Centre of Excellence for Development Impact and Learning

March 2022

Design Paper 2

Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development

Joanna Maselko, Victoria Baranov, Pietro Biroli, Sonia Bhalotra, Ashley Hagaman, Lisa M. Bates, Elizabeth L. Turner, Siham Sikander

LIDC London International Development Centre

UKaid from the British people

Oxford Policy Management

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL design paper

CEDIL design papers present the protocols of studies supported by the CEDIL programme. They describe the objectives, the methodology and the implementation plan of impact evaluations, evidence syntheses, and methodological papers supported by CEDIL.

About this design paper

This design paper was submitted to CEDIL by the "Impact of maternal depression treatment on maternal health, parental investment, and child development" Project L.260 team.

Please direct any comments or queries to the corresponding author, Joanna Maselko at jmaselko@unc.edu.

Suggested citation: Maselko, J., Baranov, V., Biroli, P., Bhalotra, S., Hagaman, A., Bates, L.M., Turner, E.L., Sikander, S. (2022) 'Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development', CEDIL Design Paper 2. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.

This project was funded by the Centre of Excellence for Development Impact and Learning (CEDIL), supported by UK aid from the UK Government. The views expressed in this research project paper do not necessarily reflect the UK Government's official policies or CEDIL.

L.260: Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development

EXECUTIVE SUMMARY

Early childhood investment in low-income countries is now widely recognized as a global development imperative.(Black et al. 2017; Britto et al. 2017) We aim to innovate and apply interdisciplinary methods to understand the mechanisms through which a five-year, low-intensity, scalable maternal depression intervention might lead to lasting improvements in maternal health, parental investments, and child development, and to identify characteristics of women for whom it is most effective. Furthermore, we seek to provide additional insights into the impacts of personalized feedback on parenting perceptions and actions as well as influences on future study participation. The results will provide important evidence around the impacts of early life adversities on maternal health and child development and optimize future research approaches to complex investigations in these fields.

Our research goals are to:

1) Evaluate the impact and provide a cost-effectiveness analysis of the randomized intervention, the Thinking Healthy Programme Peer-delivered (henceforth THPP) on maternal health and child development through age 6 years. Going beyond the estimation of average treatment effects, we plan to use novel machine learning techniques to estimate heterogeneity of treatment effects and identify baseline characteristics of women that are predictive of their responsiveness to treatment.

2) Identify behavioural and biological mechanisms through which the THPP influenced maternal health and child development. This will include an examination of the impacts of THPP on parental investment in children as well as maternal and child stress regulation measured using biomarkers.

3) Assess the impact of providing personalized feedback to parents regarding their child's performance on future parenting behaviours. We will additionally assess impacts on the family's likelihood of participating in future studies, a design feature that can inform future trials.

These questions align with CEDIL's objectives to use innovative methods to evaluate and optimize the design of programs, and of the studies evaluating them, with the ultimate aim of improving the health and development of under-resourced communities around the world. Our findings will inform the design and evaluation of future maternal and early childhood interventions and advance modelling of human capital trajectories(Cunha, Heckman, and Schennach 2010; Almond, Currie, and Duque 2018; Attanasio et al. 2019; Cunha, Attanasio, and Jervis 2016) using state of the art measurement (for instance, of stress, depression, parenting, intimate partner violence, and child development) and analytical techniques.(Chernozhukov et al. 2018)

DESCRIPTIONS OF INTERVENTIONS

Study Area Context

This project is situated in rural Punjab, Pakistan, where multiple adversities impact maternal health and child development. Over 40% of children do not attain their developmental potential, and nearly 30% of women suffer perinatal depression.(Black et al. 2017; Kieling et al. 2011; Gelaye et al. 2016; Parsons et al. 2012) Maternal depression treatment has the potential to alleviate poverty and poor health, and improve social functioning and development for mothers and children, and to empower women.(Baranov et al. 2020) The study takes place in Kallar Syedan, a rural subdistrict of Rawalpindi, Pakistan home to about 250,000 people with an average multigenerational household size of 6.2 persons. Families predominantly engage in subsistence farming and the wage earnings of one or more adult household members. Men typically perform semiskilled or unskilled labour in a nearby city, work for the government, serve in the armed forces, or participate in migrant labor. Overall, Kallar Syedan is a low-socioeconomic area. Male and female literacy rates are 80% and 50%, respectively, and infant mortality is ~84 per 1000 live births.(Bank 2013; National Institute of Population Studies - NIPS/Pakistan & ICF Intternational 2012-3)

Members of the study team have been working in this area for over 20 years, building strong relationships with the mothers and families. Data are collected by female research assistants with robust training who thoughtfully attend to the cultural context of the area. This project builds on the trust developed through this past engagement, continuing to follow families and the impacts of both the interventions (outlined below) and their broader social environments.

Main Intervention Being Evaluated: Thinking Healthy Programme- Peer-delivered Plus (THPP+)

The main intervention which our study evaluates is the Thinking Healthy Programme. The THPP is an extended duration, behavioural activation-based psychological intervention, delivered through lay-peers (other mothers) to depressed women, administered by the Human Development Research Foundation (HDRF).(World Health Organization 2015)

The first phase of this low-cost intervention was delivered from pregnancy to 3 years postnatal under NIH funding for the Bachpan study (ending in 2019, when the children were age 3) and it was shown to be cost-effective in reducing depression in the first six-months postnatal.(Sikander, Ahmad, Atif, et al. 2019) An analysis of outcomes at 36 months revealed a lack of lasting impact on maternal depression symptoms.(Maselko et al. 2020) However, the results of the intervention's combined impact at the 6 and 36 month postnatal marks were encouraging. For example, at 6 months, an analysis of the combined Pakistan and India samples revealed that significantly lower maternal depression scores among intervention arm women (adjusted mean difference on PHQ-9 0.78 (95% CI -1.47, -0.09))(Vanobberghen et al. 2019). Even as the final 36 month comparison did not reach statistical significance at that one time point, there was a statistically significant reduction in depression over the entire 3m-36m time period. Finally, our analysis from the biomarker pilot from 106 women at 12 months postpartum reveals a statistically significant difference between intervention and control arm mothers (treatment effect -0.38, p-value 0.01; manuscript in preparation). Given that the women in the intervention arm have continued to receive an intervention over time, we additionally hypothesize that an accumulation of small positive effects can add up over time for a larger observed effect.

Furthermore, a child stimulation program in Colombia had significant impacts on cognitive development.(Maselko et al. 2015; Attanasio et al. 2014) In response to this evidence, THPP was adapted by HDRF to include a new component, the WHO's Care for Development Package from the time the children are 3 until they are 5.(Atif, Bibi, Nisar, Zulfiqar, Ahmad, et al. 2019) This was designed to directly bolster child development, while continuing to address maternal mental health.(HDRF 2019)

The Thinking Healthy Program, Peer-delivered Plus (THPP+) Intervention

As part of the THPP, women who screened positive for elevated depressive symptoms in pregnancy were invited to participate in the trial. A subgroup of women who screened negative were invited to participate in the cohort only portion of the study. Women in the intervention arm received the peer-delivered psychosocial program from the third trimester of pregnancy through 36 months postpartum. The first 6 sessions were delivered in a group format monthly, then bi-monthly until 36 months. The lay-peers lived in the same community as that of the depressed women and volunteered their time. At the 3-year mark, the implementing organization leveraged the existing lay peer infrastructure to continue a further adapted version of the THP. While, in the initial period, the focus of the program was much more on reducing maternal depression symptoms, with a relatively small component focused on child development. The final stage of the Thinking Healthy Programme, which began when the child was 3 and lasted through age 5, leveraged the existing relationships between the mothers and peers and added a much larger focus on general maternal support and child development.

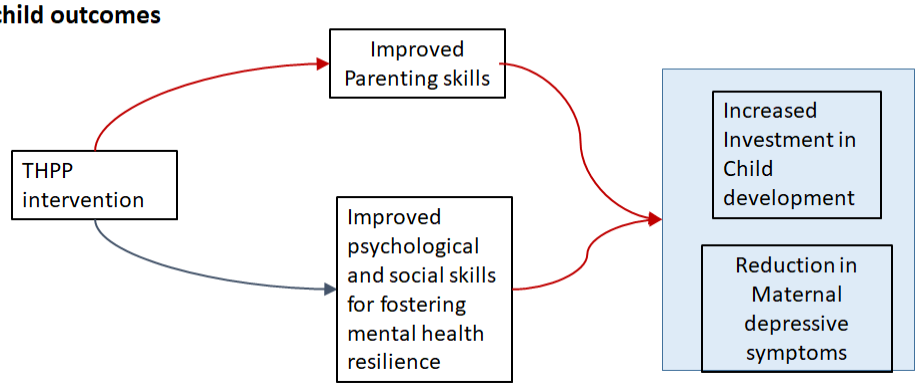
The psychosocial intervention is rooted in behavioural activation, peer-support, and problem-solving targeting maternal depression. These sessions also focused on developmental and parenting activities for children up to the 36th month.(Atif, Bibi, Nisar, Zulfiqar, Ahmed, et al. 2019) The intervention underwent robust cultural adaptation and testing and has demonstrated efficacy.(Atif, Bibi, Nisar, Zulfiqar, Ahmed, et al. 2019; Atif, Nisar, et al. 2019) Peer mothers operated within a supervisory model, clinical and community health workers of the primary care system within their village. Competence and fidelity was assessed ongoing throughout the study period.

We focus on two main mechanisms of action: through the continued support of maternal mental health and improved parenting skills. The program itself promotes school readiness as well as mothers' continued wellbeing (see table below). This phase began at child age 3 and continued through age 5. The program combines elements of the original Thinking Healthy Programme with the evidence-based intervention Learning Through Play and is delivered by the same peer volunteers through a series of 21 individual monthly sessions in the mothers' home, with the original RCT structure remaining in place(Rahman et al. 2009).

Approach	Cognitive Behavioral Therapy (CBT)
Structure of intervention	12 45-minute individual sessions (4 weekly prenatally, 8 bi-weekly postpartum); 5+ monthly community/group sessions
Areas covered	Mother's personal health, mother-child interaction, mother's interaction with others
Delivery agents	Community Health Workers (CHW) and trained Peers
Agent training	2-weeks followed by refreshers
Supervision	Monthly in groups of 10; problem discussion
Data collection	Prenatally through 6 months postpartum; main outcome depression symptoms
Intervention impacts	Remediation of maternal depression Anticipated secondary community benefits (via parenting skills, coping strategies, etc.)

The figure below illustrates our basic theoretical model.

Figure. Simplified Model of THPP program impacting maternal and child outcomes



<u>Additional Experiment on Providing Feedback to Mothers on their Child's Performance</u>

To further inform the scale and broad impacts of maternal perceptions and investments in her children, we evaluate a feedback information experiment embedded in the planned data collection protocol. Here we will focus on the following hypotheses:

1. Mothers will respond to feedback about their child's developmental progress by changing their investments in children.
2. Depressed mothers respond differently from non-depressed mothers

The operating mechanisms that we envision are as follows:

1. Mothers often have biased (incorrect) beliefs about their child's ability
2. They update their beliefs upon receiving feedback
3. Beliefs about child ability are a determinant of maternal investments in children

4. Depression can modify every step in this process- depressed women may have more biased beliefs, may update differently in response to feedback, and the manner in which their investment choices depend upon their beliefs may be different

We will test selected cognitive (literacy, numeracy) and non-cognitive (patience, altruism) skills of the child, and provide a randomly drawn half of all mothers in the sample with feedback on the child's relative performance in these tests. Before providing this information, we will elicit the mother's beliefs regarding the child's ability, the modifiability of the assessed skills, and her competence in acting to improve them. We will also ask her beliefs about the returns to child skills, and her preferences for these skills. We will then follow-up the treated and control groups to test the primary hypothesis that maternal investments in the child can be influenced by information on their child's ability, investigating the relevance of biased beliefs. We will also leverage the baseline randomization of treatment for maternal depression to test the hypotheses that depressed mothers are more likely to have biased beliefs, that the THP/depression intervention ameliorated this bias, and that depression-treated mothers are more responsive to feedback than depression-untreated mothers.

In addition to the primary goal of testing whether providing feedback improves parenting behaviour, data from the feedback experiment will allow us to test whether (1) depressed mothers are less attentive to their child's abilities, (2) the THPP intervention increased depressed mothers' attentiveness to her child's abilities, (3) there was a heterogeneous response to feedback by mother's depression status, and (4) providing feedback was complementary to the THPP intervention.

CONTRIBUTION TO LITERATURE AND POLICY RELEVANCE

We will identify the potentially protracted burdens of maternal depression and its impact on offspring. Estimates of maternal depression globally range from 10-30%(Gelaye et al. 2016; Rahman et al. 2004); in our study area, 16% of mothers of 12 month olds met criteria for depression. Exposure to maternal depression has numerous intergenerational consequences across the life course, affecting school readiness(Bono, Sy, and Kopp 2016) as well as risk of externalizing and internalizing behaviors(Murray and Cooper 1997; Goodman and Gotlib 1999; Bureau, Easterbrooks, and Lyons-Ruth 2009) and multiple neuropsychiatric and physical disorders into adulthood(Weissman et al. 2016). While intervention efforts have had some success, especially in the short term, it has been more challenging to shift the longer-term trajectories of both maternal depression and child developmental outcomes(Maselko et al. 2015; Siegenthaler, Munder, and Egger 2012; Forman et al. 2007; Gunlicks and Weissman 2008). Much also remains to be learned about the mechanistic links between maternal depression and socio-emotional and cognitive (SEC) outcomes(Weissman et al. 2016; Prince et al. 2007), which we intend to explore.

We explore risk and resiliency during the important transition to formal schooling. In low resource settings such as rural Pakistan, most children are cared for at home before entering formal schooling at age 5-6. By age 6, roughly 70% of children have transitioned to 1st grade, with over 90% in school by age 7(ASER 2016). The significant adaptation required to transition successfully to the school setting can be a source of great stress. This transition can trigger social and behavioral problems in some children, while resiliency may emerge in others.(Nelson, Kendall, and Shields 2013) Our work aims to identify family- and child-level factors leading to successful transition to the school setting.

We will provide evidence to improve the efficacy of interventions promoting child development, especially in high adversity settings. Given that 43% of children living in LMIC are at risk of not reaching their developmental potential, improving child development has become a global priority(Black et al. 2017). Our project builds on a larger consortium of projects (e.g. U19MH109998, R01MH111859) whose shared goal is to test and scale-up interventions aimed at improving the mental health and wellbeing of mothers and children. Thus, our research has immediate relevance for intervention design while also elucidating key epidemiological and mechanistic questions related to SEC development at the population level.

We leverage biological markers of stress to provide key insights into mechanistic pathways between adversity and child development. Biomarkers of stress, such as cortisol, are robust markers of HPA axis activity, and their chronic activation has been linked with several disease processes(Staufenbiel et al. 2013; Kamin and Kertes 2017). Additionally, a hypothesized mechanism through which maternal depression treatment impacts child development is through reduction of stress. Maternal depression is stressful for both the mother and her child (e.g. through impaired caregiving behaviours and negative affect) and this may cause significant dysregulations of a child's stress response, increasing risk for a multitude of health problems later in life(Garner 2013). Recent observational evidence suggests that maternal depression is associated with dysregulation of the hypothalamic–pituitary–adrenal (HPA) axis in the infant, as indexed via the stress hormone cortisol.(Karlen et al. 2013) Yet, it is not yet known whether treating maternal depression *causes* changes in child HPA axis function.

In addition to a better understanding of the role of cortisol, our project will also generate new knowledge regarding DHEA. Although extant literature has focused on cortisol, DHEA is increasingly recognized as a key component of the longer-term effects of stress on the body,(Cecconello et al. 2016; Qiao et al. 2017) as well as critical roles in brain development.(Byrne et al. 2017) DHEA may protect against the deleterious effects of heightened cortisol levels and, therefore, the cortisol/DHEA ratio is thought to be an important measure of the relative activity of the two steroids and imbalanced ratios have been identified as indicative of risk in adolescents.(Maninger et al. 2009) DHEA also plays important roles in human development (e.g., neonatal neurodevelopment, puberty), and is thought to have important neurogenerative and neuroprotective functions.(Maninger et al. 2009; Pajer et al. 2006) Therefore, analyzing multiple indicators of the HPA axis system simultaneously within the context of a randomized trial will allow us to examine transgenerational impacts of mental distress beyond a single biomarker.

This component of our project will (a) provide evidence of a biological mechanism by which treating maternal depression can impact child development, and (b) assess depression treatment effects on a set of biomarkers which can ultimately improve intervention development and evaluation in resource constrained contexts.

We explore the impacts of providing simple information to mothers on child performance as a strategy to enhance investments and future study engagement. Informational campaigns sending text messages to parents containing generalized parental suggestions have proven to be cost-effective in improving children's achievement.(Mayer et al. 2018) However, these average positive effects may mask substantial heterogeneity: each child follows a unique developmental trajectory and faces specific constraints. In order to enable each child to achieve her human capital potential, it is important to first identify her particular developmental stage and then to personalize the parental investments to her

specific needs. Parents, however, might be unaware of the child's cognitive or socioemotional situation with respect to her peers, or might not know the best investment strategy in their child's skills. Personalized feedback about the child developmental stage has the potential to enable each family to efficiently maximize their child's potential. Furthermore, receiving personalized feedback may make mothers more positive about research and increase their likelihood of participating in future studies.

Providing parents with information about their children's academic performance has been shown to be effective in changing parental beliefs about their children's abilities and improving parental investment with school children in Malawi.(Dizon-Ross 2019) However, depression might interfere with the way in which mothers respond to feedback, with evidence from neuroscience suggesting the depressed individuals place more weight on negative information.

INNOVATION

We innovate in five major areas:

1) Impact measurement with the inclusion of novel biological markers to enable early detection and mechanistic identification of treatment effects, alongside an unusually rich set of measures of child development, maternal function and parental investments, using state of the art psychometric tools.

2) Inclusion of Biomarkers. It is rare to have data on chronic HPA-axis dysregulation in large longitudinal studies with measurements at multiple time points, such as what we proposed in this study. It is unique to have such longitudinal biomarkers embedded into a randomized control trial.

3) Study design. Our project involves gathering of longitudinal data in a birth cohort study within which prenatally depressed mothers were randomized into treatment for maternal depression. We also recruit women who were pregnant but not depressed at baseline as their outcomes provide an upper bound on treatment effects under plausible assumptions.

4) Analytic approach. By using machine learning to examine heterogeneity of treatment effects, we will identify key variables that are relevant to intervention targeting.

5) Providing feedback on child performance. We incorporate an experiment testing parental responses to the delivery of personalized feedback to parents about their child's performance during developmental testing.

STAKEHOLDER ENGAGEMENT

This project reflects a strong history of stakeholder engagement which will continue. This includes:

Consortium of Institution Leading the Study in Pakistan: The consortium of organizations that will be implementing (and originally established and have been following up the Bachpan Cohort and the Peer led intervention since 2014) are the Human Development Research Foundation (www.hdrfoudation.org), Health Services Academy (www.hsa.edu.pk) and the Institute of Psychiatry, Rawalpindi Medical University (http://ioprmu.com/). This consortium is one of the most unique examples of a public-private collaboration for addressing maternal mental health and child development. The Health Services Academy, is a policy feeding institution for the Ministry of National Health Services and holds a policy guiding Health and Population Think Tank. The Human Development Research Foundation is one of the

only private sector research NGOs that holds a Memorandum of Understanding (MoU) with the Ministry of National Health Services. While Institute of Psychiatry is a WHO Collaborating Centre for mental health training and research.

Policy Relevance of the Study: Last year the President of Pakistan announced backing up the Action Plan for Mental Health in Pakistan on the 17th of Aug 2019.This action plan specifically aims at promoting maternal mental health and child wellbeing in schools. With this Plan in action – the consortium is providing inputs to the scaling up the WHO's Thinking Healthy Program for maternal mental health in Pakistan and the School Mental Health Program as well. The current study of ours with all its objectives especially the heterogeneity treatment effects will help unpack ways to preempt who will do better and who will not despite being given the psychosocial intervention. This can directly inform the scaling up of the Thinking Healthy Program which is currently underway. This study will help us understand which components of this peer delivered psychosocial intervention that started in pregnancy continued through to the fifth year of child's age needs fortification. Since this is one of the longest know psychosocial interventions delivered feasibly through lay workers takes more of a life course approach based intervention for targeting long-term maternal mental health and child development.

Continued Policy, Practice and Academic Inputs: We have a multidisciplinary team, spanning several countries with extensive experience and linkages to key stakeholders in Pakistan, the WHO, other UN organizations and NGOs. This provides excellent and diverse opportunities for getting policy traction, inputs and finally disseminating our research and impacting policy and practice both globally and nationally in Pakistan.

To help this and the visibility of our work within Pakistan, we will establish a Scientific Advisory Board (SAB). The Advisory Board will be Chaired by Dr Assad Hafeez, who is a collaborator and a network partner on this CEDIL grant. The Advisory Board will also have representation of other academics, practitioners (both from the health sector and the civil society) and policy makers; it will also have representation of the investigators from Pakistan (see table below). The Board will convene, and meet at the Health Services Academy (HSA), Islamabad. The Academy is one of the premier academic public health institutions of Pakistan and functions as the policy arm of the Ministry of National Health Services Regulations. The Board will meet annually to review the progress of the project, provide strategic inputs to improve dissemination and add new insights from the field of maternal and child wellbeing. The Board will also provide policy relevant guidelines considering the findings to promote practice change.

Members of the Scientific Advisory Board	**Brief Background/Context**

Prof Assad Hafeez	Will be the Chair of the Scientific Advisory Board Vice Chancellor, Health Services Academy. Ex-Director General of Health, Pakistan Ex-Chairman Executive Board of World Health Organization. Paediatrician and a leading public health figure both nationally and internationally.
Rep from M/o National Health Services	The Director of Programs at the Ministry will be invited. The Director also heads the NCDs & MH Program at the Ministry. The Ministry is providing the oversight to the roll-out of Universal Health Coverage and Disease Control Priorities informed essential health services package which includes mental health.
Rep from the WHO Country Office	The Mental Health Lead at the Country office will be invited.
UNICEF, Pakistan	UNICEF is one of the main UN Partners of the Ministry and the Health Services Academy. The child health and protection lead at UNICEF will be invited to be a member.
USAID	USAID is another main UN Partner of the Ministry and the Health Services Academy. It currently holds several government-to-government (G2G) grants. USAID is supporting the Health & Population Think Tank at the Academy. The Director of Health at USAID will be invited to be a member.
SPARC	Society for the Protection of the Rights of the Child (SPARC) is the leading civil society organization working for the rights of children in Pakistan. The lead for child health and wellbeing will be invited to be a member.

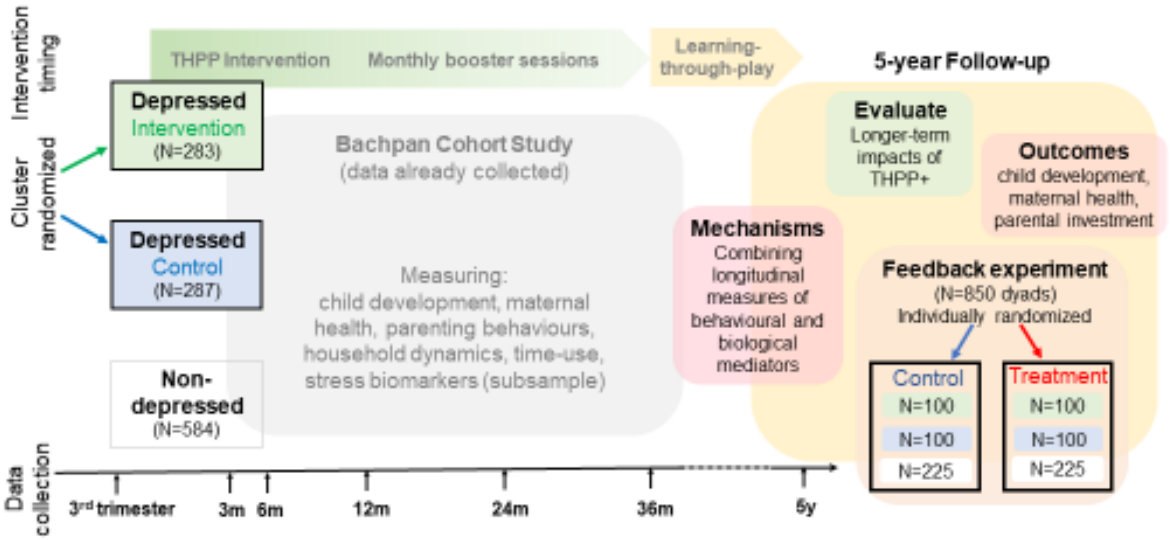
Other Civil Society reps	Representatives from other organizations that work with women and children like ROZAN, Aurat Foundation and Child Rights Movement will be invited to join the Board.
Academics	Experts from the field of public health, reproductive health, gender-based violence and paediatricians will also be included.
CEDIL Investigators/Consultants	All investigators on the grant will be de facto members of the Board.

METHODS

Sample and power

The initial phase of the THPP was implemented as a cluster-RCT in 40 rural villages outside of Rawalpindi, Pakistan (Figure 1). Depressed women were identified through systematic screening by community health workers and were enrolled in pregnancy. Mother-child dyads were assessed again at 3, 6, 12, 24, and 36 months post-partum.(Turner et al. 2016) The birth cohort (baseline n=1,154) includes an equal number of low-risk, prenatally not-depressed women,(Sikander, Ahmad, Bates, et al. 2019) enabling us to examine how much of the excess risk due to maternal depression exposure can be mitigated by the intervention. The 36-month follow-up had a high retention rate, with 91% of still eligible mother-child dyads being

Figure 1: THPP intervention timeline, Bachpan Cohort Study, sample and design



interviewed. Additionally, retention rates have been equal across intervention and control groups, which is important for the integrity of the experimental design. This bodes well for the continued follow-up.

For the analyses of main intervention effect (THPP), we anticipate a sample size of 400 prenatally depressed dyads, (11/village cluster, 40 village clusters, half intervention/control) allowing us to detect an effect size of 0.34 SD at a 5% significance level with 80% power, assuming an ICC of 0.05. Another 450 dyads from the baseline non-depressed will form the low-risk comparison group.

For the feedback experiment, we will use the full sample of 850 women from the Bachpan cohort, individually randomising half of the women to receive feedback (stratified by THPP arm and baseline non-depressed group – as shown in Figure 1). With N=850, we can detect effects of 0.19 SD (at 80% power, and 5% significance) for the main effect of providing feedback.

Data collection, key variables

We propose to re-assess mother-child dyads when the child is 5-6 years old (Figure 1, right panel). The interviews will continue to be conducted by the Pakistani assessors who have been implementing interviews since the start of the cohort in 2014, when the mothers were recruited during pregnancy. Importantly, we have seen a minimal loss-to-follow up and our cohort remains representative of the underlying study population. All of our interviews occur in the woman's home at a pre-specified day/time that is convenient for her. Over the last 3 years, the assessors have developed important relationships with key familial gatekeepers, such as mothers-in-law, as well as community health workers in the region, who maintain careful records of women in their assigned villages. Together with the community health workers, our study team has detailed contact information of our participants, including phone numbers and addresses of relatives. Based on our experience so far, we anticipate being able to successfully conduct 884 interviews at the 5-year mark.

Key Variables

Most of the measures described below have been assessed at multiple time-points from baseline through 36 months post-partum. By fielding them again in the proposed age-6 wave, we will produce an unusually long high-frequency trajectory of a rich set of child development indicators and parenting behaviours. There will also be economies from time saved on translation and piloting of measures. We are interested in evaluating both cognitive skills and socioemotional development and so will conduct a comprehensive assessment that will include the following domains. We will also explore the inclusion of additional domains.

- *Executive function and inhibitory control* assessed with the day/night task and the head-shoulders-knees-and-toes task.(McClelland et al. 2014)
- *Overall cognitive development* will be assessed using several subdomains of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV), e(Wechsler 2002). Our team has used WPPSI previously(Maselko et al. 2015) and have chosen it because of its demonstrated sensitivity to distinguish children exposed to adverse environments, including maternal depression(Cicchetti, Rogosch, and Toth 2000) and a track record of successful use in diverse environments, including low income countries.(Fernald, Kariger, and Engle 2009; Fernald, Gertler, and Neufeld 2009)

- *Socioemotional development* will be assessed with the Strength and Difficulties Questionnaire (SDQ), a culturally valid tool comprising of the 5 subscales (emotional problems, conduct problems, hyperactivity, peer problems and prosocial behaviours), and the overall total Difficulties (TD) score.(Goodman and Goodman 2009)
- *School readiness* will be assessed with the Urdu version of the (Annual Status of Education Report) ASER for math and literacy.(ASER 2016)
- *Patience and altruism:* measured using incentivized lab-in-the-field experiments developed by behavioural economists and personality psychologist.

Biomarkers

We will utilize hair cortisol, cortisone, and dehydroepiandrosterone (DHEA), as markers of chronic HPA axis activity. Hormones extract from hair are a robust measure of average hormone output (1 cm of hair represents the past month's average cortisol output), that is acceptable and feasible to collect in the field. We will collect hair samples from all 884 mother and child dyads. Of these, 104 will be the same children from whom we have 12- and 36-month hair biomarker data. We will use standard protocol for hair sample collection and collect 2-3cm of hair, and analyse using state of the art processing techniques.(Short et al. 2016; Russell et al. 2012; Stalder et al. 2017; Wennig 2000; Gao et al. 2013)(Kirschbaum et al. 2009; Davenport et al. 2006) Hair samples will be processed and analytes will be extracted and measured using LC-MS/MS by Dresden LabService.

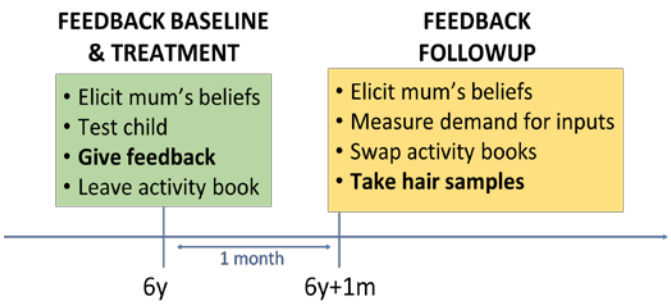
Other important variables

Detailed demographic and socioeconomic data is available from earlier waves, and will be collected again for time-varying variables. These also include domains such as Maternal Depression (PHQ-9,(Kroenke, Spitzer, and Williams 2001; Gallis et al. 2018) SCID(Spitzer et al. 1992)), anxiety (GAD(Spitzer et al. 2006)), functionality (WHO-DAS(Federici et al. 2017)), as well as intimate partner violence(WHO 2001) and women's empowerment/autonomy.(Phan 2016)

Combining the previously collected data with the proposed data collection will enable multi-dimensional analyses of trajectories for many outcome domains and allow us to study longer term intervention effects, identify mechanistic pathways, and any fading or persistence.

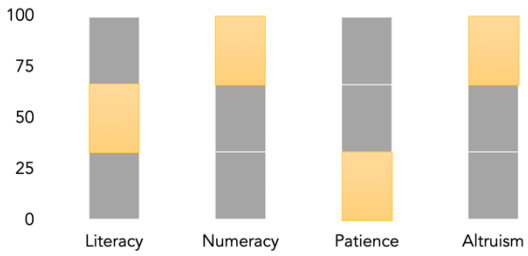
Feedback experiment details

We expect about 850 families (200 from control arm, 200 from intervention, and 450 from prenatally non-depressed) to participate in the experiment. Among half the families (n=425), at the end of the interview, women will be given specific feedback about their child's performance on the following 4 developmental domains.



1. literacy(ASER 2016): based on ASER literacy questions
2. numeracy (ASER 2016): based on ASER numeracy questions
3. patience: incentivized measure using the "marshmallow test"(Falk, Kosse, and Pinger 2020)

4. altruism/pro-social behaviours(Goodman and Goodman 2009; Sutter, Zoller, and Glätzle-Rützler 2019): incentivized measure using dictator game



The feedback will include giving the mother a pictorial report card with her child's progress in the various developmental domains (reported as terciles: e.g., your child in the top third in reading, bottom third in numeracy, see example below) and a colouring book with additional activities meant to build on the domains assessed for the feedback. The feedback will only be given to mothers individually randomized into the feedback arm, while the activity/colouring book will be given to all mothers.

To assess the direct impact of the information on maternal decision-making process, prior to giving the feedback and a month after the intervention we will elicit the **mother's beliefs** about the four developmental domains. Specifically, we will ask the mother about her beliefs on the child's skills, the modifiability of these skills, her efficacy in improving these skill, her expectation of the pecuniary and non-pecuniary returns to skills, and her preferences over them. These beliefs will be asked on a probabilistic scale, following the state of the art in the literature(Delavande 2014; Manski 2004), as in earlier Bachpan cohort waves.(Bhalotra et al. 2020) Each of these beliefs will allow us to understand the potential effectiveness of the intervention and the mechanisms though which the provided feedback can translate into improved maternal behaviours. They will also enable us to study whether depression is an important moderator of belief updating, as has been hypothesized in the neurological literature.(Harmer and Browning 2020) Specifically, we will investigate the following maternal beliefs:

1. *Her beliefs about her child's skills, and her confidence in her beliefs*. Using our objective measurements of the child's development in the specified domains, we will learn whether the mother under or over-estimates her child's ability at baseline. Our post-intervention (post-feedback) surveys will reveal the extent to which feedback leads the mother to update her beliefs in each domain.
2. Her beliefs about the *modifiability of these skills (in general)*, and her confidence in her beliefs. If a mother thinks that skills like numeracy or patience are entirely determined at birth and cannot be modified, then she may not consider it important to act to enhance these skills in her child. By asking her, we can identify the extent to which these beliefs act as a barrier to her investing in the child.
3. Her beliefs about *her* ability to modify these skills (self-efficacy) and her confidence in her beliefs. Even if a mother thinks a child's skills can be developed, she may not sufficiently increase her investments in the child if she is not confident of her ability to effect improvements.
4. Her beliefs about *the returns* to investment in these skills, or the production technology for each skill. If a mother believes in her efficacy to modify these skills, but does not perceive any benefit from improving them, then there might still be no observable change in behaviour. We will ask pecuniary and non-pecuniary returns, as previous research indicates a role for both in influencing investments.
5. Finally, we are interested in her preferences over the skills that we provide feedback on. Other things equal, women who have stronger preferences for a particular skill will tend to make greater investments in developing it.

One month after the intervention, all families will be revisited. During this short follow up interview, we will re-assess frequency of parenting stimulation behaviours and also inquire about women's reported likelihood of participating in future studies. We will collect the activity/colouring book (to measure engagement and investment in stimulating activities) and provide a new one. Finally, we will give the mother vouchers to receive a participation gift. The gifts on offer will include learning materials as well as other gifts such as toys or snacks. We will use the mother's choice of gift as a measure of demand for learning materials, testing whether feedback stimulated the mother's demand for such investments.

Ethical considerations

Regarding the ethics of withholding feedback from the control women, there are three points to note. First, to mitigate concerns the control women did not receive any feedback at all, we will provide control mothers a growth chart that shows her child's physical growth. Since the intervention is individually randomized, it was important that we provide something to the control mothers so that they don't feel they have missed out if they learn that treatment women received feedback. In essence, both groups will receive feedback: the control women will get a "report card" for physical growth, while the treatment women will get a "report card" for cognitive and non-cognitive skills. Second, the control women would be experiencing the survey much in the same way that most Early Child Development interventions would operate under business as usual: extensive testing of the child is conducted without providing personalized feedback on how the child has done on the tests. For example, the study team has been conducting assessments of children from birth until 3-years without providing any feedback to the mother on how the child is doing. Third, we will provide all mothers with activity books and opportunities to choose gifts that would help build the child's skills. So while the control mothers might not receive personalized feedback, they would still receive information and materials to help reinforce these skills.

Regarding whether receiving feedback might cause harm to the intervention women, we were similarly concerned that providing feedback to the mother might be distressing if the feedback shows the child to be in very low percentile. To this end, we have designed the feedback to be primarily focused on how well the child does relative to the mother's expectations rather than giving exact percentiles on where the child falls in the distribution compared to other children. In terms of percentiles, we will only report if the child falls in top half vs bottom half. Furthermore, it is important to note that these children are of school age and mothers are already routinely receiving report cards, and hence feedback on child ability along some of the domains we also measure (e.g. literacy and numeracy). Additionally, by providing the activity book and opportunities to choose learning materials as gifts mitigates concerns that we provide negative feedback without any opportunity for the mother to help the child improve. Finally, we will be providing the feedback with a clear caveat to the mother that some children develop along domains at different rates, and that is normal.

Analysis

Goal 1 – Main intervention evaluation and heterogeneity of treatment effects using machine learning

To assess the impact of the intervention, we will estimate intention to treat (ITT) effects for multiple outcomes, following established best-practice statistical methods in impact evaluation (e.g., randomization inference accounting for clustering, adjustment for multiple comparisons, attrition bounds).(Duflo, Glennerster, and Kremer 2007) We will investigate if attrition is differential by treatment

status (although prior follow-ups have thus far been balanced) and provide attrition bounds to assess robustness of the estimates to any differences.

The longitudinal nature of the data offers us a rare opportunity of estimating the evolution of the treatment effect over 5 years. First, we will estimate the cumulative effect of the intervention by constructing a summary index over all the time periods. This will give us greater statistical power to understand whether a marginal positive effect at any given point in time translates into a more sizeable cumulative effect. Secondly, analyses of trajectories will allow us to study the emergence of effects, to identify critical stages for different indicators of development and to test for any fading or persistence of treatment effects. Approaches for the trajectories analyses will include a data-driven group-based multi-trajectory model (GBTM)(Nagin 2005) as well as a theory-driven model that corresponds with exposure to depression during key developmental periods.(Goodman and Gotlib 1999)

Estimating heterogeneous treatment effects of an intervention is fundamental to improve targeting and eligibility of future treatments. Having a precise estimate of an average treatment effect is oftentimes insufficient, particularly when the underlying population has very diverse characteristics. A more precise description of heterogeneity of treatment effects will enable us to identify the subsample of mothers and children who benefitted the most from the intervention as well as the potential subset for whom the intervention was not cost-effective. Identifying these groups can improve *personalised interventions*, i.e. targeting an intervention to specific individuals based on their characteristics, and can be extremely valuable to establish a multi-layered policy approach which targets some mothers directly with this low-cost and scalable intervention, and leaves more funds for other mothers who might require more cost- and time-intensive treatments.

Given that the current sample size might be too small to detect an interaction effect between the intervention and all of the baseline maternal characteristics that we can consider. Therefore, we decided to tie our hands by using predefined machine learning algorithms that automatically avoid p-hacking and control for multiple hypothesis testing. To the best of our knowledge, there are no established methods to perform power-calculations when using these type of ML methods. However, ML methods have been developed precisely to overcome the curse of dimensionality and the issue of insufficient degrees of freedom to estimate the fully interacted model (the so-called “short, fat data” problem, or $p >> n$). The same ML methods we intend to leverage have been successfully used by other studies with similar sample size (e.g. (Carlana, Ferrara, and Pinotti 2018).

Since previous work does not clearly suggest which observable maternal characteristic can be used to identify the subset of mothers who are more likely to benefit from our intervention, we decided to take a data-driven, atheoretical approach to estimate heterogeneity of treatment effects. Recently developed algorithmic solutions, generally referred to as supervised machine learning (ML) methods, for estimating heterogeneous effects have the potential to offer better estimations than traditional statistical methods. ML methods can discover complex, semi-parametric or non-parametric structures in the data without having to precisely model them in advance. Furthermore, using pre-established algorithms ties our hands and limits the problem of p-hacking, HARKING, and overfitting thanks to the use of regularisation and sample-splitting in order to have a similar performance on out-of-sample data points (Athey and Imbens 2019).

Despite the growing literature on ML methods for treatment effect estimation, there is no clear guideline for which methods to apply under different circumstances. Therefore, we decided to use three different ML algorithms--based on either random forests of gradient boosting ensemble methods--which have been already used in the literature in settings with similar effect sizes as ours (Carlana, Ferrara, and Pinotti 2018) and are readily accessible via a series of convenient R packages. These algorithms are causal forests, the R-Learner, and Double Machine Learning.

To estimate heterogeneous treatment effects, we plan to use the same set of outcomes and covariates across the three ML methods to maintain comparability. Notably, we plan to use the following covariates: maternal age, education of mother and husband, maternal employment and empowerment status, religiosity, existing children and their gender, the presence of the child's grandmother, the number of people in and size of the house, the life satisfaction score, and the score on the Life Events Checklist questionnaire. The set of covariates is used unscaled with causal forests and scaled with the R-Learner and Double ML procedure.

In terms of outcomes, we are interested in the following: the SCID, PHQ-9, WHO-DAS and PSS-10 are mental-health related. The MSPSS measures perceived social support and MSES is a measure for parenting capabilities. As child outcomes, we consider the ASQ from the six months to the 36 months follow-up and the SDQ available at the 36 month follow-up survey. Further details are available in Turner et al. (2016) and Sikander et al. (2019)(Turner et al. 2016; Sikander, Ahmad, Bates, et al. 2019).

Cost-effectiveness analyses will provide information for policy makers on the viability of introducing similar interventions elsewhere. We have precise knowledge of the costs as our team implemented the intervention, and we will be able to measure not only the immediate benefits in terms of improved mental health of the mothers but also dynamic benefits in terms of her longer term mental health and function, as well as child development. This is important as the widespread tendency to look only at the immediate targeted benefits of programmes can lead to severe under-estimation of their cost-effectiveness, and hence under-investments in such programmes.

Goal 2 – mechanisms and biomarkers

We will use the full longitudinal data from baseline through the 5-year follow-up to assess mechanisms. Parental investments will cover a wide range of behaviours aimed at both cognitive as well as socio-emotional development: HOME inventory(Linver, Brooks-Gunn, and Cabrera 2004), education expenditure, Observation of a Mother-Child Interaction (OMCI, live coding of mother-child interactions viewing picture books(Scherer et al. 2019; Rasheed and Yousafzai 2015)), parental stimulation scale (covering activities such as reading, telling stories, and playing with the child).(Jeong et al. 2016)

Mechanisms will initially be assessed through a Preacher and Hayes multilevel mediation approach.(Preacher and Hayes 2004; Preacher and Hayes 2008) In a second step we will estimate mediation analysis using propensity score matching, leveraging a machine learning algorithm to select the matching covariates.(Huber, Lechner, and Strittmatter 2018; Imai, Keele, and Yamamoto 2010; Linden and

Yarnold 2016) We note that we cannot do a causal mediation analysis using these approaches. We will look to add more structure by modelling child development using the production function approach.(Attanasio et al. 2019) As discussed above, we will generalize the standard model of human capital development by incorporating maternal depression and the different biological and behavioural pathways through which it may influence child development.(Baranov et al. 2017)

We will use regression analysis to assess the impact of the THPP intervention on child HPA axis hormones. Specifically, the dependent variables are Hair Concentrations (HC) of the HPA axis hormones (DHEA, cortisol, cortisone, and the cortisol/DHEA ratio). These will likely be log-transformed to account for skewed distributions. We will pool the baseline depressed (experimental) sample along with the baseline non-depressed and estimate the following model:

$$\ln(HC)_{ic} = \alpha_1 + \delta_1 DepPrenat_i + \delta_2 THPP_c + \gamma' X_{ic} + \epsilon_{ic}$$

Where $DepPrenat_i$ is an indicator for women who were depressed at baseline and $THPP_{ic}$ is an indicator for receiving the THPP intervention. Standard errors will be adjusted for the clustered design at the village level, and inference will be conducted using p-values computed using randomization inference and adjusting for multiple hypothesis testing following Westfall and Young.(Young 2019) Control variables, X_{ic}, will include any observables at baseline that were imbalanced, weight of the hair sample, child gender and age. The coefficient δ_1 captures the (non-causal) difference in child HPA axis activity between children of mothers who were depressed and did not receive the THPP intervention and children of mothers who were not depressed at baseline. δ_2 captures the causal effect of being randomized into the THPP intervention on child HPA axis activity.

In evaluating the impact of THPP on child HPA axis activity, given our sample size of 850, we are powered to detect a minimum effect size of 0.34 SD difference in outcomes with 80% power at a 5% level of significance (assuming an ICC of 0.05, which is the upper bound of the ICCs estimated from our data on maternal and child outcomes at 36 months). To understand whether these effect sizes are realistic, we have several bases of comparison. First, previous studies measuring infant cortisol from urine found that infants of prenatally depressed mothers had 0.7 SD (or 34%) higher cortisol than infants of non-depressed mothers(Field et al. 2004). Second, we are able to examine the association in our own study measuring hair hormones at 12m postpartum in a sub-sample of N=104. Mothers who were depressed at baseline and untreated had 0.7 SD (p<0.01) higher levels of hair cortisol than the low-risk comparison group. Furthermore, we find that the intervention led to a reduction in *mother's* cortisol by 0.5 SD (p<0.01).

Goal 3 – feedback experiment

First, we will test the efficacy of providing feedback on parenting behaviour and intended participation in future studies. Randomisation of the personalized feedback intervention will be at the individual level, giving us greater power to estimate the causal impact of providing feedback. The outcomes will be measured at the 1-month follow-up, so given the short timeframe between the feedback and the follow-up, we do not anticipate attrition to be a major concern for this experiment. Furthermore, as the mother are already accustomed to us testing her child from previous waves, we do not expect issues of consent/non-participation. Finally, the feedback experiment does not involve any continued engagement from the mother, so we will not have issues of adherence or non-compliance. Therefore, the estimated effect will yield the Average Treatment Effect of providing feedback.

The primary outcomes of interest for studying the impacts of providing feedback are mother's belief updating about her child's ability and her parenting behaviours (self-reported activities as well the objectively measured inputs such as the use of the activity book and types of gifts purchased with the voucher). The secondary outcome is the stated likelihood of participation in future studies. Two important dimensions of heterogeneity which we will explore is whether feedback had differential effects by the depression status of the mother (at the time of feedback) and the gender of the child. We will also test the interaction between the THPP intervention and feedback:

$$y_{ic} = \alpha_1 + \beta_1 Feedback_i + \beta_2 THPP_c + \delta Feedback_i \times THPP_c + \gamma' X_{ic} + \epsilon_{ic}$$

Conclusion

The anticipated milestones, publications and other dissemination plans, as well as impact of COVID-19 on the project are described in the accompanying Inception Report.

Bibliography

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. *Childhood Circumstances and Adult Outcomes: Act II*.

ASER. 2016. "Annual Status of education Report (ASER) Pakistan." In. Islamabad: South Asian Forum for Education Development (SAFED).

Athey, Susan, and Guido W. Imbens. 2019. 'Machine Learning Methods That Economists Should Know About', *Annual Review of Economics*, 11: 685-725.

Atif, N., A. Bibi, A. Nisar, S. Zulfiqar, I. Ahmad, K. LeMasters, A. Hagaman, S. Sikander, J. Maselko, and A. Rahman. 2019. 'Delivering maternal mental health through peer volunteers: A five-year report', *International Journal of Mental Health Systems*, 13: 62.

Atif, N., A. Nisar, A. Bibi, S. Khan, S. Zulfiqar, I. Ahmad, S. Sikander, and A. Rahman. 2019. 'Scaling-up psychological interventions in resource-poor settings: training and supervising peer volunteers to deliver the 'Thinking Healthy Programme' for perinatal depression in rural Pakistan', *Glob Ment Health (Camb)*, 6: e4.

Atif, Najia, Amina Bibi, Anum Nisar, Shaffaq Zulfiqar, Ikhlaq Ahmed, Katherine LeMasters, Ashley Hagaman, Siham Sikander, Joanna Maselko, and Atif Rahman. 2019. 'Delivering maternal mental health through peer volunteers: a 5-year report', *International Journal of Mental Health Systems*, 13: 62.

Attanasio, OP, S Cattan, E Fitzsimons, C Meghir, and M Rubio-Codina. 2019. "Estimating the production function for human capital: Results from a randomized control trial in Colombia." In *Working Ppaer*.

Attanasio, Orazio P, Camila Fernández, Emla O A Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2014. 'Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial', *BMJ : British Medical Journal*, 349: g5785.

Bank, World. 2013. 'World development indicators 2013 (English)', World Bank Group. http://documents.worldbank.org/curated/en/449421468331173478/World-development-indicators-2013

Baranov, V., S. Bhalotra, P Biroli, and J. Maselko. 2017. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Large Randomized Control Trial " In *Working Paper*.

———. 2020. 'Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Control Trial', *American Economic Review*, 110: 824-59.

Bhalotra, S., Adeline Delavande, Paulino Gilabert, and Joanna Maselko. 2020. 'Maternal Investments in Children: The Role of Expected Effort and Returns', *IZA Discussion*, Paper No. 13056.

Black, Maureen M., Susan P. Walker, Lia C. H. Fernald, Christopher T. Andersen, Ann M. DiGirolamo, Chunling Lu, Dana C. McCoy, Günther Fink, Yusra R. Shawar, Jeremy Shiffman, Amanda E. Devercelli, Quentin T. Wodon, Emily Vargas-Barón, and Sally Grantham-McGregor. 2017. 'Early childhood development coming of age: science through the life course', *The Lancet*, 389: 77-90.

Bono, Katherine E., Susan R. Sy, and Claire B. Kopp. 2016. 'School readiness among low-income black children: family characteristics, parenting, and social support', *Early Child Development and Care*, 186: 419-35.

Britto, Pia R., Stephen J. Lye, Kerrie Proulx, Aisha K. Yousafzai, Stephen G. Matthews, Tyler Vaivada, Rafael Perez-Escamilla, Nirmala Rao, Patrick Ip, Lia C. H. Fernald, Harriet MacMillan, Mark Hanson, Theodore D. Wachs, Haogen Yao, Hirokazu Yoshikawa, Adrian Cerezo, James F. Leckman, and Zulfiqar A. Bhutta. 2017. 'Nurturing care: promoting early childhood development', *The Lancet*, 389: 91-102.

Bureau, J. F., M. A. Easterbrooks, and K. Lyons-Ruth. 2009. 'Maternal depressive symptoms in infancy: Unique contribution to children's depressive symptoms in childhood and adolescence?', *Development and Psychopathology*, 21: 519-37.

Byrne, M. L., S. Whittle, N. Vijayakumar, M. Dennison, J. G. Simmons, and N. B. Allen. 2017. 'A systematic review of adrenarche as a sensitive period in neurobiological development and mental health', *Dev Cogn Neurosci*, 25: 12-28.

Carlana, Michela, Eliana La Ferrara La Ferrara, and Paolo Pinotti. 2018. 'Goals and Gaps: Educational Careers of Immigrant Children', *SSRN Electronic Journal*.

Cecconello, Ana Lúcia, Iraci L. S. Torres, Carla Oliveira, Priscila Zanini, Gabriela Niches, and Maria Flávia Marques Ribeiro. 2016. 'DHEA administration modulates stress-induced analgesia in rats', *Physiology & Behavior*, 157: 231-36.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. 'Double/debiased machine learning for treatment and structural parameters', *The Econometrics Journal*, 21: C1-C68.

Cicchetti, D., F. A. Rogosch, and S. L. Toth. 2000. 'The efficacy of toddler-parent psychotherapy for fostering cognitive development in offspring of depressed mothers', *Journal of Abnormal Child Psychology*, 28: 135-48.

Cunha, Flavio, OP Attanasio, and P Jervis. 2016. "Parental Beliefs and Investments in Human Capital." In *Working Paper*.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. 'Estimating the Technology of Cognitive and Noncognitive Skill Formation', *Econometrica*, 78: 883-931.

Davenport, Matthew D, Stefan Tiefenbacher, Corrine K Lutz, Melinda A Novak, and Jerrold S Meyer. 2006. 'Analysis of endogenous cortisol concentrations in the hair of rhesus macaques', *General and comparative endocrinology*, 147: 255-61.

Delavande, Adeline. 2014. 'Probabilistic Expectations in Developing Countries', *Annual Review of Economics*, 6: 1-20.

Dizon-Ross, Rebecca. 2019. 'Parents' Beliefs about Their Children's Acadmic Ability: Impications for Educational Investments', *American Economic Review*, 109: 2728-65.

Duflo, E., R Glennerster, and M Kremer. 2007. 'Chapter 61: Using Randomization in Development Economics Research.' in E. Duflo, R Glennerster and Kremer M (eds.), *Handbook of Development Economics*.

Falk, A., F. Kosse, and P. Pinger. 2020. 'Re-Revisiting the Marshmallow Test: A Direct Comparison of Studies by Shoda, Mischel, and Peake (1990) and Watts, Duncan, and Quan (2018)', *Psychol Sci*, 31: 100-04.

Federici, S., M. Bracalenti, F. Meloni, and J. V. Luciano. 2017. 'World Health Organization disability assessment schedule 2.0: An international systematic review', *Disabil Rehabil*, 39: 2347-80.

Fernald, L. C. H., P. J. Gertler, and L. M. Neufeld. 2009. '10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study', *Lancet*, 374: 1997-2005.

Fernald, L. C. H., Patricia Kariger, and P. Engle. 2009. "Examining early child development in low-income countries: A toolkit for the assessment of children in their first five years of life." In. Washington, DC: World Bank.

Field, Tiffany, Miguel Diego, John Dieter, Maria Hernandez-Reif, Saul Schanberg, Cynthia Kuhn, Regina Yando, and Debra Bendell. 2004. 'Prenatal depression effects on the fetus and the newborn', *Infant behavior and development*, 27: 216-29.

Forman, D. R., M. W. O'Hara, S. Stuart, L. L. Gorman, K. E. Larsen, and K. C. Coy. 2007. 'Effective treatment for postpartum depression is not sufficient to improve the developing mother-child relationship', *Development and Psychopathology*, 19: 585-602.

Gallis, J., J. Maselko, Karen O'Donnell, K. E. Song, K Saqib, E. L. Turner, and S. Sikander. 2018. 'Criterion-related validity and reliability of the Urdu version of the patient health questionnaire in community-based women in Pakistan', *Peerj*, 6.

Gao, W., T. Stalder, P. Foley, M. Rauh, H. Deng, and C. Kirschbaum. 2013. 'Quantitative analysis of steroid hormones in human hair using a column-switching LC-APCI-MS/MS assay', *J Chromatogr B Analyt Technol Biomed Life Sci*, 928: 1-8.

Garner, Andrew S. 2013. 'Home visiting and the biology of toxic stress: opportunities to address early childhood adversity', *Pediatrics*, 132: S65-S73.

Gelaye, Bizu, Marta B. Rondon, Ricardo Araya, and Michelle A. Williams. 2016. 'Epidemiology of maternal depression, risk factors, and child outcomes in low-income and middle-income countries', *The Lancet Psychiatry*, 3: 973-82.

Goodman, A., and R. Goodman. 2009. 'Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health', *Journal of the American Academy of Child and Adolescent Psychiatry*, 48: 400-03.

Goodman, S. H., and I. H. Gotlib. 1999. 'Risk for psychopathology in the children of depressed mothers: A developmental model for understanding mechanisms of transmission', *Psychological Review*, 106: 458-90.

Gunlicks, M. L., and M. M. Weissman. 2008. 'Change in child psychopathology with improvement in parental depression: A systematic review', *Journal of the American Academy of Child and Adolescent Psychiatry*, 47: 379-89.

Harmer, C. J., and M. Browning. 2020. 'Can a Predictive Processing Framework Improve the Specification of Negative Bias in Depression?', *Biol Psychiatry*, 87: 382-83.

HDRF. 2019. 'Thinking Healthy Program - Preschool', HDRF. http://hdrfoundation.org/thinking-healthy-program-preschool-thp-preschool.

Huber, Martin, Michael Lechner, and Anthony Strittmatter. 2018. 'Direct and indirect effects of training vouchers for the unemployed', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181: 441-63.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. 'Identification, Inference and Sensitivity Analysis for Causal Mediation Effects', *Statist. Sci.*, 25: 51-71.

Jeong, J., D. C. McCoy, A. K. Yousafzai, C. Salhi, and G. Fink. 2016. 'Paternal Stimulation and Early Child Development in Low- and Middle-Income Countries', *Pediatrics*, 138: 10.

Kamin, H. S., and D. A. Kertes. 2017. 'Cortisol and DHEA in development and psychopathology', *Horm Behav*, 89: 69-85.

Karlen, J., A. Frostell, E. Theodorsson, T. Faresjo, and J. Ludvigsson. 2013. 'Maternal Influence on Child HPA Axis: A Prospective Study of Cortisol Levels in Hair', *Pediatrics*, 132: e1333-e40.

Kieling, C., H. Baker-Henningham, M. Belfer, G. Conti, I. Ertem, O. Omigbodun, L. A. Rohde, S. Srinath, N. Ulkuer, and A. Rahman. 2011. 'Child and adolescent mental health worldwide: evidence for action', *Lancet*, 378: 1515-25.

Kirschbaum, Clemens, Antje Tietze, Nadine Skoluda, and Lucia Dettenborn. 2009. 'Hair as a retrospective calendar of cortisol production—increased cortisol incorporation into hair in the third trimester of pregnancy', *Psychoneuroendocrinology*, 34: 32-37.

Kroenke, K., R. L. Spitzer, and J. B. W. Williams. 2001. 'The PHQ-9 - Validity of a brief depression severity measure', *Journal of General Internal Medicine*, 16: 606-13.

Linden, Ariel, and Paul R Yarnold. 2016. 'Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments', *Journal of Evaluation in Clinical Practice*, 22: 875-85.

Linver, M. R., J. Brooks-Gunn, and N. Cabrera. 2004. 'The Home Observation for Measurement of the Environment (HOME) Inventory: The Derivation of Conceptually Designed Subscales', *Parenting-Science and Practice*, 4: 99-114.

Maninger, N., O. M. Wolkowitz, V. I. Reus, E. S. Epel, and S. H. Mellon. 2009. 'Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS)', *Front Neuroendocrinol*, 30: 65-91.

Manski, Charles F. 2004. 'Measuring Expectations', *Econometrica*, 72: 1329-76.

Maselko, J., S. Sikander, S. Bhalotra, O. Bangash, N. Ganga, S. Mukherjee, H. Egger, L. Franz, A. Bibi, R. Liaqat, M. Kanwal, T. Abbasi, M. Noor, N. Ameen, and A. Rahman. 2015. 'Effect of an early perinatal depression intervention on long-term child development outcomes: follow-up of the Thinking Healthy Programme randomised controlled trial', *Lancet Psychiatry*, 2: 609-17.

Maselko, Joanna, Siham Sikander, Elizabeth L. Turner, Lisa M. Bates, Ikhlaq Ahmad, Najia Atif, Victoria Baranov, Sonia Bhalotra, Amina Bibi, Tayyaba Bibi, Samina Bilal, Pietro Biroli, Esther Chung, John A. Gallis, Ashley Hagaman, Anam Jamil, Katherine LeMasters, Karen O'Donnell, Elissa Scherer, Maria Sharif, Ahmed Waqas, Ahmed Zaidi, Shaffaq Zulfiqar, and Atif Rahman. 2020. 'Effectiveness of a peer-delivered, psychosocial intervention on maternal depression and child development at 3 years postnatal: a cluster randomised trial in Pakistan', *The Lancet Psychiatry*, 7: 775-87.

Mayer, SE, A Kalil, P Oreopoulos, and S Gallegos. 2018. 'Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together Intervention', *Journal of Human Resources*: 0617-8835R.

McClelland, Megan M., Claire E. Cameron, Robert Duncan, Ryan P. Bowles, Alan C. Acock, Alicia Miao, and Megan E. Pratt. 2014. 'Predictors of early growth in academic achievement: the head-toes-knees-shoulders task', *Frontiers in Psychology*, 5: 599-99.

Murray, L., and P. J. Cooper. 1997. 'Effects of postnatal depression on infant development', *Archives of Disease in Childhood*, 77: 99-101.

Nagin, Daniel. 2005. *Group-based modeling of development* (Harvard University Press: Cambridge, Mass.).

National Institute of Population Studies - NIPS/Pakistan & ICF Intternational. 2012-3. "Pakistan Demographic and Health Survey 2012-2013." In. Islamabad: NIPS/Pakistan and ICF International.

Nelson, H., G. Kendall, and L. Shields. 2013. 'Children's social/emotional characteristics at entry to school: implications for school nurses', *J Child Health Care*, 17: 317-31.

Pajer, K., R. Tabbah, W. Gardner, R. T. Rubin, R. K. Czambel, and Y. Wang. 2006. 'Adrenal androgen and gonadal hormone levels in adolescent girls with conduct disorder', *Psychoneuroendocrinology*, 31: 1245-56.

Parsons, C. E., K. S. Young, T. J. Rochat, M. L. Kringelbach, and A. Stein. 2012. 'Postnatal depression and its effects on child development: a review of evidence from low- and middle-income countries', *British Medical Bulletin*, 101: 57-79.

Phan, Ly. 2016. 'Measuring Women's Empowerment at Household Level Using DHS Data of Four Southeast Asian Countries', *Social Indicators Research*, 126: 359-78.

Preacher, K, and A Hayes. 2004. 'SPSS and SAS procedures for estimating indirect effects in simple mediation models', *Behavior Research Methods, Instruments, & Computers*, 36: 717-31.

Preacher, K. J., and A. F. Hayes. 2008. 'Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models', *Behavior Research Methods*, 40: 879-91.

Prince, M., V. Patel, S. Saxena, M. Maj, J. Maselko, M. R. Phillips, and A. Rahman. 2007. 'Global mental health 1 - No health without mental health', *Lancet*, 370: 859-77.

Qiao, Shan, Xiaoming Li, Samuele Zilioli, Zheng Chen, Huihua Deng, Juxian Pan, and Weigui Guo. 2017. 'Hair Measurements of Cortisol, DHEA, and DHEA to Cortisol Ratio as Biomarkers of Chronic

Stress among People Living with HIV in China: Known-Group Validation', *PLoS One*, 12: e0169827-e27.

Rahman, A., Z. Iqbal, J. Bunn, H. Lovel, and R. Harrington. 2004. 'Impact of maternal depression on infant nutritional status and illness - A cohort study', *Archives of General Psychiatry*, 61: 946-52.

Rahman, A., Z. Iqbal, C. Roberts, and N. Husain. 2009. 'Cluster randomized trial of a parent-based intervention to support early development of children in a low-income country', *Child Care Health and Development*, 35: 56-62.

Rasheed, M. A., and A. K. Yousafzai. 2015. 'The development and reliability of an observational tool for assessing mother-child interactions in field studies- experience from Pakistan', *Child Care Health Dev*, 41: 1161-71.

Russell, Evan, Gideon Koren, Michael Rieder, and Stan Van Uum. 2012. 'Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions', *Psychoneuroendocrinology*, 37: 589-601.

Scherer, E., A. Hagaman, E. Chung, A. Rahman, K. O'Donnell, and J. Maselko. 2019. 'The relationship between responsive caregiving and child outcomes: evidence from direct observations of mother-child dyads in Pakistan', *Bmc Public Health*, 19: 252.

Short, Sarah J, Tobias Stalder, Kristine Marceau, Sonja Entringer, Nora K Moog, Elizabeth A Shirtcliff, Pathik D Wadhwa, and Claudia Buss. 2016. 'Correspondence between hair cortisol concentrations and 30-day integrated daily salivary and weekly urinary cortisol measures', *Psychoneuroendocrinology*, 71: 12-18.

Siegenthaler, E., T. Munder, and M. Egger. 2012. 'Effect of Preventive Interventions in Mentally Ill Parents on the Mental Health of the Offspring: Systematic Review and Meta-Analysis', *Journal of the American Academy of Child and Adolescent Psychiatry*, 51: 8-17.

Sikander, S., I. Ahmad, N. Atif, A. Zaidi, F. Vanobberghen, H. A. Weiss, A. Nisar, H. Tabana, Q. U. Ain, A. Bibi, S. Bilal, T. Bibi, R. Liaqat, M. Sharif, S. Zulfiqar, D. C. Fuhr, L. N. Price, V. Patel, and A. Rahman. 2019. 'Delivering the Thinking Healthy Programme for perinatal depression through volunteer peers: a cluster randomised controlled trial in Pakistan', *Lancet Psychiatry*, 6: 128-39.

Sikander, S., I. Ahmad, L. M. Bates, J. Gallis, A. Hagaman, K. O'Donnell, E. L. Turner, A. Zaidi, A. Rahman, and J. Maselko. 2019. 'Cohort Profile: Perinatal depression and child socioemotional development ; the Bachpan cohort study from rural Pakistan', *BMJ Open*, 9: e025644.

Spitzer, R. L., K. Kroenke, J. B. Williams, and B. Lowe. 2006. 'A brief measure for assessing generalized anxiety disorder: the GAD-7', *Archives of Internal Medicine*, 166: 1092-7.

Spitzer, R. L., J. B. Williams, M. Gibbon, and M. B. First. 1992. 'The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description', *Archives of General Psychiatry*, 49: 624-9.

Stalder, Tobias, Susann Steudte-Schmiedgen, Nina Alexander, Tim Klucken, Annika Vater, Susann Wichmann, Clemens Kirschbaum, and Robert Miller. 2017. 'Stress-related and basic determinants of hair cortisol in humans: a meta-analysis', *Psychoneuroendocrinology*, 77: 261-74.

Staufenbiel, Sabine M., Brenda W. J. H. Penninx, Anne T. Spijker, Bernet M. Elzinga, and Elisabeth F. C. van Rossum. 2013. 'Hair cortisol, stress exposure, and mental health in humans: A systematic review', *Psychoneuroendocrinology*, 38: 1220-35.

Sutter, Matthias, Claudia Zoller, and Daniela Glätzle-Rützler. 2019. 'Economic behavior of children and adolescents – A first survey of experimental economics results', *European Economic Review*, 111: 98-121.

Turner, E. L., S. Sikander, O. Bangash, A. Zaidi, L. Bates, J. Gallis, N. Ganga, K. O'Donnell, A. Rahman, and J. Maselko. 2016. 'The effectiveness of the peer-delivered Thinking Healthy PLUS (THPP+) Program for maternal depression and child socioemotional development in Pakistan: study protocol for a randomized controlled trial', *Trials*, 17: 442.

Vanobberghen, Fiona, Helen A. Weiss, Daniela C. Fuhr, Siham Sikander, Ejma Afonso, Ikhlaq Ahmad, Najia Atif, Amina Bibi, Tayyaba Bibi, Samina Bilal, Aveena De Sa, Ethel D'Souza, Akankasha Joshi, Priya Korgaonkar, Revathi Krishna, Anisha Lazarus, Rakshanda Liaqat, Maria Sharif, Benedict Weobong, Ahmed Zaidi, Shaffaq Zuliqar, Vikram Patel, and Atif Rahman. 2019. 'Effectiveness of the Thinking Healthy Programme for perinatal depression delivered through peers: Pooled analysis of two randomized controlled trials in India and Pakistan', *Journal of Affective Disorders*.

Wechsler, D. 2002. "Wechsler Preschool and Primay Scale of Intelligence - Third Edition (WPPSI-III) Technical and Interpretive Manual." In. San Antonio, TX: The Psychological Corporation.

Weissman, M. M., O. O. Berry, V. Warner, M. J. Gameroff, J. Skipper, A. Talati, D. J. Pilowsky, and P. Wickramaratne. 2016. 'A 30-Year Study of 3 Generations at High Risk and Low Risk for Depression', *JAMA Psychiatry*, 73: 970-7.

Wennig, R. 2000. 'Potential problems with the interpretation of hair analysis results', *Forensic Sci Int*, 107: 5-12.

WHO. 2001. "Multi-Country Study on Women's Health and Domestic Violence Study Protocol " In. Geneva: WHO.

World Health Organization. 2015. "Thinking Healthy: A manual for psychological managament of perinatal depression." In *(WHO generic field-trial version 1.0)*. Geneva: WHO.

Young, Alwyn. 2019. 'Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*', *The Quarterly Journal of Economics*, 134: 557-98.

www.CEDILprogramme.org