.

To carefully address this issue, we measured our respondent's general tendency to provide socially desirable responses using a 13-item Marlowe-Crowne scale at baseline (Crowne & Marlowe, 1960; Dhar et al., 2022). This scale was developed by psychologists and has been validated in various contexts. The questionnaire asks whether respondents have various too-good-to-be-true personality traits—such as whether respondents are excellent listeners or never hurting anyone's feelings on purpose—to create a social desirability bias or SDB scale ($0 \leq SDB \leq 13$). The higher the scale, the higher is the respondent's tendency to give socially desirable answers. Using this scale, we carry out a heterogeneity analysis to check whether people that score higher on the Marlowe-Crowne scale are more likely to experience stronger treatment effects (evidence for experimenter demand effects). We report this result in Table 5, where we find that the coefficients on the interaction term never reach statistical significance at conventional levels. More importantly, among the respondents with low SDB score (hence less likely to give socially desirable responses), we find that our treatment effects remain sizable and statistically significant at conventional levels (coefficients on 'Treatment'). This robustness check, therefore, suggests that our main results are less likely to be a product of experimenter demand effects.

Anthropometrics at endline. A major challenge in this project was collecting anthropometric outcomes during the coronavirus pandemic. However, expert anthropometric enumerators from BRAC Bangladesh successfully collected the heights and weights of children at endline over mobile phones. Mothers were carefully instructed by enumerators to use 'hand' and 'finger' units to report children's heights, and weights after weighing their child and comparing it to weights of 'rice sacks' (more details in section 3.3.1). Although this was very carefully executed by the enumerators, it might still raise three main concerns among readers: (i) how accurate is this measure; (ii) whether this measure is correlated with mothers' opinion about children's growth, rather than the actual growth; and, (iii) whether this is also subject to experimenter demand effects.

We cannot directly check whether heights and weights were incorrectly measured by mothers. However, a close alternative would be by checking the change in heights from baseline to endline (i.e., endline minus baseline). Here a large-negative difference would possibly mean incorrect reporting by mothers, as heights are unlikely to fluctuate heavily. In our case, only 3% of mothers reported height being lower than 5 millimeters at endline relative to baseline and about 20% reported lower height (of any value) at endline than baseline. As these measures were collected remotely, during a pandemic, we consider these error fractions to be very low. That is, about 80-97% of mothers were successful at following enumerators' guidance and measuring the

height of their children.[19] Next, to check if remote measures of height and weight are correlated with mothers' opinions about children's improvements in height and weight, we regress mothers' opinions (=1 if they think children's height/weight have improved) on height and weight measures. These results are reported in columns 1 and 2, Table A7 in Appendix A. We do not find opinions to be correlated with measures of children's height and weight, suggesting mothers' opinions about growth did not play any role while reporting children's height and weight to enumerators over mobile phones. Finally, we test for heterogeneous treatment effects on height-for-age, weight-for-age, and weight-for-height z-scores using the Marlowe-Crowne scale (Table A8 in Appendix A). As the interaction terms are insignificant and coefficients on 'Treatment' are statistically significant at 1% level, it suggests that treatment effects can be observed even among mothers that scored low on the social desirability bias scale. Therefore, remote measures of height and weight were possibly not subject to experimenter demand effects.

Judgment of mothers. Our next concern is whether results on child development—which are carefully observed and then reported by mothers—are influenced by mothers' judgment. This is because being mentally unwell can impair a person's attention to detail or their short-term memory, which can eventually affect their judgment (Zuckerman et al., 2018; Keller et al., 2019). For instance, a depressed mother might not have the mental strength to carefully observe their child or remember various incidents that are strong indicators for child development, while a non-depressed might not face such problems. While this is potentially an important issue, we are confident that our results are unlikely to be a product of such 'judgment bias'. First, enumerators from BRAC are highly trained with several years of experience surveying respondents from low-income households. They were also specifically trained to be very patient with our respondents to allow them ample time to recall and carefully answer questions. Second, many of the child development questions were validated by enumerators during the interview, for example, how quickly child grabs mother's finger, whether child follows a toy when moved around, can jump, responds to mother's calling, arrange toys vertically/horizontally or beads in a string, etc.[20] Therefore, we can rule out the concerns of memory or attention affecting mothers' answers to child questionnaires. Finally, to empirically address this concern, we re-estimate the treatment effects reported in Table 2 by excluding mothers who experienced an *improvement/change* in their mental health. The main assumption is that mothers that remained mentally the same at endline as they were at baseline (i.e., depressed at baseline remained depressed at endline or non-depressed at baseline remained non-depressed at endline), their attention to details and, hence, judgment should remain fairly constant. If among this sample we observe statistically significant treatment effects on child development outcomes, then mothers' bias in judgment must not be explaining our findings. We report these conservative estimates in Table A9, which shows that our main results on child development outcomes are robust to such extreme adjustments.[21]

Contamination check. Another challenge during the intervention was the possibility of contaminating control group women (e.g., possibly by sharing advice learned during sessions, playing games, sharing weekly homework, etc.). This is because camps are overcrowded and

[19] If we drop the "errors" from our sample, then our treatment effects on anthropometric outcomes increase. Note that we do not do a similar check using weight because it is very common for weights to fluctuate or fall.

[20] Since endline was conducted over-the-phone, mothers were asked to check these during interviews while the enumerator was on the phone.

[21] This also implies that child development outcomes were not solely affected via mothers' improvements in mental health, as other mediators might possibly explain some impacts. We discuss this issue in section 5.

women might socialize with other women in their neighborhood. We address this concern in the following ways. First, we used a cluster-RCT to randomize treatment at the block (i.e., neighborhood) level. Second, the average distance from treatment to control blocks was about 70 meters, with multiple non-intervention blocks (neither part of treatment or control) in between. However, as treatment assignment was done randomly, there were some control blocks with adjacent treatment blocks and vice versa. This allows us to empirically test whether women in control blocks with adjacent treatment blocks experienced any improvements in their mental health (i.e., contamination).[22] We test this by carrying out a heterogeneity analysis, where we interact the treatment dummy with another dummy that captures whether a block has any adjacent treatment block or not. This result is reported in columns 1 and 6, Table A10 in Appendix A. We do not find any statistically significant evidence that suggests having any adjacent treatment blocks improved the mental health of women in the control block. Moreover, we also do not find any evidence for augmented treatment effects among women in treatment blocks with adjacent treatment blocks. Next, we again repeat this exercise with a categorical 'adjacent' variable (columns 2 and 7), where we interact the treatment dummy with a categorical variable that captures the number of adjacent treatment blocks that each block has (this variable has four categories, between 0 and 3, where 3 corresponds to having 3 adjacent treatment blocks). Again, we do not find any evidence that having more adjacent blocks improved the mental health of control group women. Using the *proportion* of adjacent treatment blocks to total adjacent blocks also does not change this result (columns 3 and 8).

Finally, instead of adjacent blocks, we use the number of treatment blocks within the 200 and 400 meters radius of each block to check whether having more treatment blocks within 200/400 meters radius improved the mental health of control group women (or augmented the mental health impact among the treatment group women). These results are reported in columns 4-5 and 9-10 in Table A10 in Appendix A. We again do not find any evidence for contamination in our camps. One plausible reason is that male household-heads are overly protective and conservative in this culture (Beech, 2017; Tayeb, 2021b), which might have discouraged/prevented women to leave their own blocks or allow them to take frequent walks to nearby blocks that could have caused contamination. Moreover, social distancing rules were implemented after 24 sessions, which also restricted the socialization of women across blocks during the intervention.

4.5 Differential attrition and treatment effects

As highlighted in section 3.6, there is significantly higher attrition in the control group relative to the treatment group (T-test: $p < 0.01$). Thus, to check whether differential attrition might have biased our estimated treatment effects in sections 4.1 and 4.2, we use four different approaches. First, we use inverse probability weighting (IPW) to estimate the treatment effects. For this, respondents are weighted by the inverse of their response-probability, which implies that women with characteristics similar to women that are missing at endline are up-weighted in the analysis, whereas those with a high probability to respond at endline are given low weights in the analysis. These attrition-adjusted estimates are almost identical to the unadjusted estimates,

[22] This data is only available for about 1,800 respondents, as the distance data was collected from this interactive map of the camps in mid-2021, and many block ID numbers from our dataset could not be matched with that in the map. One possible reason is that many block ID numbers have changed since 2017 and the map might be showing the updated ID numbers. Note that BRAC does not have this distance information.

which are presented in Table 7 (unadjusted effects in column 1 and IPW-adjusted effects in column 2). Second, following Lee (2009), we conduct a trimming bounds analysis. For this, outcomes are first sorted from better to worse within treatment and control groups, then trims the sample from above and below in the treatment group (since 'excess observations' are in the treatment arm) to get lower and upper bounds. Our conclusions remain largely consistent with Lee (2009) bounds (columns 3-4, Table 7), where most of the treatment effects survive.

Third, following Kling et al. (2007); Karlan & Valdivia (2011), we impute the missing outcome-observations in the treatment arm using the following equation:

$$Missing\ values^T = \overline{Y}^T + \delta \tag{4}$$

where $\overline{Y}^T$ is the mean of mental health outcomes (Y) in the treatment group (T), and $\delta = 0.05$, 0.10, or 0.25 standard deviations. In other words, we first generate the averages of mental health outcome variables in the treatment arm ($\overline{Y}^T$) and then create three new variables by adding 0.05, 0.10, and 0.025 standard deviations (δ) to the averages of the outcomes (i.e., $\overline{Y}^T + \delta$), respectively. Finally, we impute these newly generated values to the mental health outcomes of attritors (or non-responders) in the treatment group. On the other hand, instead of subtracting 0.05, 0.10, and 0.025 SD to the averages in the control arm, we impute zeros to missing observations in the control arm. This is because, we make these adjustments to control-standardized outcome indices, where the control group has mean 0 already. Since negative values for mental health variables correspond to favorable outcomes, imputing $\overline{Y}^T + \delta$ to missings in the treatment arm creates three lower bounds. In contrast, positive values for subjective well-being and child development outcomes correspond to favorable outcomes. Thus, for these outcomes, we impute $\overline{Y}^T - \delta$ to missings in the treatment arm and 0 to that in the control arm to generate their lower bounds. Finally, higher z-scores for anthropometrics is also associated with favorable outcomes, but these z-scores are not control group-standardized. Therefore, to create the lower bounds, we impute $\overline{Y}^T - \delta$ to missings in the treatment arm and $\overline{Y}^C + \delta$ to missings in the control arm, where $\overline{Y}^C$ is the mean of the outcome in the control arm (C).

Results using these newly generated lower bounds is presented in Table A11 in Appendix A, where columns 2-4 report estimates with $\delta = 0.05$ SD (column 2), $\delta = 0.10$ (column 3), and $\delta = 0.25$ SD (column 4). These three bounds show that our main results would hold even if the outcomes of the attrited sample in the treatment group were 0.25 SD worse on average than that in the control group. In fact, except for mothers' happiness and children's trauma, all other results remain similar to the unadjusted effects (column 1) even for the more extreme $\delta = 0.25$ adjustments (column 4).

Finally, although based on extreme assumptions about attrition, we follow Horowitz & Manski (2000)'s version in Karlan & Valdivia (2011) to create two additional extreme bounds (both lower and upper). For this, we impute on the basis of minimal and maximal possible values to missing information. For instance, the lower (upper) bound was obtained by imputing missing data with the minimum (maximum) value in the observed treatment distribution to attritors in the treatment group and maximum (minimum) value in the observed control distribution to attritors in the control group. This gives us the most extreme lower and upper bounds. In a similar manner, instead of imputing minimal and maximal values, we replace missing data with the mean value of the lowest (highest) 10% observations in the observed treatment distribution

to attritors in the treatment group and highest (lowest) 10% observations in the observed control distribution to attritors in the control group for the lower (upper) bound. This gives us the 2^{nd}-most extreme lower and upper bounds. We report treatment effects using these bounds in columns 5-9 in Table A11 in Appendix A. We find that Horowitz & Manski (2000) bounds yield very wide bounds due to imputing extreme values. This is because, this bounds analysis is suitable when outcomes are discrete and attrition is very low (Ozler, 2017). In fact, Karlan & Valdivia (2011) also finds these bounds to be very wide due to imputing extreme values.

In summary, although we observe some degree of sensitivity while incorporating extreme bounds, our estimated treatment effects are not sensitive to trimming observations from above and below or to imputing missing information with up to 0.25 SD. According to column 1 in Table A11 (same as column 2 in Table 2), the largest effect size for mental health outcomes is for trauma, which is -0.23 or 0.23 SD below the control group mean (recall negative coefficient implies improvement in mental health). Thus, imputing attrited sample in the treatment group with +0.25 SD and that in the control group with 0—implying attrited mothers in the treatment group were much worse-off than attrited mothers in the control group—only changes the effect size by roughly 0.03 SD (from -0.23 to -0.20). This suggests that the mental health of attritors in the treatment arm would have to be extremely poor than non-attritors to change our main conclusions.

5 Possible mechanisms

Next, we turn to understanding the mechanisms. While results in section 4.3 provide some indications that children's mental health might have improved via mothers' mental well-being, there might be various other channels through which the outcomes were affected. Thus, the purpose of this section is to explore some potential mechanisms, which we do in two steps. First, we examine the direct effect of our multifaceted psychosocial intervention on several intermediate outcomes (or potential mediators), and then, second, we use a causal mediation analysis to examine how much of the treatment effects on child outcomes are mediated through mothers' mental well-being and how much of the effects are due to unobservable mediators.

5.1 Effect on intermediate outcomes

Mothers. We consider several potential mechanisms for improvements in outcomes of mothers, such as (i) mothers' improvement in physical health, as several pieces of advice given in the "my well-being" part focus on healthy diet, exercise, and various physical activities that could have improved mothers' physical health; (ii) relationship with husbands, as various advice on communication or sharing might have improved relationships or spending an hour away from home on a weekly basis might have deteriorated relationships with husbands; (iii) seeking more help for household chores, as less burden from household chores might be affecting mental well-being; and, (iv) staying connected with friends and family during the coronavirus lockdown period, as it might have improved the mental health of mothers. We explore these four potential channels in Panel A, Table 6. Treatment effects on these potential mediators are statistically insignificant, suggesting these are unlikely to be potential channels. Moreover, the program might have improved the mental health of mothers either directly or through other mediators

that are unobserved. The 'direct channel' explanation is plausible in this context because, through psychoeducation, mothers directly participated in various activities and conducted various mental exercises that might have improved their mental well-being.

Channel (ii) serves another important purpose. Given the overly protective and conservative nature of male household-heads in this culture (Tayeb, 2021b), one concern is that participation in weekly sessions might have caused mothers' relationships to deteriorate with their intimate partners. However, a statistically insignificant treatment effect on intimate partner relationships suggests that our program did not cause relationship frictions or, possibly, initiate intimate partner violence. One potential reason is that sessions took place within their neighborhoods at another (neighbor) Rohingya woman's home, whom they possibly trust and are familiar with. Besides, humanitarian organizations, such as BRAC Bangladesh, play an important role in refugee well-being within the camps; thus, male household-heads or intimate partners chose not to impose any restrictions on participating in activities organized by BRAC.

Finally, sociological theory of social ties and mental health suggest that socialization can help improve the psychological well-being of people in emotional hardship ('stress-buffering' mechanism) as well as in distress-free conditions ('main effects' mechanism) (Cohen et al., 2000; Kawachi & Berkman, 2001). Given mothers in the control group also attended unstructured social gatherings (i.e., there were no structured activities, curriculum, or mother-volunteers to administer gatherings), it is possible that the effect of social interaction on mental health remained constant across treatment arms. In contrast, Kawachi & Berkman (2001) also suggests that social gatherings can also backfire and lead to the poor mental health of attendees; however, we do not believe that the women in our control group experienced any substantial negative effects on their mental health. This is because such negative impacts of social engagement only occur when engagement is obligatory or required by traditional norms, for instance, attending mosque/church gatherings every week.[23]

Children. We next consider several potential mechanisms for improvements in child development outcomes: (i) mothers' and fathers' time input on children, as the intervention might have encouraged both mothers and fathers to spend more quality time with their children; (ii) until what age mothers breastfed their children, as more caring mothers might have breastfed their children longer; (iii) number of times children are fed per day; (iv) reduced negative parenting (such as scolding, beating, etc.); (v) asking others to spend time with children, as spending more supervised time with other adults from the household implies more daily care for children; (vi) less indoor smoking by fathers, as mothers might have become more careful about children's health and, thus, asking fathers to smoke outdoors more often; and, (vii) being more careful about children walking/playing barefoot, which prevents hookworm infections and exposure to various bacterial and fungal organisms. Panel B in Table 6 reports these results. We find strong support for mothers' time input being a potential mediator ($p < 0.01$), where treated mothers spend about 1.5 hours more every day on their children relative to control group mothers (16% more relative to the control group); however, this effect does not differ by child's gender, suggesting mothers spent similar time on girl and boy children. In addition to being a mechanism for child development and growth, this is also a possible mechanism for reducing the mental health

[23]Note that control arm mothers were encouraged to attend such gatherings; thus, it was not obligatory/compulsory.

gap between mothers and children (in section 4.3), as spending more time on children must have allowed ample time for children to absorb and integrate the mental health of their mothers. Of course, in a non-refugee, non-camp context, this mechanism might be somewhat problematic, as accommodating additional 1.5 hours everyday for children might mean sacrificing leisure time or less participation in income-generating activities by mothers. However, such opportunity costs for mothers are very small or close to zero in this context because—as mentioned in section 2—refugees cannot leave their designated camps or be employed outside. As a result, mothers mostly spend idle time in their homes after finishing household chores. Therefore, another important contribution of the home-based HPL program is that it improved mothers' decisions regarding time allocation and helped them prioritize time-input on children over leisure. Moreover, ample discretionary time can also lead to lower subjective well-being and develop mental health problems (Sharif et al., 2021); thus, reductions in 'too-much' discretionary time could also be a possible mechanism for mothers' mental health improvements.

Next, we also find that treated mothers are less likely to allow their children to play or walk barefoot ($p < 0.05$) and engage in negative parenting ($p < 0.10$), suggesting improvements in mothers' health behaviors toward their children are other potential mechanisms for children's development and growth. While we do not see any gender bias in negative parenting, we observe that mothers are only careful about whether their sons walk or play barefoot, but not about their daughters. This is not surprising because 'son-preference' in South and Southeast Asian countries is widespread (Barcellos et al., 2014; Kabeer et al., 2014). On the other hand, fathers' time input on children, mothers' breastfeeding time or the frequency of feeding children, seeking help for babysitting, or discouraging fathers to smoke indoors do not appear to be potential channels.

These results together with the reduced mental health gap between mothers and children (as reported and discussed in section 4.3) show that maternal mental health and time-input (possibly affected by mental health as well) are important for children's mental, physical, nutritional, and cognitive development. But, to what extent mothers' mental health and children's development are responsible for the mental health improvements among children? We quantify these mediated impacts through a formal mediation analysis in the following subsection.

5.2 Mediation analysis to understand impacts on children

We closely follow Heckman et al. (2013) and Heckman & Pinto (2015) to estimate what proportion of the impact on children's mental health (Table 2) are due to improvements in mothers' mental health and children's development (*mediated effects*) and what proportion is the residual—a combination of *direct effect* of the program and effects from unobserved mediators. For this, we assume that children's mental health outcome (Y_{ijc}) is a linear function of potential observed mediators (M_{ijc}) and several individual- and household-level characteristics that also include the mental health of children at baseline (X_{ijc}). In other words, we need a production function that reflects that mothers' mental well-being and children's own socioemotional, physical, and cognitive development are important determinants of children's mental health. Therefore, we can write the production function with observed mediators and baseline characteristics mapping

into children's mental health outcomes as follows:

$$Y_{ijc} = \alpha^{res} Treat_{jc} + \sum_{a=1}^{3} \beta_a M_{aijc} + \mathbf{\Gamma}' \boldsymbol{X}_{ijc} + \theta_c + \epsilon_{ijc} \tag{5}$$

where α^{res} is the residual effect, as it cannot be explained by improvements in either mothers' mental health or children's development. On the other hand, if we reject the null hypothesis $\beta_a = 0$ (where $a \in [1, 3]$) then that would imply M_a affects children's mental well-being. Also, the share of treatment effect explained by all observed mediators combined can be given by $1 - (\alpha^{res}/\beta_1)$ (where β_1 is from equation 1, our main regression model). Moreover, in model 5, $\sum M_{aijc}$ includes mothers' trauma and depression levels, and children's composite development index that aggregates communication, gross-motor, fine-motor, problem-solving, and social skills development together. X_{ijc} includes all controls (as in equation 1) and children's trauma and depression levels at baseline.

Figure 3 shows that all three of our mediators jointly and significantly affected children's depression, where about 30% of this effect can be explained by mothers' reductions in depression ($p < 0.01$), 7% by mothers' reductions in trauma ($p = 0.04$), and 18% by children's improvements in socioemotional, physical, and cognitive development ($p < 0.01$). In total, about 55% of the impact on children's depression can be explained by these three mediators jointly, whereas the remaining 45% are residual. On the other hand, 83% of the total effect on children's trauma can be jointly explained by the three mediators, where mothers' trauma explains 49% ($p < 0.01$), mothers' depression explains 9% ($p = 0.05$), and children's development explains 24% ($p < 0.01$) of the total impact documented. This implies that not only mothers' mental health but also children's development are important determinants of children's mental well-being. As children had direct engagement through play activities during sessions, we believe some of the estimated residual effects, $\hat{\alpha}^{res}$, can be explained by the direct impact of the intervention. However, this is only speculative as we cannot disentangle the two using equation 5.

For completeness, in Appendix A, we also report results from a similar exercise that checks if improvements in children's socioemotional, mental, and cognitive development, and anthropometric outcomes are mediated through improvements in mothers' and children's mental health (i.e., M_a includes four mental health mediators, $a \in [1, 4]$). We report these results in Figure A9 in Appendix A. We find that roughly 20% of the total impact on skills development can be explained by improvements in mental health.[24] On the other hand, although 5% of the total impact on anthropometrics seems to be explained by mental health mediators, these are not statistically significant at conventional levels.

Note that this mediation analysis requires an additional assumption that observed (M_a) and unobserved mediators (captured in the error term, possibly also affected by the treatment) are statistically independent. This is a strong assumption that we cannot directly test using our data and a violation of this assumption would lead to biased estimates of β_a in equation 5. Nevertheless, we take a naive approach, for exploratory purposes, to check whether β_a fluctuates when we use statistically significant mediators reported in section 5.1 to augment equation 5. Therefore, in the augmented equation, M_a also includes mothers' time-input on children, negative parenting,

[24] Except for mothers' trauma ($p = 0.11$), all other mental health mediators are statistically significant at 5% level.

and mothers' concern about children playing/walking barefoot as additional mediators that were previously considered unobserved. This result is reported in Figure A10 in Appendix A. Adding these only slightly changes β_a, but the three additional mediators do not explain children's mental health improvements (all $p > 0.10$); therefore, our conclusions reported in Figure 3 do not change if these previously unobserved mediators are added to the model.[25]

6 Heterogeneous treatment effects using machine learning

To understand who benefited the most versus the least from this program, we use a machine learning method developed by Chernozhukov et al. (2020) to examine the heterogeneity in impacts. First, it splits the sample into two equal parts, 'auxiliary' and 'main' sample. From the 'auxiliary' sample, it then generates proxy predictors, $S(Z)$, using machine learning algorithms (in this case, Random Forest) for the conditional average treatment effect (CATE) denoted by:

$$s_0(Z) = E[Y_1|Z] - E[Y_0|Z] \tag{6}$$

where Z is a vector of covariates, Y_1 is the outcome for participants in the treatment group and Y_0 is for control group. Using $S(Z)$, it then generates predictions for the main sample to extract three important properties of $s_0(Z)$: (i) the best linear predictor or BLP—reports the average treatment effect estimates (ATE) and conducts a joint test for the presence of heterogeneity with respect to Z and that machine learning methods can detect it (HET); (ii) group average treatment effects or GATES—averages of $s_0(Z)$ by quintiles, where the first quintile corresponds to those who were *least* affected by the treatment and the fifth quintile corresponds to those who were *most* affected by the treatment; and, (iii) classification analysis or CLAN—reports and tests the differences between the average characteristics of participants in the *least* versus the *most* affected groups. To economize on space, we report BLP and GATES results in Table A12 in Appendix A and CLAN results in Table 8 in the paper. Table A12 in Appendix A shows that none of the HET parameters are statistically significant at conventional levels (columns 2 and 5), suggesting machine learning algorithms cannot detect the presence of heterogeneity with respect to our set of covariates. Moreover, the differences between the most and least affected quintile groups are also statistically insignificant (columns 3 and 6 for GATES). Even though there is no heterogeneity by characteristics jointly, we are interested in exploring if there is any heterogeneity by baseline characteristics individually. This is beneficial for two reasons: it allows us to better understand the treatment effects reported in section 4 (i.e., whether effects are lower-/upper-bounds) and it helps policymakers decide whom to target during scale-up, as it can maximize the benefits of the program at the same implementation cost.

In Table 8, we report CLAN results for the following baseline covariates: baseline trauma and depression, age of mother/child, gender of child, households' exposure to violent conflict in Myanmar, and mothers' exposure to camp-based abuse.[26] In terms of mothers' improvements

[25] For instance, in explaining the mediated impact on children's trauma, β on mothers' trauma changes ($\beta_{old} - \beta_{new}$) by -0.0006, mothers' depression changes by -0.0009, and composite child development index by 0.0001 due these additional mediators. Similarly, in explaining the mediated impact on children's depression, β on mothers' trauma changes by 0.0013, mothers' depression changes by -0.0009, and composite child development index by 0.0001 due these additional mediators.

[26] Note that we cannot examine heterogeneity by household income, expenditure, employment status, etc. because

in mental health, mothers that had poor mental health at baseline had high exposure to violent conflict in Myanmar, and have experienced relatively more abuse in refugee camps have significantly benefited more from the program (all statistically significant at 1% level). In addition, mothers that are less educated significantly benefited more in terms of trauma ($p < 0.01$), and mothers that are relatively older significantly benefited more in terms of improvements in depression ($p < 0.01$). Similarly, in terms of children's improvements in mental health, we do not observe any heterogeneity in improvements in their psychological trauma. But, in terms of depression, children that had more depressive symptoms at baseline and are older benefited greatly from the program (both $p < 0.01$). However, there appears to be no observed heterogeneity by children's gender, household's exposure to violent conflict in Myanmar, and mothers' experiences with camp-based abuse.

Next, we turn to children's skills development and anthropometric outcomes. With respect to skills development, only children that are older, from households with less exposure to violence in Myanmar, and mothers with less experience of camp-based abuse have benefited the most (both $p < 0.01$).[27] Moreover, children that had more depressive symptoms at baseline appear to have experienced more development following the program, but this is only marginally significant at 10% level. In terms of improvements in stunting, underweight, and wasting, older children benefited the most from the program. Also, in terms of reductions in the incidence of underweight and wasting, boy children benefited the most in reductions in underweight while girl children benefited the most in reductions in wasting. However, almost 97% of the most affected children in terms of improvements in underweight are boys, whereas only 56% of the most affected children in terms of improvements in wasting are girls (and no heterogeneity in stunting by gender), implying more boys on average experienced nutritional benefits from the program than girls. Finally, we also find that children with fewer trauma symptoms at baseline benefited the most in terms of improvements in wasting and that households with more exposure to violence in Myanmar benefited the most in terms of improvements in stunting.

We also examine heterogeneity using the more traditional 'interaction' approach, where we interact the treatment dummy with the covariates and test if the coefficients on the interactions are statistically significant at conventional levels. We report these results in Appendix D. Compared to the machine learning approach, we find these results to be less sensitive in terms of capturing heterogeneity. However, some results, such as heterogeneity by exposure to violence, are robust to using the 'interaction' approach.

7 Lessons from the field and cost-effectiveness

Fieldwork challenges. We faced various challenges during the intervention, which the readers and interested policymakers should pay attention to. The first was training our session facilitators (mother-volunteers) who were mostly illiterate. Therefore, their training had to be carefully conducted by experienced psycho-social experts, and session materials had to have pictorial directions to help them run the sessions (see some examples in Figures A1, A2, and A3

Rohingyas cannot work or earn inside/outside the camp.

[27] When we plot treatment effects over children's age for the disaggregated development outcomes (Figure A11 in Appendix A), we find that the treatment-control gap gradually widens with children's age in case of communication, gross-motor, and problem-solving skills.

in Appendix A). Moreover, mother-volunteers never had any prior experience in running group discussion/activity sessions. Therefore, rigorous capacity-building training was also provided and mother-volunteers were also supervised by psycho-social experts when needed.[28] The second major challenge was when the country and refugee camps went into the COVID-19 pandemic lockdown. As sessions had to continue for 20 additional weeks, shifting from in-person to mobile phone delivered sessions was very challenging. BRAC initially had mobile numbers of 42% of refugee households; therefore, field operations staff, camp managers, and block-*Majhis* (leaders of blocks) had to carefully collect phone numbers from the remaining households. Households without any mobile access were lent mobile phones every week by camp managers and block-*Majhis*. Finally, participating mothers were not always able to participate in over-the-phone sessions at the scheduled time or day. Therefore, mother-volunteers often had to reschedule sessions over the phone to a different time or day that was suitable for participants. As a result, it increased the workload of mother-volunteers during the lockdown to some extent. However, each mother-volunteer was only responsible for roughly 7 mother-child pairs and, we believe, this workload (i.e., calling 7 mothers once or twice a week) did not generate additional mental burden for mother-volunteers during the lockdown. In fact, a recent study by Hussam et al. (2021) on the Rohingya refugees suggests that employment within refugee camps can itself improve the mental well-being of refugees. We, however, did not measure the mental health of mother-volunteers during or after the intervention; therefore, our take on the mental cost of mother-volunteers is only speculative.

We rapidly overcame our fieldwork challenges because of the capacity and already-installed research infrastructure of BRAC Bangladesh in the refugee camps. Moreover, BRAC is well-respected and trusted by refugees in the camps, which, we believe, played an important role in attracting initial enrolment and weekly participation of these women coming from a very conservative background.

Cost-effectiveness. There are two important features of this program that interests policymakers in low- and middle-income countries. First, it is a peer-delivered program, where peers work as volunteers (i.e., mother-volunteers) and do not need to undergo complex or lengthy training. Moreover, mother-volunteers do not need to be educated to administer sessions, as all session materials are fairly simple and available in pictorial forms if required. This is essential for resource-poor environments as, previously, a major barrier to providing mental support to refugees in camps was the shortage of trained professionals in host countries with adequate knowledge on the culture, language, and religion of the refugees in need. Even if such professionals are hired by host countries, providing such support at scale can be very costly and unsustainable in the medium to long term in low-income countries.

Second, is the overall cost of the intervention. On average, it cost USD 45 per mother-child pair to provide this support for a year. In Table A13 in Appendix A, we break down this cost, where 50% of our costs were associated with over-the-phone sessions that include costs of phone calls, payments for block-*Majhis* and camp managers, mobile phones for mother-volunteers, and revising session materials. In its absence, our cost per mother-child pair would reduce to USD 23. Another possible cost that does not incur in our case is the cost of setting up and maintaining

[28] For instance, if mother-volunteers had difficulty understanding or recalling pictorial directions, they could contact the psycho-social experts for advice and support.

session locations. This is because mother-volunteers provided their homes to be used for the weekly sessions, which certainly helped us keep the total cost low. Future interventions can also try supporting new mother-volunteers through old mother-volunteers to keep the entire support system as peer-delivered and subsequently test if it can generate comparable impacts. Nevertheless, our USD 45 per mother-child dyad (or $19 for 0.1 SD reductions in trauma and $31 for 0.1 SD reductions in depressive symptoms) is still quite low compared to various existing studies.

Since we cannot quantify the benefits of our intervention (mental health, child development, and nutritional) in dollar terms, we compare our effect sizes to studies involving cash transfers to understand the benefits of our intervention in monetary units. On mental health, a recent review on the impact of cash transfers shows that an unconditional transfer worth USD 540 can reduce depressive symptoms among the recipients by 0.163 SD (or $330 to reduce depression by 0.1 SD) (Romero et al., 2020). On anthropometric outcomes, Carneiro et al. (2021) finds that cash transfers ($22 per month for 2 years or $528) combined with parenting information improved the heights of treated children by 0.62cm in Nigeria. Next, Attanasio et al. (2022) finds that intervention involving the training of mothers ($115) and providing nutritional supplements ($209) can be effective in reducing stunting among children in Colombia (by 5.8pp compared to the control group). Also, it cost USD 14 per woman (in 2006) to provide 16 cognitive-behavioral therapy sessions in rural Pakistan (Baranov et al., 2020), which led to significant reductions in depressive symptoms among new mothers by 0.62 SD (their most immediate impact) but had no impact on children's development. In comparison, our intervention was relatively low-cost with significant mental health benefits to mothers and children, and early childhood development and nutritional benefits to children. However, our context and population are very different from that of these existing studies, which might explain some differences in costs-benefits that we observe. There are various other possible benefits and impacts—such as on labor market activities, productivity, income, savings, and so on—that we could not explore in this paper. The primary reason is the context, as the legal status of Rohingya refugees in Bangladesh does not allow their social inclusion, participation in employment, or mobility out of camps.

8 Concluding remarks

Using a cluster randomized controlled trial on the Rohingya refugees, we show in this paper that a multifaceted psychoeducation, parenting support, and play activity program can be successfully set up in resource-poor settings, such as in refugee camps in developing countries. Through such programs, both mothers and young children can be simultaneously targeted and provided weekly support. We find that this multifaceted program was successful in many ways. First, it improved the mental health of both mothers and children and caused their mental health to align with each other. Second, it significantly reduced the prevalence of stunting, underweight, and wasting among children. Third, it improved children's socioemotional, physical, and cognitive development after the intervention. Finally, it was a very low-cost intervention availing human resources from the refugee community (costs only USD 45 per dyad), which makes it highly scalable and attractive to policymakers.

The Bangladeshi government wishes to repatriate the Rohingya people to Myanmar, but

the state of affairs in the Rakhine State remains unstable as of 2022, making the future of these refugees uncertain. Moreover, limited financial capacity and resources in a lower-income country, such as Bangladesh, can easily divert the priority from the refugees to other policies targeting the welfare of the locals. This, as a result, can severely affect the human capital accumulation of child refugees with various detrimental economic consequences later in their lives. Therefore, our low-cost and easily scalable program can be an effective solution to support the health and well-being of refugee children and their primary-caregiver mothers. In fact, our intervention is gradually being scaled up in refugee camps in Bangladesh by BRAC, where over 13 thousand mother-child dyads currently benefit from it. Considering the humanitarian situation and urgency, mother-child dyads in the control group also started receiving psychosocial support as part of the scale-up. Since about 30 thousand infants are born in Rohingya refugee camps every year (Tayeb, 2021a), international donors could consider sharing the costs and resources with NGOs and humanitarian agencies for a rapid scale-up and ensure the coverage of all future newborns by the program.

As the world is currently experiencing another large conflict in Ukraine and already millions, mostly women and children, have been forced out from their country, our findings could also offer important insights on the immediate psychosocial needs of these vulnerable refugees during their resettlement.

References

Ackard, D. M., Neumark-Sztainer, D., Story, M., & Perry, C. (2006). Parent–child connectedness and behavioral and emotional health among adolescents. *American Journal of Preventive Medicine*, *30*(1), 59–66.

Adhvaryu, A., Fenske, J., & Nyshadham, A. (2019). Early life circumstance and adult mental health. *Journal of Political Economy*, *127*(4), 000–000.

Ahlburg, D. (1998). Intergenerational transmission of health. *American Economic Review Papers & Proceedings*, *88*(2), 265–270.

Ahmed, B., Orcutt, M., Sammonds, P., Burns, R., Issa, R., Abubakar, I., & Devakumar, D. (2018). Humanitarian disaster for rohingya refugees: impending natural hazards and worsening public health crises. *The Lancet Global Health*, *6*(5), e487–e488.

Akresh, R., Lucchetti, L., & Thirumurthy, H. (2012). Wars and child health: Evidence from the Eritrean–Ethiopian conflict. *Journal of Development Economics*, *99*(2), 330–340.

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021). Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *Quarterly Journal of Economics*, *136*(4), 2147–2194.

Almond, D., Edlund, L., & Palme, M. (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden. *Quarterly Journal of Economics*, *124*(4), 1729–1772.

Amadu, S., Attanasio, O. P., Caeyers, B., Cattan, S., Sosa, L. C., Krutikova, S., Leighton, P., Masselus, L., & Yakubu, M. (2019). Improving early childhood development in rural Ghana

through scalable low-cost community-run play schemes: Programme impact evaluation report. Technical report, IFS Report.

American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association, Washington DC.

American Psychological Association (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, *48*(1).

Andrews, G. & Slade, T. (2001). Interpreting scores on the kessler psychological distress scale (k10). *Australian and New Zealand Journal of Public Health*, *25*(6), 494–497.

Angelucci, M. & Bennett, D. (2021). The Economic Impact of Depression Treatment in India. *IZA Discussion Paper No. 14393*.

Asadujjaman, M., Molla, M. B. A., & Al Noman, S. N. (2019). Stature estimation from hand anthropometric measurements in Bangladeshi population. *Journal of Forensic and Legal Medicine*, *65*, 86–91.

Attanasio, O., Baker-Henningham, H., Bernal, R., Meghir, C., Pineda, D., & Rubio-Codina, M. (2022). Early stimulation and nutrition: The impacts of a scalable intervention. *Journal of European Economic Association (forthcoming)*.

Attanasio, O., Meghir, C., & Nix, E. (2020). Human capital development and parental investment in India. *Review of Economic Studies*, *87*(6), 2511–2541.

Baranov, V., Bhalotra, S., Biroli, P., & Maselko, J. (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review*, *110*(3), 824–59.

Barcellos, S. H., Carvalho, L. S., & Lleras-Muney, A. (2014). Child gender and parental investments in India: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, *6*(1), 157–89.

Barker, N., Bryan, G. T., Karlan, D., Ofori-Atta, A., & Udry, C. R. (2021). Mental health therapy as a core strategy for increasing human capital: Evidence from Ghana. *NBER working paper no. 29407*.

Beech, H. (2017). 'I'm Struggling to Survive': For Rohingya Women, Abuse Continues in Camps. https://www.nytimes.com/2017/12/23/world/asia/rohingya-women-abuse-myanmar.html. Online; Accessed September 12, 2020.

Bellows, J. & Miguel, E. (2009). War and local collective action in Sierra Leone. *Journal of Public Economics*, *93*(11-12), 1144–1157.

Beyrer, C. & Kamarulzaman, A. (2017). Ethnic cleansing in Myanmar: The Rohingya crisis and human rights. *The Lancet*, *390*(10102), 1570–1573.

Bhalotra, S. R., Delavande, A., Gilabert, P., & Maselko, J. (2020). Maternal investments in children: The role of expected effort and returns. *IZA Discussion Paper NO. 13056*.

Bharadwaj, P., Løken, K. V., & Neilson, C. (2013). Early life health interventions and academic achievement. *American Economic Review*, *103*(5), 1862–91.

Blattman, C., Jamison, J. C., & Sheridan, M. (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia. *American Economic Review*, *107*(4), 1165–1206.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the American South. *Quarterly Journal of Economics*, *122*(1), 73–117.

Bolton, P., Bass, J., Neugebauer, R., Verdeli, H., Clougherty, K. F., Wickramaratne, P., Speelman, L., Ndogoni, L., & Weissman, M. (2003). Group interpersonal psychotherapy for depression in rural Uganda: A randomized controlled trial. *JAMA*, *289*(23), 3117–3124.

Breierova, L. & Duflo, E. (2004). The impact of education on fertility and child mortality: Do fathers really matter less than mothers? *NBER working paper no. 10513*.

Cappelen, A., List, J., Samek, A., & Tungodden, B. (2020). The effect of early-childhood education on social preferences. *Journal of Political Economy*, *128*(7), 2739–2758.

Carneiro, P., Kraftman, L., Mason, G., Moore, L., Rasul, I., & Scott, M. (2021). The impacts of a multifaceted prenatal intervention on human capital accumulation in early life. *American Economic Review*, *111*(8), 2506–49.

Chambers, R. (1986). Hidden losers? the impact of rural refugees and refugee programs on poorer hosts. *International Migration Review*, *20*(2), 245–263.

Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2020). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. *NBER working paper no. 24678*.

Cheung, S. (2011). Migration Control and the Solutions Impasse in South and Southeast Asia: Implications from the Rohingya Experience. *Journal of Refugee Studies*, *25*(1), 50–70.

Cohen, S., Underwood, L. G., & Gottlieb, B. H. (2000). *Social support measurement and intervention: A guide for health and social scientists*. Oxford University Press.

Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349.

Currie, J. (2009). Healthy, wealthy, and wise: Is there a causal relationship between child health and human capital development? *Journal of Economic Literature*, *47*(1), 87–122.

Currie, J. & Stabile, M. (2006). Child mental health and human capital accumulation: The case of ADHD. *Journal of Health Economics*, *25*(6), 1094–1118.

Dhar, D., Jain, T., & Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *American Economic Review*, *112*(3), 899–927.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, *79*(2), 645–677.

Doyle, O. (2020). The first 2,000 days and child skills. *Journal of Political Economy*, *128*(6), 2067–2122.

Dyregrov, A., Gupta, L., Gjestad, R., & Mukanoheli, E. (2000). Trauma exposure and psychological reactions to genocide among rwandan children. *Journal of Traumatic Stress*, *13*(1), 3–21.

Eyal, K. & Burns, J. (2019). The parent trap: cash transfers and the intergenerational transmission of depressive symptoms in South Africa. *World Development*, *117*, 211–229.

Fischbach, R. L. & Herbert, B. (1997). Domestic violence and mental health: correlates and conundrums within and across cultures. *Social Science & Medicine*, *45*(8), 1161–1176.

Fortify Rights (2020). The Right to Mental Health for Rohingya Survivors of Genocide in Myanmar and Bangladesh. https://www.fortifyrights.org/mya-inv-rep-2020-12-10/. Online; Accessed Janaury 04, 2022.

Gonçalves, H., Pearson, R., Horta, B., González-Chica, D., Castilho, E., Damiani, M., Lima, R., Gigante, D., Barros, F., Stein, A., et al. (2016). Maternal depression and anxiety predicts the pattern of offspring symptoms during their transition to adulthood. *Psychological Medicine*, *46*(2), 415–424.

Habib, M., Jubb, C., Ahmad, S., Rahman, M., & Pallard, H. (2018). *Forced migration of Rohingya: an untold experience*. Ontario International Development Agency: Ontario, Canada.

Hanna, R. & Oliva, P. (2016). Implications of climate change for children in developing countries. *The Future of Children*, 115–132.

Haushofer, J., Mudida, R., & Shapiro, J. (2020). The comparative impact of cash transfers and psychotherapy on psychological and economic well-being. *NBER working paper no. 28106*.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–86.

Heckman, J. J. & Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews*, *34*(1-2), 6–31.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*(3), 411–482.

Herman, J. L. (2015). *Trauma and recovery: The aftermath of violence–from domestic abuse to political terror*. Basic Books, New York.

Hodal, K. (2017). Rohingya children close to starvation amid 'health crisis on an unimaginable scale'. https://www.theguardian.com/global-development/2017/nov/10/rohingya-kids-starvation-health-crisis-unimaginable-scale-malnutrition-myanmar-bangladesh. Online; Accessed September 12, 2020.

Horowitz, J. L. & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, *95*(449), 77–84.

Hossain, A., Ahmed, S., Shahjalal, M., & Ahsan, G. U. (2019). Health risks of Rohingya children in Bangladesh: 2 years on. *The Lancet*, *394*(10207), 1413–1414.

Hussam, R. N., Kelley, E. M., Lane, G. V., & Zahra, F. T. (2021). The psychosocial value of employment. *NBER working paper no. 28924*.

Imbens, G. W. & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Jacobsen, K. (2005). *The economic life of refugees*. Kumarian Press.

Johnston, D. W., Schurer, S., & Shields, M. A. (2013). Exploring the intergenerational persistence of mental health: Evidence from three generations. *Journal of Health Economics*, *32*(6), 1077–1089.

Kabeer, N., Huq, L., & Mahmud, S. (2014). Diverging stories of "missing women" in South Asia: Is son preference weakening in Bangladesh? *Feminist Economics*, *20*(4), 138–163.

Karlan, D. & Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and Statistics*, *93*(2), 510–527.

Kawachi, I. & Berkman, L. F. (2001). Social ties and mental health. *Journal of Urban Health*, *78*(3), 458–467.

Keller, A. S., Leikauf, J. E., Holt-Gosselin, B., Staveland, B. R., & Williams, L. M. (2019). Paying attention to attention in depression. *Translational Psychiatry*, *9*(1), 1–12.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Kumsa, M. K. (2006). 'No! I'm not a refugee!' The poetics of be-longing among young Oromos in Toronto. *Journal of Refugee Studies*, *19*(2), 230–255.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, *76*(3), 1071–1102.

List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*(4), 773–793.

Lukens, E. P. & McFarlane, W. R. (2004). Psychoeducation as evidence-based practice: Considerations for practice, research, and policy. *Brief Treatment and Crisis Intervention*, *4*(3), 205–225.

Mahmood, S. S., Wroe, E., Fuller, A., & Leaning, J. (2017). The Rohingya people of Myanmar: Health, human rights, and identity. *The Lancet*, *389*(10081), 1841–1850.

Mensah, F. K. & Kiernan, K. E. (2010). Parents' mental health and children's cognitive and social development. *Social Psychiatry and Psychiatric Epidemiology*, *45*(11), 1023–1035.

Miguel, E. & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159–217.

Minoiu, C. & Shemyakina, O. (2012). Child health and conflict in Côte d'Ivoire. *American Economic Review*, *102*(3), 294–99.

Motlagh, J. (2018). The Survivors of the Rohingya Genocide. https://www.rollingstone.com/politics/politics-features/rohingya-genocide-myanmar-701354/. Online; Accessed October 9, 2019.

Neugebauer, R., Fisher, P. W., Turner, J. B., Yamabe, S., Sarsfield, J. A., & Stehling-Ariza, T. (2009). Post-traumatic stress reactions among Rwandan children and adolescents in the early aftermath of genocide. *International Journal of Epidemiology*, *38*(4), 1033–1045.

Nezlek, J. B., Imbrie, M., & Shean, G. D. (1994). Depression and everyday social interaction. *Journal of Personality and Social Psychology*, *67*(6), 1101.

Ozler, B. (2017). Dealing with attrition in field experiments. https://blogs.worldbank.org/impactevaluations/dealing-attrition-field-experiments. Online; Accessed October 23, 2020.

Patel, V., Rahman, A., Jacob, K., & Hughes, M. (2004). Effect of maternal mental health on infant growth in low income countries: New evidence from South Asia. *BMJ*, *328*(7443), 820–823.

Porter, M. & Haslam, N. (2005). Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: a meta-analysis. *JAMA*, *294*(5), 602–612.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401.

Rahman, A., Malik, A., Sikander, S., Roberts, C., & Creed, F. (2008). Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: A cluster-randomised controlled trial. *The Lancet*, *372*(9642), 902–909.

Rahman, A., Patel, V., Maselko, J., & Kirkwood, B. (2008). The neglected 'm'in mch programmes–why mental health of mothers is important for child nutrition. *Tropical Medicine & International Health*, *13*(4), 579–583.

Raley, S. & Bianchi, S. (2006). Sons, daughters, and family processes: Does gender of children matter? *Annual Review of Sociology*, *32*, 401–421.

Ridley, M., Rao, G., Schilbach, F., & Patel, V. (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, *370*(6522).

Romero, J., Esopo, K., McGuire, J., & Haushofer, J. (2020). The effect of economic transfers on psychological well-being and mental health. *NBER working paper no. 28106*.

Save the Children (2018). Nearly 40 percent of Rohingya children in Cox's Bazar are stunted: new study. https://www.savethechildren.org/us/about-us/media-and-news/2018-press-releases/children-stunted-coxs-bazaar. Online; Accessed October 01, 2020.

Save the Children (2019). 1 in 5 Rohingya Child Refugees Suffer Severe Mental Health Issues. https://www.voanews.com/a/south-central-asia_1-5-rohingya-child-refugees-suffer-severe-mental-health-issues/6178558.html. Online; Accessed October 01, 2020.

Schilbach, F., Schofield, H., & Mullainathan, S. (2016). The psychological lives of the poor. *American Economic Review P&P*, *106*(5), 435–40.

Sharif, M. A., Mogilner, C., & Hershfield, H. E. (2021). Having too little or too much time is linked to lower subjective well-being. *Journal of Personality and Social Psychology*.

Singhal, S. (2019). Early life shocks and mental health: The long-term effect of war in Vietnam. *Journal of Development Economics*, *141*, 102244.

Squires, J. & Bricker, D. (2009). *Ages & Stages Questionnaires: A Parent-Completed Child Monitoring System*. Paul H. Brookes Publishing Co: Baltimore, USA.

Squires, J., Bricker, D., & Twombly, E. (2015). *Ages & stages questionnaires: Social-emotional (A parent-completed child monitoring system for social-emotional behaviors)*. Paul H. Brookes Publishing Co: Baltimore, USA.

Steel, Z., Chey, T., Silove, D., Marnane, C., Bryant, R. A., & Van Ommeren, M. (2009). Association of torture and other potentially traumatic events with mental health outcomes among populations exposed to mass conflict and displacement: a systematic review and meta-analysis. *JAMA*, *302*(5), 537–549.

Summerfield, D. (2000). War and mental health: a brief overview. *BMJ*, *321*(7255), 232–235.

Tayeb, T. (2021a). Four years on, no end in sight to the plight of the Rohingya. https://www.thedailystar.net/opinion/closer-look/news/four-years-no-end-sight-the-plight-the-rohingya-2159946. Online; Accessed Janaury 05, 2022.

Tayeb, T. (2021b). Why are Rohingya women and girls so unsafe in refugee camps? https://www.thedailystar.net/opinion/closer-look/news/

why-are-rohingya-women-and-girls-so-unsafe-refugee-camps-2911316. Online; Accessed Janaury 04, 2022.

Taylor, J. E., Filipski, M. J., Alloush, M., Gupta, A., Valdes, R. I. R., & Gonzalez-Estrada, E. (2016). Economic impact of refugees. *Proceedings of the National Academy of Sciences*, *113*(27), 7449–7453.

The Lancet (2019). The rohingya people: past, present, and future. *The Lancet*, *394*(2202).

UNHCR (2021a). Improving Refugee Health Worldwide. https://www.unhcr.org/uk/partners/partners/3fcb53882/improving-refugee-health-worldwide.html. Online; Accessed Janaury 10, 2022.

UNHCR (2021b). Refugee Population Statistics Database. United Nations High Commissioner for Refugees, Geneva. Accessed on February 26, 2022: https://www.unhcr.org/refugee-statistics/.

UNHCR Camp Profiles (2019). Camp Profiles: Rohingya Refugee Response Bangladesh. United Nations High Commissioner for Refugees, Geneva.

UNHCR Population Factsheet (2019). Population Factsheet: Rohingya Refugee Response Bangladesh. United Nations High Commissioner for Refugees, Geneva.

Vlassopoulos, M., Siddique, A., Rahman, T., Pakrashi, D., Islam, A., Ahmed, F., et al. (2021). Improving women's mental health during a pandemic. *Munich Papers in Political Economy no. 02/2021*.

Westfall, P. H. & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons.

WHO (2006). WHO Child Growth Standards: Methods and Development. https://www.who.int/childgrowth/standards/Technical_report.pdf. Online; Accessed December 06, 2019.

WHO (2009). WHO Child Growth Standards: Growth Velocity Based on Weight, Length and Head Circumference: Methods and Development. Geneva, Swtizerland: WHO Department of Nutrition for Health and Development.

WHO Situation Report (2017). WHO Weekly Situation Report no. 8. https://www.who.int/docs/default-source/searo/bangladesh/bangladesh---rohingya-crisis---pdf-reports/sitreps/2017/weekly-sitrep08-cxbban.pdf?sfvrsn=dc203c83_2. Online; Accessed October 01, 2020.

Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, *134*(2), 557–598.

Zuckerman, H., Pan, Z., Park, C., Brietzke, E., Musial, N., Shariq, A. S., Iacobucci, M., Yim, S. J., Lui, L. M., Rong, C., et al. (2018). Recognition and treatment of cognitive dysfunction in major depressive disorder. *Frontiers in Psychiatry*, 655.

9 Main Figures and Tables

Figure 1: Program timeline

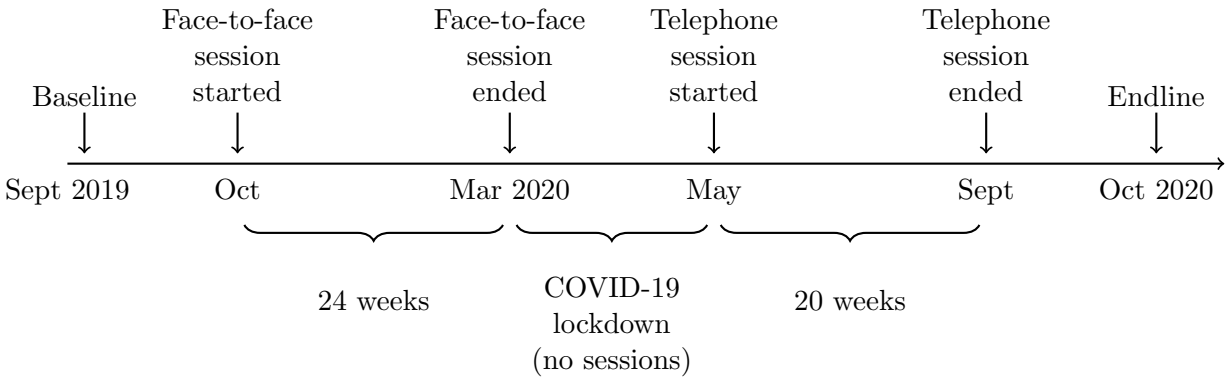
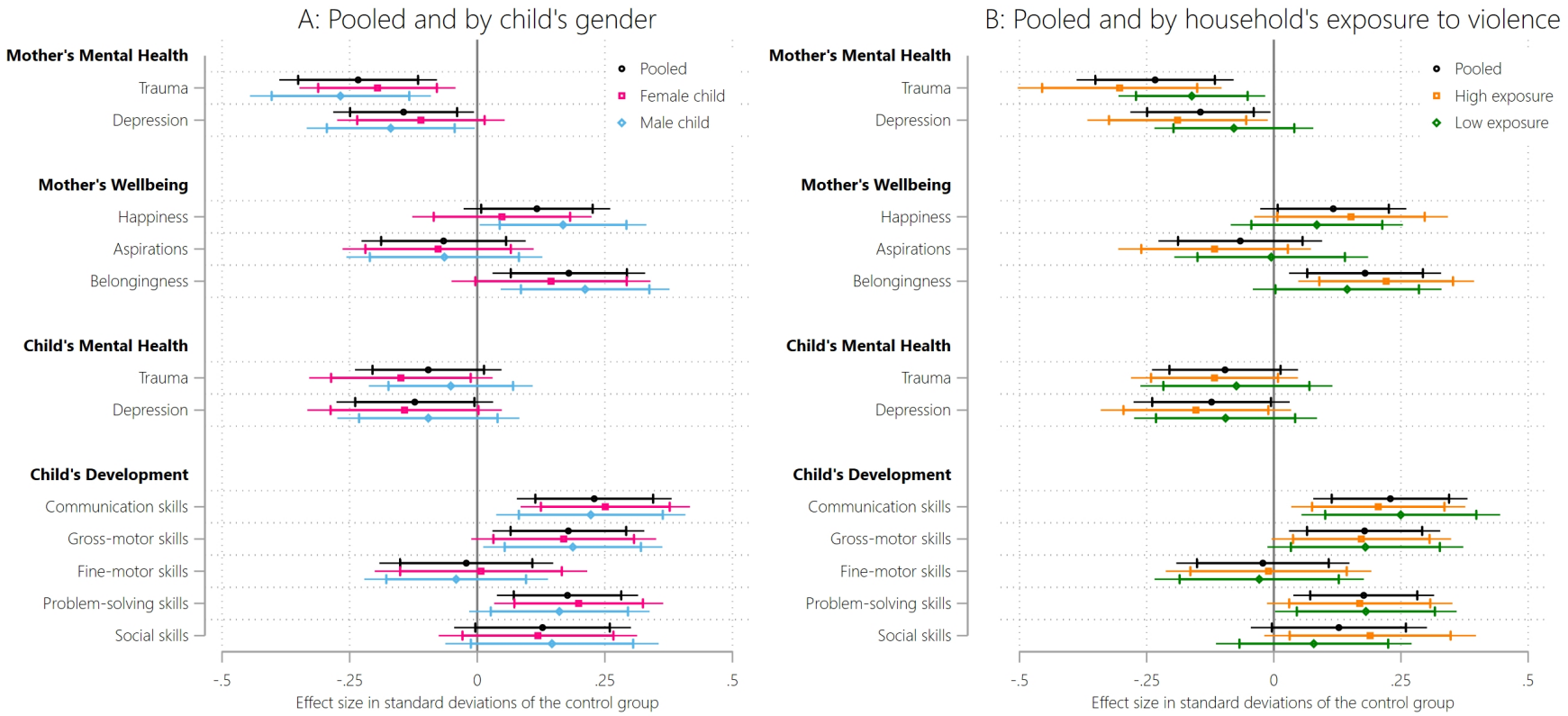
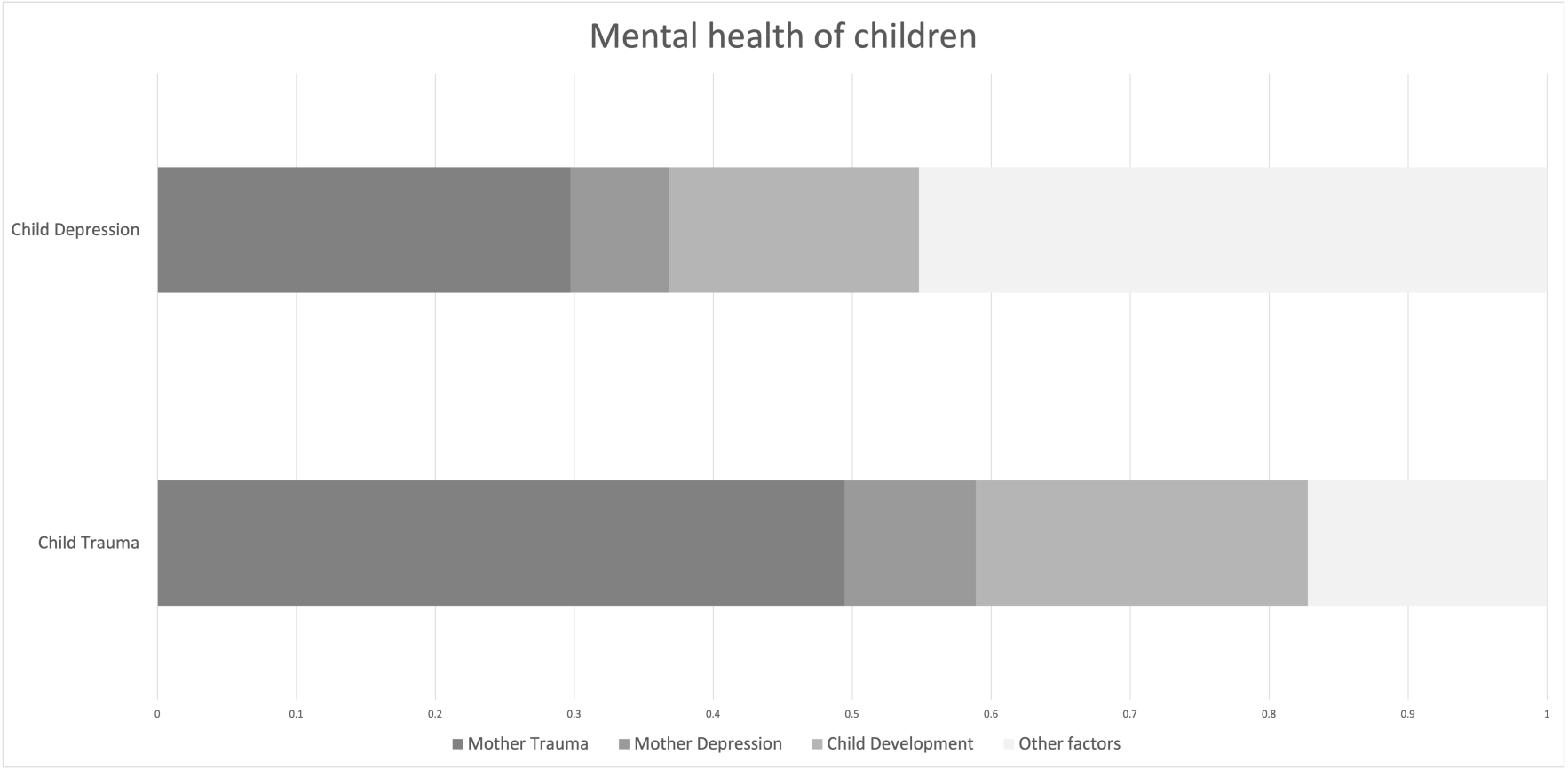


Figure 2: Treatment effects in standard deviations



Note: This figure shows estimated treatment effects in standard deviation units, where the control group has mean 0 and standard deviation 1. Effects reported with 99% and 95% confidence intervals.

Figure 3: Mediated effects on children's mental health



Note: This figure reports the mediated and residual effects on children's depression and trauma outcomes. Each shade of a bar corresponds to the proportion of the total effect that is mediated.

Table 1: Baseline characteristics and balance checks

VARIABLES	**Treatment** (*Std. Dev.*)	N_T	**Control** (*Std. Dev.*)	N_C	**T-test** ***p*-values**	**RI** ***p*-values**
Age of mother	25.70	1,909	25.25	1,586	0.04	0.03
	(5.76)		(5.72)			
Mother receives food voucher (=1 if true)	0.51	1,909	0.50	1,586	0.50	0.49
	(0.50)		(0.50)			
Household size	5.30	1,911	5.19	1,586	0.11	0.10
	(2.05)		(1.90)			
Mother employed (=1 if true)	0.02	1,909	0.03	1,586	0.89	0.90
	(0.15)		(0.17)			
Monthly income of mother (=1 if $> 5,000$)	0.41	46	0.49	45	0.65	0.67
	(0.50)		(0.51)			
Husband is alive (=1 if true)	0.97	1,911	0.97	1,586	0.28	0.29
	(0.18)		(0.16)			
Number of children	2.93	1,911	2.90	1,586	0.56	0.55
	(2.00)		(1.89)			
Mother attended school (=1 if true)	0.73	1,910	0.73	1,586	0.83	0.84
	(0.44)		(0.44)			
Months living in the camp	25.00	1,911	26.41	1,586	0.13	0.15
	(8.61)		(18.28)			
Mother is the household head (=1 if true)	0.22	1,911	0.21	1,586	0.53	0.55
	(0.41)		(0.40)			
Mother victim of conflict abuse (=1 if true)	0.87	1,911	0.86	1,586	0.96	0.97
	(0.34)		(0.34)			
Mother victim of camp abuse (=1 if true)	0.16	1,911	0.16	1,586	0.96	0.97
	(0.36)		(0.36)			
Age of child	14.59	1,911	14.23	1,588	0.11	0.11
	(6.44)		(6.50)			
Gender of child	0.50	1,911	0.52	1,588	0.29	0.28
	(0.50)		(0.50)			
Child victim of camp abuse (=1 if true)	0.03	1,911	0.05	1,588	0.38	0.40
	(0.17)		(0.21)			

Note: *Treatment* and *Control* columns show mean of the corresponding variables; all variables with "=1 if true" are dummies and are self explanatory; *Age* is in years; *Household Size* is the number of household members who eat together; *Monthly Income* is a dummy variable that equals 1 if the employed mother earns more than 5,000 Taka per month and 0 if earns less than 5,000 Taka per month (please note that only 91 mothers are employed within the camp); *Months living in the camp* is the number of months the mother have been living in the refugee camp; *Mother victim of conflict abuse* is a dummy variable that equals to 1 if the mother or any household member has experienced at least one type of conflict induced abuse/violence (i.e. either physical, sexual, or verbal abuse, or any harm to the house or the village) and 0 otherwise; *Mother victim of camp abuse* is a dummy variable that equals to 1 if the mother has experienced at least one type of abuse in refugee camps (i.e. either physical, sexual, or verbal abuse); *Child victim of camp abuse* is a dummy variable that equals to 1 if the child has experienced at least one type of abuse in refugee camps (i.e. either physical, sexual, or verbal abuse). T-test p-values are derived from linear regressions, where the dependent variable is from the list above and the independent variable is a dummy that equals 1 if belongs to the treatment group and 0 if not, with camp fixed effects and robust standard errors clustered at the block level; RI p-values are randomization inference p-values (with 1,000 replications) (Young, 2019). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 2: Treatment effects on mental health and child development

	Treatment effects				
	Without covariates	**With covariates**	**Tr./Dep. at baseline**	**(2)-RI *p*-values**	**(2)-FWER *p*-values**
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
A1. Mothers' mental health‡					
Trauma severity	-0.233***	-0.233***	-0.255***	0.00	0.00
	(0.055)	(0.051)	(0.068)		
Depression severity	-0.146**	-0.144***	-0.288***	0.00	0.02
	(0.057)	(0.054)	(0.095)		
Composite mental health index	-0.223***	-0.223***	-0.276***	0.00	0.00
	(0.059)	(0.054)	(0.072)		
A2. Mothers' well-being					
Happiness	0.108*	0.117**	-	0.04	0.04
	(0.057)	(0.056)			
Aspirations	-0.068	-0.066	-	0.32	0.69
	(0.062)	(0.062)			
Belongingness	0.180***	0.179***	-	0.00	0.00
	(0.058)	(0.057)			
Composite SWB index	0.116**	0.119**	-	0.04	0.02
	(0.057)	(0.055)			
B1. Children's mental health‡					
Trauma severity	-0.117**	-0.096*	-0.127*	0.08	0.02
	(0.057)	(0.055)	(0.074)		
Depression severity	-0.128**	-0.122**	-0.239**	0.03	0.02
	(0.061)	(0.059)	(0.098)		
Composite mental health index	-0.139**	-0.123**	-0.153**	0.03	0.01
	(0.061)	(0.059)	(0.073)		
B2. Children's development					
Communication skills	0.251***	0.229***	-	0.00	0.00
	(0.061)	(0.059)			
Gross-motor skills	0.197***	0.179***	-	0.00	0.00
	(0.061)	(0.058)			
Fine-motor skills	0.006	-0.021	-	0.76	0.89
	(0.071)	(0.066)			
Problem-solving skills	0.195***	0.177***	-	0.00	0.00
	(0.058)	(0.055)			
Social skills	0.125*	0.128*	-	0.05	0.01
	(0.067)	(0.067)			
Composite child development index	0.203***	0.182***	-	0.00	0.00
	(0.072)	(0.069)			
Observations	2,845	2,840	$1{,}240^{T}/508^{D}$	-	-

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Column (1): treatment effect estimated without any baseline covariates. Column (2): treatment effect estimated with all baseline covariates (as in equation 1). Column (3): treatment effect only on mothers that were found to be traumatized ($N = 1,240$)/depressed ($N = 508$) at the baseline, with all covariates. All outcomes are standardized indices, such that the control group has mean 0 and standard deviation 1. The composite indices aggregate the individual outcome indices under each panel. For mental health outcomes (under A1 and B1), lower values correspond to improvement in mental health. For other outcomes (under A2 and B2), higher values correspond to more favorable outcomes. Covariates include baseline measures of age (mother's and child's), whether mother attend school, household size, monthly household spending, months lived in the camp, whether mother receives monthly food voucher, whether child's father is alive, any family member stranded in Myanmar, gender of the child, number of children, household victimization (based on household's experience during conflict in Myanmar), mothers' camp-victimization (based on abuse in the camp), and children's camp-victimization (based on abuse in the camp). Standard errors, clustered at the block level (251 clusters), are in parentheses. Column (4) reports RI *p*-values for the full model (column 2), which are randomization inference *p*-values (with 1,000 replications) (Young, 2019). Column (5) reports FWER *p*-values for the full model (column 2), which are the List-Shaikh-Xu familywise error rate adjusted p-values (with 3,000 replications) based on 12 tests (List et al., 2019).

Table 3: Treatment effects on child anthropometrics

		Treatment effects				
	Control mean	**Without covariates**	**With all covariates**	**Girl child**	**Boy child**	**Diff (5)-(4)**
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
A. Underweight						
Weight-for-age z-score (WAZ)	-3.17	0.630***	0.641***	0.764***	0.551***	-0.115
	[2.34]	(0.093)	(0.092)	(0.120)	(0.125)	(0.162)
Weight (in kg)	8.75	0.700***	0.635***	0.778***	0.519***	-0.172
	[3.23]	(0.130)	(0.123)	(0.158)	(0.168)	(0.220)
Underweight (=1 if $WAZ < -2$)	0.73	-0.070***	-0.071***	-0.107***	-0.046*	0.042
	[0.44]	(0.019)	(0.019)	(0.026)	(0.026)	(0.035)
Severe underweight (=1 if $WAZ < -3$)	0.60	-0.150***	-0.158***	-0.185***	-0.134***	0.033
	[0.49]	(0.020)	(0.020)	(0.028)	(0.027)	(0.037)
B. Stunting						
Height-for-age z-score (HAZ)	-2.66	0.647***	0.515***	0.645***	0.417**	0.015
	[3.77]	(0.153)	(0.139)	(0.192)	(0.193)	(0.256)
Height (in cm)	80.5	2.366***	1.576***	2.090***	1.156*	-0.185
	[13.91]	(0.625)	(0.454)	(0.640)	(0.628)	(0.855)
Stunting (=1 if $HAZ < -2$)	0.69	-0.081***	-0.070***	-0.081***	-0.063**	-0.015
	[0.46]	(0.019)	(0.018)	(0.028)	(0.026)	(0.038)
Severe stunting (=1 if $HAZ < -3$)	0.60	-0.143***	-0.130***	-0.132***	-0.132***	-0.038
	[0.49]	(0.020)	(0.019)	(0.028)	(0.028)	(0.039)
C. Wasting						
Weight-for-height z-score (WHZ)	-2.43	0.423***	0.508***	0.464***	0.566***	0.019
	[3.23]	(0.135)	(0.125)	(0.156)	(0.182)	(0.213)
Wasting (=1 if $WHZ < -2$)	0.57	-0.059***	-0.081***	-0.056**	-0.102***	-0.029
	[0.49]	(0.022)	(0.018)	(0.028)	(0.025)	(0.035)
Severe wasting (=1 if $WHZ < -3$)	0.43	-0.108***	-0.128***	-0.089***	-0.159***	-0.053*
	[0.50]	(0.022)	(0.018)	(0.024)	(0.026)	(0.032)
Observations	1,166	2,845	2,840	1,400	1,440	2,840

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Column (1): control group average at endline with standard deviations in brackets; Column (2): treatment effect estimated without any baseline covariates. Column (3): treatment effect estimated with all baseline covariates (as in equation 1). Column (4): treatment effect on girl child. Column (5): treatment effect on boy child. Column (6): difference between column (4) and (5), which is the coefficient on the interaction between treatment dummy and child's gender dummy. Average age of child at endline was 27 months. For z-scores, higher values correspond to more favorable outcomes. For indicators, lower values correspond to more favorable outcomes. Standard errors, clustered at the block level (251 clusters), are in parentheses. Column (4) reports RI p-values for the full model (column 2), which are randomization inference p-values (with 1,000 replications) (Young, 2019). Column (5) reports FWER p-values for the full model (column 2), which are the List-Shaikh-Xu familywise error rate adjusted p-values (with 3,000 replications) based on 12 tests (List et al., 2019).

Table 4: Treatment effect on transmission of mental health, pooled and by gender of child

	Treatment effects			
	Without covariates	**With covariates**	**(2)-RI *p*-values**	**(2)-FWER *p*-values**
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**
A1. Trauma, pooled				
Difference	-0.188***	-0.177***	0.00	0.00
	(0.056)	(0.054)		
A2. Trauma, by child's gender				
Difference, if girl	-0.157**	-0.147**	0.03	-
	(0.066)	(0.066)		
Difference, if boy	-0.221***	-0.216***	0.00	-
	(0.070)	(0.069)		
B1. Depression severity, pooled				
Difference	-0.157**	-0.155**	0.03	0.00
	(0.072)	(0.069)		
B2. Depression severity, by child's gender				
Difference, if girl	-0.167*	-0.167**	0.05	-
	(0.086)	(0.084)		
Difference, if boy	-0.141*	-0.134*	0.09	-
	(0.081)	(0.079)		
Observations	2,803	2,798	-	-

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Column (1): treatment effect estimated without any baseline covariates. Column (2): treatment effect estimated with all baseline covariates (as in equation 3). Covariates are listed under Table 2. Dependent variables are absolute differences in mental health indices (trauma under A1 and A2, and depression under B1 and B2) between mothers and children. That is, $Difference = |Mother - Child|$'s mental health index. Standard errors, clustered at the block level (251 clusters), are in parentheses. Column (3) reports RI p-values for the full model (column 2), which are randomization inference p-values (with 1,000 replications) (Young, 2019). Column (4) reports FWER p-values for the full model (column 2), which are the List-Shaikh-Xu (LSX) familywise error rate adjusted p-values (with 3,000 replications) (List et al., 2019). Since LSX does not allow covariates or conditions (e.g., by gender of the child), we report FWER-adjusted p-values only for the pooled observations. Observations with girl child is 1,387 in column (1) and 1,382 in column (2). Observations with boy child is 1,416 in both columns (1) and (2). Correlation of mother-child mental health is reported in Table A6 in Appendix A.

Table 5: Social desirability bias check

	Mother outcomes					Child outcomes						
VARIABLES	**Trauma (1)**	**Dep. (2)**	**Happ. (3)**	**Aspr. (4)**	**Belong. (5)**	**Trauma (6)**	**Dep. (7)**	**Comm. (8)**	**Gross. (9)**	**Fine. (10)**	**Prob. (11)**	**Social. (12)**
Treatment	-0.229***	-0.123**	0.114**	-0.060	0.200***	-0.068	-0.134**	0.208***	0.158**	-0.035	0.144**	0.144*
	(0.055)	(0.061)	(0.057)	(0.067)	(0.067)	(0.061)	(0.067)	(0.070)	(0.070)	(0.071)	(0.063)	(0.080)
High SDB	0.083	0.042	-0.024	0.014	-0.009	0.100	0.014	0.025	-0.041	-0.019	-0.039	0.063
	(0.050)	(0.058)	(0.054)	(0.057)	(0.062)	(0.061)	(0.063)	(0.056)	(0.063)	(0.057)	(0.058)	(0.058)
Treatment×High SDB	-0.008	-0.050	0.005	-0.013	-0.048	-0.065	0.027	0.051	0.049	0.032	0.076	-0.036
	(0.061)	(0.068)	(0.068)	(0.075)	(0.077)	(0.077)	(0.076)	(0.069)	(0.078)	(0.073)	(0.073)	(0.076)
Observations	2,798	2,798	2,798	2,798	2,798	2,798	2,798	2,840	2,840	2,840	2,840	2,840
R-squared	0.040	0.026	0.028	0.063	0.062	0.032	0.017	0.081	0.054	0.093	0.081	0.026

Robust standard errors clustered at the block level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: All outcomes are standardized indices such that the control group has mean zero and SD one. Outcomes in columns 1-5 are of mothers: (1) trauma, (2) depression, (3) happiness, (4) future aspirations, and (5) belongingness. Outcomes in columns 6-12 are of children: (6) trauma, (7) depression, (8) communication skills, (9) gross-motor skills, (10) fine-motor skills, (11) problem-solving skills, and (12) social skills. Treatment is a dummy that equals to 1 if respondents are in the treatment arm and 0 otherwise. High SDB is a dummy that equals to 1 if the social desirability bias (SDB) score is above 8 (which is the median value) and 0 if below. All specifications include the usual set of controls and camp fixed effects as in Table 2.

Table 6: Potential mechanisms

		Treatment effects			
	Control mean	Pooled	Girl child	Boy child	Diff (4)-(3)
Intermediate outcomes	(1)	(2)	(3)	(4)	(5)
A. Mental health of mothers					
Doctor visits (0-4)	1.88	0.004	0.014	-0.011	-0.027
	[0.79]	(0.034)	(0.045)	(0.045)	(0.059)
Disagreements/arguments with spouse (0-4)	1.04	-0.054	-0.070	-0.038	0.022
	[0.90]	(0.034)	(0.053)	(0.045)	(0.068)
Seek help for household chores (0-4)	1.05	-0.016	0.004	-0.041	-0.030
	[0.95]	(0.039)	(0.058)	(0.055)	(0.078)
Communication during lockdown (0-4)	1.93	-0.011	0.011	-0.023	0.005
	[0.78]	(0.029)	(0.041)	(0.043)	(0.055)
B. Children's development					
Mother's time input per day (0-24)	9.15	1.498***	1.915***	1.113***	-0.684
	[5.83]	(0.244)	(0.324)	(0.331)	(0.436)
Father's time input per day (0-24)	5.14	0.066	-0.053	0.144	0.215
	[3.01]	(0.114)	(0.168)	(0.160)	(0.226)
Age stopped breastfeeding	20.83	0.161	-0.161	0.414*	0.653*
	[5.04]	(0.173)	(0.267)	(0.250)	(0.361)
Times feeding child per day	3.97	0.011	0.041	-0.017	-0.074
	[1.47]	(0.057)	(0.080)	(0.074)	(0.104)
Negative parenting (0-4)	0.67	-0.022*	-0.027	-0.016	0.004
	[0.33]	(0.011)	(0.017)	(0.014)	(0.022)
Ask others to babysit (0-4)	0.87	0.011	0.035	-0.007	-0.060
	[0.94]	(0.038)	(0.058)	(0.052)	(0.071)
Prevalence of indoor smoking (0-4)	0.32	0.036	0.067	0.006	-0.028
	[0.76]	(0.030)	(0.044)	(0.041)	(0.059)
Let child walk/play barefoot (0-4)	0.65	-0.069**	-0.029	-0.117***	-0.056
	[0.83]	(0.032)	(0.046)	(0.042)	(0.059)
Observations	1,166	2,840	1,400	1,440	2,840

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Column (1): control group average at endline with standard deviations in brackets; Column (2): treatment effect estimated without any baseline covariates. Column (3): treatment effect estimated with all baseline covariates (as in equation 1). Standard errors, clustered at the block level (251 clusters), are in parentheses. Column (4) reports RI p-values for the full model (column 2), which are randomization inference p-values (with 1,000 replications) (Young, 2019). Column (5) reports FWER p-values for the full model (column 2), which are the List-Shaikh-Xu familywise error rate adjusted p-values (with 3,000 replications) based on 12 tests (List et al., 2019).

Table 7: Treatment effects: Inverse Probability Weighting & Lee bounds

	Treatment effects		Lee (2009) bounds	
	Unadjusted	IPW	Lower	Upper
Dependent variables	(1)	(2)	(3)	(4)
A1. Mothers' mental health				
Trauma	-0.233***	-0.234***	-0.470***	-0.160***
	(0.051)	(0.049)	(0.035)	(0.037)
Depression	-0.144***	-0.144***	-0.330***	-0.104***
	(0.054)	(0.052)	(0.030)	(0.034)
A2. Mothers'wellbeing				
Happiness	0.117**	0.124**	0.011	0.523***
	(0.056)	(0.054)	(0.040)	(0.044)
Aspirations	-0.066	-0.073	-0.295***	0.242***
	(0.062)	(0.061)	(0.040)	(0.046)
Belongingness	0.179***	0.190***	0.076*	0.490***
	(0.057)	(0.055)	(0.043)	(0.044)
B1. Children's mental health				
Trauma	-0.096*	-0.094*	-0.380***	-0.024
	(0.055)	(0.054)	(0.036)	(0.038)
Depression	-0.122**	0.117**	-0.343***	-0.059
	(0.059)	(0.057)	(0.029)	(0.038)
B2. Children's development				
Communication skills	0.229***	0.232***	0.139***	0.609***
	(0.059)	(0.058)	(0.042)	(0.049)
Gross-motor skills	0.179***	0.189***	0.175***	0.482***
	(0.058)	(0.056)	(0.042)	(0.044)
Fine-motor skills	-0.021	-0.017	-0.271***	0.289***
	(0.066)	(0.064)	(0.051)	(0.045)
Problem-solving skills	0.177***	0.172***	-0.027	0.489***
	(0.055)	(0.054)	(0.047)	(0.046)
Social skills	0.128*	0.148**	-0.135***	0.410***
	(0.067)	(0.066)	(0.042)	(0.043)
B3. Children's anthropometrics				
Height-for-age z-score	0.515***	0.521***	-0.512***	1.487***
	(0.139)	(0.137)	(0.150)	(0.161)
Weight-for-age z-score	0.641***	0.635***	0.037	1.142***
	(0.092)	(0.092)	(0.089)	(0.094)
Weight-for-height z-score	0.508***	0.493***	-0.336***	1.054***
	(0.125)	(0.127)	(0.114)	(0.128)

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Column (1) reports unadjusted/unweighted treatment effects, same as in Table 2. Column (2) reports the Inverse Probability Weight (IPW) adjusted treatment effects. Columns (3)-(4) report the lower and upper bound treatment effects using Lee (2009) bounds.

Table 8: Heterogeneity Using Random Forest: Classification Analysis (CLAN)

COVARIATES	**Most** (1)	**Least** (2)	**Difference** (3)	COVARIATES	**Most** (4)	**Least** (5)	**Difference** (6)
Outcome: mothers' trauma				**Outcome: children's trauma**			
Mother's trauma at baseline	0.602	0.361	0.232	Child's trauma at baseline	0.493	0.488	0.004
	(0.545,	(0.305,	(0.151,		(0.434,	(0.429,	(-0.078,
	0.659)	0.417)	0.314)		0.552)	0.546)	0.087)
	-	-	[0.000]***		-	-	[1.000]
Age of mother	26.38	25.55	0.815	Age of child	13.95	14.97	0.973
	(25.70,	(24.86,	(-0.142,		(13.19,	(14.23,	(-2.037,
	27.06)	26.27)	1.773)		14.71)	15.72)	0.090)
	-	-	[0.191]		-	-	[0.154]
Attended primary	0.664	0.843	-0.188	Gender of child	0.498	0.522	-0.019
	(0.615,	(0.794,	(-0.257,		(0.439,	(0.464,	(-0.102,
	0.716)	0.892)	-0.119)		0.557)	0.581)	0.064)
	-	-	[0.000]***		-	-	[1.000]
Victimization in Myanmar	0.197	0.100	0.094	Victimization in Myanmar	0.148	0.145	0.002
	(0.182,	(0.086,	(0.073,		(0.134,	(0.130,	(-0.019,
	0.211)	0.115)	0.115)		0.163)	0.160)	0.024)
	-	-	[0.000]***		-	-	[1.000]
Abuse in camp	0.027	0.006	0.022	Abuse in camp	0.013	0.014	0.000
	(0.021,	(0.000,	(0.014,		(0.008,	(0.008,	(-0.007,
	0.033)	0.011)	0.030)		0.019)	0.019)	0.008)
	-	-	[0.000]***		-	-	[1.000]
Outcome: mothers' depression index				**Outcome: children's depression index**			
Mother depressed at baseline	0.457	0.052	0.400	Child depressed at baseline	0.282	0.071	0.216
	(0.412,	(0.007,	(0.336,		(0.240,	(0.028,	(0.158,
	0.502)	0.097)	0.463)		0.324)	0.113)	0.276)
	-	-	[0.000]***		-	-	[0.000]***
Age of mother	26.40	24.90	1.446	Age of child	15.61	13.25	2.435
	(25.73,	(24.17,	(0.505,		(14.86,	(12.51,	(1.346,
	27.07)	25.57)	2.396)		16.37)	13.99)	3.529)
	-	-	[0.006]***		-	-	[0.000]***
Attended primary	0.731	0.719	0.011	Gender of child	0.524	0.496	0.015
	(0.680,	(0.666,	(-0.060,		(0.465,	(0.438,	(-0.068,
	0.783)	0.773)	0.084)		0.583)	0.555)	0.099)
	-	-	[1.000]		-	-	[1.000]
Victimization in Myanmar	0.192	0.098	0.094	Victimization in Myanmar	0.153	0.134	0.015
	(0.177,	(0.083,	(0.074,		(0.138,	(0.120,	(-0.006,
	0.207)	0.113)	0.115)		0.168)	0.150)	0.036)
	-	-	[0.000]***		-	-	[0.313]
Abuse in camp	0.024	0.007	0.018	Abuse in camp	0.018	0.012	0.006
	(0.018,	(0.001,	(0.009,		(0.012,	(0.005,	(-0.001,
	0.030)	0.013)	0.026)		0.023)	0.018)	0.014)
	-	-	[0.000]***		-	-	[0.219]
Outcome: Children's composite development index				**Outcome: Children's stunting**			
Child's trauma at baseline	0.516	0.450	0.073	Child's trauma at baseline	0.484	0.447	0.052
	(0.458,	(0.392,	(-0.009,		(0.426,	(0.388,	(-0.029,
	0.574)	0.508)	0.154)		0.542)	0.505)	0.134)
	-	-	[0.165]		-	-	[0.418]
Child depressed at baseline	0.180	0.118	0.065	Child depressed at baseline	0.153	0.164	-0.015
	(0.139,	(0.076,	(0.005,		(0.110,	(0.122,	(-0.076,
	0.222)	0.161)	0.128)		0.195)	0.207)	0.045)
	-	-	[0.066]*		-	-	[1.000]
Age of child	18.81	10.640	8.073	Age of child	15.97	13.840	2.039
	(18.17,	(10.02,	(7.215,		(15.22,	(13.08,	(1.001,
	19.43)	11.25)	8.931)		16.71)	14.58)	3.102)
	-	-	[0.000]***		-	-	[0.000]***
Gender of child	0.477	0.495	-0.035	Gender of child	0.510	0.492	0.016
	(0.419,	(0.437,	(-0.117,		(0.452,	(0.434,	(-0.067,
	0.535)	0.553)	0.047)		0.568)	0.550)	0.098)
	-	-	[0.803]		-	-	[1.000]
Victimization in Myanmar	0.131	0.167	-0.036	Victimization in Myanmar	0.157	0.135	0.024

	(0.116,	(0.152,	(-0.058,		(0.143,	(0.121,	(0.005,
	0.146)	0.181)	-0.015)		0.171)	0.150)	0.045)
	-	-	[0.002]***		-	-	[0.027]**
Abuse in Camp	0.007	0.021	-0.014	Abuse in Camp	0.016	0.011	0.004
	(0.002,	(0.015,	(-0.021,		(0.010,	(0.005,	(-0.003,
	0.012)	0.027)	-0.006)		0.021)	0.016)	0.012)
	-	-	[0.001]***		-	-	[0.449]
Outcome: Children's underweight				**Outcome: Children's wasting**			
Child's trauma at baseline	0.504	0.459	0.044	Child's trauma at baseline	0.413	0.550	-0.147
	(0.445,	(0.401,	(-0.037,		(0.356,	(0.492,	(-0.230,
	0.562)	0.517)	0.126)		0.471)	0.609)	-0.065)
	-	-	[0.504]		-	-	[0.001]***
Child depressed at baseline	0.165	0.148	0.013	Child depressed at baseline	0.139	0.175	-0.028
	(0.122,	(0.106,	(-0.046,		(0.097,	(0.133,	(-0.090,
	0.209)	0.190)	0.072)		0.180)	0.218)	0.032)
	-	-	[1.000]		-	-	[0.718]
Age of child	20.87	7.844	13.000	Age of child	15.79	13.09	2.812
	(20.52,	(7.498,	(12.52,		(15.06,	(12.36,	(1.795,
	21.21)	8.188)	13.51)		16.51)	13.81)	3.829)
	-	-	[0.000]***		-	-	[0.000]***
Gender of child	0.970	0.032	0.923	Gender of child	0.442	0.552	-0.118
	(0.949,	(0.009,	(0.891,		(0.384,	(0.494,	(-0.200,
	0.990)	0.055)	0.954)		0.501)	0.610)	-0.035)
	-	-	[0.000]***		-	-	[0.011]**
Victimization in Myanmar	0.150	0.147	0.001	Victimization in Myanmar	0.151	0.135	0.016
	(0.135,	(0.132,	(-0.020,		(0.138,	(0.121,	(-0.004,
	0.166)	0.162)	0.023)		0.166)	0.148)	0.036)
	-	-	[1.000]		-	-	[0.215]
Abuse in camp	0.011	0.016	-0.004	Abuse in camp	0.013	0.012	0.001
	(0.006,	(0.011,	(-0.013,		(0.008,	(0.007,	(-0.005,
	0.017)	0.022)	0.003)		0.018)	0.016)	0.008)
	-	-	[0.541]		-	-	[1.000]

*** p<0.01, ** p<0.05, * p<0.1

Note: This table reports CLAN results using Random Forest. 90% confidence interval are in parenthesis; p-values for the hypothesis that the parameter is equal to zero are in brackets. 'Most' and 'Least' are the 20% most (top quintile) and 20% least (bottom quintile) affected groups; 'Difference' is the difference in average characteristics between 'Most' and '"Least' affected groups (i.e., most minus least). Outcome of each panel is mentioned at the top. Outcomes that are indices have been control group-standardized. Stunting, Underweight, and Wasting outcomes are dummies where 1 equals stunted, underweight, or wasted growth and 0 otherwise.

Forced Displacement, Mental Health, and Child Development: Evidence from the Rohingya Refugees

Online Appendix

By Asad Islam[1], Tanvir Ahmed Mozumder[2], Tabassum Rahman[3], Tanvir Shatil[4], Abu Siddique[5]

List of Figures

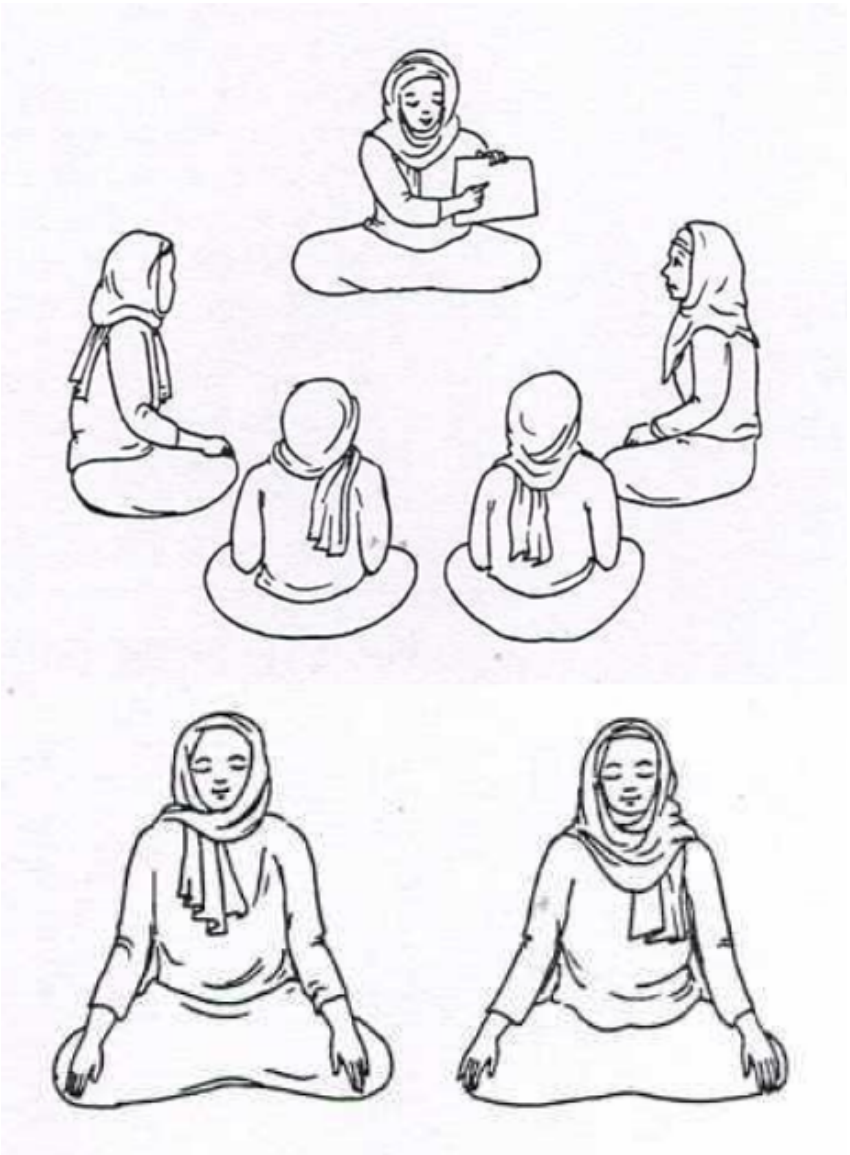
[1]Centre for Development Economics & Sustainability (CDES) and Department of Economics, Monash University, Australia; J-PAL. Email: asadul.islam@monash.edu
[2]BRAC Institute of Governance & Development, BRAC University, Bangladesh. Email: tanvir.mozumder@bracu.ac.bd
[3]School of Medicine & Public Health, University of Newcastle, Australia. Email: tabassum.rahman@uon.edu.au
[4]BRAC Institute of Governance & Development, BRAC University, Bangladesh. Email: tanvir.shatil@bracu.ac.bd
[5]Department of Political Economy, King's College London, UK; CDES; TUM. Email: abu.b.siddique@kcl.ac.uk

List of Tables

A Appendix: Additional Tables and Figures

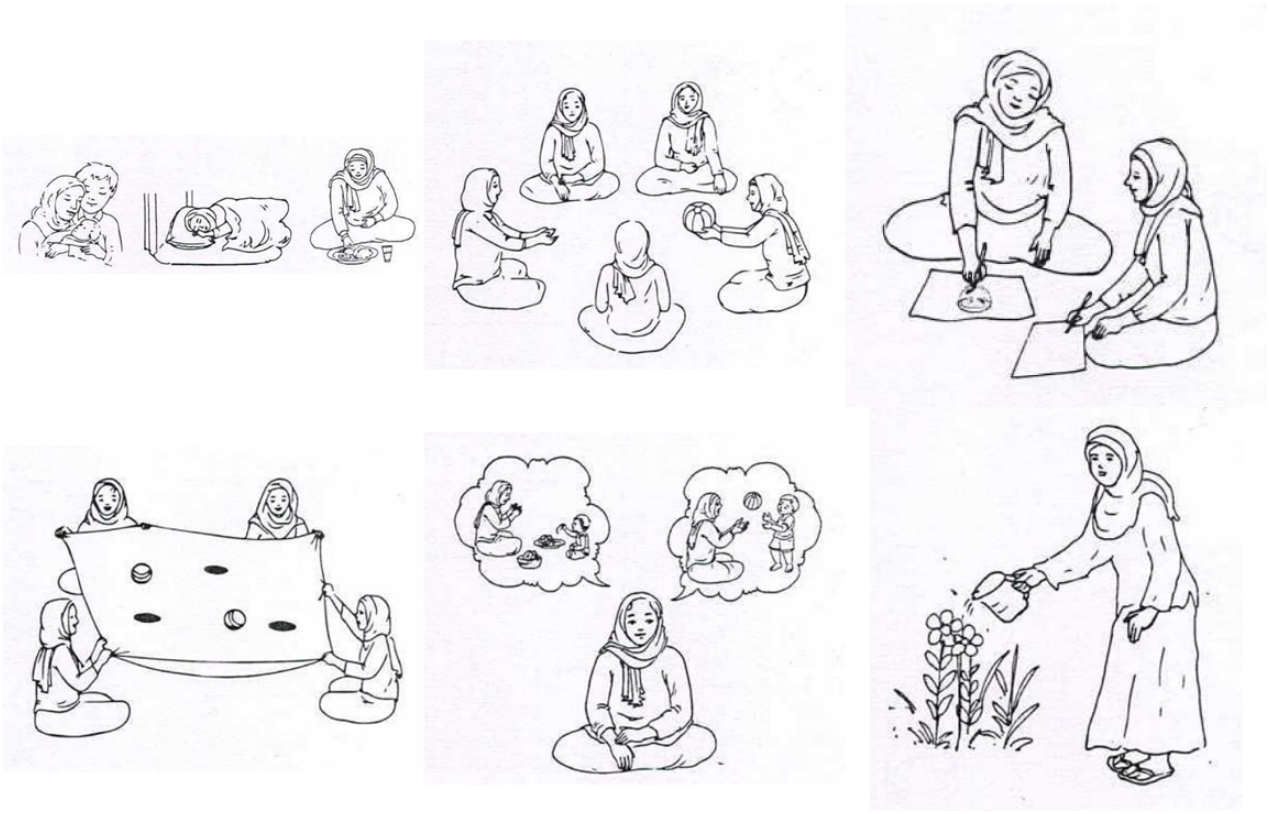
A.1 Figures

Figure A1: Session: step 1



Note: In step 1, mothers greet each other, do breathing exercises, and provide feedback on the previous week's homework. These pictures are taken from the session materials (pictorial directions) for illiterate mother-volunteers.

Figure A2: Session: step 2

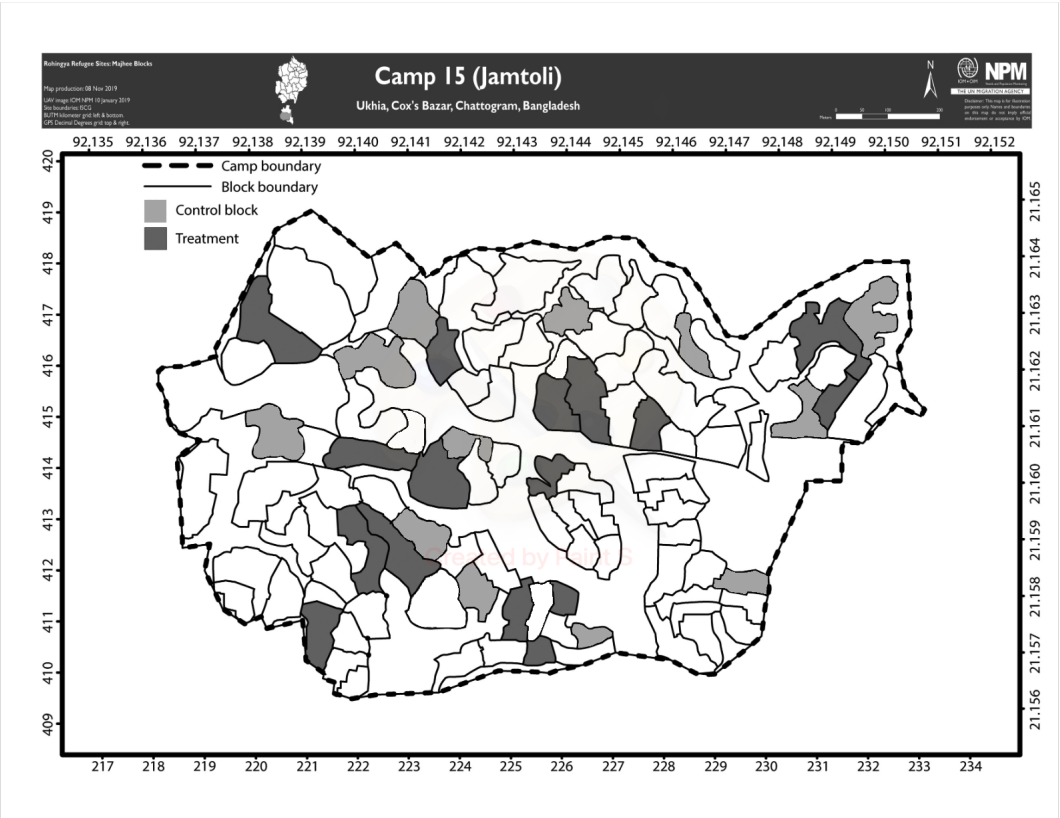


Note: In step 2, mothers engage in discussions on their personal well-being and play games with other mothers. These pictures are taken from the session materials (pictorial directions) for illiterate mother-volunteers.

Figure A3: Session: step 3

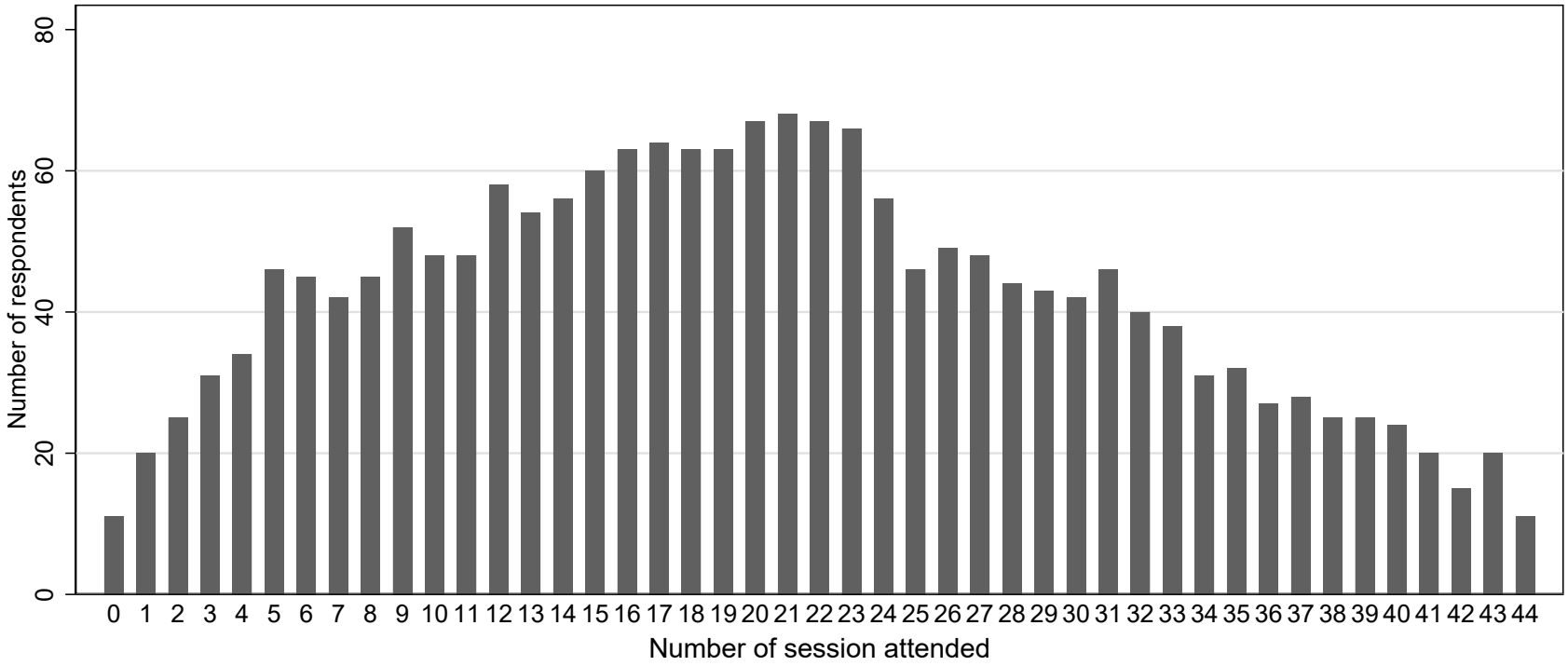
Note: In step 3, mothers receive advice on childcare and engage in play activities with their children during the session. These pictures are taken from the session materials (pictorial directions) for illiterate mother-volunteers.

Figure A4: Map of a Rohingya camp



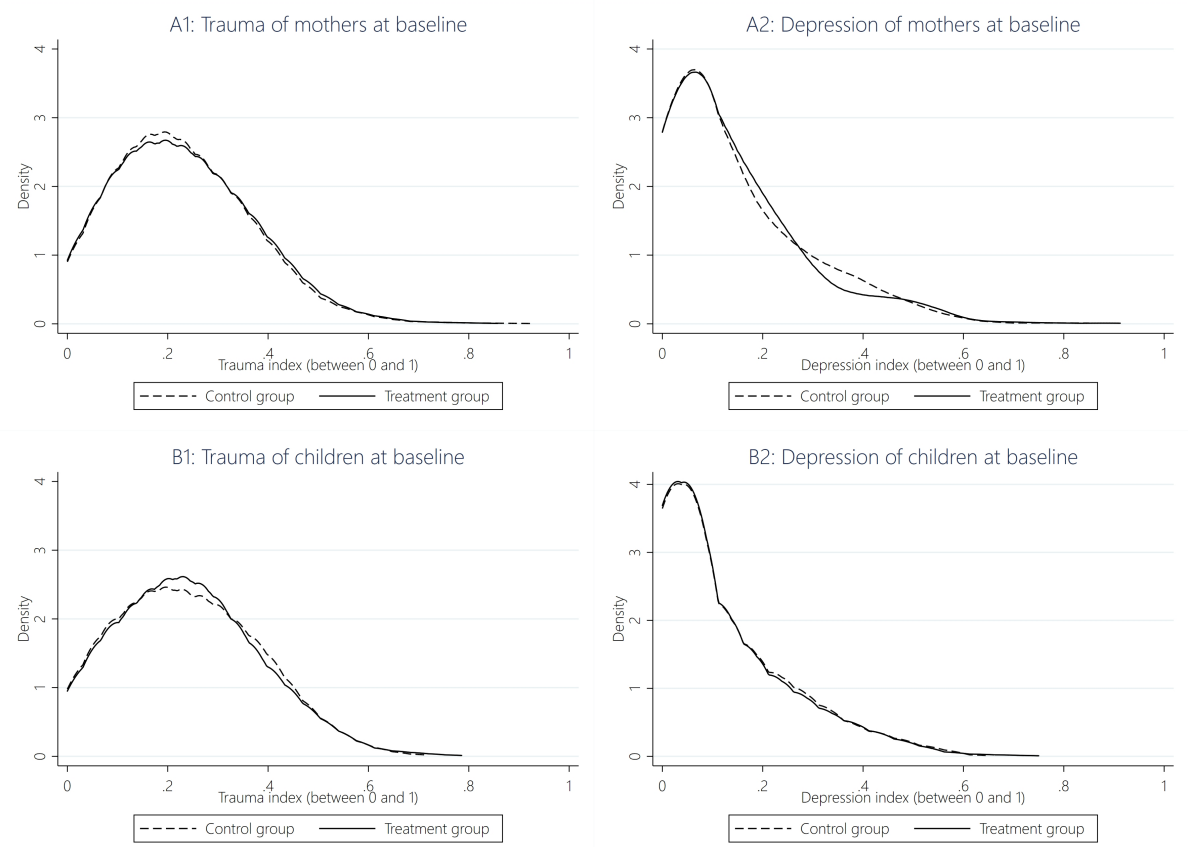
Note: This is a map of Camp 15, showing the treatment and control blocks, and boundaries.

Figure A5: Attendance in treatment sessions



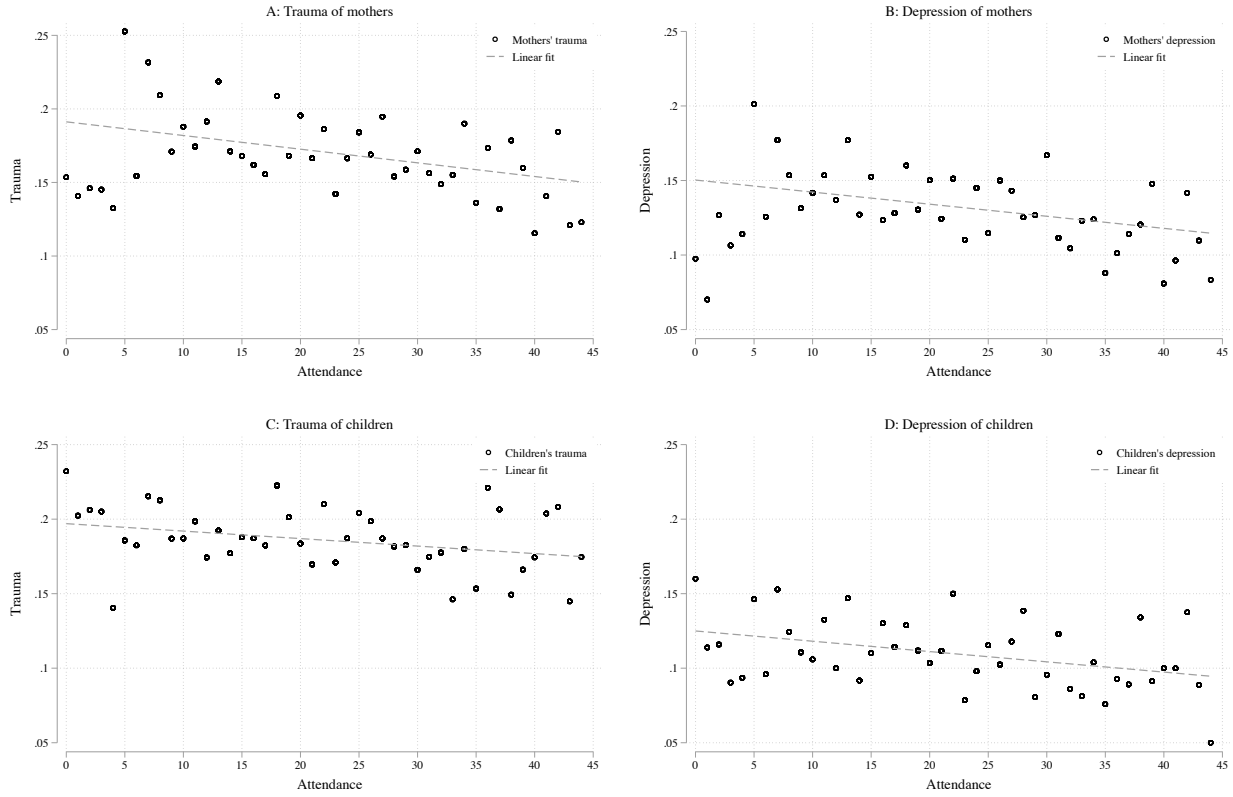
Note: This figure shows the distribution of attendance in treatment group sessions. 0 in the x-axis corresponds to the number of participants that never attended any sessions and 44 corresponds to the number of participants that attended all sessions.

Figure A6: Mental health of mothers and children at baseline



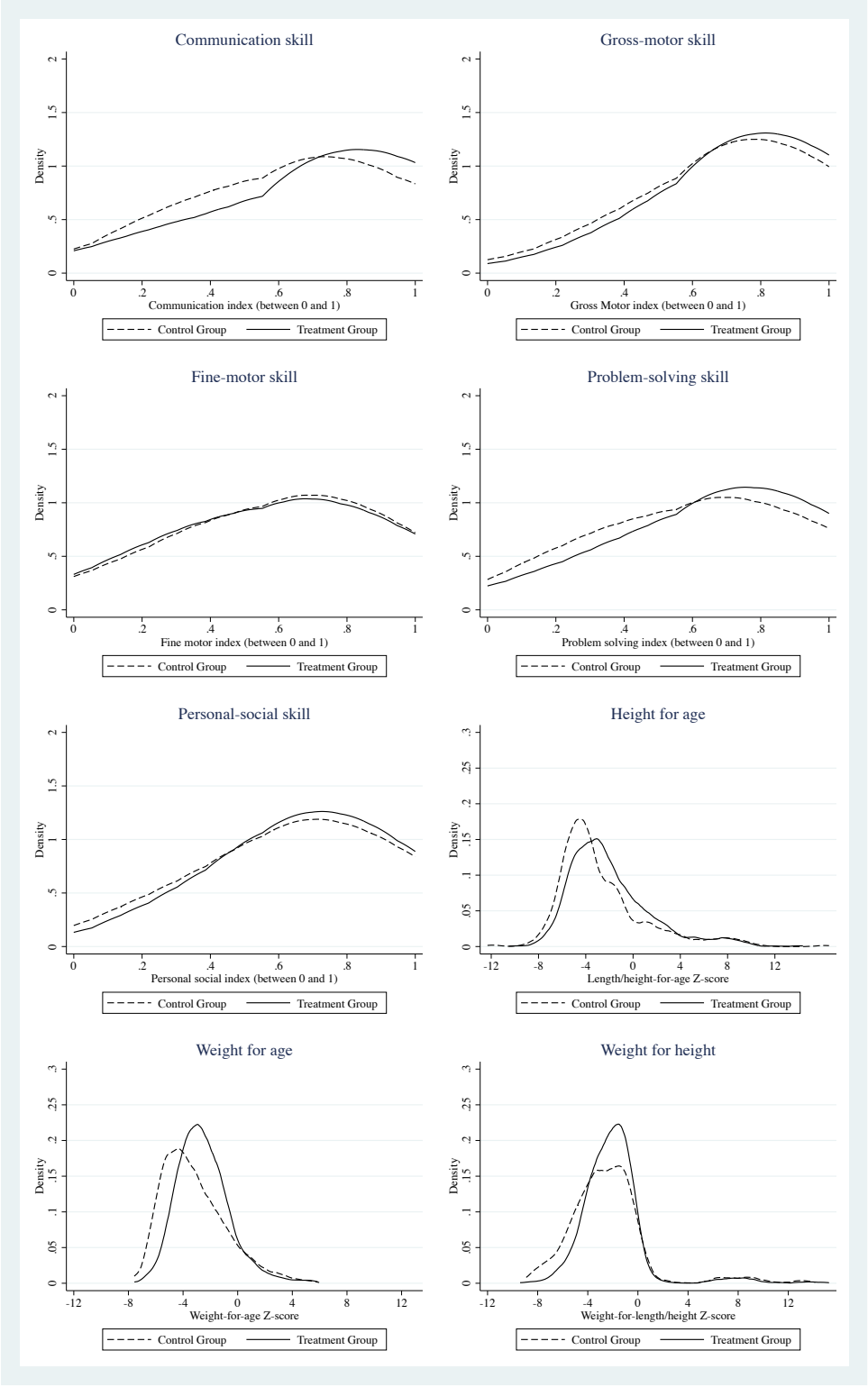
Note: This figure shows the distribution of mental health of mothers (A1 and A2) and children (B1 and B2) at the baseline (estimated from kernel density estimation). Trauma and depression indices are averages of responses to trauma and depression questions, where higher values correspond to more severe mental health conditions. For details on how these two indices are constructed, see Appendix B.

Figure A7: Correlation between mental health and session attendance



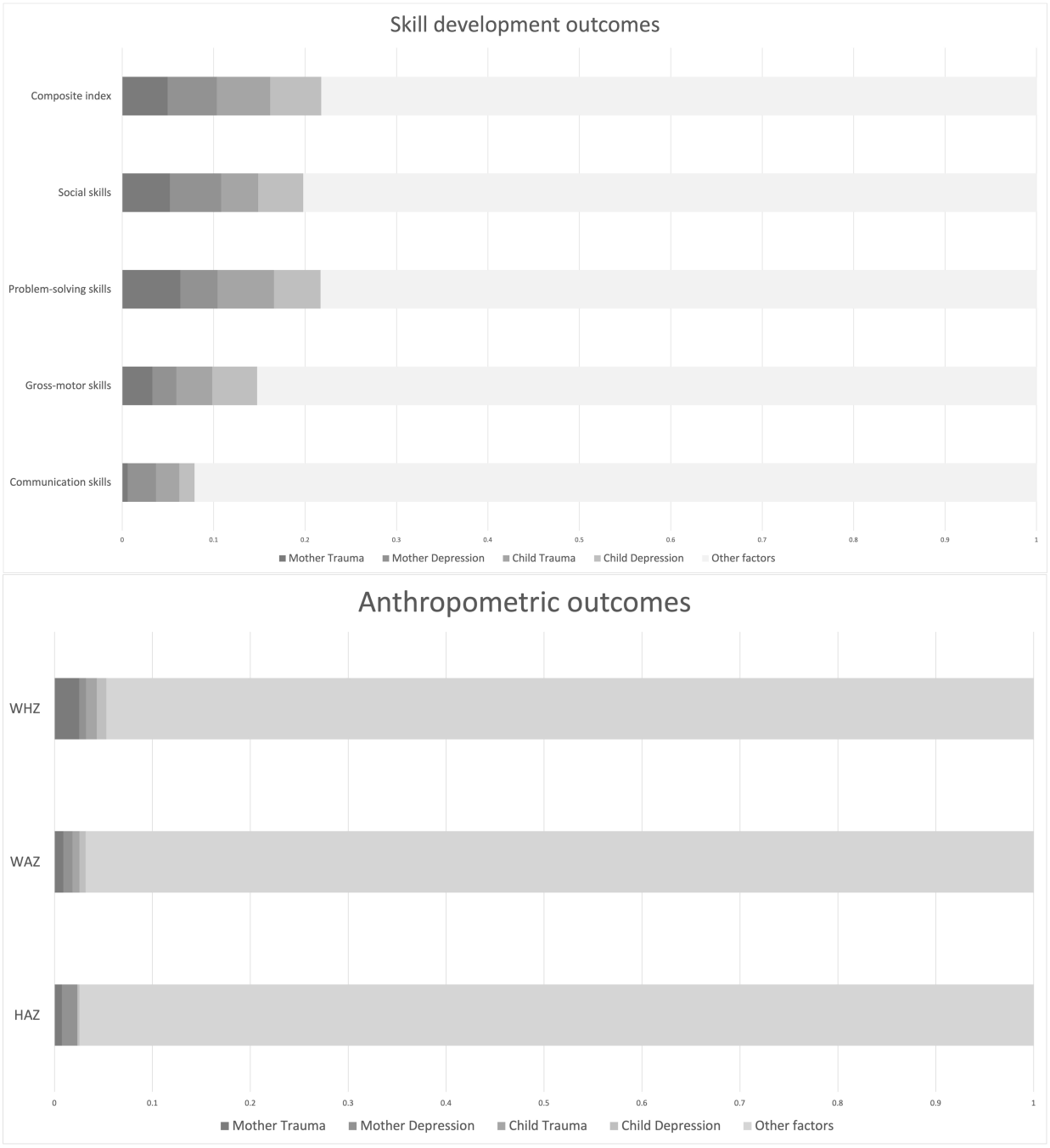
Note: This figure shows the correlation between mental health (y-axis) and session attendance (x-axis). All mental health outcomes have been normalized to be between 0 and 1, where higher value corresponds to poor mental health. Attendance is between 0 and 44, where 44 corresponds to those who attended all 44 sessions.

Figure A8: Skills development and anthropometric distributions at endline, by treatment arm



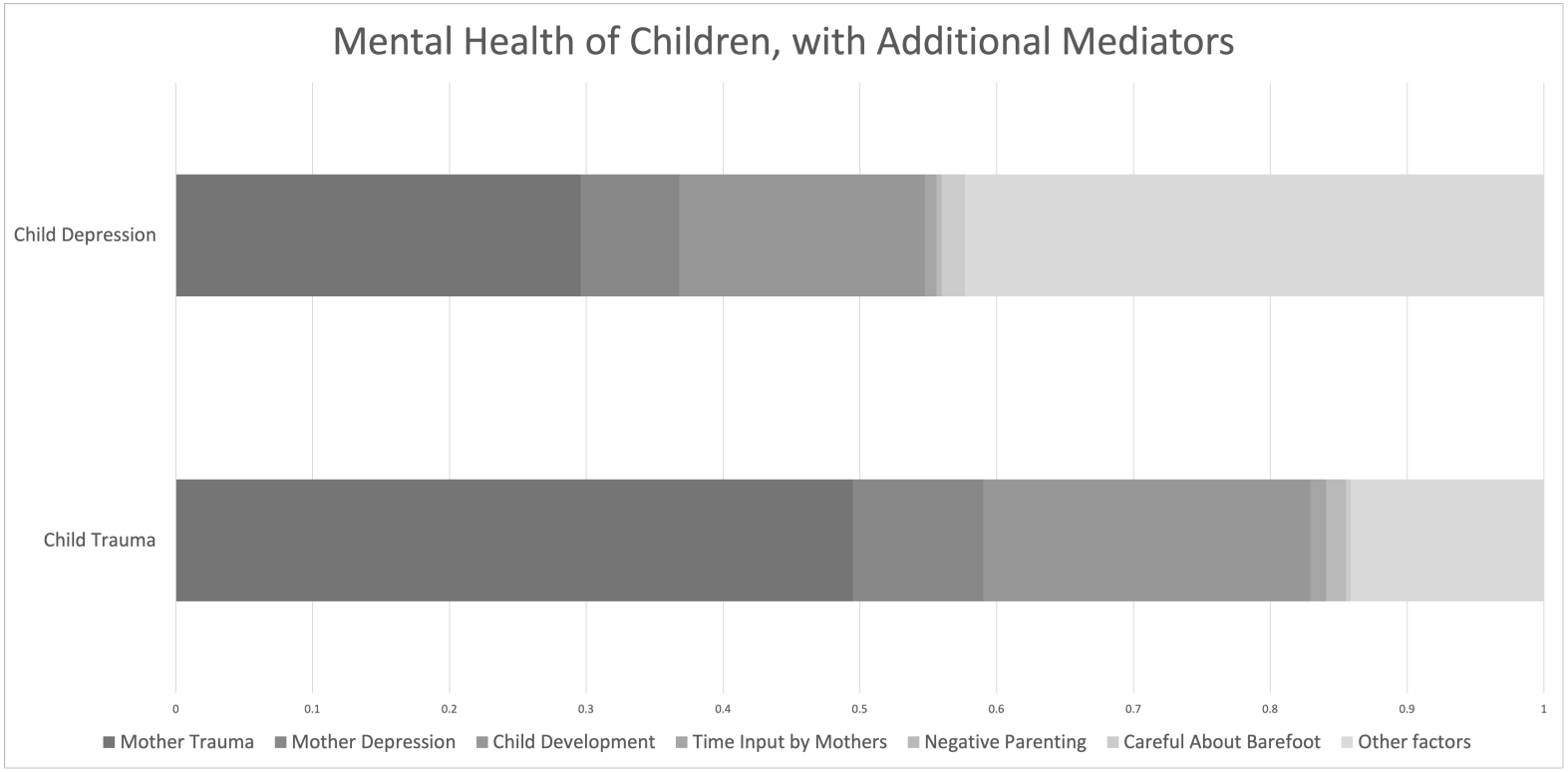
Note: This figure shows the distribution, by treatment, of the five skills and three anthropometric development outcomes. All outcomes are measured at endline.

Figure A9: Mediated effects on skill development and anthropometric outcomes



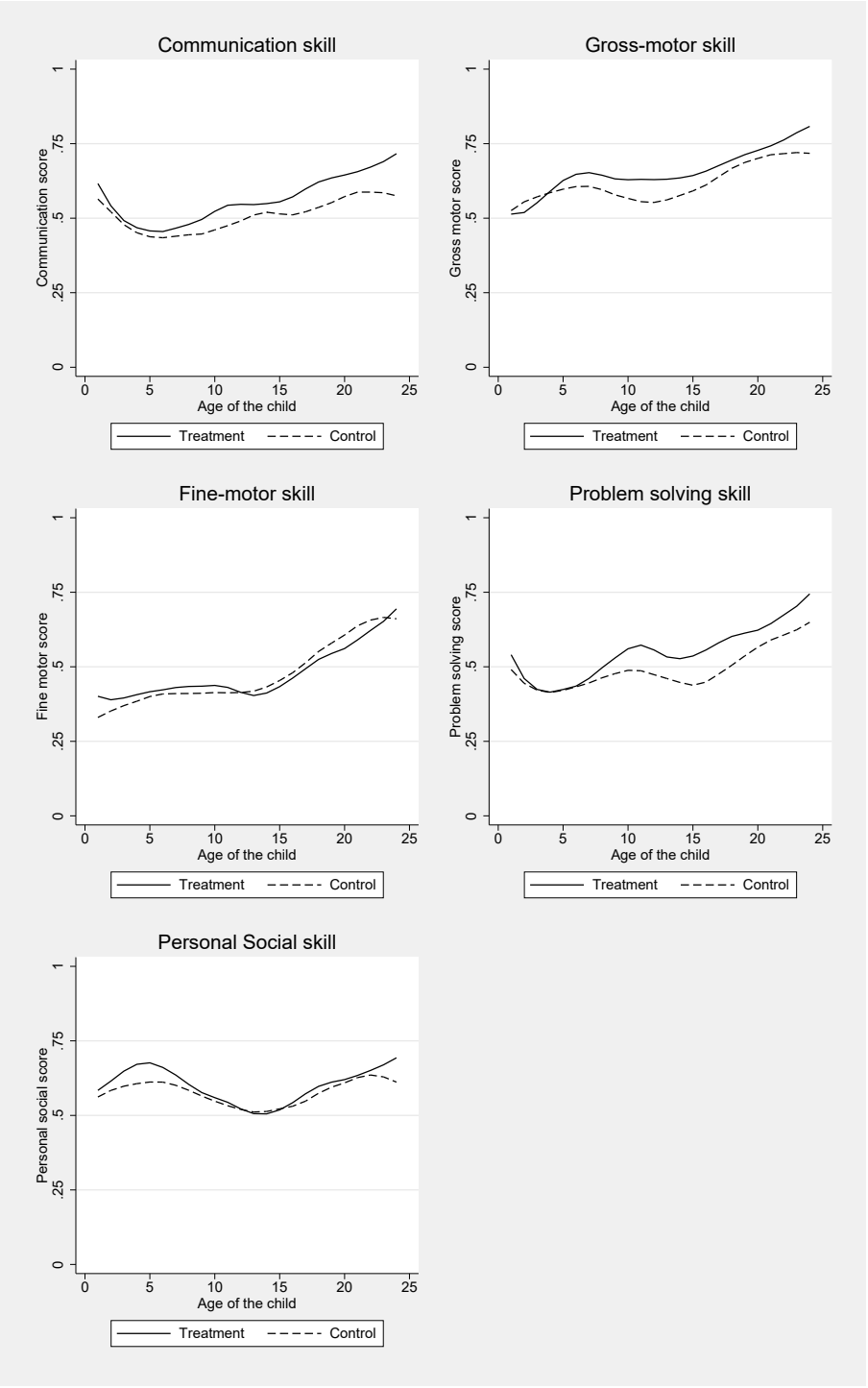
Note: This figure reports the mediated and residual effects on child development and anthropometric outcomes. Each shade of a bar corresponds to the proportion of the total effect that is mediated.

Figure A10: Mediated effects on children's mental health, with additional mediators



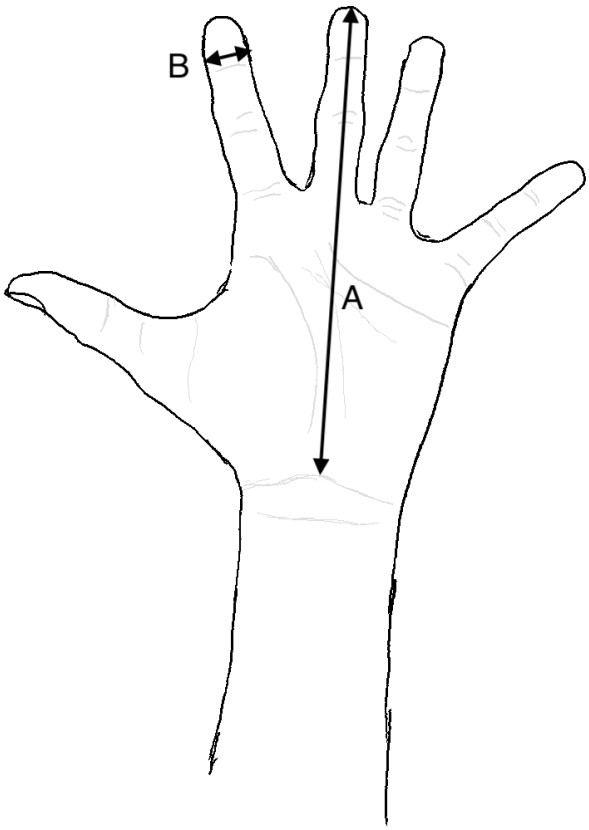
Note: This figure reports the mediated and residual effects on children's mental health outcomes with additional mediators that were previously considered unobserved. Each shade of a bar corresponds to the proportion of the total effect that is mediated.

Figure A11: Treatment effects on skills development, by children's age



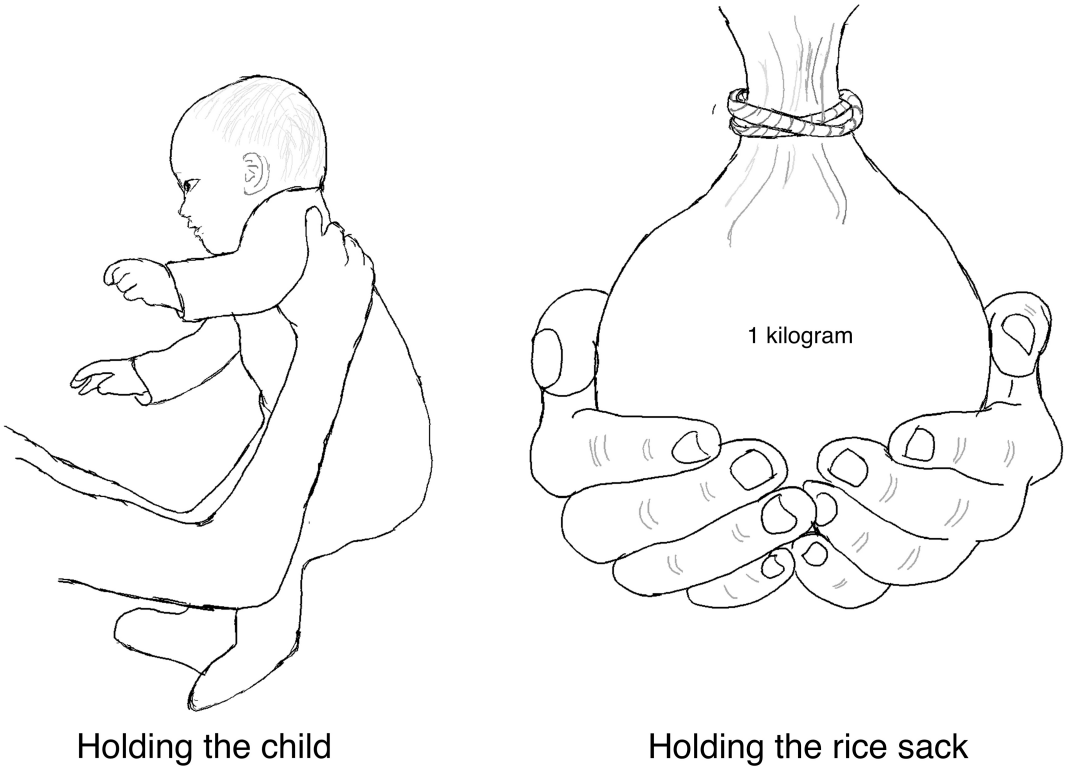
Note: This figure shows the treatment effects on skills development by children's age (between 0-24 months). All outcomes are measured at endline.

Figure A12: Over-the-phone measures of length



Note: We used two obsolete anthropic unit of length—hand and finger—to measure children's height over the phone. Here 'hand' length is the distance from the tip of the middle finger to the mid-points of the distal transverse crease of the wrist (i.e., length of A); 'finger' is the width of the index finger (i.e., length of B). All measures were carried out using the right hand, and mothers reported lengths to enumerators in 'hand' and 'finger' units (also, 'half-hand' or 'half-finger' units were considered). Later, using Asadujjaman et al. (2019), we converted these two units into centimeters (cm): hand length= 16 cm and finger= 2 cm.

Figure A13: Over-the-phone measures of weight



Note: To measure weight, enumerators asked mothers to first hold their child using both hands and then hold a sack of 1 kilogram of rice (which they receive from camp authorities through food vouchers) using both hands, and then repeat the process again. After weighing child and rice sack, mothers were asked to report the weight of the child in 'rice sack' units (also, half, quarter, fifth, and tenth of the rice sack units were considered). Later, 'rice sack' unit was converted to kilograms.

A.2 Tables

Table A1: Baseline outcomes and balance checks

VARIABLES	**Treatment** (*Std. Dev.*)	N_T	**Control** (*Std. Dev.*)	N_C	**T-test** ***p*-values**	**RI** ***p*-values**
			Panel A: Mother outcomes			
Traumatized (=1 if true)	0.45	1,911	0.44	1,586	0.69	0.70
	(0.50)		(0.50)			
Depressed (=1 if true)	0.17	1,911	0.20	1,586	0.87	0.88
	(0.37)		(0.40)			
Happiness index ($0 \leq index \leq 1$)	0.77	1,911	0.78	1,586	0.52	0.53
	(0.17)		(0.17)			
Aspiration index ($0 \leq index \leq 1$)	0.61	1,911	0.62	1,586	0.21	0.21
	(0.11)		(0.11)			
Belongingness index ($0 \leq index \leq 1$)	0.79	1,911	0.79	1,586	0.23	0.20
	(0.15)		(0.16)			
			Panel B: Child outcomes			
Traumatized (=1 if true)	0.49	1,911	0.48	1,588	0.57	0.57
	(0.50)		(0.50)			
Depressed (=1 if true)	0.17	1,911	0.18	1,588	0.97	0.97
	(0.37)		(0.38)			
Communication skills index ($0 \leq index \leq 1$)	0.56	1,911	0.56	1,588	0.75	0.73
	(0.30)		(0.31)			
Gross-motor skills index ($0 \leq index \leq 1$)	0.63	1,911	0.63	1,588	0.86	0.85
	(0.32)		(0.33)			
Fine-motor skills index ($0 \leq index \leq 1$)	0.50	1,911	0.48	1,588	0.44	0.43
	(0.31)		(0.31)			
Problem-solving skills index ($0 \leq index \leq 1$)	0.47	1,911	0.48	1,588	0.60	0.96
	(0.31)		(0.33)			
Social skills index ($0 \leq index \leq 1$)	0.58	1,911	0.59	1,588	0.72	0.64
	(0.28)		(0.29)			
Underweight for age (=1 if true)	0.23	1,911	0.24	1,588	0.52	0.49
	(0.42)		(0.42)			
Severely underweight for age (=1 if true)	0.08	1,911	0.08	1,588	0.97	0.96
	(0.27)		(0.26)			
Stunted for age (=1 if true)	0.27	1,911	0.27	1,588	0.56	0.58
	(0.44)		(0.45)			
Severely stunted for age (=1 if true)	0.13	1,911	0.12	1,588	0.80	0.80
	(0.33)		(0.32)			
Wasted for age (=1 if true)	0.15	1,911	0.18	1,588	0.04	0.04
	(0.36)		(0.39)			
Severely wasted for age (=1 if true)	0.06	1,911	0.07	1,588	0.45	0.43
	(0.24)		(0.25)			

Note: *Treatment* and *Control* columns show mean of the corresponding variables. Variables that are indices are averages of responses to survey questions associated with the outcomes, such that the value of each variable is between 0 and 1. For instance, *Communication skills* is measured using 6 questions and each question is answered as either 'yes' (=1) or 'no' (=0). So, the *Communication skills* variable under Panel B simply adds up responses and divides the total by 6 (the highest total score). All index variables have been generated in this way. Therefore, these variables simply show the averages. All variables with "=1 if true" are dummies and are self explanatory; T-test p-values are derived from linear regressions, where the dependent variable is from the list above and the independent variable is a dummy that equals to 1 if belongs to the treatment group and 0 if belongs to the control group with camp fixed effects and robust standard errors clustered at the block level; RI p-values are randomization inference p-values (with 1,000 replications) (Young, 2019). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A2: Attrition and baseline characteristics

VARIABLES	**Only Baseline** (*Std. Dev.*)	N_{OB}	**Baseline & Endline** (*Std. Dev.*)	N_{BE}	**T-test** ***p*-values**	**RI** ***p*-values**
A: Mother & household characteristics						
Age	25.57	653	25.49	2,842	0.75	0.73
	(5.89)		(5.73)			
Whether receives voucher	0.49	653	0.51	2,842	0.48	0.48
	(0.50)		(0.51)			
Household size	5.22	654	5.26	2,845	0.70	0.72
	(2.00)		(1.98)			
Employed	0.03	653	0.03	2,842	0.83	0.81
	(0.16)		(0.16)			
Monthly income	0.44	18	0.45	74	0.99	0.96
	(0.51)		(0.50)			
Husband alive	0.96	654	0.97	2,845	0.11	0.14
	(0.20)		(0.17)			
Number of children	2.98	654	2.91	2,845	0.41	0.41
	(1.99)		(1.94)			
Attended school	0.71	654	0.74	2,844	0.0.26	0.25
	(0.46)		(0.44)			
Months in camp	25.07	654	25.75	2,845	0.06*	0.05*
	(10.85)		(14.58)			
Mother is the HH head	0.26	654	0.20	2,845	0.04**	0.03**
	(0.44)		(0.40)			
Household victimization (conflict)	0.15	654	0.16	2,845	0.99	0.99
	(0.13)		(0.12)			
Mother's victimization (camp)	0.01	654	0.01	2,845	0.93	1.00
	(0.05)		(0.04)			
HH victim of at least one conflict abuse	0.87	654	0.87	2,845	0.58	0.70
	(0.33)		(0.34)			
Mother victim of at least one camp abuse	0.15	654	0.16	2,845	0.74	0.71
	(0.36)		(0.37)			
B: Child characteristics						
Age	14.54	654	14.38	2,845	0.67	0.72
	(6.48)		(6.45)			
Gender	0.53	654	0.51	2,845	0.31	0.21
	(0.50)		(0.50)			
Whether elder siblings attend HPL	0.02	654	0.03	2,845	0.23	0.20
	(0.15)		(0.18)			
Child's victimization (camp)	0.01	654	0.01	2,845	0.45	0.49
	(0.06)		(0.05)			
Child victim of at least one camp abuse	0.05	654	0.04	2,845	0.39	0.39
	(0.21)		(0.19)			
Weight (kg)	8.75	654	8.60	2,845	0.19	0.23
	(2.26)		(2.15)			
Height (cm)	75.07	654	74.17	2,845	0.09*	0.11
	(10.08)		(9.74)			

Note: Column 'Only Baseline' reports averages of mothers/children that only took part in the baseline and column N_{OB} reports the corresponding sample size. Column 'Baseline & Endline' reports averages of mothers/children that took part in both baseline and endline surveys, and column N_{BE} reports the corresponding sample size. See the note under Table 3.4 for all variable descriptions. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A3: Attrition, by treatment

	Treatment	Control	Interaction
VARIABLES	**(1)**	**(2)**	**(3)**
Treatment			-0.014
			(0.213)
Age of mothers	0.000	0.001	0.001
	(0.002)	(0.003)	(0.003)
Treatment×Age of mothers			-0.000
			(0.004)
Household Size	-0.004	-0.025	-0.025*
	(0.006)	(0.015)	(0.015)
Treatment×Household Size			0.021
			(0.016)
Mother attended school	0.011	-0.058*	-0.058*
	(0.023)	(0.034)	(0.033)
Treatment×Mother attended school			0.069*
			(0.040)
Household spending	-0.000	0.000**	0.000**
	(0.000)	(0.000)	(0.000)
Treatment×Household spending			-0.000**
			(0.000)
Duration in the camp	0.000	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Treatment×Duration in the camp			0.001**
			(0.001)
Mother receives voucher	-0.012	-0.038	-0.038
	(0.024)	(0.036)	(0.036)
Treatment×Mother receives voucher			0.026
			(0.044)
Husband is alive	-0.019	-0.056	-0.056
	(0.052)	(0.078)	(0.078)
Treatment×Husband is alive			0.037
			(0.093)
Family member stranded	-0.034	0.013	0.013
	(0.025)	(0.046)	(0.046)
Treatment×Family member stranded			-0.047
			(0.053)
HH victimization (conflict)	-0.090	0.080	0.080
	(0.068)	(0.155)	(0.154)
Treatment×HH victimization			-0.170
			(0.169)
Mothers' victimization (camp abuse)	-0.076	0.071	0.071
	(0.193)	(0.305)	(0.304)
Treatment×Mothers' victimization			-0.147
			(0.360)
Mother is the HH head	0.047*	0.068	0.068
	(0.028)	(0.045)	(0.045)
Treatment×Mother is the HH head			-0.021
			(0.053)
Number of children	0.005	0.016	0.016
	(0.007)	(0.015)	(0.015)
Treatment×Number of children			-0.011
			(0.017)
Age of children	-0.000	-0.006	-0.006
	(0.003)	(0.004)	(0.004)
Treatment×Age of children			0.006
			(0.005)
Gender of children	0.002	0.015	0.015
	(0.016)	(0.021)	(0.021)
Treatment×Gender of children			-0.013

			(0.026)
Weight of children (kg)	-0.001	0.007	0.007
	(0.007)	(0.011)	(0.011)
Treatment×Weight of children (kg)			-0.008
			(0.013)
Height of children (cm)	0.001	0.004**	0.004**
	(0.002)	(0.002)	(0.002)
Treatment×Height of children (cm)			-0.003
			(0.003)
Child's victimization (camp)	0.032	0.037	0.037
	(0.172)	(0.208)	(0.207)
Treatment×Child's victimization (camp)			-0.005
			(0.269)
Observations	1,907	1,586	3,493
R-squared	0.007	0.032	0.056
Attrition rate	0.12	0.27	-
Joint p-value on individual/household characteristics	0.83	0.02	-
Joint p-value on interactions	-	-	0.19

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: All columns present estimates using a linear probability model, where the dependent variable is attrition, a dummy variable that equals 1 if a mother did not participate in the endline survey and 0 if she participated in both baseline and endline surveys. The sample in column 1 is mothers/children in the treatment group and the sample in column 2 is mothers/children in the control group. Column 3 pools all sample together. We do not interact the treatment dummy with 'HH victim of at least one conflict abuse', 'Mother victim of at least one camp abuse', and 'Child victim of at least one camp abuse' because these indicators were derived from the 3 victimization indices that we already use. All variables were measured at the baseline. Overall attrition rate is roughly 19% (654 out of 3,499 mothers did not participate in the endline).

Table A4: Mobile phone ownership, by treatment

	Treatment	Control	Interaction
VARIABLES	**(1)**	**(2)**	**(3)**
Treatment			0.238
			(0.163)
Age of mothers	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.002)
Treatment×Age of mothers			0.000
			(0.003)
Household size	-0.004	-0.003	-0.003
	(0.006)	(0.010)	(0.010)
Treatment×Household size			-0.001
			(0.011)
Mother attended school	-0.012	-0.002	-0.002
	(0.015)	(0.021)	(0.021)
Treatment×Mother attended school			-0.009
			(0.026)
Household spending	0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Treatment×Household spending			0.000**
			(0.000)
Duration in the camp	0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)
Treatment×Duration in the camp			0.000
			(0.000)
Mother receives voucher	-0.017	-0.021	-0.021
	(0.015)	(0.016)	(0.016)
Treatment×Mother receives voucher			0.004
			(0.022)
Husband is alive	0.029	0.036	0.036
	(0.047)	(0.058)	(0.058)
Treatment×Husband is alive			-0.007
			(0.075)
Family member stranded	0.020	0.021	0.021
	(0.022)	(0.029)	(0.029)
Treatment×Family member stranded			-0.001
			(0.036)
HH victimization (conflict)	-0.047	-0.073	-0.073
	(0.070)	(0.063)	(0.063)
Treatment×HH victimization			0.026
			(0.094)
Mothers' victimization (camp abuse)	0.179	-0.024	-0.024
	(0.169)	(0.159)	(0.159)
Treatment×Mothers' victimization			0.203
			(0.232)
Mother is the HH head	-0.001	0.021	0.021
	(0.020)	(0.021)	(0.021)
Treatment×Mother is the HH head			-0.022
			(0.029)
Number of children	0.007	0.013	0.013
	(0.006)	(0.009)	(0.009)
Treatment×Number of children			-0.006
			(0.011)
Age of children	-0.001	-0.006**	-0.006**
	(0.002)	(0.002)	(0.002)
Treatment×Age of children			0.006*
			(0.003)
Gender of children	-0.008	0.000	0.000
	(0.014)	(0.016)	(0.016)
Treatment×Gender of children			-0.008

			(0.022)
Weight of children (kg)	0.013*	0.008	0.008
	(0.007)	(0.007)	(0.007)
Treatment×Weight of children (kg)			0.005
			(0.010)
Height of children (cm)	-0.004**	0.001	0.001
	(0.002)	(0.002)	(0.002)
Treatment×Height of children (cm)			-0.005**
			(0.002)
Child's victimization (camp)	0.147	0.022	0.022
	(0.152)	(0.090)	(0.090)
Treatment×Child's victimization (camp)			0.125
			(0.177)
Observations	1,907	1,586	3,493
R-squared	0.008	0.010	0.009
Mobile ownership	0.8702	0.8690	-
Joint p-value on individual/household characteristics	0.053	0.290	-
Joint p-value on interactions	-	-	0.603

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: All columns present estimates using a linear probability model, where the dependent variable is mobile ownership, a dummy variable that equals 1 if a mother (or any household member) has a mobile phone and 0 if she does not. The sample in column 1 is mothers/children in the treatment group and the sample in column 2 is mothers/children in the control group. Column 3 pools all sample together. We do not interact the treatment dummy with 'HH victim of at least one conflict abuse', 'Mother victim of at least one camp abuse', and 'Child victim of at least one camp abuse' because these indicators were derived from the 3 victimization indices that we already use. All variables were measured at the baseline. Overall mobile phone ownership is roughly 87%.

Table A5: Mentally unwell in treatment arm versus mentally healthy in control arm: Are the treated catching up?

	X: Trauma		Y: Depression	
	Without covariates	**With covariates**	**Without covariates**	**With covariates**
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**
***A1. Mother's mental health*‡**				
Trauma severity	-0.190***	-0.200***	-0.131	-0.136
	(0.067)	(0.073)	(0.100)	(0.122)
Depression severity	-0.093	-0.106	0.010	0.041
	(0.063)	(0.067)	(0.085)	(0.106)
A2. Mother's well-being				
Happiness	0.107	0.117	0.195**	0.243**
	(0.073)	(0.076)	(0.094)	(0.107)
Aspirations	-0.069	-0.075	-0.026	-0.049
	(0.079)	(0.078)	(0.096)	(0.102)
Belongingness	0.204***	0.207***	0.308***	0.351***
	(0.074)	(0.072)	(0.096)	(0.093)
***B1. Child's mental health*‡**				
Trauma severity	-0.142**	-0.122*	-0.118	-0.110
	(0.072)	(0.071)	(0.087)	(0.091)
Depression severity	-0.161**	-0.162**	-0.233***	-0.297***
	(0.080)	(0.082)	(0.088)	(0.091)
B2. Child's development				
Communication skills	0.210***	0.158**	0.285***	0.277***
	(0.078)	(0.076)	(0.108)	(0.104)
Gross-motor skills	0.216***	0.190**	0.285***	0.327***
	(0.078)	(0.079)	(0.105)	(0.108)
Fine-motor skills	0.092	0.043	0.162	0.161
	(0.091)	(0.085)	(0.126)	(0.114)
Problem-solving skills	0.258***	0.211***	0.282***	0.230**
	(0.073)	(0.071)	(0.098)	(0.095)
Social skills	0.096	0.100	0.216*	0.281**
	(0.086)	(0.087)	(0.116)	(0.115)
Observations	1,405	1,405	852	852

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Vertical panel X (trauma) includes mothers from the treatment arm that were traumatized at baseline (or the mentally unwell) and mothers from the control arm that did not have trauma at baseline (or the mentally well). Similarly, vertical panel Y (depression) includes mothers from the treatment arm that were depressed at baseline (or the mentally unwell) and mothers from the control arm that did not have depression at baseline (or the mentally well). Columns (1) and (3): treatment effects estimated without any baseline covariates. Columns (2) and (4): treatment effect estimated with all baseline covariates (as in equation 3). Covariates are listed under Table 2. For outcomes with ‡, negative coefficients imply more favorable outcomes.

Table A6: Correlation of mental health between mothers and children

	Trauma of Children			Depression of Children		
	Pooled	**Girls**	**Boys**	**Pooled**	**Girls**	**Boys**
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Panel A: At baseline					
Trauma of Mothers	0.188***	0.172***	0.201***			
	(0.027)	(0.033)	(0.033)			
Depression of Mothers				0.190***	0.186***	0.200***
				(0.048)	(0.058)	(0.072)
All Controls	Yes	Yes	Yes	Yes	Yes	Yes
Camp FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,493	1,705	1,788	3,493	1,705	1,788
R-squared	0.094	0.104	0.094	0.048	0.050	0.057
	Panel B: At endline					
Trauma of Mothers	0.246***	0.277***	0.215***			
	(0.028)	(0.038)	(0.039)			
Depression of Mothers				0.157***	0.173***	0.140***
				(0.031)	(0.041)	(0.044)
All Controls	Yes	Yes	Yes	Yes	Yes	Yes
Camp FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,798	1,382	1,416	2,798	1,382	1,416
R-squared	0.083	0.110	0.081	0.034	0.038	0.043

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS estimates reported. Dependent variables are standardized trauma (columns 1-3) and depression (columns 4-6) indices (same as in A1 and B1 panels in Table 2). Columns 1 and 4 report estimates of the entire sample, whereas the remaining columns report estimates by child's gender. Controls are listed under Table 2.

Table A7: Growth opinions and height/weight measures

	Height↑	**Weight↑**
VARIABLES	**(1)**	**(2)**
Height (in cm)	0.000	-
	(0.000)	
Weight (in kg)	-	-0.000
		(0.002)
All Controls	Yes	Yes
Camp FE	Yes	Yes
Observations	2,840	2,840
R-squared	0.025	0.033

Robust SE clustered at the block level
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS estimates reported. Dependent variables are mothers' opinions about children's improvement in height (Height↑) and weight (Weight↑). All outcomes are dummy variables, where 1 means improved and 0 means did not improve. Independent variables Height (in cm) and Weight (in kg) are measures of height and weight at endline. Controls are listed under Table 2.

Table A8: Social desirability bias check: HAZ, WAZ, and WHZ

	HAZ	**WAZ**	**WHZ**
VARIABLES	**(1)**	**(2)**	**(3)**
Treatment	0.562***	0.727***	0.633***
	(0.175)	(0.117)	(0.158)
High SDB	0.383*	0.311**	0.196
	(0.217)	(0.137)	(0.195)
Treatment×High SDB	-0.102	-0.198	-0.292
	(0.272)	(0.166)	(0.226)
All Controls	Yes	Yes	Yes
Camp FE	Yes	Yes	Yes
Observations	2,840	2,840	2,761
R-squared	0.106	0.048	0.093

Robust SE clustered at the block level
*** p<0.01, ** p<0.05, * p<0.1

Note: Dependent variables are height-for-age/HAZ (column 1), weight-for-age/WAZ (column 2), and weight-for-height/WHZ (column 3). Treatment is a dummy that equals to 1 if respondents are in the treatment arm and 0 otherwise. High SDB is a dummy that equals to 1 if the social desirability bias (SDB) score is above 8 (which is the median value) and 0 if below. All specifications include the usual set of controls and camp fixed effects as in Table 2.

Table A9: Judgment of mothers

	X: Trauma		Y: Depression	
	Without covariates	With covariates	Without covariates	With covariates
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**
A. Child's mental health ‡				
Trauma severity	0.024	0.025	-0.056	-0.039
	(0.058)	(0.058)	(0.062)	(0.060)
Depression severity	-0.025	-0.030	-0.034	-0.039
	(0.060)	(0.061)	(0.064)	(0.062)
B. Child's development				
Communication skills	0.211**	0.195**	0.218***	0.199***
	(0.085)	(0.081)	(0.072)	(0.070)
Gross-motor skills	0.213**	0.207**	0.158**	0.147**
	(0.084)	(0.081)	(0.074)	(0.071)
Fine-motor skills	-0.037	-0.055	-0.058	-0.073
	(0.094)	(0.087)	(0.082)	(0.077)
Problem-solving skills	0.214**	0.194**	0.167**	0.154**
	(0.082)	(0.078)	(0.069)	(0.066)
Social skills	0.131	0.134	0.081	0.080
	(0.091)	(0.091)	(0.075)	(0.075)
C. Child's anthropometrics				
Height-for-age z-score	0.500**	0.437**	0.447**	0.362**
	(0.223)	(0.210)	(0.190)	(0.173)
Weight-for-age z-score	0.567***	0.594***	0.466***	0.481***
	(0.127)	(0.127)	(0.107)	(0.108)
Weight-for-height z-score	0.478***	0.509***	0.278*	0.322**
	(0.169)	(0.156)	(0.155)	(0.150)
Observations	1,311	1,308	1,893	1,891

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: All panels include women whose mental health remained unchanged from baseline to endline. Columns (1) and (3): treatment effects estimated without any baseline covariates. Columns (2) and (4): treatment effect estimated with all baseline covariates (as in equation 3). Covariates are listed under Table 2. For outcomes with ‡, negative coefficients imply more favorable outcomes.

Table A10: Contamination check

	Mother's trauma					Mother's depression				
	Adj	**Adj No.**	**Adj-%**	**200m**	**400m**	**Adj**	**Adj No.**	**Adj-%**	**200m**	**400m**
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Treat	-0.346***	-0.365***	-0.312***	-0.303***	-0.351***	-0.103	-0.157	-0.074	-0.147	-0.158
	(0.127)	(0.109)	(0.117)	(0.100)	(0.110)	(0.123)	(0.107)	(0.114)	(0.102)	(0.105)
Adjacent	-0.129					-0.066				
	(0.143)					(0.135)				
Treat×adjacent	0.159					-0.042				
	(0.149)					(0.149)				
No. of adjacent		-0.073					-0.027			
		(0.071)					(0.074)			
Treat×No. of adjacent		0.135*					0.026			
		(0.080)					(0.092)			
% of treat adjacent			-0.119					0.153		
			(0.335)					(0.333)		
Treat×% of treat adjacent			0.332					-0.246		
			(0.375)					(0.400)		
Treated in 200m radius				-0.039					-0.025	
				(0.065)					(0.070)	
Treat×Treated in 200m radius				0.079					0.019	
				(0.075)					(0.093)	
Treated in 400m radius					-0.047					-0.007
					(0.069)					(0.073)
Treat×Treated in 400m radius					0.118					0.032
					(0.079)					(0.089)
All Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Camp FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,801	1,801	1,788	1,801	1,801	1,801	1,801	1,788	1,801	1,801
R-squared	0.050	0.052	0.049	0.049	0.051	0.038	0.036	0.037	0.036	0.036

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS estimates reported. The outcome variable in columns 1-5 is trauma and that in columns 6-10 is depression severity. Both outcomes are standardized indices, such that the control group has mean 0 and standard deviation 1. 'Treat' is a dummy that equals 1 if the block is treated and 0 if control; 'Adjacent' is a dummy that equals 1 if a block has at least 1 adjacent block that is treatment and 0 otherwise; 'No. of adjacent' is the number of adjacent treatment blocks; '% of treat adjacent' is the number of adjacent divided by the total number of adjacent blocks; 'Treated in 200m radius' and 'Treated in 400m radius' are the number of treatment blocks within the 200 and 400 meter radius of each block. This information is only available on roughly 1,800 individuals, which explains the smaller sample sizes.

Table A11: Treatment effects: Additional bounds analysis

	Unadjusted	Kling et al. (2007) Bounds			Most Extreme Bounds		2^{nd}-Most Extreme Bounds	
	Treatment Effects	$\delta = 0.05$	$\delta = 0.10$	$\delta = 0.25$	Lower	Upper	Lower	Upper
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
A1. Mothers' mental health								
Trauma	-0.233***	-0.227***	-0.221***	-0.200***	-1.778***	0.792***	-1.069***	0.212***
	(0.051)	(0.038)	(0.038)	(0.038)	(0.148)	(0.095)	(0.085)	(0.050)
Depression	-0.144***	-0.126***	-0.120***	-0.099**	-2.420***	0.914***	-1.020***	0.233***
	(0.054)	(0.040)	(0.040)	(0.040)	(0.217)	(0.101)	(0.090)	(0.049)
A2. Mothers' well-being								
Happiness	0.117**	0.095**	0.089**	0.068	-0.426***	0.933***	-0.102*	0.579***
	(0.056)	(0.043)	(0.043)	(0.043)	(0.061)	(0.078)	(0.058)	(0.059)
Aspirations	-0.066	-0.054	-0.061	-0.082*	-1.095***	1.091***	-0.864***	0.641***
	(0.062)	(0.049)	(0.049)	(0.049)	(0.087)	(0.098)	(0.080)	(0.068)
Belongingness	0.179***	0.176***	0.169***	0.148***	-0.414***	0.984***	-0.001	0.499***
	(0.057)	(0.046)	(0.046)	(0.046)	(0.063)	(0.081)	(0.057)	(0.058)
B1. Children's mental health								
Trauma	-0.096*	-0.094**	-0.087**	-0.066	-2.089***	0.823***	-1.166***	0.315***
	(0.055)	(0.042)	(0.042)	(0.042)	(0.189)	(0.090)	(0.106)	(0.053)
Depression	-0.122**	-0.111**	-0.105**	-0.084*	-2.574***	0.893***	-1.226***	0.486***
	(0.059)	(0.045)	(0.045)	(0.046)	(0.235)	(0.101)	(0.113)	(0.069)
B2. Children's development								
Communication skills	0.229***	0.226***	0.220***	0.201***	-0.454***	0.867***	-0.398***	0.762***
	(0.059)	(0.048)	(0.048)	(0.048)	(0.069)	(0.070)	(0.068)	(0.063)
Gross-motor skills	0.179***	0.174***	0.168***	0.149***	-0.408***	0.850***	-0.048	0.657***
	(0.058)	(0.046)	(0.046)	(0.046)	(0.062)	(0.072)	(0.058)	(0.068)
Fine-motor skills	-0.021	-0.015	-0.021	-0.040	-0.608***	0.581***	-0.205***	0.403***
	(0.066)	(0.053)	(0.053)	(0.053)	(0.068)	(0.072)	(0.064)	(0.072)
Problem-solving skills	0.177***	0.178***	0.172***	0.153***	-0.400***	0.720***	-0.020	0.577***
	(0.055)	(0.044)	(0.044)	(0.044)	(0.061)	(0.062)	(0.055)	(0.062)
Social skills	0.128*	0.113**	0.107**	0.088*	-0.542***	0.829***	-0.087	0.574***
	(0.067)	(0.052)	(0.052)	(0.052)	(0.072)	(0.075)	(0.068)	(0.069)
B3. Children's anthropometrics								
Height-for-age z-score	0.515***	0.511***	0.505***	0.486***	-5.898***	5.338***	-2.391***	2.793***
	(0.139)	(0.133)	(0.133)	(0.133)	(0.550)	(0.368)	(0.263)	(0.199)
Weight-for-age z-score	0.641***	0.637***	0.631***	0.612***	-2.565***	2.916***	-1.066***	1.985***
	(0.092)	(0.089)	(0.089)	(0.089)	(0.276)	(0.175)	(0.160)	(0.120)
Weight-for-height z-score	0.508***	0.506***	0.499***	0.478***	-5.610***	4.807***	-2.071***	2.536***
	(0.125)	(0.120)	(0.120)	(0.120)	(0.485)	(0.291)	(0.220)	(0.162)

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Column (1) reports unadjusted treatment effects, same as in Table 2. Columns 2-4 report treatment effects with moderate bounds following Kling et al. (2007) and Karlan & Valdivia (2011). Columns 5-9 report treatment effects with extreme bounds following Horowitz & Manski (2000). All specifications control for baseline characteristics and standard errors are clustered at the unit of randomization.

Table A12: Heterogeneity using Random Forest: BLP and GATES results

	(1)	(2)	(3)		(4)	(5)	(6)
	Outcome: mothers' trauma				**Outcome: children's trauma**		
	ATE	**HET**			**ATE**	**HET**	
BLP	-0.226	0.485	-	BLP	-0.113	0.164	-
	(-0.353,-0.107)	(-0.487,1.437)	-		(-0.245,0.019)	(-1.231,1.720)	-
	[0.001]***	[0.654]	-		[0.185]	[1.000]	-
	Most	**Least**	**Difference**		**Most**	**Least**	**Difference**
GATES	-0.365	-0.234	-0.123	GATES	-0.130	-0.096	-0.046
	(-0.624,-0.110)	(-0.457,-0.013)	(-0.469,0.209)		(-0.353,0.104)	(-0.334,0.147)	(-0.353,0.274)
	[0.010]***	[0.077]*	[0.919]		[0.529]	[0.810]	[1.000]
	Outcome: mothers' depression index				**Outcome: children's depression index**		
	ATE	**HET**			**ATE**	**HET**	
BLP	-0.135	0.626	-	BLP	-0.122	0.327	-
	(-0.261,-0.007)	(-0.179,1.462)	-		(-0.261,0.013)	(-0.446,1.135)	-
	[0.077]*	[0.270]	-		[0.155]	[0.894]	-
	Most	**Least**	**Difference**		**Most**	**Least**	**Difference**
GATES	-0.258	-0.069	-0.194	GATES	-0.199	-0.062	-0.143
	(-0.497,-0.019)	(-0.143,0.276)	(-0.116,0.491)		(-0.442,0.038)	(-0.290,0.157)	(-0.458,0.178)
	[0.070]*	[1.000]	[0.453]		[0.208]	[1.000]	[0.729]
	Outcome: Children's composite development index				**Outcome: Children's stunting**		
	ATE	**HET**			**ATE**	**HET**	
BLP	0.196	0.261	-	BLP	-0.070	0.021	-
	(0.042,0.345)	(-0.250,0.708)	-		(-0.120,-0.021)	(-0.296,0.348)	-
	[0.027]**	[0.675]	-		[0.010]***	[1.000]	-
	Most	**Least**	**Difference**		**Most**	**Least**	**Difference**
GATES	0.263	0.095	0.168	GATES	-0.057	-0.056	-0.005
	(0.008,0.507)	(-0.174,0.359)	(-0.182,0.499)		(-0.162,0.040)	(-0.167,0.055)	(-0.148,0.146)
	[0.087]*	[0.984]	[0.675]		[0.487]	[0.641]	[1.000]
	Outcome: Children's underweight				**Outcome: Children's wasting**		
	ATE	**HET**			**ATE**	**HET**	
BLP	-0.072	0.889	-	BLP	-0.090	0.155	-
	(-0.123,-0.022)	(-0.277,2.146)	-		(-0.140,-0.039)	(-0.201,0.497)	-
	[0.011]**	[0.267]	-		[0.001]***	[0.772]	-
	Most	**Least**	**Difference**		**Most**	**Least**	**Difference**
GATES	0.016	-0.116	0.134	GATES	-0.069	-0.117	0.049
	(-0.091,0.127)	(-0.229,-0.002)	(-0.019,0.288)		(-0.174,0.038)	(-0.226,0.000)	(-0.101,0.202)
	[1.000]	[0.092]*	[0.170]		[0.402]	[0.101]	[1.000]

*** p<0.01, ** p<0.05, * p<0.1

Note: This table reports BLP and GATES results using Random Forest. 90% confidence interval are in parenthesis; *p*-values for the hypothesis that the parameter is equal to zero are in brackets. ATE is the average treatment effect and HET is the heterogeneity loading parameter. 'Most' and 'Least' are the 20% most (top quintile) and 20% least (bottom quintile) affected groups; 'Difference' is the difference in average characteristics between 'Most' and '"Least' affected groups (i.e., most minus least). Outcome of each panel is mentioned at the top. Outcomes that are indices have been control group-standardized. Stunting, Underweight, and Wasting outcomes are dummies where 1 equals stunted, underweight, or wasted growth and 0 otherwise.

Table A13: Program cost

Cost details	**Cost in BDT**	**Cost in USD**
Salary and benefits of Senior Psycho-Social Counselors	432,507	5,088.32
Salary and benefits of Psycho-Social Counselors	1,505,851	17,715.89
Session material development workshop	1,223,543	14,394.62
Hiring, training, and refreshers for mother-volunteers	206,835	2,433.35
Session materials and printing	241,641	2,842.84
Training on play pedagogy for all staff	27,260	320.71
Mobile phone cost and support	3,657,051	43,024.13
Total cost	**7,294,688**	**85,819.86**
Cost per treated mother-child pair (N=1,911)	**3,817.21**	**44.90**

Note: USD 1 = 85 Bangladeshi Taka (BDT).

B Appendix: Data

B.1 Mental health outcome indices

Trauma Index (Mothers). The primary symptoms of psychological trauma are fear that is usually caused by re-experiencing traumatic memories and unusual emotional outbursts. Survey questions on trauma symptoms are asked on a 5-point scale, between 1-5 (never (= 1); less often (= 2); sometimes (= 3); very often (= 4); always (= 5)):

1. After conflict, did you hurt yourself out of anger or sadness?
2. Are you scared of going outside the house at night, for example fetching water from tubewell or using the toilet (in this camp)?
3. Do you live in fear for or are you worried about your life (in this camp)?
4. Do you live in constant fear of leaving this camp?
5. In the last month, did you feel so nervous that nothing could calm you down?
6. In the last month, did you feel so restless that you could not sit still?
7. In the last month, did you feel so sad that nothing could cheer you up?

Trauma Index (Children). Trauma questions for children were answered on a 5-point scale, between 1-5 (never (= 1); less often (= 2); sometimes (= 3); very often (= 4); always (= 5)):

Your child...

1. Is scared easily?
2. Is startled easily?
3. Gets angry without any reason?
4. Is easily scared if someone is not around (family members)?
5. Has nightmares and wakes up crying?
6. Becomes terrified if something specific happens such as loud noise or looking at strangers or anything else?
7. Talks about traumatic memories?

Depression Index (Mother). The main symptoms of depression include insomnia, oversleep, fatigue, weight loss, feeling sad or hopeless, feelings of worthlessness, difficulty in concentrating, and contemplating self-harm. To create this index, we use depression questions from the 'Center for Epidemiologic Studies Depression Scale', which consists of a self-reported measure of depressive symptoms. Respondents had to answer about different depression symptoms experienced in the past two weeks on a 5-point scale, between 0-4 (not at all or less than 1 day in last week (= 0); 1-2 days in the last week (= 1); 3-4 days in last week (= 2); 5-7 days in last week (= 3); almost everyday in last two weeks (= 4)):

1. My appetite was poor.
2. I could not shake off the blues.
3. I had trouble keeping my mind on what I was doing.
4. I felt depressed.
5. My sleep was restless.
6. I felt sad.

7. I could not focus on any work.
8. Nothing made me happy.
9. I felt like a bad person.
10. I lost interest in my usual activities.
11. I slept much more than usual.
12. I felt like I was moving too slowly.
13. I felt fidgety.
14. I wish I was dead.
15. I wanted to hurt myself.
16. I was tired all the time.
17. I was not liking myself.
18. I lost a lot of weight without trying.
19. I had difficulty falling asleep.
20. I could not focus on any important work.

Depression Index (Child): Depression questions for children were answered on a 5-point scale (never (= 1); less often (= 2); sometimes (= 3); very often (= 4); always (= 5)):

1. Cries without any reason.
2. Is always sad or down.
3. Is always annoyed.
4. Never gets excited about anything.
5. Cannot concentrate on anything.

B.2 Mothers' subjective well-being indices

Happiness Index (Mother). We explore happiness of mothers both in absolute and relative terms. We combine two survey responses that ask mothers to report their level of happiness based on their current state of lives and that relative to their friends, relatives, or neighbors. Happiness questions were answered on a 5-point scale, between 1 and 5, where 5 corresponds to very happy:

1. How happy are you with the current state of your life?
2. How happy are you with your life, compared to others' who had to leave Myanmar like yourself, for example, your relatives, friends, and neighbors?

Aspiration Index (Mother). We explore different aspects of future aspirations of mothers. Specifically, we measure their hopefulness about their family's future and children's formal education, possibilities of their own and husbands' employment and earnings, and the possibility of their relocation and return to their homeland. Aspiration questions were answered on a 5-point scale, between 1 and 5, where 5 corresponds to high aspirations:

1. How hopeful are you about yourself and your family's future?
2. I hope to educate (formal education) my children in the future.
3. I hope to start working and earning my living in the future (in the next 6 months).
4. I hope that to my husband or the main income earner will start earning income in future (in the next 6 months).

5. I hope to relocate outside the camp in the future (within 12 months).
6. How likely it is to return to your homeland?

Belongingness Index (Mother). People in the diasporas often struggle to 'belong' to a particular country and, hence, go through the psychological stress of searching for identity and the need to belong (Kumsa, 2006). Therefore, belongingness is related to general social identity and well-being in general. To measure mothers' belongingness to their new home in Bangladesh, we ask about their feelings of assurance about daily living and their feelings on safety and security relative to their homeland. Belongingness questions were answered on a 5-point scale, between 1 and 5, where 5 corresponds to high belongingness:

1. How assured do you feel about the daily living (e.g. food, shelter, clothes, etc.) for yourself and your family?
2. How safe and secure do you feel compared to your life in your homeland?

B.3 Children's skills development indices

All five indices were created using responses to the ASQ-3 questionnaire (available here). All questions are age-group specific. That means, for each age-group (groups have 2 months interval in between), questions on different skills development vary to match with the concerned age-group:

Communication Skills Index (Child). Communication skills exhibit a child's ability to perform age specific communication with mothers or others, such as making noises, chuckling and smiling, responding with sounds, constructing small sentences, following simple instructions without repetition, describe things, and so on. This widely covers questions on verbal skills development and effective communication with the mother. There are 6 questions in total, answered on a 3-point scale (Not yet= 0, Sometimes= 1, Yes= 2).

Motor Skills Indices (Child). Both skills define the physical performance of age-specific activities. Gross-motor skills explore the development of large body muscles and eye-hand coordination skills, such as standing, walking, jumping, throwing, catching, kicking, etc. Fine-motor skills, on the other hand, explore the development of small body muscles using wrists, fingers, toes, etc., such as holding toys, grabbing a person's finger, turning book pages, vertically arrange small boxes, etc. Using gross and fine motor skills development survey questions together, we create two motor skills index. Gross motor skill questions (6 in total) widely focus on large muscle or whole-body movements, whereas fine motor skill questions (6 in total) focus on small muscle or body-part movements. All questions were answered on a 3-point scale (Not yet= 0, Sometimes= 1, Yes= 2).

Problem Solving Skills Index (Child). Solving simple problems such as observe and follow (hand or object) movements, drawing lines, observing daily activities, putting in and taking out stones from bottles, rearrange items, etc. This widely covers questions on attention, inspection, and small problem solving skills in their daily lives. There are 6 questions in total, answered on a 3-point scale (Not yet= 0, Sometimes= 1, Yes= 2).

Personal-Social Skills Index (Child): Achieving age-specific abilities about self-regulation, compli- ance, adaptive functioning, autonomy, interaction with people, etc. For instance, smile back if smiled at, smiling when looking at a mirror, eating by her/himself, copying what mothers do, caring for toys, etc. Or social skills widely cover questions on social etiquette, personal care, and interaction with others. There are 6 questions in total, answered on a 3-point scale (Not yet= 0, Sometimes= 1, Yes= 2).

B.4 Control variables that are indices

Victimization Index (based on conflict in Myanmar). Following Bellows & Miguel (2009), we created a victimization index using responses on both physical and non-physical attacks on family members that had happened during the conflict. The baseline victimization index aggregates the following eight survey questions about physical and non-physical attack experiences of family members and then divides the aggregated value by the total, so that this index is normalized between 0 and 1 . Answers are on a 5-point scale (never (= 0); less often (= 1); sometimes (= 2); very often (= 3); always (= 4):
Did someone...

1. Physically abuse you? (physical attack)
2. Sexually abuse you? (physical attack)
3. Verbally abuse you or your family member? (non-physical attack)
4. Threaten to kill you or your family member? (non-physical attack)
5. Kill your family member(s)? (physical attack)
6. Loot and/or take over your assets illegally? (non-physical attack)
7. Set your house on fire? (non-physical attack)
8. Set your village on fire? (non-physical attack)

Camp Abuse Index of Mothers. Similarly, we created a victimization index using responses on both physical and non-physical attacks that had happened/happening during their time in the refugee camp. This index aggregates the following seven survey questions about mothers' physical and non-physical attack experiences in the camp. Answers are on a 5-point scale (never (= 0); less often (= 1); sometimes (= 2); very often (= 3); always (= 4)):
Did...

1. Your neighbors in the camp (other refugees) verbally abuse you? (non-physical attack)
2. Your neighbors in the camp (other refugees) physically abuse you? (physical attack)
3. Camp authorities verbally abuse you? (non-physical attack)
4. Camp authorities physically abuse you? (physical attack)
5. Your family member(s) experience physical abuse? (physical attack)
6. Your neighbors in the camp (other refugees) sexually exploit you? (physical attack)
7. Camp authorities sexually exploit you? (physical attack)

Camp Abuse Index of Children. We created a camp abuse index of children using responses on physical attacks on a child or whether the child observed camp abuse of family members. This index aggregates the following five survey questions, all answered on a 5-point scale (never (= 0); less often (= 1); sometimes (= 2); very often (= 3); always (= 4)): Whether

the child...

1. Was physically abused by people who are not family members.
2. Witnessed mother to be physically abused or experiencing any other violence.
3. Witnessed father to be physically abused or experiencing any other violence.
4. Witnessed siblings to be physically abused or experiencing any other violence.
5. Witnessed other people who are not family members to be physically abused or experiencing any other violence.

B.5 Social desirability bias questions

Using the following 13 questions (short version of Crowne & Marlowe (1960), as used in Dhar et al. (2022)), we create a social desirability score or SDB score. Each question has two answer choices: disagree= 0 or agree= 1:

1. It is sometimes hard for me to go on with my work if I am not encouraged.
2. I sometimes feel resentful when I don't get my way.
3. On a few occasions, I have given up doing something because I thought too little of my ability.
4. There have been times when I felt like rebelling against people in authority even though I knew they were right.
5. No matter who I'm talking to, I'm always a good listener.
6. There have been occasions when I took advantage of someone.
7. I'm always willing to admit it when I make a mistake.
8. I sometimes try to get even rather than forgive and forget.
9. I am always courteous, even to people who are disagreeable.
10. I have never been irked when people expressed ideas very different from my own.
11. There have times when I was quite jealous of the good fortune of others.
12. I am sometimes irritated by people who ask favors of me.
13. I have deliberately said something that hurt someone's feelings.

C Appendix: Program details

C.1 Mother volunteers

For the intervention, mother-volunteer (MV) positions were filled out by approaching adult women in blocks that satisfied the following criteria: (i) she is a mother of a young child herself; (ii) age was between 18 and 40 years; (iii) willing to let participants come to her home for sessions every week; (iv) willing work voluntarily and attend trainings, had good rapport building skills, and were well respected in the blocks; and, (v) although educational qualifications were not mandatory, preference was given to mothers that had reading and writing skills. Prior to approaching potential MVs, camp managers contacted block-*Majhis* (leaders of blocks) and provide them with the criteria for MVs, and sought his assistance in finding potential candidates. Based on *Majhi*'s candidate list an interview was arranged and the MVs were hired to be volunteers. These hiring were approved by the camp in charge (CIC), following the Bangladeshi government regulation that the selected MVs would need to be approved by the CIC. We hired 256 MVs, one for each session.

C.2 Over-the-phone sessions

To accommodate over-the-phone sessions, session duration and structure were carefully revised by experts at the BRAC Institute of Education and Development. We explain these changes in detail below:

- **Over-the-phone sessions:** Over-the-phone sessions replaced the usual face-to-face sessions after 24 weekly sessions. Initially, BRAC only had mobile phone numbers of 42% of enroled mothers or their spouses. To collect contact numbers from the remaining households, mother volunteers (MV), project assistants, and community leaders went door-to-door and gathered contact numbers from an additional 45% of enroled mothers.[6] If a mother did not have any mobile phone, they collected the contact number of another household member (which were typically their spouses). In total, 87% of enroled mothers had mobile phones and mobile phone ownership does not differ by treatment groups (T-test: $p > 0.10$). Contact numbers from 13% of enroled mothers (similar across treatment groups) could not be collected because neither them nor their household members owned any mobile phones. Therefore, 13% of the mother-child pairs in the treatment group were offered mobile phones that belonged to either the camp managers or block-*Majhis* (leaders of each block). During the preparation period (between March and May), MVs were also remotely trained for over-the-phone treatment delivery by the same experts that trained them for face-to-face treatment delivery. Eventually, over-the-phone sessions started in mid-May 2020 and ended in late-September 2020, following the completion of 20 weekly sessions.
- **Duration:** The duration of the over-the-phone sessions had to be adjusted to make it feasible for telephone conversations. Thus, hour-long sessions had to be curtailed to 20-minutes-long sessions. Analogous to face-to-face sessions, telephone sessions were also conducted on a weekly basis. However, over-the-phone sessions were conducted on a one-to-one basis.
- **Content:** Face-to-face sessions were more interactive, as it had a combination of discussion and sharing with MVs and other participants, and play activities for both mothers

[6] Field staff strictly followed COVID-19 health guidelines while collecting mobile phone numbers from households.

and children. However, group play activities and various group discussions/sharing could not be conducted during over-the-phone sessions. Therefore, over-the-phone sessions excluded these activities and instead focused on: (1) Step 1—feedback on homework for 3 minutes; (2) Step 2—discussion on well-being and sharing with MVs for 8 minutes; (3) Step 3—childcare discussions and practices, and playing with children for 8 minutes; and (4) Step 4—homework for 1 minute. MVs also encouraged mothers to continue doing the physical activities and play activities with children that they had learned from face-to-face sessions. Finally, during Steps 2 and 3, information on COVID-19 preventive measures, such as social distancing, hand-washing, and coughing/sneezing etiquette, was also added to the telecounseling modules to help the mothers and their families keep safe and informed during the pandemic.

D Appendix: Heterogeneity analysis using interactions

To estimate whether treatment effects vary by children's gender, household's exposure to violence during the conflict, mother's experiences of abuse in the camp, mother education, and age of mothers and children, we estimate the following interaction model:

$$Y_{1ijc} = \beta_0 + \beta_1 Treat_{jc} + \beta_2 G_{ijc} + \beta_3 Treat_{jc} \times G_{ijc} + \beta_4 Y_{0ijc} + \mathbf{\Gamma}' \boldsymbol{X}_{ijc} + \theta_c + \epsilon_{ijc} \quad \text{(D.1)}$$

where G_{ijc} is either children's gender (an indicator for male), an indicator for high exposure to violence during the conflict in Myanmar (=1 if the household victimization index is above the median value and 0 otherwise), an indicator for more experience of abuse in the camp (=1 if more and 0 otherwise), and an indicator for mother that attended primary school (=1 if true and 0 otherwise), and an indicator for old (=1 if mother/child's age is above the median value and 0 otherwise).

We did not formulate any specific hypothesis with regards to how this intervention might affect mental health of mothers with male versus female children under 2 years. It is possible that the treatment had stronger effects on the mental health of mothers with sons as parents tend to be happier and optimistic in general if they have sons versus daughters (Raley & Bianchi, 2006). Also, prevalence of son-preference might induce mothers to be more attentive and engaging during counseling sessions, which might have affected outcomes differently. Similarly, mothers/children from households that were exposed to more violent conflict in Myanmar might have been strongly affected by our intervention relative to mothers/children from households that were exposed to less violent conflicts, as recurring memories from traumatic events should be more common among the highly exposed. Since we cannot claim violence exposure of households to be entirely exogenous, we do not claim any statistically significant heterogeneous effects by violence exposure as causal.

Table D1 presents the heterogeneity results by children's gender and household exposure to violence in Myanmar. Column 1 reports the pooled effects (same as column 2 in Table 2 in the main paper), while columns 2 and 3 disaggregate the effects by child's gender. Column 4 reports the coefficient on the interaction between child's gender and the treatment indicator, showing the difference-in-differences. We find that, on most occasions, treatment effects appear to be relatively larger among women with male children (Panels A1 and A2). Whereas, in terms of children's mental health (Panel B1), development (Panel B2), and anthropometric (Panel B3) outcomes, female children appear to have benefited more than male children. However, differences between effects reported in column 2 and column 3 are not statistically significant at conventional levels, as suggested by all insignificant coefficients in column 4. Thus, we do not find any evidence for heterogeneous treatment effects by child's gender.

We now turn to examining heterogeneity by violence exposure. Column 5 reports estimates among the highly exposed individuals while column 6 reports estimates among the least exposed. Column 7 reports the coefficients on the interaction term. In terms of mental health outcomes (Panels A1 and B1), we find that treatment effects are larger among the highly exposed relative to the least exposed, and these differences are only marginally significant for mothers' mental health. This suggests that the improvement in mental health of highly exposed mothers are larger then the improvement observed among mothers that had low exposure to violent conflict. Surprisingly, we also find that treatment effect on mother's aspirations are negative and vary

by violence exposure, where aspirations of mothers that had high exposure to violent conflict deteriorated relatively more following the intervention. However, this difference is marginally significant ($p < 0.10$). In terms of children's development and anthropometric outcomes, we do not find any evidence for heterogeneity by violence exposure.

We also conduct additional heterogeneity analysis by experiences of abuse by mothers in the refugee camp and mothers' education level. These results are reported in Table D2. We do not find any heterogeneity in mental health impacts (neither of mothers nor children) by camp-based abuse and education level. In case of development, we find that mothers that did not encounter any camp abuse, their children experienced a significant improvement in problem-solving skills than children of mothers that encountered at least one camp abuse (column 4, Panel B2). In addition, children of uneducated mothers (i.e., never went to primary school) benefited the most in terms of improvements in communication and personal-social skills (column 7, Panel B2). Finally, in terms of anthropometrics, children of mothers that experienced abuse in refugee camps benefited most in terms of reductions in the incidence of underweight and wasting.

Finally, we explore heterogeneous treatment effects by age of mothers and children in Table D3. We only find that older children benefited the most in terms improvements in communication skills. Moreover, although marginally significant, we also find that younger children benefited more in terms of trauma reductions and younger mothers benefited more in terms of improvements in their sense of belongingness. For the rest, we do not observe any heterogeneity by age.

Table D1: Heterogeneous treatment effects, by gender and violence exposure

		by child's gender			by violence exposure		
	Pooled	**Girl**	**Boy**	**Difference (β_3)**	**High**	**Low**	**Difference (β_3)**
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
A1. Mothers' mental health							
Trauma	-0.233***	-0.195***	-0.268***	-0.068	-0.303***	-0.161***	-0.157*
	(0.051)	(0.061)	(0.064)	(0.070)	(0.072)	(0.056)	(0.082)
Depression	-0.144***	-0.110*	-0.170***	-0.056	-0.189***	-0.079	-0.130*
	(0.054)	(0.063)	(0.063)	(0.062)	(0.068)	(0.060)	(0.070)
A2. Mothers' wellbeing							
Happiness	0.117**	0.048	0.168***	0.102	0.152**	0.085	0.064
	(0.056)	(0.068)	(0.063)	(0.065)	(0.073)	(0.065)	(0.082)
Aspirations	-0.066	-0.077	-0.064	0.006	-0.116	-0.005	-0.145*
	(0.062)	(0.072)	(0.074)	(0.074)	(0.073)	(0.074)	(0.080)
Belongingness	0.179***	0.145*	0.211***	0.084	0.221***	0.144**	0.058
	(0.057)	(0.075)	(0.062)	(0.073)	(0.065)	(0.072)	(0.082)
B1. Children's mental health							
Trauma	-0.096*	-0.150**	-0.052	0.065	-0.117*	-0.074	-0.010
	(0.055)	(0.069)	(0.062)	(0.069)	(0.063)	(0.073)	(0.079)
Depression	-0.122**	-0.142*	-0.096	0.006	-0.153**	-0.095	-0.029
	(0.059)	(0.074)	(0.069)	(0.073)	(0.072)	(0.070)	(0.079)
B2. Children's development							
Communication skills	0.229***	0.251***	0.222***	-0.007	0.205***	0.250***	-0.083
	(0.059)	(0.070)	(0.071)	(0.074)	(0.066)	(0.076)	(0.081)
Gross-motor skills	0.179***	0.169**	0.187***	0.015	0.172**	0.180**	-0.048
	(0.058)	(0.070)	(0.068)	(0.075)	(0.068)	(0.074)	(0.083)
Fine-motor skills	-0.021	0.007	-0.041	-0.063	-0.010	-0.029	-0.016
	(0.066)	(0.081)	(0.070)	(0.075)	(0.078)	(0.080)	(0.084)
Problem-solving skills	0.177***	0.199***	0.161**	-0.010	0.169**	0.181***	-0.041
	(0.055)	(0.062)	(0.068)	(0.069)	(0.071)	(0.069)	(0.084)
Social skills	0.128*	0.119	0.146*	0.011	0.189**	0.079	0.044
	(0.067)	(0.075)	(0.081)	(0.077)	(0.080)	(0.074)	(0.080)
B3. Children's anthropometrics							
Height-for-age z-score	0.515***	0.645***	0.417**	0.015	0.530***	0.541***	-0.064
	(0.139)	(0.192)	(0.193)	(0.256)	(0.195)	(0.195)	(0.281)
Weight-for-age z-score	0.641***	0.764***	0.551***	-0.115	0.590***	0.710***	-0.163
	(0.092)	(0.120)	(0.125)	(0.162)	(0.138)	(0.123)	(0.175)
Weight-for-height z-score	0.508***	0.464***	0.566***	0.019	0.431***	0.570***	-0.140
	(0.125)	(0.156)	(0.182)	(0.213)	(0.164)	(0.183)	(0.229)
Observations	2,798	1,382	1,416	2,798	1,457	1,341	2,798

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS estimates reported. Columns 2-3 and 5-6 report estimates from split samples. For instance, estimates in column 2 are derived from the sample with only female children and column 3 are from male children sample. High exposure=1 when households' exposure to violence in Myanmar is higher than the median value and 0 if low. Columns 4 and 7 report the coefficients on the interaction term from equation D.1.

Table D2: Heterogeneous treatment effects, by mothers' camp abuse and education

		Victim of at least one camp abuse			Attended primary school		
	Pooled	Yes	No	Difference (β_3)	Yes	No	Difference (β_3)
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A1. Mothers' mental health							
Trauma	-0.233***	-0.344***	-0.214***	-0.071	-0.218***	-0.267***	0.089
	(0.051)	(0.120)	(0.052)	(0.101)	(0.056)	(0.078)	(0.079)
Depression	-0.144***	-0.206**	-0.133**	-0.015	-0.153**	-0.116	-0.008
	(0.054)	(0.104)	(0.053)	(0.084)	(0.061)	(0.072)	(0.069)
A2. Mothers' wellbeing							
Happiness	0.117**	0.090	0.118**	0.016	0.157**	0.006	0.095
	(0.056)	(0.105)	(0.058)	(0.101)	(0.063)	(0.086)	(0.084)
Aspirations	-0.066	-0.160*	-0.047	-0.072	-0.106	0.038	-0.168*
	(0.062)	(0.096)	(0.068)	(0.105)	(0.066)	(0.100)	(0.097)
Belongingness	0.179***	0.322***	0.153***	0.175	0.192***	0.171*	0.032
	(0.057)	(0.110)	(0.058)	(0.111)	(0.060)	(0.101)	(0.094)
B1. Children's mental health							
Trauma	-0.096*	-0.057	-0.107*	0.130	-0.095	-0.086	0.028
	(0.055)	(0.110)	(0.058)	(0.101)	(0.063)	(0.077)	(0.089)
Depression	-0.122**	-0.210*	-0.103*	-0.074	-0.109	-0.164**	0.043
	(0.059)	(0.119)	(0.061)	(0.111)	(0.069)	(0.081)	(0.094)
B2. Children's development							
Communication skills	0.229***	0.253**	0.230***	-0.025	0.177***	0.412***	-0.233***
	(0.059)	(0.100)	(0.062)	(0.090)	(0.064)	(0.092)	(0.090)
Gross-motor skills	0.179***	0.156	0.186***	-0.087	0.142**	0.276***	-0.138
	(0.058)	(0.101)	(0.061)	(0.101)	(0.062)	(0.103)	(0.097)
Fine-motor skills	-0.021	0.035	-0.023	-0.055	-0.059	0.101	-0.127
	(0.066)	(0.118)	(0.069)	(0.107)	(0.068)	(0.117)	(0.105)
Problem-solving skills	0.177***	0.037	0.206***	-0.249**	0.148**	0.268**	-0.098
	(0.055)	(0.115)	(0.056)	(0.110)	(0.059)	(0.104)	(0.100)
Social skills	0.128*	0.209*	0.120*	-0.022	0.068	0.330***	-0.225**
	(0.067)	(0.121)	(0.071)	(0.120)	(0.069)	(0.116)	(0.103)
B3. Children's anthropometrics							
Height-for-age z-score	0.515***	0.865**	0.461***	0.427	0.571***	0.456*	0.195
	(0.139)	(0.393)	(0.152)	(0.386)	(0.167)	(0.261)	(0.287)
Weight-for-age z-score	0.641***	1.151***	0.566***	0.557**	0.648***	0.706***	0.010
	(0.092)	(0.246)	(0.099)	(0.237)	(0.108)	(0.171)	(0.187)
Weight-for-height z-score	0.508***	1.064***	0.424***	0.579**	0.497***	0.669***	-0.071
	(0.125)	(0.242)	(0.139)	(0.273)	(0.148)	(0.217)	(0.245)
Observations	2,840	449	2,391	2,840	1,445	1,395	2,840

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS estimates reported. Columns 2-3 and 5-6 report estimates from split samples. Columns 4 and 7 report the coefficients on the interaction term from equation D.1.

Table D3: Heterogeneous treatment effects, by age

		Mothers' age			Children's age		
	Pooled	Old	Young	Difference (β_3)	Old	Young	Difference (β_3)
Dependent variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
A1. Mothers' mental health							
Trauma	-0.233***	-0.249***	-0.210***	-0.037	-0.229***	-0.225***	0.014
	(0.051)	(0.062)	(0.060)	(0.068)	(0.063)	(0.059)	(0.063)
Depression	-0.144***	-0.184***	-0.093	-0.091	-0.102	-0.187***	-0.022
	(0.054)	(0.065)	(0.062)	(0.066)	(0.069)	(0.065)	(0.061)
A2. Mothers' wellbeing							
Happiness	0.117**	0.063	0.179***	-0.115	0.125*	0.105	0.001
	(0.056)	(0.067)	(0.067)	(0.071)	(0.064)	(0.068)	(0.073)
Aspirations	-0.066	-0.051	-0.071	0.018	-0.105	-0.033	-0.078
	(0.062)	(0.073)	(0.078)	(0.081)	(0.076)	(0.074)	(0.080)
Belongingness	0.179***	0.094	0.281***	-0.127*	0.142**	0.209***	-0.030
	(0.057)	(0.067)	(0.069)	(0.073)	(0.070)	(0.065)	(0.072)
B1. Children's mental health							
Trauma	-0.096*	-0.086	-0.114	-0.005	-0.040	-0.155**	0.132*
	(0.055)	(0.061)	(0.074)	(0.076)	(0.062)	(0.069)	(0.069)
Depression	-0.122**	-0.106	-0.144**	0.039	-0.137*	-0.100*	-0.001
	(0.059)	(0.069)	(0.070)	(0.071)	(0.077)	(0.060)	(0.071)
B2. Children's development							
Communication skills	0.229***	0.291***	0.153**	0.108	0.371***	0.089	0.221**
	(0.059)	(0.064)	(0.076)	(0.073)	(0.061)	(0.085)	(0.088)
Gross-motor skills	0.179***	0.189***	0.170**	0.022	0.246***	0.118	0.074
	(0.058)	(0.064)	(0.075)	(0.070)	(0.066)	(0.081)	(0.087)
Fine-motor skills	-0.021	0.010	-0.053	0.063	-0.041	-0.004	-0.080
	(0.066)	(0.072)	(0.081)	(0.076)	(0.084)	(0.074)	(0.089)
Problem-solving skills	0.177***	0.190***	0.164**	0.043	0.254***	0.109	0.117
	(0.055)	(0.067)	(0.069)	(0.076)	(0.071)	(0.069)	(0.087)
Social skills	0.128*	0.163**	0.085	0.049	0.244***	0.005	0.125
	(0.067)	(0.077)	(0.076)	(0.074)	(0.076)	(0.087)	(0.098)
B3. Children's anthropometrics							
Height-for-age z-score	0.515***	0.630***	0.379*	0.311	0.213	0.871***	-0.434
	(0.139)	(0.188)	(0.203)	(0.259)	(0.186)	(0.204)	(0.273)
Weight-for-age z-score	0.641***	0.632***	0.650***	0.072	0.498***	0.779***	-0.196
	(0.092)	(0.129)	(0.123)	(0.171)	(0.134)	(0.111)	(0.166)
Weight-for-height z-score	0.508***	0.472***	0.550***	-0.041	0.571***	0.391***	0.109
	(0.125)	(0.178)	(0.170)	(0.244)	(0.206)	(0.126)	(0.226)
Observations	2,840	449	2,391	2,840	1,445	1,395	2,840

Robust standard errors clustered at the block level are in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS estimates reported. Columns 2-3 and 5-6 report estimates from split samples. Columns 4 and 7 report the coefficients on the interaction term from equation D.1. Old=1 if age is higher than the median (25 years of mothers and 14 months for children) and 0 otherwise.