Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Control Trial

By VICTORIA BARANOV AND SONIA BHALOTRA AND PIETRO BIROLI AND JOANNA MASELKO *

We evaluate the medium-term impacts of treating maternal depression on women's mental health, financial empowerment, and parenting decisions. We leverage variation induced by a cluster-randomized control trial that provided psychotherapy to 903 prenatally depressed mothers in rural Pakistan. It was one of the world's largest psychotherapy interventions, and it dramatically reduced postpartum depression. Seven years after psychotherapy concluded, we returned to the study site to find that impacts on women's mental health had persisted, with a 17% reduction in depression rates. The intervention also improved women's financial empowerment and increased both time- and money-intensive parental investments by between 0.2 and 0.3 standard deviations.
JEL: I15, I30, O15
Keywords: mental health, maternal depression, women's labor supply, empowerment, early life, parenting, child development, randomized controlled trial, Pakistan

Major Depressive Disorder is the single most important contributor to Years Lived With Disability, estimated to affect 13 percent of the global population (Vos et al., 2012). The burden of depression is about twice as high for women

(DeRubeis, Siegle and Hollon, 2008), and women of child-bearing age are especially at risk due to higher rates of perinatal depression. Prevalence of perinatal depression ranges from around 10 percent in high-income countries to 20 percent in low- and middle-income countries, where the condition often goes undiagnosed and hence untreated (Gelaye et al., 2016).

Little is known about the causal effects of depression or depression treatment on economic decision-making. Depression is characterized by symptoms including sadness, pessimism, and loss of agency, which create the potential for it to impair productivity, hamper economic decision-making, and impact women's financial empowerment (Kessler and Frank, 1997; Currie and Madrian, 1999). Women tend to have low levels of control over household spending, especially in low-income settings. Depression may exacerbate this by reducing labor supply (De Quidt and Haushofer, 2016) or else by blunting the bargaining power of women. Furthermore, since women's share of household income influences expenditure on children (Lundberg, Pollak and Wales, 1997), perinatal depression may disadvantage children. While a growing literature highlights the importance of parental investments in early childhood for later life outcomes (Almond and Mazumder, 2013), there is no causal evidence that treating perinatal depression can influence parenting decisions.

We leverage experimental variation from one of the largest psychotherapy trials in the world to investigate the impacts of treating depression on decision-making. The intervention provided cognitive behavioral therapy (CBT) to perinatally depressed women in rural Pakistan and was shown to significantly reduce depression in the first year postpartum (Rahman et al., 2008). The program has been hailed as evidence that a low-cost, community-based intervention can have substantial impacts on depression, and it was recently adopted by the World Health Organization (WHO) as a model for other countries. We investigate whether the reduction in postpartum depression placed women in the treatment arm on a better trajectory, generating future improvements in their mental health, their economic status, and their parenting decisions. Seven years after the end of the intervention, we returned to the study site and gathered rich data with multiple indicators of women's mental health, financial empowerment, parenting choices, fertility, investments in children, and child developmental outcomes. In addition to providing some of the first causal estimates of economic impacts of treating depression, our study allows us to evaluate impacts over a longer horizon.

The psychotherapy intervention we study, called the Thinking Healthy Program, was implemented across 40 communities in rural Pakistan. Half of the communities were randomly assigned to the treatment arm, and the remaining half served as controls. All pregnant women in these communities were clinically assessed for depression, and all 903 women diagnosed as depressed were included in the trial. Women in both arms received 16 home visits from community health workers, starting in the third trimester of pregnancy and ending when the children were ten months old. In the control communities, women received routine

maternal and health services. In the treated communities, women additionally received the CBT component of the intervention which focused on identifying and modifying cognitive distortions common in depression, using techniques of active listening, collaboration with the family, and guided discovery of healthy thinking.

We successfully located 585 (65%) of the original trial participants enrolled at baseline. In addition, we recruited 300 mother-child dyads from among the women excluded from the trial because they did not suffer prenatal depression at baseline. This sample allows us to evaluate the depression gap in outcomes between untreated prenatally depressed women and women who were not depressed in pregnancy. We also compare the magnitudes of treatment effects to these descriptive gaps.

We first demonstrate that the CBT intervention had persistent impacts on mental health. At the 7-year mark, the intervention reduced rates of depression by 5 percentage points, a clinically significant size, although smaller than at the 1-year mark, when the reduction was 30 percentage points. The narrowing of the gap between treated and control women is driven by spontaneous recovery in the control group, as the share of women in the treated arm who were assessed as having no depression was similar in the 1- and 7-year follow-ups. This is the first evidence of the persistence of the mental health benefits of a one-off CBT intervention as far out as seven years, and it is important given the chronicity of depression. Importantly, we also find that the mental health benefits of treatment, particularly in the longer run, are significantly larger for mothers who had been pregnant with a girl rather than a boy at the start of the intervention.

Turning to outcomes related to women's decision-making, the intervention had lasting positive impacts on an index of the financial empowerment of women, which we estimate was 0.29 of a standard deviation (henceforth SD) higher among intervention arm women. In particular, they were 9 percentage points more likely to have control over household spending, a 17% increase over the control group. We also identify a 0.2–0.3 SD improvement in indices of time- and money-intensive parental investments. For example, we find that children in the intervention arm were more likely to attend better quality and private schools, and to have more learning materials in the home, all of which were objectively measured. Mothers also reported higher expected grade attainment. However, we find no discernible average differences in parenting style or fertility between women in the treated and control communities. The intervention substantially narrowed or closed the gaps in empowerment and parental investments associated with being depressed. As there is negative selection into depression, the descriptive gaps are likely to be larger than causal differences. This is a testament to the power of the intervention.

Treatment effects are consistently larger for mothers who were pregnant with a girl at the time of the intervention: 0.46 SD increase in financial empowerment, 0.47 SD for monetary investments, 0.26 SD for time-intensive investments, and 0.21 SD for parenting style, compared with estimates indistinguishable from zero for boys. With the exception of time investments, the coefficients are statistically

significantly different by gender. This pattern of heterogeneity is not surprising in light of the fact that the longer-run mental health benefits of CBT were concentrated in mothers of girls. Control group means in the 7-year follow-up reveal that mothers who were pregnant with girls were disempowered and more likely to be depressed relative to mothers of boys, and monetary investments in girls were smaller. The intervention narrowed these gender gaps.

Our finding that treatment improves women's control over household spending, along with increasing investments in children, lines up with evidence that additional resources in the hands of women tend to be directed toward children (Duflo, 2003; Bhalotra and Clots-Figueras, 2014). Moreover, increased control over spending was much larger and only statistically significant for mothers of girls, which is likely why girls received higher parental investments.

Since the intervention empowered mothers and increased investments in children, we expected it to generate improvements in child development. However, effects on cognitive or socio-emotional development at age 7 are small and imprecisely estimated (confirming some results shown in Maselko et al. (2015)) and effects on physical health and child survival are only suggestive. Our investigations confirm that, while the indicators we use appear to be appropriate for the context, there are no sizeable effects in any part of the distribution, and there is no evidence of compensating investments in the control arm. Correlating parental investments to children outcomes, and comparing children of depressed women in the control arm to children of women who were not prenatally depressed, indicates that the expected effect sizes might be too small to detect at this age, and might require larger samples. It may be that returns to increased parental investments appeared and then faded (Andrew et al., 2018) or that they will emerge at a later stage of the life-cycle (Heckman, Stixrud and Urzua, 2006).

Finally, we study the impacts of the intervention on all other available domains of behavior with a view to illuminating the mechanisms at play. We find little evidence that the treatment influenced the woman's physical health or her husband's income over time. There are indications that it improved the quality of the relationship with the husband and the probability of co-residence with the grandmother of the child, but these effects were small and were not sustained in the longer run. We do find that the intervention led to a sustained increase in perceived social support, a measure closely linked with mental health. While it is difficult to identify a single mechanism that explains the persistent effects on empowerment and parenting, improved mental health appears to be a likely channel.

Our results provide the first evidence that positive impacts of CBT on mental health persist seven years after the end of treatment.[1] For more than a quarter of patients, depression is chronic (DeRubeis, Siegle and Hollon, 2008), and an

---

[1]We identified six trials with follow-up beyond 24 months. The longest follow-up was up to 6 years after completion of CBT with a sample size of 40 (Fava et al., 1998). The largest trial had 248 with a follow-up of 3.5 years (Cuijpers et al., 2013; Wiles et al., 2016). None of these trials studied CBT alone, they included pharmacological treatment.

episode of maternal depression raises the chances of experiencing another by about 50% (Shi and Altshuler, 2004) making it important to identify treatments that generate sustained recovery.

The main contribution of this study lies in providing the first causal evidence that treating depression can have large impacts on economic decision-making and, remarkably, impacts that are evident seven years after the end of a one-off intervention.[2] Our results are in line with the findings of two recent studies which have reported large benefits of CBT-based interventions. Targeted at specific populations with impulse control problems, such as crime-prone youth in Chicago (Heller et al., 2017) and Liberia (Blattman, Jamison and Sheridan, 2017), these interventions were effectively "non-cognitive" skills training and did not aim to improve mental health.

This study also ties into other strands of the literature. First, we contribute to a growing literature on women's empowerment (Duflo, 2012; Almas et al., 2015), offering the first evidence that treating perinatal depression may improve women's financial empowerment in the medium term. Many successful interventions for women's empowerment (education, fertility planning, cash transfers, savings accounts) target adolescent girls (Bandiera et al., 2017), but have often failed to generate persistent effects (Baird and Özler, 2016), possibly because of social constraints (Buvinic and Furst-nichols, 2016; Field et al., 2016). Our results suggest that interventions aimed at reducing maternal depression might be an effective policy tool to increase women's empowerment even in difficult contexts, for example among adult women or when social constraints are binding.

Second, we provide causal evidence that treating maternal depression impacts investments in children. A large literature has explored the later-life consequences of shocks to the physical health of pregnant women or infants (Almond and Currie, 2011; Almond, Currie and Duque, 2017; Bhalotra, Karlsson and Nilsson, 2017). Although a few recent studies have analyzed maternal stress (Aizer, Stroud and Buka, 2009) or bereavement during pregnancy (Persson and Rossin-Slater, 2017; Black, Devereux and Salvanes, 2016), much less is known about the impacts of maternal mental health. Recent studies have emphasized the importance of parental investments (Almond and Mazumder, 2013; Cunha, 2017) and early childhood stimulation (Walker et al., 2005; Attanasio et al., 2014; Doyle et al., 2017), both of which may be compromised by the mother's depression. The potential costs of failing to treat maternal depression are large, because early childhood skills influence adult health and earnings.

---

[2] A large literature links depression with financial and labor market outcomes (see Mintz et al. (1992); Kessler (2012) for a meta-analysis of evidence based on associations) and a handful of randomized controlled trials show short-term effects of psychotherapy interventions on days worked or the ability to perform labor (Patel et al., 2017; Buttorff et al., 2012; Bolton et al., 2003). Several studies in the economics literature have estimated the impacts of mental health on economic decision-making or labor market outcomes, primarily using instrumental variables (e.g., death of a friend, religiosity) for identification. See Frijters, Johnston and Shields (2014) for one such example and a review.

I. Conceptual Background

In order to better understand the mechanisms through which depression treatment can influence maternal decision-making, we outline a framework that serves the dual purpose of systematizing our thinking and helping us summarize the literature on depression as it relates to decision-making. A sketch of a model reflecting the discussion here is in the working paper version of this paper. In view of the symptoms that characterize depression, it may impact decisions through preferences, expectations or constraints. Potential changes in *preferences* include time discounting and the disutility of effort. Depression or negative affect may encourage time discounting (Lempert and Pizzagalli, 2010). Depression has been linked to impulsivity and inconsistent intertemporal choices (Takahashi et al., 2008), and to present-biased behaviors such as drinking, smoking, or suicide (Dennhardt and Murphy, 2011; Imhoff et al., 2014). A second channel is a higher cost of effort (Cohen et al., 1982; Den Hartog et al., 2003)–associated with stress and fatigue–which may, for instance, increase the mental cost of simple tasks such as interacting with the child (Bhalotra et al., 2017) or advocating for oneself, thus reducing bargaining power (Mirowsky, 1985). Depression may also impact the marginal utility of consumption or leisure (Kung, Johnston and Shields, 2018), anhedonia being a classic symptom of depression.

Depression may affect *expectations*, or subjective beliefs over the distribution of shocks, as posited by Beck's cognitive theory (1967; 1979). It may induce a pessimistic view in the mother of her ability to bargain or influence future outcomes such as her child's skills (De Quidt and Haushofer, 2016; MacLeod and Salaminiou, 2001). Alternatively, depression may make mothers more pessimistic about the returns to parenting investments, for instance by making them believe that negative shocks are more likely. By distorting beliefs, depression might lead to lower aspirations and, thereby, lower child investments (Dalton, Ghosal and Mani, 2010).

Finally, depression may impact women's decision-making through tightening *constraints*, for instance, it could increase sick days and reduce disposable time (Grossman, 1972) or lower the effectiveness of maternal investments and tighten productivity constraints. Depression has been linked to lower productivity in the labor force (Lerner et al., 2004; Lerner and Henke, 2008). In home production, depression could reduce maternal productivity in combining investments to improve child quality, holding fixed the level of inputs. This might occur if depression affects cognitive function (Den Hartog et al., 2003), or the quality of parent-child interactions (Ronda, 2016).

II. The Intervention

The Thinking Healthy Program (THP) was a cluster randomized trial addressing perinatal depression in rural Punjab, Pakistan. Perinatal depression is defined as a depressive disorder with onset at any time during pregnancy or within the

first year of birth. The Diagnostic and Statistical Manual of Mental Disorders, IV (DSM-IV) criteria for diagnosis of perinatal depression are the same as for depression at any other stage of life. They include at least five of these nine symptoms, present nearly every day within a two-week period: (1) Feelings of sadness, emptiness, or hopelessness, nearly every day, for most of the day or the observation of a depressed mood made by others; (2) Loss of interest or pleasure in activities; (3) Weight loss or decreased appetite; (4) Changes in sleep patterns; (5) Feelings of restlessness; (6) Loss of energy; (7) Feelings of worthlessness or guilt; (8) Loss of concentration or increased indecisiveness; (9) Recurrent thoughts of death, with or without plans of suicide (American Psychiatric Association, 2013).

The trial was randomized across 40 clusters, these being Union Councils, the smallest geopolitical unit. Twenty clusters were randomized into receiving the intervention and twenty to the control arm. Women were enrolled during April 2005-March 2006. All women in their third trimester of pregnancy (married, ages 16-45, no other significant illness) who met the DSM-IV criteria for Major Depressive Episode were invited to participate in the study. A total of 3898 women were identified, with 8% refusing before any assessment, and 2% were not found. Rates were not different by treatment status, Online Appendix Table A.3 shows sample numbers by treatment cluster through time.

A total of 3518 women were assessed for clinical depression, with 903 (26%) identified as prenatally depressed, a prevalence consistent with previous estimates for this region (Rahman, Iqbal and Harrington, 2003). There were 463 depressed mothers in the clusters randomized to the intervention, and 440 in the control arm. Only women diagnosed as depressed completed the baseline survey. To our knowledge, this is the largest psychotherapy intervention to treat mild to moderate depression evaluated to date.[3]

All women who were offered participation in the study accepted, and those who were not offered participation were unable to receive the treatment or other therapies. There were no psychologists in the public sector, and only three psychiatrists (based in Rawalpindi city) for the whole district. The field team confirmed that antidepressant drugs were not readily available to this population.

The THP intervention was based on principles of cognitive behavioral therapy (CBT), a class of psychosocial interventions that are the most widely used evidence-based practice for treating mental disorders (Field, Beeson and Jones, 2015). CBT focuses on the development of personal coping strategies that target solving current problems and changing unhelpful patterns in cognition (thoughts, beliefs, and attitudes), behaviors, and emotional regulation. Meta-analyses of CBT indicate that it is often at least as effective as pharmacotherapy (Bolier et al., 2013; Tolin, 2010; Cuijpers et al., 2008a). Through extensive piloting (Rahman, 2007), the original study team designed an intervention which could

---

[3]See Cuijpers et al. (2008a,b) for meta-analyses. The authors have compiled a database (www.evidencebasedpsychotherapies.org) which includes interventions to January 1, 2013. From 2013, we searched all completed psychotherapy interventions posted on clinicaltrials.gov.

be delivered by primary health workers based in the village (the essential features are summarized in Online Appendix Table A.1). They developed a manual with step-by-step instructions for each session to train the health workers and for them to keep for reference. The manual can be download from the WHO website: `www.who.int/mental_health/maternal-child/thinking_healthy`.

The intervention was delivered by Lady Health Workers (LHWs), appointed by the federal government to deliver community health services. During the CBT sessions, LHWs focused on identifying and modifying cognitive distortions common in depression, such as how the mother viewed her own health, her relationship with the baby, and the people around her (changing "unhealthy thinking" to "healthy thinking"). Mothers received health education and supporting materials with pictorial and verbal key messages to facilitate the discovery of alternative health beliefs. The intervention was based on a psychosocial model and not presented as a treatment for a mental health problem. While other studies have provided CBT to perinatally depressed mothers in developing countries (Cooper et al., 2009; Gao et al., 2010; Lara, Navarro and Navarrete, 2010), the component of the intervention that provided guided discovery of healthy behavior is unique to this study. The CBT intervention was delivered at a relatively low cost of USD10 per woman.
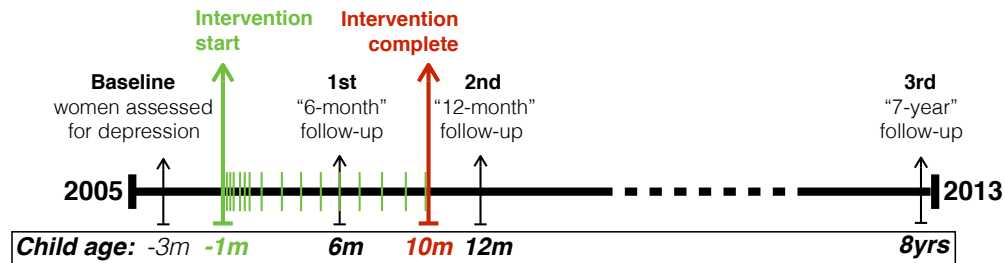
FIGURE 1. TIMELINE OF INTERVENTION AND FOLLOW-UPS

The timeline for the intervention and all follow-ups is summarized in Figure 1. Every woman in the trial received 16 home visits from a LHW. These were delivered as weekly sessions for 4 weeks in the last pregnancy month, three sessions in the first postnatal month, and monthly sessions for the following 9 months. Mothers in the control arm received enhanced routine care with an equal number of visits, enhanced not in content but because the frequency of visits was greater than the usual, which was once a month. We have no data on the duration of each session but it is likely they were longer in the treated clusters on account of the additional content. The standard health visits included advice on infant health issues such as tetanus, immunizations, and breastfeeding.[4] Each LHW was

---

[4]This enhanced frequency appears to be associated with better child outcomes for the control sample

responsible for approximately 1000 women in her catchment area. There were a total of 40 LHWs who visited either treatment or control mothers. Thus, the catchment areas of LHWs were nested within clusters to avoid contamination.

A. Follow-Up Surveys

The THP study conducted follow-up surveys at 6 and 12 months postpartum to evaluate maternal mental health, infant outcomes, and parenting behavior. The intervention led to large (approximately 30 percentage point) reductions in depression rates at 6 and 12 months, increased the frequency of interaction with the infant for both mothers and fathers (based on self-report), but no discernible effects on infant growth (Rahman et al., 2008).

There had been no additional contact with participants since the 12-month follow-up until 2013. We worked with the same NGO that conducted the intervention to implement a follow-up study to assess whether the reduction in postpartum depression had lasting effects on mothers and children. Five field supervisors, blind to the woman's trial status, worked with LHWs to contact participants who had completed the 12-month follow-up. Additional queries with neighbors or relatives, as well as local hospital record checks assisted this. A field team of 9 assessors, also blind to treatment status, worked mid 2013 to early 2014. Each dyad interview had two parts, one in the woman's home and the second in the child's school or the LHW's house, so as to administer the cognitive function tests to the child in a quiet and more standardized environment than the home.

We enrolled 300 new mother-child dyads from the sample of women who were assessed for the original THP study but did not pass the DSM-IV criteria for major depression. As we had limited baseline data for these women, we used each participant's village and LHW assignment to identify a prenatally non-depressed woman to contact for re-enrollment. A full follow-up interview was completed by this additional sample, but the only baseline characteristic available for them is their depression status.

III. Data

We provide a brief description of the data and outcomes below. A detailed description of the data, analysis of sample flow, balance, and attrition are found in the Online Appendix (Section D).

Sample.

Our analytical sample comprised the experimental group of mothers randomized into treatment or control arms, and the non-experimental group who were not depressed at baseline. The starting experimental sample consisted of 463 treated

relative to the overall region. We find an infant mortality rate of 4.6 percent, approximately two-thirds of that of the Punjab region. Rates of breastfeeding and completing immunization in our control sample were also higher than in the 2012-2013 Pakistan Demographic and Health Surveys (DHS).

and 440 control mothers. At the 7-year follow-up, we located 295 of the control dyads and 289 of the intervention arm dyads, together corresponding to 64.8% of the baseline sample. A third of the attrition was due to child mortality which was not different by treatment status. The non-experimental group interviewed at the 7-year follow-up consisted of 300 mother-child dyads, 150 from each arm.

Baseline Balance.

Baseline balance across intervention arm was achieved for all observable characteristics other than household composition (Table 1). Treated women were more likely to have their mother-in-law or mother present (each marginally significant). Nevertheless, an overall joint test of balance yields a p-value of 0.13.

Table 1—Balance

	Baseline Sample N=903				1-year Sample N=704		7-year Sample N=585	
	(1) Control Mean	(2) SD	(3) T-C Diff	(4) p-val	(5) T-C Diff	(6) p-val	(7) T-C Diff	(8) p-val
Mother's age	27.00	5.2	−0.51	0.14	−0.39	0.30	−0.32	0.44
Mother's height (m)	1.56	0.1	0.00	0.32	0.01	0.17	0.01	0.18
Mother's BMI	23.13	4.1	0.24	0.37	0.15	0.61	0.36	0.28
Mother's education	3.87	4.0	0.37	0.17	0.75	0.01	0.71	0.03
Empowered	0.55	0.5	−0.03	0.35	−0.05	0.19	−0.04	0.36
Mother usually works	0.03	0.2	−0.01	0.25	−0.01	0.43	−0.01	0.38
Parity	2.33	1.8	−0.17	0.15	−0.29	0.03	−0.34	0.02
First child	0.19	0.4	0.01	0.82	0.01	0.61	0.01	0.65
Index child is female	0.49	0.5	0.03	0.34	0.05	0.18	0.06	0.13
Share children female	0.52	0.3	0.02	0.44	0.00	0.93	0.01	0.84
Depression score	14.43	4.0	0.40	0.15	0.25	0.41	0.31	0.35
Disability score	8.30	2.7	−0.17	0.34	−0.16	0.45	−0.03	0.88
General functioning	62.05	5.3	−0.01	0.97	0.35	0.37	0.58	0.18
Perceived social support	44.49	16	1.08	0.32	1.82	0.13	2.63	0.05
Joint/extended family	0.57	0.5	0.04	0.27	0.05	0.17	0.06	0.18
Mother-in-law present	0.40	0.5	0.06	0.05	0.08	0.02	0.09	0.04
Mother's mother present	0.05	0.2	0.03	0.07	0.04	0.05	0.03	0.11
Father's education	7.09	3.9	−0.13	0.61	−0.10	0.73	−0.22	0.48
Father employed	0.91	0.3	−0.01	0.53	−0.02	0.50	−0.00	0.88
Not in manual labor	0.29	0.5	−0.01	0.80	0.01	0.86	0.00	0.93
SES (0=poor, 4=rich)	1.38	1.0	0.01	0.91	0.11	0.15	0.12	0.13
Wealth index[a]	−0.03	2.0	0.06	0.66	0.24	0.11	0.19	0.24
LTFU from baseline	0.33	0.5	0.05	0.14				
Joint test (p-value)				0.13		0.12		0.01

Notes: This table tests for balance in the baseline, 1-year follow-up (Rahman et al., 2008), and 7-year follow-up samples. Columns 1 and 2 show the mean and standard deviation in the control group in the full baseline sample, and Columns 3, 5, and 7 show the difference in means between treatment and control group, in the three different samples. The respective p-values are reported in Columns 4, 6, and 8.. LTFU from baseline = share of baseline respondents who are lost to follow-up. [a] The wealth score is a 19-item PCA-weighted index of assets (radio, TV, refrigerator, washing machine, air conditioning), house and roofing materials (brick walls, metal roof), and water and waste infrastructure (type of drinking water, flush toilet or any type of latrine).

Imbalance in baseline characteristics was mostly driven by mothers of boys, while the sample of mother of girls appears more balanced: a joint test of balance yields a *p*-value of 0.04 for mothers of boys, and a *p*-value of 0.81 for mother of girls (see Appendix Table A2).

The few imbalances already present at baseline were exacerbated in the follow-up samples at 1 and 7 years. Besides persistent differences in the presence of the grandmother in the household, treated group women reported about 0.7 more years of education, and 0.3 fewer children. In the 7-year follow-up sample, treated mothers also reported higher perceived social support at baseline. A joint test of balance for the 7-year follow-up sample yields a *p*-value of 0.01. We discuss how we address this imbalance in the next section.

Attrition.

Seven years after the end of the intervention, we located 585 (65%) of the baseline sample, 289/463 (62%) in the intervention arm and 296/440 (67%) in the control arm. Attrition was not significantly different by treatment status. A two-sample Wilcoxon rank-sum test of difference in means cannot reject the null of no difference, with a *p*-value of 0.12. Although fairly similar along most characteristics (Appendix Table A1), attritors (lost-to-follow-up, LTFU) were less empowered, perceived less social support, and were more likely to work than mothers re-enrolled at year 7. There were no significant differences at the 5% level in baseline characteristics between treated and control LTFU mothers, and we fail to reject the joint test that characteristics of attritors in treatment were different to controls (p=0.59).[5] Attrition patterns differed by the gender of the child in the womb at the start of the intervention. For mothers of girl, attrition rates were very similar across trial arms, with 65% of the treatment arm located after 7 years and 66% of the control arm. By contrast, for mothers of boys, attrition rates were significantly higher in the treatment arm (just 60% located, compared to 69% of controls, see Online Appendix Table D.10).

Estimated treatment effects on 6- and 12-month mental health outcomes are the same regardless of whether we use the full sample or the 7-year follow-up subsample (Online Appendix Table D.11), suggesting that attrition was not systematically related to improvements in mental health. Across all the range of mental health outcomes, a joint test of whether treatment effects are different for the 7-year subsample yields a *p*-value=0.60 for the 6-month outcomes and 0.98 for 12-month outcomes. Differences in treatment effects across the different samples range between 2 and 5 percent of a standard deviation.

Nevertheless, we also assess the robustness of our results to accounting for attrition in two ways (details are in Online Appendix D.3). First, we calculate treatment effects using Inverse Probability Weighting, where the weights are cal-

---

[5]The average magnitude of the difference between attritors and non-attritors is about 0.07 standard deviations (of the control group), and the difference between treatment and control among attritors is 0.08 SD.

culated as the predicted probability of being in the 7-year follow-up sample based on the available baseline controls. Second, we calculate attrition bounds based on Lee (2009), which sorts the outcomes from best to worst within each treatment arm and then trims the sample from above and below to construct groups of equal size. Our conclusions are, in general, robust to these corrections.

A. Outcomes

Maternal mental health

The THP intervention was designed to reduce the incidence of depression among prenatally depressed mothers. As such, it was careful in measuring clinical depression. Maternal depression was assessed using the Structured Clinical Interview (SCID) for DSM-IV diagnosis. Evaluations to diagnose major depressive episodes at baseline, 6-months, and 1-year were done by a psychiatrist. At the 7-year follow-up, depression was again determined by the SCID, but administered by trained assessors. In addition to the binary indicator for major depression, the baseline, 6-month, and 1-year surveys also contained other indicators of mental health including the Hamilton Depression Rating (a measure of depression severity), Brief Disability Questionnaire (measure of how disabling symptoms are), the Generalized Assessment of Functioning (assessor-determined measure of functioning incorporating severity of symptoms and their effect on functioning). More detail with references to the validation and timing of the measures is in Online Appendix J.

Mother's decision-making

We observe several measures related to maternal decisions and we group them into five categories: the mother's financial empowerment, fertility trajectory, and parenting behaviors. Parenting is grouped into 3 domains: time-intensive investment, money-intensive investment, and parenting style. For brevity, we label these categories "mother's decision making" since they are related to choices that mothers made. However, we acknowledge that these outcomes emerge from a more complex interplay between external constraints and household bargaining, and they are not solely dependent on mothers' decision-making.

We define financial empowerment as control over resources (as in Karlan et al. (2017); Lavy, Lotti and Yan (2016)). In Pakistan, women's empowerment is low, with strictly defined gender norms. Most women are confined to their homes, do housework for the extended family, and are excluded from decision-making (Ahmad and Khan, 2016). In our sample, only 3% of mothers reported usually working at baseline (recall they are pregnant at baseline), none were working at the 6 or the 12-month follow-ups, and only 12 percent worked 7 years after the end of the intervention.

We evaluate treatment effects on *trajectories* of empowerment and fertility, leveraging data from earlier and later follow-up. For parenting inputs we rely

on the extensive measures collected at the 7-year follow-up, consisting of both mother reported and direct interviewer observations but we present results for parenting in infancy in the Online Appendix (Table H.44). As we have numerous indicators within each domain, we generate summary indices. We describe the indicators succinctly here and provide detailed definitions and summary statistics in the Online Appendix (B.3–B.5, and Section J).

1) **Financial empowerment** was measured using whether the woman was employed, her monthly earnings, and whether she had control over spending. We use all three measures from the 7-year follow-up but include only control over spending from earlier waves as no women worked in the first year postpartum.[6]

2) **Monetary investments** included the school quality index, whether the index child attended private school, the mother's expected grade attainment for the index child, the family's education expenditures in the past month, as well as the learning material and physical environment sub-scales of the HOME inventory. Since expected grade attainment is related to schooling, we include it in the monetary index alongside other questions pertaining to schooling. Mothers answered a detailed module on the home environment, which was a cultural adaptation of the short HOME inventory (Caldwell and Bradley, 1984) similar to the HOME-SF used, for instance, in the National Longitudinal Survey of Youth, 1979. A number of the HOME questions were based on interviewer observations (rather than mother reports)–see Online Appendix J. School quality was objectively assessed by visiting the child at school and recording information on school amenities (playground, computers, library, etc), number of teachers and classrooms, and classroom amenities.

3) **Time-intensive parental investments** included the frequency that the mother and father played with the index child, whether someone helped with his or her studies, as well as enrichment, family companionship, and family integration sub-scales of HOME (all measures are directed at the index child).

4) **Parenting style** captured parenting behavior that did not have explicit time or monetary costs. For example, how the mother spoke to the child, or the style of discipline she used. It was measured using the Parenting Practices Inventory (PPI) (Webster-Stratton, Reid and Hammond, 2001), capturing the disciplinary style of the parents such as the extent of harsh or inconsistent disciplining; and the responsivity, encouragement of maturity, and emotional climate sub-scales of the HOME score.

5) **Fertility trajectory** was assessed between the beginning of the intervention and the 7-year follow-up. At the 1-year follow-up, mothers were asked if they were pregnant. At the 7-year follow-up, mothers completed a child roster, with ages (but not genders) of children. From this we can infer how many surviving

---

[6]Short term effects (at 6 months) on control over spending were documented in (Rahman et al., 2012) when the intervention was still ongoing.

children the mother had since the intervention began. First, to summarize results in a style consistent with other outcomes, we create an index with the total number of siblings of the index child (the child in womb at the start of the intervention) born since treatment started, whether the mother was pregnant again at the 1-year follow-up, whether the index child is the last child, and the ideal number of children (asked only at the 7-year follow-up). Second, based on the child roster data reported by the mother at the 7-year follow-up, we also plot the fertility trajectory as a function of time since treatment.

<div align="center">Child outcomes</div>

We report results for four groups of child outcomes at the 7-year follow-up, referring to the physical, cognitive, socio-emotional development of the index child, and the survival of siblings. For infancy we had the length and weight of the child, diarrheal episodes and acute respiratory infections (ARIs); these results are in Online Appendix Table H.45. Although the women had other children that could have been affected by the intervention, most outcomes analyzed are for the index child, i.e. the child in the womb at the start of the intervention. Similarly, most of the questions about parenting refer specifically to the index child. Some items of the HOME inventory, such as the physical environment subscale measured the overall quality of the household environment, such as cleanliness and safety.

1) **Physical development** was measured using growth, fine motor skills, and illness. Interviewers measured height and weight, and motor skills were assessed using the Grooved Pegboard Test. Mothers reported any severe illness, hospitalizations, eye and hearing problems of the child.

2) **Cognitive development** was assessed with the Wechsler Preschool and Primary Scale of Intelligence, designed for children 2.5-7.5 years old (WPPSI-IV). It provides primary index scales for verbal comprehension, visual-spatial, fluid reasoning, working memory, and processing speed. Executive functioning was assessed using a Stroop-like test, which gauges inhibition and working memory. Basic literacy and numeracy tests were administered by us, providing math and Urdu scores based on the number of correct answers out of 16 and 12 questions respectively. The school grade of the child was based on teacher reports.

3) **Socio-emotional development** was measured along two broad domains: behavioral and emotional problems, assessed with the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997)); and anxiety, assessed with the Spence Children's Anxiety Scale (SCAS) (Spence, 1998)). Both measures were based on questions answered by the mother.

4) **Sibling survival**. Survival of the index child's siblings is the only sibling outcome we are able to evaluate. We use the mother's reports of child mortality since the intervention, and the sex ratio of surviving children at the 7-year

follow-up to assess treatment effects.[7]

IV.   Econometric Specification

Since treatment assignment was random, our principal estimating equation is:

$$(1) \qquad Y_{ic} = \alpha + \beta T_c + \mathbf{\Gamma}' \mathbf{X_{ic}} + \varepsilon_{ic}$$

where Y_{ic} is the outcome for the mother or child i living in cluster c. As discussed, we create outcome indices. These are standardized within the control group to have mean zero and standard deviation one, and positive values are always associated with more favorable outcomes.

T_c is a dummy equal to one if the mother is in the intervention group, which by the cluster design varies only at the Union Council level, c. $\mathbf{X_{ic}}$ is a vector of controls. We first show a parsimonious model controlling only for interview date and interviewer fixed effects. We then include the full set of baseline characteristics, demeaned and interacted with the treatment indicator. This accounts for the fact that the follow-up sample was not balanced along all observable baseline characteristics, and the interaction with treatment allows for differing impacts of these characteristics on outcomes. The characteristics include mother's age and its square, parity, family structure, presence of grandmother (mother or mother-in-law of depressed mother), mother's education, father's education, if the mother was employed, if mother was empowered, wealth index, depression severity (Hamilton score), and the Multidimensional Scale of Perceived Social Support (MSPSS). Our results are not sensitive to adding controls.[8]

We report standard errors clustered at the unit of randomization. However, p-values are computed using randomization inference, adjusting for multiple hypothesis testing and with randomization permuted at the cluster level (Young, 2018).

The parameter of interest, β, represents the Intention-to-Treat (ITT) effect. Since attrition was not different by treatment status, treatment was not available to control mothers, and all women offered the treatment accepted it, β would be the average treatment effect on the treated (ATT) if we could assume that all treated women participated in all sessions. Although we had full compliance we do not observe how many sessions the women actually received.

We identify the causal effect of the CBT intervention treating depression rather than the causal effect of depression. Not all mothers recovered from depression in the treatment arm, and many mothers in the control arm spontaneously recovered. We do not use the randomized assignment as an instrument for depression, since it

---

[7]While attrition due to mortality was scrupulously reported between baseline and 1-year, between the 1-year and 7-year follow-up child survival was not reported for study participants who moved or refused at any time, and only a few instances of child deaths were specifically reported.

[8]Child age is not in the controls as it is potentially endogenous and the range is limited, the children being of a single birth cohort born 2005-2006. However treatment did not affect the age of the child (by lengthening pregnancy, for example), and the results are nearly identical if we control for age.

is possible that the intervention may have had direct impacts on some outcomes above and beyond affecting maternal depression, through encouraging healthy thinking and bonding with the child.

Multiple Inference and Power.

As we gathered numerous indicators within each domain, following O'Brien (1984); Kling, Liebman and Katz (2007); Anderson (2008), we construct summary indices as the most efficient weighted average of a set of outcomes.[9] This addresses the problem of multiple inference by reducing the number of hypotheses tested and also improves the power of our statistical tests for whether the intervention had effects for groups of related outcomes.

We further account for multiple hypothesis testing across the indices by calculating p-values using a step-down procedure with a non-parametric permutation test which controls the family-wise error rate (FWER), following Westfall and Young (1993); Efron and Tibshirani (1994). The p-values used in this procedure use the p-values generated by randomization inference (Young, 2018).

Power calculations for the 7-year follow-up were structured around the WPPSI-III full-scale IQ measure for children. Calculations were based on re-enrollment numbers that were slightly optimistic with N of 328 in the intervention arm (actual 289) and 314 in the control arm (actual 296) and an inter-cluster correlation (ICC) of 0.05. The ICC was based on the observed ICC in the same clusters for the maternal mental health variables in the original study (Rahman et al., 2008). With these parameters, the study had 80% power to detect a difference of 0.36 standard deviations in standardized scores.

Heterogeneous Treatment Effects.

We report heterogeneous treatment effects, estimated separately for each gender of the index child, since in South Asia son preference leads to investments and outcomes often favoring sons (Sathar et al., 2015; Miller, 1981; Bhalotra et al., 2016) and mothers of sons (Milazzo, 2014; Bhalotra et al., 2018). Additionally, males are biologically more sensitive to stress *in utero* (Trivers and Willard, 1973), and male and female children respond differently to early life shocks (Autor et al., 2016; Bhalotra and Venkataramani, 2011). Additionally, we test for treatment effect heterogeneity by socio-economic status, demographic indicators, and baseline depression severity (see Online Appendix Section G).

---

[9]The index weights outcomes using the inverse of their variance-covariance matrix. This procedure is akin to estimating a Seemingly Unrelated Regression model of all (standardized) outcomes on the treatment indicator jointly, while constraining the coefficients to be equal within each grouping. As this is a Generalized Least Squares estimator it provides the most efficient estimation of the treatment effect across related outcomes. As an alternative, we compute factor scores, a method more suited when the measures are proxies of an underlying one-dimensional latent factor, measured with noise. The results (Online Appendix Table E.14) are qualitatively similar. The factor score drops individuals with any missing data, whereas the GLS-weighted index allows for item non-response, weighting outcomes more where more data are available.

V.   Results

We find persistent benefits of the intervention on women's mental health at the 7-year mark. We also identify meaningful and statistically significant intervention impacts on mother's control over household resources and parental investments in children (including parenting style for girls). Yet we find no detectable average effects on fertility, and only small and imprecisely estimated effects across multiple domains of child development.

Our main results are presented in Figure 2 and in Tables 2, 3, 4, and 6. The tables display treatment effects with parsimonious and enriched controls and report the randomization inference *p*-values from the richer model, adjusted to control for the FWER (FWER adjustment is not performed in Table 2 since we test the effect of the intervention on only one construct, clinical depression, which is the main target of the intervention). We also report treatment effects estimated separately for each gender of the index child. Our findings are robust to numerous sensitivity checks (Appendix Section E), and generally survive both parametric and non-parametric corrections for attrition, reported in Table 8.
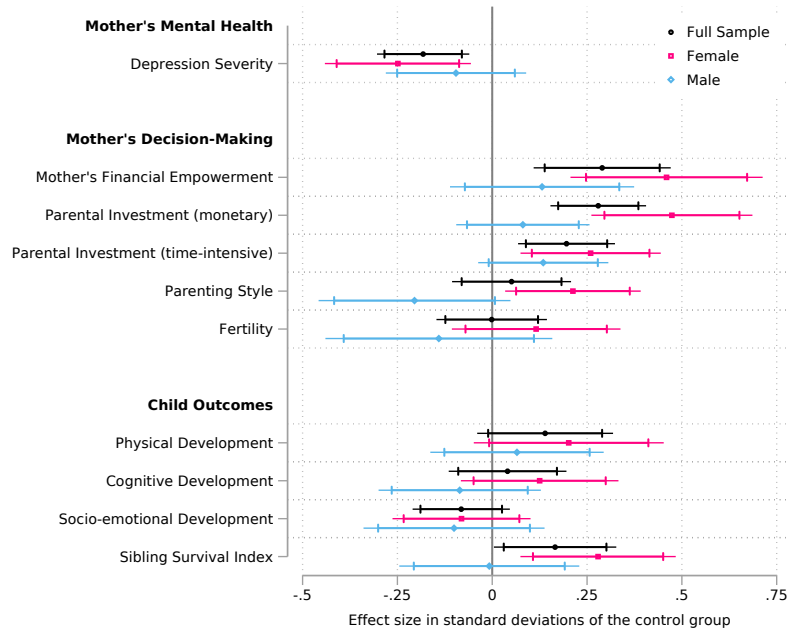

FIGURE 2.   INTERVENTION EFFECTS ON MAIN OUTCOMES, POOLED AND BY GENDER

Notes: Figure shows the estimated intervention effects in standard deviations, pooled and separately by gender (reported in Tables 2-9), along with 99% and 95% confidence intervals calculated using randomization inference (Young, 2018). Details of the estimation procedure appear in the Section IV and Tables notes.

A. Depression Trajectories

The CBT intervention was highly successful in reducing depression, not only at 6-months and 1-year, as shown by Rahman et al. (2008), but also 7 years after the intervention concluded. Figure 3 shows the rapid recovery in the treatment arm, with depression rates stabilizing around 25% after just 6 months. By contrast, the recovery rate in the control arm was much more gradual, with 58% of mothers in the control group suffering from depression after 1 year postpartum. This spontaneous recovery in the control group, or mean-reversion, is consistent with the literature showing that major depression episodes subside over time even without treatment (Cuijpers et al., 2014; Fisher et al., 2012; Gelaye et al., 2016). Nevertheless, even after 7 years, recovery rates in the control group did not catch up with treatment: the rate of depression among women in the control arm was 30% (n=90/296), compared to 24% (n=69/289) in the intervention arm.
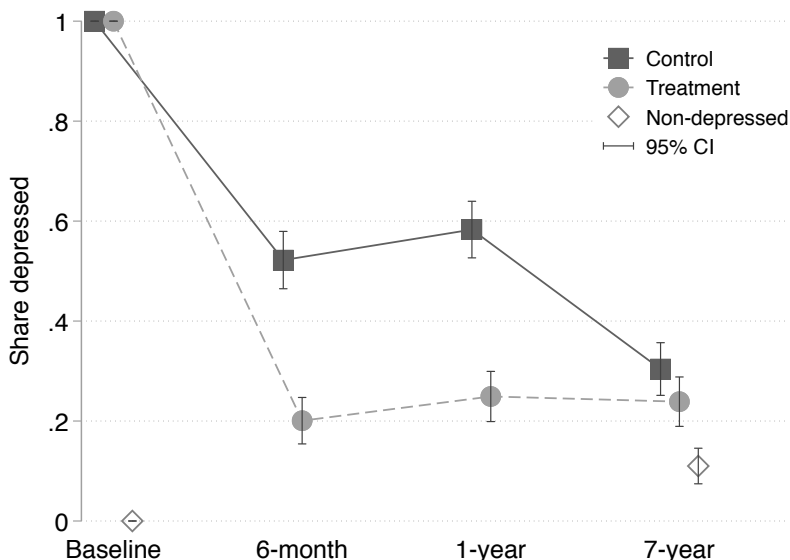
FIGURE 3. TREATMENT EFFECTS ON DEPRESSION

Notes: Figure shows the share of women clinically diagnosed with major depression at each follow-up wave by trial arm. 95 percent confidence intervals are based on conventional standard errors of the raw (unadjusted) means.

Table 2 reports the regression-adjusted estimates of the treatment effects on depression and its severity. At the 7-year follow-up, women in the treated arm were 5 percentage points less likely to be clinically depressed (p=0.076) and experienced a 0.18 SD reduction in depression severity (p=0.004) relative to control women (column 3). This longer-run reduction in severity is about one-third of the

reduction in depression severity identified in the short-run (0.6-0.7 SD). Importantly, the diminishing treatment effect over time does not arise because treated women relapse, but rather due to the spontaneous recovery of the control women.

The quick and stable recovery brought about by CBT is remarkable. As can be seen in the last panel of Table 2, women in the treatment arm were 23 percentage points more likely to be not depressed at both post-CBT follow-ups, a sign of permanent recovery, and 12 percentage points less likely to have never recovered (depressed at both 1- and 7-year follow-ups). The latter result is important as it indicates that intervention was effective against chronic depression.

TABLE 2—TRAJECTORY OF MATERNAL MENTAL HEALTH

	Intervention Effects - Full sample			
	(1) Control group mean	(2) β (s.e.)	(3) Adjusted β (s.e.)	(4) RI p-value
At the 6-month followup				
Depressed	0.52	−0.32	−0.32	0.000
	(0.50)	(0.05)	(0.04)	
Depression severity	0.00	−0.62	−0.60	0.000
	(1.00)	(0.10)	(0.08)	
At the 1-year followup				
Depressed	0.58	−0.32	−0.30	0.000
	(0.49)	(0.04)	(0.05)	
Depression severity	0.00	−0.70	−0.66	0.000
	(1.00)	(0.09)	(0.09)	
At the 7-year followup				
Depressed	0.30	−0.06	−0.05	0.076
	(0.46)	(0.03)	(0.03)	
Depression severity	−0.00	−0.22	−0.18	0.004
	(1.00)	(0.06)	(0.06)	
Recovery trajectory				
Recovered permanently[a]	0.34	0.25	0.23	0.000
	(0.47)	(0.04)	(0.04)	
Never recovered[a]	0.23	−0.14	−0.12	0.000
	(0.42)	(0.03)	(0.03)	

Notes: N=585. Depression severity is an index of all depression measures available in the data at each time point (see Online Appendix J and Table B.2 for detailed definitions). Standard errors, clustered at the level of randomization, in parentheses. Column (4) reports randomization inference p-values, clustered at the level of randomization (Young, 2018). All models control for interview date and interviewer. Specifications with all controls additionally adjust for baseline characteristics (all centered and interacted with the treatment indicator). The set of baseline characteristics include mother's age and its square, parity, family structure, presence of grandmother (mother or mother-in-law of depressed mother), mother's education, father's education, if mother was employed, if mother empowered, PCA-weighted wealth index, depression severity (Hamilton score), and perceived social support (MSPSS).Attrition-adjusted estimates are presented in Table 8.
[a] "Never recovered" is defined as being clinically diagnosed with depression at both follow-ups after the therapy ended (1-year and 7-year). "Recovered permanently" is defined as being assessed as not clinically depressed at both post-therapy follow-ups.

Treatment effects on depression, particularly in the longer-run, were significantly larger for mothers of girls with an estimated reduction in depression rate of 10 percentage points (Table 3). Interestingly, at 6 months postpartum, treatment benefited mothers of boys slightly more, but by 1 year postpartum, the treatment effect was 8 pp larger for mothers of girls. The benefits of treatment thus shifted over time toward mothers of girls. In particular, treatment was more likely to reduce chronic depression in mothers of girls versus boys (by 16 percentage points, $p=0.002$).

TABLE 3—DIVERGENT TRAJECTORIES OF MATERNAL MENTAL HEALTH BY CHILD GENDER

	Intervention Effects - By Child Gender				
	(1) Girl control mean	(2) Boy control mean	(3) β^{Girl} (s.e.)	(4) β^{Boy} (s.e.)	(5) $\beta^{Girl} = \beta^{Boy}$ p-value
At the 6-month followup					
Depressed	0.48	0.56	−0.29 (0.05)	−0.37 (0.07)	0.287
Depression severity	−0.08	0.07	−0.51 (0.10)	−0.75 (0.12)	0.097
At the 1-year followup					
Depressed	0.58	0.58	−0.36 (0.06)	−0.28 (0.05)	0.171
Depression severity	−0.01	0.01	−0.73 (0.12)	−0.63 (0.10)	0.402
At the 7-year followup					
Depressed	0.33	0.28	−0.10 (0.04)	0.05 (0.04)	0.011
Depression severity	−0.01	0.01	−0.25 (0.09)	−0.10 (0.09)	0.229
Recovery trajectory					
Recovered permanently[a]	0.34	0.34	0.28 (0.06)	0.21 (0.04)	0.224
Never recovered[a]	0.26	0.20	−0.18 (0.04)	−0.02 (0.04)	0.002

Notes: N=585. Depression severity is an index of all depression measures available in the data at each time point (see Online Appendix J and Table B.2 for detailed definitions). Columns 3 and 4 reports treatment effects separately by gender of the index child (controlling for all baseline characteristics described in Table 2). Standard errors, clustered at the level of randomization, in parentheses. Column 5 reports the test of equality in treatment effects between the two samples. Attrition-adjusted treatment effect estimates by child gender are reported in Online Appendix Table D.12.

[a] "Never recovered" is defined as being clinically diagnosed with depression at both follow-ups after the therapy ended (1-year and 7-year). "Recovered permanently" is defined as being assessed as not clinically depressed at both post-therapy follow-ups.

While the longer term effects on depression are not sensitive to attrition correction using inverse probability weighting, they are somewhat sensitive to attrition bounds, especially when bounds are calculated by grouping the mothers of girls

and boys together (Table 8). Separately estimating the bounds by child gender, the 95% confidence interval for the treatment effect on depression at the 7-year mark for mothers of girls ranges from -0.22 to 0.03.

In order to interpret the results on women's decision-making that follow, we note that the total effect on women's depression stemming from the intervention is the cumulative difference in depression throughout all these years (rather than just the difference in prevalence at the 7-year mark).

B.   *Women's Decision-Making*

The CBT intervention significantly increased mother's financial empowerment by 0.29 SD, monetary parental investments by 0.28 SD, and time-intensive parental investments by 0.20 SD, as summarized in Figure 2 and shown in Table 4.[10] On average, we find no effects on parenting style or fertility, with very small estimated treatment effects of 0.04 SD and 0.01 SD respectively.

Treatment effects are driven by mothers who were pregnant with a girl at the start of the intervention. This is reassuring since baseline characteristics and patterns of attrition were more balanced in the subsample of mothers with girls. The estimates for mothers of boys are consistently smaller and often indistinguishable from zero. This is true both for the indices and their individual components. Mothers of girls saw a 0.46 SD increase in financial empowerment, a 0.47 SD increase in monetary investments, and a 0.21 SD improvement in parenting style. All of these effects are statistically stronger for mother of girls than mothers of boys. Time-intensive investments in girls increased by 0.26 SD, twice as much as for boys, but in this case the gender difference is not significant.

That the mother's gaining greater control over household resources goes hand in hand with greater investments in children is consistent with previous research, for example (Lundberg, Pollak and Wales, 1997). This result is reinforced by the fact that increases in financial empowerment and in parental investments were larger for mothers of girls. Notice that this gender difference in effects does not necessarily require that mothers had stronger preferences for girls or for equality between boys and girls, although this may have contributed.[11]

To put the magnitudes of the estimated treatment effects into context, Table 5 reports the mean differences in outcomes between control arm women who were

---

[10]We have used data from every follow-up to construct the indices of maternal financial empowerment and fertility. Rahman et al. (2012) reported the impact of the same intervention on mother's control over spending while the intervention was ongoing (at 6 months). Repeating the analysis using only the outcomes measured at 7 years, the magnitude of the effect on financial empowerment falls from 0.29 to 0.18 SD, and the effect for fertility goes from 0.01 to 0.03, but the results are qualitatively similar (Online Appendix Tables E.15 and H.34.

[11]We considered whether differential effects by gender might reflect selective processes of gender-differentiated abortion or fertility over time, resulting in treated girls being more "wanted". However, sex-selective abortion is very unlikely as prenatal sex detection diagnostics were unavailable to this population (Zaidi and Morgan, 2016), Muslims appear not to conduct female feticide (Almond, Edlund and Milligan, 2013; Bhalotra, Clots-Figueras and Iyer, 2017), and we find no significant evidence of gender differences in survival (Table 1) or fertility (Online Appendix Table H.33.)

TABLE 4—INTERVENTION EFFECTS ON MOTHER'S DECISION-MAKING

	Full Sample			By child gender		
	(1) β / (s.e.)	(2) Adjusted β / (s.e.)	(3) FWER p-val spec (2)	(4) β^{Girl} (s.e.)	(5) β^{Boy} (s.e.)	(6) $\beta^{Girl} =$ β^{Boy} p-value
Mother's financial empowerment	0.29 (0.11)	0.29 (0.09)	0.019	0.46 (0.12)	0.13 (0.12)	0.031
Parental investment (monetary)	0.35 (0.07)	0.28 (0.06)	0.012	0.47 (0.11)	0.08 (0.09)	0.003
Parental investment (time-intensive)	0.20 (0.07)	0.20 (0.06)	0.028	0.26 (0.09)	0.13 (0.08)	0.280
Parenting style	0.04 (0.08)	0.05 (0.08)	0.807	0.21 (0.09)	-0.20 (0.13)	0.001
Fertility trajectory	0.01 (0.10)	-0.00 (0.09)	0.989	0.12 (0.11)	-0.14 (0.15)	0.092

Notes: N=585.   Summary indices are normalized to be mean 0 and SD 1 in the control group, with positive values are associated with more favorable outcomes. Index construction is described in the text (Section III.A). Standard errors, clustered at the level of randomization, in parentheses. All models control for interview date and interviewer. Specifications with all controls additionally adjust for baseline characteristics (all centered and interacted with the treatment indicator). The set of baseline characteristics include mother's age and its square, parity, family structure, presence of grandmother (mother or mother-in-law of depressed mother), mother's education, father's education, if mother was employed, if mother empowered, PCA-weighted wealth index, depression severity (Hamilton score), and perceived social support (MSPSS). Inference is conducted using randomization inference p-values, clustered at the level of randomization (Young, 2018). RI p-values are adjusted to control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 4 and 5 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 6 reports the test of equality in treatment effects between the two samples.

depressed at baseline and the women who were not depressed at baseline. The magnitudes of the estimated treatment effects are very similar to these descriptive differences, which represent the association between perinatal depression and outcomes 7 years later in the absence of treatment, which we call the "depression gap." These gaps show significantly lower financial empowerment for the perinatally depressed woman, lower fertility, and lower time and monetary investments in children 7 years later. Parenting style does not seem to be strongly associated with depression. Gender differences in these gaps are slightly larger for girls, except for time-intensive investments, but none of these gender differences are statistically significant. CBT closed or considerably narrowed these depression gaps, bringing medium-term outcomes of perinatally depressed women close to those of their non-depressed counterparts.

The magnitude of the estimated treatment effects can also be compared to the gender gap in the control group means (Online Appendix Table E.19). Gender gaps in the control group are either small, as in the case of parenting style and time intensive investments, or in favor of boys. The intervention acted to narrow or close the gender gap for mother's empowerment and monetary investment, while it did not close the gender gap in fertility.

Table 5—Depression gap in mother's outcomes

	Depressed controls − Non-Depressed			
	(1) Pooled	(2) Girls	(3) Boys	(4) p-value
Depression index (7y)	0.54 (0.07)	0.51 (0.10)	0.57 (0.09)	0.64
Mother's financial empowerment (7y)	−0.31 (0.08)	−0.45 (0.12)	−0.18 (0.12)	0.12
Parental investment (monetary)	−0.32 (0.08)	−0.35 (0.13)	−0.29 (0.11)	0.72
Parental investment (time-intensive)	−0.39 (0.08)	−0.25 (0.12)	−0.51 (0.11)	0.12
Parenting style	−0.09 (0.08)	−0.16 (0.12)	−0.04 (0.11)	0.46
# kids born past 7yrs	−0.24 (0.08)	−0.26 (0.13)	−0.23 (0.11)	0.82

Notes: Table reports raw difference in means between prenatally non-depressed mothers (N=300) and prenatally depressed controls (N=296) for key outcome indices and variables at the 7-year follow-up. Since prenatally non-depressed mothers were interviewed only at the 7-year follow-up, maternal financial empowerment index does not contain mother controls spending (6m) and mother gets pocket money (12m), and only the number of kids born in the past 7 years are reported in lieu of the fertility index. Columns 2 and 3 report the depression gap by child gender, column 4 reports the *p*-value of the test that the gap differs by child gender.

To better understand these results, we look at treatment effects for each component of these indices (Online Appendix Tables H.29–H.33). For financial empowerment, we find consistent improvements in treated women getting higher control over spending (about 10 percentage points at 6-month, 1-year, and 7-year follow-ups). Treatment effects are greater for mothers of girls, even in the short run, but the difference is statistically significant only at the 7-year mark, mirroring the results on mother's depression trajectories. Women were also more likely to be working, but this is imprecisely estimated in part because of very low labor force participation among women (only 10% of control women work).

For monetary parental investment, treatment induced a higher probability of sending children to private schools (12 percentage points) and better quality schools (0.22 SD), as well as higher educational expectations (one-third of a grade) and better learning materials in the home (0.21 SD). For time-intensive parental investments, the intervention improved by about 0.20 SD all of the subscales of the HOME inventory which indicate enrichment, family companionship, and family integration, as well as the probability that someone helped with studies (6 percentage points). For parenting style, none of the individual components of the index were robustly improved by the treatment: we find only a marginal improvement in the emotional climate, responsivity, and not being harsh, especially for girls. For fertility, we find a precise null effect: none of the individual outcomes were shifted by the intervention, and no differences in number or timing of subsequent births (Online Appendix Figures I.1 and I.2).

Unpacking the indices further reveals significant changes in both mother-reported

and interviewer assessed measures (such as school quality, presence of learning materials in the home), suggesting that our findings are not driven by experimenter demand effects.

C.   Child Outcomes

The estimated effects of the intervention on indicators of child development in the cognitive, socio-emotional, or physical domains at the 7-year follow-up are small and noisily estimated. Effects for sibling survival are slightly larger, but not significantly different from zero once adjusting for multiple hypothesis testing. This supports the conclusions of Maselko et al. (2015), who estimate the impact of this program on a subset of the child development outcomes that we consider.[12] Besides broadening the set of child outcomes, we created indices with more statistical power and conducted additional robustness checks. Still, larger samples will be necessary to detect effects of reasonable magnitudes.

Estimated average effects range from -0.08 to 0.14 SD, as shown in Table 6. We are able to reject effects bigger than 0.19 SD in socio-emotional, 0.33 SD in cognitive, and 0.40 SD in physical development, as shown by the attrition-corrected upper 95% confidence intervals in Table 8. We explore whether treatment effect heterogeneity may be masked in the average effects, but find little evidence to suggest that is the case. Quantile treatment effects are not significantly different from zero in any part of the distribution, though treatment effects on physical development show larger effects in the lower two-thirds of the distribution (Online Appendix Figure I.4). There is no evidence of substantial positive treatment effects in any of the subpopulations identified by child gender, mother's education, mother's age, parity, wealth, and presence of grandmother at baseline (Online Appendix Table G.24).

For completeness, we analyze treatment effects on sub-components of each child development index (Online Appendix Tables H.35–H.38). Even at this more granular level, we find that the only few positive effects on specific indicators are concentrated among girls, especially in the domains of physical and cognitive development. For example, of the eight measures of physical development, all average effects are smaller than 0.06 SD, sometimes negative. Differentiated by gender, only BMI-for-age shows a treatment effect bigger than 0.15 SD, positive for girls but negative for boys. Of the nine components of cognitive development, only processing speed (+2.52 points or 0.26 SD) and fluid reasoning (+1.43 points or 0.13 SD) improve in the whole treated group by more than one IQ point, and these results are driven by girls. Of the eleven components of socio-emotional development, only two increases more than 0.15 SD: treated girls display fewer conduct problems but higher obsessive-compulsive subscale scores. The sibling survival index is 0.17 standard deviations larger in the treated group, and this is

---

[12]They analyze only anthropometrics, verbal and visual WPPSI, SDQ and SCAS. See Online Appendix Table B.4 for more details.

TABLE 6—INTERVENTION EFFECTS ON CHILD OUTCOMES AT AGE SEVEN

	Full Sample			By child gender		
	(1)	(2)	(3)	(4)	(5)	(6)
		Adjusted	FWER	β^{Girl}	β^{Boy}	$\beta^{Girl} = \beta^{Boy}$
	β / (s.e.)	β / (s.e.)	p-val spec (2)	(s.e.)	(s.e.)	p-value
Physical development	0.15	0.14	0.435	0.20	0.07	0.358
	(0.09)	(0.09)		(0.13)	(0.12)	
Cognitive development	0.09	0.04	0.652	0.13	−0.09	0.092
	(0.08)	(0.08)		(0.11)	(0.11)	
Socio-emotional development	−0.11	−0.08	0.456	−0.08	−0.10	0.896
	(0.07)	(0.07)		(0.09)	(0.12)	
Sibling survival index	0.19	0.17	0.260	0.28	−0.01	0.019
	(0.08)	(0.08)		(0.10)	(0.11)	

Notes: N=585.   Summary indices are normalized to be mean 0 and SD 1 in the control group, with positive values are associated with more favorable outcomes. Index construction is described in the text (Section III.A). Standard errors, clustered at the level of randomization, in parentheses. All models control for interview date and interviewer. Specifications with all controls additionally adjust for baseline characteristics (all centered and interacted with the treatment indicator). The set of baseline characteristics include mother's age and its square, parity, family structure, presence of grandmother (mother or mother-in-law of depressed mother), mother's education, father's education, if mother was employed, if mother empowered, PCA-weighted wealth index, depression severity (Hamilton score), and perceived social support (MSPSS). Inference is conducted using randomization inference p-values, clustered at the level of randomization (Young, 2018). RI p-values are adjusted to control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993).

driven by improved survival of sisters. Survival of the index child does not seem to be impacted by the treatment, so these results indicate sibling spillovers.[13]

Overall, we are underpowered to detect reasonable effect sizes on child development at this age in our sample. So as to benchmark potential treatment effects, we perform two calculations. First, we compare child development indicators for women who were not prenatally depressed with those for prenatally depressed women in the control group (Table 7). Perinatal depression is descriptively associated with worse children outcomes in the 7-year follow-up, although for cognitive development the depression gap is smaller and noisily estimated. Differences in physical development appeared for both boys and girls (driven by mother-reported illness rather than child growth), while the depression gap is slightly bigger for girls in cognitive development and sibling survival, but for boys in socio-emotional development, although none of these differences are statistically significant. Second, we use the correlational relationship between parental investments and child outcomes (shown in Online Appendix Table E.16) to calculate the change in child development indicators that is to be expected as a consequence of the treatment effect on parental investments.[14]   These expected changes are 0.006 for socio-

---

[13]Appendix Table A1 reports differences in known death rates for attritors between treatment and control groups. Overall, 10 percent of the index children died or were severely ill after birth (explaining 28 percent of attrition), a lower mortality rate than the Punjab region. Of known child deaths, rates were not different by treatment status.

[14]We thank one of the referees for suggesting this calculation. See Appendix Section E.2 for more

emotional development, 0.048 for physical development, and 0.077 for cognitive development; all these changes are quite small, and fall within the confidence intervals of our estimated effects. We discuss alternative explanations for these small effects on child development further in Section V.E.

TABLE 7—DEPRESSION GAP IN CHILD OUTCOMES

| | Depressed controls − Non-Depressed | | | |
	(1) Pooled	(2) Girls	(3) Boys	(4) p-value
Physical development	−0.23 (0.08)	−0.21 (0.11)	−0.25 (0.11)	0.81
Cognitive development	−0.12 (0.09)	−0.24 (0.12)	−0.01 (0.12)	0.17
Socio-emotional development	−0.17 (0.08)	−0.05 (0.12)	−0.28 (0.11)	0.16
Sibling survival index	−0.16 (0.08)	−0.25 (0.12)	−0.07 (0.09)	0.22

Notes: Table reports raw difference in means between children of prenatally non-depressed mothers (N=300) and prenatally depressed controls (N=296) for key outcome indices and variables at the 7-year follow-up. Columns 2 and 3 report the depression gap by child gender, column 4 reports the *p*-value of the test that the gap differs by child gender.

D. Potential Mechanisms

The CBT intervention may have improved the outcomes through several mechanisms. For example, treatment may have led to better physical health among women as a result of better mental health (Ferrari et al., 2013), or as a direct result of the positive thinking therapy; husbands who reduced working hours in order to take care of the depressed wife or aid in child rearing may have been more likely to return to work in the treated group; the treatment may have directly engendered a better relationship with the husband or other members of the household (as this was part of the content of the intervention); or the child's grandmother might have been more likely to have moved in to help care for the mother and the child in the treated group, and this might have contributed to the outcomes we document. We delve into these mechanisms first by looking at direct effects of the intervention on intermediate outcomes, and then assessing heterogeneity of treatment effects based on observed baseline characteristics.

INTERMEDIATE OUTCOMES

In order to shed light on which of these mechanisms are more plausible, we investigate the effect of the intervention on several intermediate outcomes between birth and the 7-year follow-up. We generate indices for the trajectories of

details.

TABLE 8—ATTRITION-CORRECTED TREATMENT EFFECTS

	Treatment Effect β / (s.e.)		Attrition Bounds
	(1) Unweighted	(2) IPW	(3) 95% CI
Mother's mental health			
Depression index (7y)	−0.18	−0.19	[-0.43 , 0.19]
	(0.06)	(0.06)	
Depressed (7y)	−0.05	−0.05	[-0.16 , 0.08]
	(0.03)	(0.03)	
Never recovered	−0.12	−0.13	[-0.21 , 0.01]
	(0.03)	(0.03)	
Recovered permanently	0.23	0.23	[0.16 , 0.41]
	(0.04)	(0.04)	
Mother's decision-making			
Mother's financial empowerment (7y)	0.18	0.18	[0.06 , 0.56]
	(0.07)	(0.07)	
Mother's financial empowerment	0.29	0.29	[0.13 , 0.65]
	(0.09)	(0.09)	
Parental investment (monetary)	0.28	0.27	[0.01 , 0.68]
	(0.06)	(0.06)	
Parental investment (time-intensive)	0.20	0.20	[-0.00 , 0.67]
	(0.06)	(0.06)	
Parenting style	0.05	0.06	[-0.32 , 0.36]
	(0.08)	(0.08)	
Fertility trajectory	−0.00	0.00	[-0.35 , 0.39]
	(0.09)	(0.09)	
Child development outcomes			
Physical development	0.14	0.14	[-0.26 , 0.40]
	(0.09)	(0.09)	
Cognitive development	0.04	0.04	[-0.33 , 0.33]
	(0.08)	(0.08)	
Socio-emotional development	−0.08	−0.07	[-0.50 , 0.19]
	(0.07)	(0.07)	
Sibling survival index	0.17	0.16	[-0.22 , 0.43]
	(0.08)	(0.08)	
Potential mediators			
Grandmother present (7y)	0.06	0.07	[0.01 , 0.25]
	(0.04)	(0.04)	
Perceived social support (7y)	2.85	2.73	[-0.93 , 7.52]
	(0.84)	(0.86)	

Notes: Column 1 reproduces the treatment effect estimates without attrition correction. Column 2 shows attrition-corrected treatment effect estimates using IPW (Inverse Probability Weighting). Estimates in Columns 1 and 2 control for the full set of baseline characteristics as described in previous tables. Standard errors, clustered at the level of randomization, in parentheses. Column 3 shows 95% confidence intervals for the treatment effect using attrition bounds (Lee, 2009; Imbens and Manski, 2004) using the baseline starting sample with $N = 903$.

husbands' labor earnings (at 6-month, 1-year, and 7-year follow-ups), mothers' physical health (self-reported measures at 7-year, weight at 6-months), relationship quality with husband and mother in law (at 1 and 7-year follow-ups), grandmother trajectory (presence in the household in each of the three follow-ups), and perceived social support (at each of the three follow-ups); with all indices coded such that higher values correspond to more favorable outcomes.

Table 9 shows no evidence of effects on women's physical health (0.07 SD) or husbands' earnings (-0.04 SD), but some evidence of a more supportive household environment. Specifically, the intervention improved both relationship quality index and presence of grandmothers in the household by 0.16 SD. Grandmothers were 6 percentage points more likely to be present at each follow-up, while relationship quality improvements occurred primarily at 1-year but do not persist (Online Appendix Tables H.39–H.43).[15] Treated mothers reported a large (0.52 SD) and persistent improvement in the trajectory of perceived social support. Even though the magnitude of the treatment effect diminishes by 60% by the 7-year mark, perceived social support remains 0.23 SD higher among treated women 7 years after the intervention.

TABLE 9—POTENTIAL MEDIATORS

	Full Sample			By child gender		
	(1) β / (s.e.)	(2) Adjusted β / (s.e.)	(3) FWER p-val spec (2)	(4) β^{Girl} (s.e.)	(5) β^{Boy} (s.e.)	(6) $\beta^{Girl} = \beta^{Boy}$ p-value
Mother's physical health	0.07 (0.07)	0.07 (0.08)	0.700	0.15 (0.09)	−0.03 (0.13)	0.185
Husband's income trajectory	−0.02 (0.10)	−0.04 (0.10)	0.715	−0.14 (0.12)	0.04 (0.13)	0.199
Relationship quality	0.14 (0.09)	0.16 (0.09)	0.234	0.16 (0.13)	0.22 (0.12)	0.678
Grandmother trajectory	0.34 (0.08)	0.16 (0.07)	0.145	0.13 (0.10)	0.19 (0.11)	0.664
Social support trajectory	0.58 (0.10)	0.52 (0.09)	0.021	0.53 (0.10)	0.57 (0.11)	0.731

Notes: N=585. Summary indices are normalized to be mean 0 and SD 1 in the control group, with positive values are associated with more favorable outcomes. Index construction is described in the text (Section III.A). Standard errors, clustered at the level of randomization, in parentheses. All models control for interview date and interviewer. Specifications with all controls additionally adjust for baseline characteristics (all centered and interacted with the treatment indicator). The set of baseline characteristics include mother's age and its square, parity, family structure, presence of grandmother (mother or mother-in-law of depressed mother), mother's education, father's education, if mother was employed, if mother empowered, PCA-weighted wealth index, depression severity (Hamilton score), and perceived social support (MSPSS). Inference is conducted using randomization inference p-values, clustered at the level of randomization (Young, 2018). RI p-values are adjusted to control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993).

[15]These results should interpreted with the caveat that grandmother presence was not balanced at baseline, particularly for mothers of boys. We do, however, control for it.

These results suggest that improved social support within the household, either through a better relationship with the husband or asking grandmothers for help, might be a mechanism underlying the effectiveness of this CBT intervention. However, there is no clear evidence of heterogeneity in the presence of grandmothers or in perceived social support by gender of the index child.

HETEROGENEITY IN TREATMENT EFFECTS BY OTHER CHARACTERISTICS

Another way of shedding light onto potential mechanisms is to look for subgroups displaying stronger treatment effects. As mentioned above, we find evidence that the average treatment effects are driven by mothers who were pregnant with a girl at the start of the intervention. This is consistent with the fact that mothers of sons in the control group might have a more supportive household environment, and the CBT intervention acts to bridge this gap. Here we also consider heterogeneity by the education of the mother, her age, whether the index child was the first child, a household wealth index, and the presence of the grandmother at baseline (Online Appendix Tables G.22–G.24).

First, we explore heterogeneity in women's depression trajectories. Treatment effects at both 1- and 7-year follow-ups are significantly larger among women who, at baseline, did not have a grandmother of the index child in the household. Among women who had grandmothers present at baseline, control mothers caught up to the treated mothers by the 7-year follow-up. That means the treatment only sped up their recovery. By contrast, the intervention led to a 10 percentage point reduction in depression rates at the 7-year follow-up among women in households where the grandmother was absent. This suggests that CBT may have been more effective for women who lack support within the household, especially in the long run. There is no significant heterogeneity of treatment effects on depression by the other characteristics.

Turning to other outcomes, the pattern is less clear. Treatment effects on women's financial empowerment are significantly larger among women without a grandmother in the household at baseline, but also among older women and women with more education. There are no significant differences in treatment effects by wealth, or by whether the child in the womb at baseline was the first child. Treatment effects on money and time investments, parenting style, and fertility do not vary significantly with any of the stated characteristics, with the exception that treatment impacts on parenting style are significantly larger in wealthier households.

E.   Discussion: Exploring Results for Child Development

In light of the growing evidence of the longer term benefits of an improved early life environment (Campbell et al., 2014; Almond, Currie and Duque, 2017), we find it surprising that the intervention did not significantly influence child development, especially considering its effects on maternal depression, financial

empowerment, and parental investments. Here we investigate several potential hypotheses for this puzzling result.

First, improvements in child development could have occurred earlier, but have faded by the 7-year follow-up. Fading has been noted in other early childhood interventions, for instance in Chetty et al. (2011), Heckman, Pinto and Savelyev (2013) and Andrew et al. (2018). Fading seems unlikely in our setting: looking at data from the 6-month and 1-year follow-ups, Rahman et al. (2008) find very little evidence of better child development (no sizable effect on anthropometrics, slightly slower probability of diarrhea episodes). There is also no evidence consistent with catch-up or compensating investments on the part of the control group: if anything, we find higher parental investments and support from the grandmother in the treatment arm during infancy and at 7 years, consistent with reinforcing behavior rather than compensation.

Second, differential survival might bias the estimates toward zero if the marginal child surviving due to the treatment is negatively selected and accumulates lower human capital. However, we find no evidence of differential survival due to the treatment, or fatter left-tails in the distributions of skills for treated children (Online Appendix Tables H.38 and I.3).[16]

Third, the measurements of child development we use might not be sensible constructs for this population. However, we find positive and significant correlations of different items within each index, and across indices of child development. More importantly, regressions of each index on baseline family characteristics reveal the expected associations with wealth and education (Online Appendix Tables B.6 and B.7), as well as with measures of parental inputs (Online Appendix Table E.16). These positive and statistically significant correlations indicate that these indices have the potential to capture relevant variation in this setting. Another possible concern is that not all of the relevant domains of child development were measured. However we measured health, cognitive, and socio-emotional skills using state-of-the-art inventories with numerous questions, so this seems unlikely.

Another hypothesis is that experimenter demand effects or self-reporting issues might bias our results: treated mothers might have reported better outcomes to satisfy the experimenters, or might have been more attuned to the disposition of their child and been better at detecting negative outcomes. However, short-term depression was assessed by clinic psychiatrists trained to ask questions in a way such that Hawthorne effects do not mask the true condition. Additionally, although socio-emotional development was mother-reported, the indices of cognitive and physical development include several objective measurements, and we are able to detect significant effects for a number of interviewer-measured investment outcomes like school quality, learning material, and mother-child interactions in the HOME inventory (noting that interviewers were blind to the treatment status

---

[16]We found some evidence of intervention effects only on sibling survival, which did not stand up the multiple inference adjustment, and it was restricted to girls (Online Appendix Table H.38). In contrast, our failure to find intervention effects on child development is similar for boys and girls.

of mothers).

In summary, fading of treatment effects, differential survival of weaker children, imprecise measures of child development, and experimenter demand effects are unlikely to explain the null findings on child development as these explanations are undermined by features of the data. We are unable to rule out two possible explanations for the lack of treatment effects on child development. One is that, although additional time and money investments were made by mothers in the treated group, these changes induced by the intervention were not effective at promoting child development. For instance, time spent together may not have included sufficient stimulation (Attanasio et al., 2017), or monetary investments including learning materials may not have been age-appropriate.[17]

A second explanation is the lack of statistical power. While we can confidently rule out impacts larger than a quarter of a standard deviation for cognitive development or larger than a third of a standard deviation for physical development, we are not powered to detect smaller changes.[18] It is possible that child outcomes between treated and control arm will diverge over time (Cunha and Heckman, 2007; Heckman and Mosso, 2014; Doyle et al., 2017; Almond, Currie and Duque, 2017), and thus might become evident at later ages even if not yet evident at age 7. Verifying this explanation requires longer-term follow-up data.

VI. Conclusion

We evaluate the impact of treating perinatal depression on women's mental health and economic decision-making by leveraging exogenous variation in depression treatment generated by a cluster randomized control trial. The intervention, one of the largest psychotherapy trials to date, provided cognitive behavioral therapy (CBT) to 903 clinically depressed pregnant women in rural Pakistan. Public sector health workers were trained to deliver CBT, so the intervention was scalable and cost only US$10 per woman. Seven years after the CBT intervention concluded, we identify large and persistent impacts on women's mental health, financial empowerment, and their parenting investments, especially for mothers of girls.

To our knowledge, this is the first evidence that a low-cost, scalable, and one-off CBT-based intervention offers significant protection against the chronicity of

---

[17]Analysis of the Preschool Curriculum Evaluation Research Initiative Study (PCER, 2008) by Duncan et al. (2017) shows that targeting skills is more effective for learning outcomes than the usual "whole child" approach. Attanasio et al. (2017) find cognitive impacts of an intervention in Colombia that was designed to stimulate children and their estimates suggest that the parenting intervention was key.

[18]Other studies using administrative data to asses child developmental gains from early life interventions have found smaller effects than this. For example, Bharadwaj et al. (2013) find that improved neonatal care for low birth weight babies increases their academic performance by 0.15-0.22 SD in Chile and Norway respectively. Bharadwaj, Lundborg and Rooth (2017) estimate that a 10% difference in birth weight between twins in Chile increases outcomes in math and language scores by 0.04-0.06 SD. Examining twins' birth weight in Florida, Figlio et al. (2014) estimate that the heavier twin scores on average 0.05 SD better than the lighter twin. Bhalotra, Karlsson and Nilsson (2017) estimate the impacts of a postnatal health intervention in 1930s Sweden on cognitive performance in primary school of 0.11 SD.

depression. Depression is commonly addressed using antidepressants. Given the evidence that patients who recover using antidepressants are at risk of relapse if they stop, drug therapy needs to be maintained. This requires continuing expenditure and tolerance of side effects which may be of particular concern for women in the active childbearing years. Furthermore, antidepressants may not be available in many low resource settings. The CBT-based intervention we discuss was a one-off process, delivered at a relatively low cost, and with few known adverse side effects. Our finding that impacts persist well beyond the end of therapy implies that the benefit-cost ratio for such interventions is higher than is commonly recognized.

The value of an effective treatment for depression is further reinforced by its positive influence on women empowerment and economic decision making. Since depression has been linked to worse economic outcomes, which in turn may induce stress and exacerbate depression, our results suggest that this vicious cycle might be broken. There is extraordinary potential for CBT-based interventions to improve not just psychological well-being but also economic well-being.

Although our findings are of particular interest for developing countries where fertility, poverty, and the incidence of depression are high, and women's financial empowerment is low, the behavioral effects are likely to have wider relevance. Our findings suggest that treating maternal depression may have persistent economic impacts on women and be a factor in the intergenerational transmission of inequality. The results in this paper provide new evidence to motivate greater policy investment in recognizing and addressing maternal depression.

REFERENCES

Ahmad, Nuzhat, and Huma Khan. 2016. "Measuring women's disempowerment in agriculture in Pakistan." *IFPRI discussion paper.*

Aizer, Anna, Laura Stroud, and Stephen Buka. 2009. "Maternal Stress and Child Well-Being: Evidence from Siblings." 1–49.

Almas, Ingvild, Alex Armand, Orazio Attanasio, and Pedro Carneiro. 2015. "Measuring and Changing Control: Women's Empowerment and Targeted Transfers." *NBER working paper*, 250415.

Almond, Douglas, and Bhashkar Mazumder. 2013. "Fetal Origins and Parental Responses." *Annual Review of Economics*, 5(1): 37–56.

Almond, Douglas, and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives*, 25(3): 153–172.

Almond, Douglas, Janet Currie, and Valentina Duque. 2017. "Childhood Circumstances and Adult Outcomes: Act II." *NBER Working Paper.*

Almond, Douglas, Lena Edlund, and Kevin Milligan. 2013. "Son Preference and the Persistence of Culture: Evidence from South and East Asian

Immigrants to Canada." *Population and Development Review*, 39(March): 75–95.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders.* American Psychiatric Association.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

Andrew, Alison, Orazio Attanasio, Emla Fitzsimons, Sally Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina. 2018. "Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia." *PLoS Medicine*, 15(4).

Attanasio, O. P., C. Fernandez, E. O. a. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. 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*, 349(sep29 5): g5785–g5785.

Attanasio, Orazio P, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2017. "Estimating the production function for human capital Results from a randomized control trial in Colombia." *Working Paper.*

Autor, David H, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2016. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *NBER Working Paper.*

Baird, Sarah, and Berk Özler. 2016. "Sustained Effects on Economic Empowerment of Interventions for Adolescent Girls: Existing Evidence and Knowledge Gaps CGD Background Paper." *CGD Background Paper*, November: 22.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa." *Working Paper.*

Beck, Aaron T. 1967. *Depression: Causes and Treatment.* Philadelphia:University of Pennsylvania Press.

Beck, Aaron T. 1979. *Cognitive therapy and the emotional disorders.* Penguin.

Bhalotra, Sonia, Adeline Delavande, Paulino Font, and Joanna Maselko. 2017. "The impact of early cognitive and non-cognitive skills on later outcomes." *Mimeo.*

Bhalotra, Sonia, and Irma Clots-Figueras. 2014. "Health and the Political Agency of Women." *American Economic Journal: Economic Policy*, 6(2): 164–197.

Bhalotra, Sonia, I Clots-Figueras, and L Iyer. 2017. "Religion and Abortion: The Role of Politician Identity." *Mimeo.*

Bhalotra, Sonia, Martin Karlsson, and Therese Nilsson. 2017. "Infant Health and Longevity: Evidence from A Historical Intervention in Sweden." *Journal of the European Economic Association*, 29(5): 208–230.

Bhalotra, Sonia Radhika, Abhishek Chakravarty, Selim Gulesci, Sonia Bhalotra, Abhishek Chakravarty, and Selim Gulesci. 2016. "The Price of Gold: Dowry and Death in India." *IZA Discussion Paper.*

Bhalotra, Sonia R., and Atheendar Venkataramani. 2011. "The Captain of the Men of Death and His Shadow: Long-Run Impacts of Early Life Pneumonia Exposure." *IZA Discussion Papers.*

Bhalotra, Sonia, Uma Kambhampati, Samantha Rawlings, and Zahra Siddique. 2018. "Intimate Partner Violence and the Business Cycle." *IZA Discussion Paper Series*, , (11274).

Bharadwaj, Prashant, Katrine Vellesen Løken, Christopher Neilson, and Katrine Vellesen Løken. 2013. "Early life health interventions and academic achievement." *American Economic Review*, 103(6864): 1862–1891.

Bharadwaj, Prashant, Petter Lundborg, and Dan-Olof Rooth. 2017. "Birth Weight in the Long Run." *Journal of Human Resources*, 0715–7235R.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2016. "Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes." *American Economic Journal: Applied Economics*, 8(1): 193–223.

Blattman, Christopher, Julian C Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review*, 107(4): 1165–1206.

Bolier, Linda, Merel Haverman, Gerben Westerhof, Heleen Riper, Filip Smit, and Ernst Bohlmeijer. 2013. "Positive psychology interventions: a meta-analysis of randomized controlled studies." *BMC public health*, 13: 119.

Bolton, Paul, Judith Bass, Richard Neugebauer, Helen Verdeli, Kathleen F Clougherty, Priya Wickramaratne, Liesbeth Speelman, Lincoln Ndogoni, and Myrna Weissman. 2003. "Group Interpersonal Psychotherapy for Depression in Rural Uganda. A Randomized Controlled Trial." *JAMA*, 289(23): 3117.

Buttorff, Christine, Rebecca Hock, Helen Weiss, Smita Naik, Ricardo Araya, Betty Kirkwood, Daniel Chisholm, and Vikram Patel. 2012. "Economic evaluation of a task-shifting intervention for common mental disorders in India." *Bulletin of the World Health Organization*, 90(11): 813–821.

Buvinic, Mayra, and Rebecca Furst-nichols. 2016. "Promoting Women's Economic Empowerment." *World Bank Policy Research Working Paper*, 7087.

Caldwell, B., and R. Bradley. 1984. *Home Observation for Measurement of the Environment (HOME)-revised edition.* Little Rock, AR:University of Arkansas, Little Rock.

Campbell, Frances A, Gabriella Conti, James J Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth P Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science*, 343(6178): 1478–1485.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from project star." *Quarterly Journal of Economics*, 126(November): 1593–1660.

Cohen, Robert M, Herbert Weingartner, Sheila A Smallberg, David Pickar, and Dennis L Murphy. 1982. "Effort and Cognition in Depression." *Archives of General Psychiatry*, 39(5): 593.

Cooper, Peter J, Mark Tomlinson, Leslie Swartz, Mireille Landman, Chris Molteno, Alan Stein, Klim McPherson, and Lynne Murray. 2009. "Improving quality of mother-infant relationship and infant attachment in socioeconomically deprived community in South Africa: randomised controlled trial." *BMJ (Clinical research ed.)*, 338: b974.

Cuijpers, Pim, Annemieke van Straten, Gerhard Andersson, and Patricia van Oppen. 2008*a*. "Psychotherapy for depression in adults: A meta-analysis of comparative outcome studies." *Journal of Consulting and Clinical Psychology*, 76(6): 909–922.

Cuijpers, Pim, Annemieke van Straten, Lisanne Warmerdam, Gerhard Andersson, C Munizza, and IW Miller. 2008*b*. "Are individual and group treatments equally effective in the treatment of depression in adults?: A meta-analysis." *The European Journal of Psychiatry*, 22(1): 38–51.

Cuijpers, Pim, Eirini Karyotaki, Erica Weitz, Gerhard Andersson, Steven D Hollon, and Annemieke van Straten. 2014. "The effects of psychotherapies for major depression in adults on remission, recovery and improvement: A meta-analysis." *Journal of Affective Disorders*, 159: 118–126.

Cuijpers, Pim, Steven D Hollon, Annemieke van Straten, Claudi Bockting, Matthias Berking, and Gerhard Andersson. 2013. "Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis." *BMJ Open*, 3(4): e002542.

Cunha, Flavio. 2017. "Gaps in Early Investments in Children." *Working Paper.*

Cunha, Flavio, and James J Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97(2): 31–47.

Currie, Janet, and Brigitte C Madrian. 1999. "Health, health insurance and the labor market." *Handbook of labor economics*, 3: 3309–3416.

Dalton, Patricio, Sayantan Ghosal, and Anandi Mani. 2010. "Poverty and aspirations failure: A theoretical framework." Working paper, Tilburg University and University of Warwick.

Den Hartog, H M, M M A Derix, A L Van Bemmel, B Kremer, and J Jolles. 2003. "Cognitive functioning in young and middle-aged unmedicated out-patients with major depression: testing the effort and cognitive speed hypotheses." *Psychological Medicine*, 33(8): 1443–1451.

Dennhardt, Ashley A., and James G. Murphy. 2011. "Associations between depression, distress tolerance, delay discounting, and alcohol-related problems in European American and African American college students." *Psychology of Addictive Behaviors*, 25(4): 595–604.

De Quidt, Jonathan, and Johannes Haushofer. 2016. "Depression for Economists." In *The Economics of Asset Accumulation and Poverty Traps.*, ed. Christopher B Barrett, Michael Carter and Jean-Paul Chavas, Chapter 3. University of Chicago Press.

DeRubeis, Robert J, Greg J Siegle, and Steven D Hollon. 2008. "Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms." *Nature Reviews Neuroscience*, 9(10): 788–796.

Doyle, Orla, Colm Harmon, James J Heckman, Caitriona Logue, and Seong Hyeok Moon. 2017. "Early skill formation and the efficiency of parental investment: A randomized controlled trial of home visiting." *Labour Economics*, 45: 40–58.

Duflo, Esther. 2003. "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *The World Bank Economic Review*, 17(>1): 1–25.

Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature*, 50(4): 1051–1079.

Duncan, Greg J, Jade Marcus Jenkins, Anamarie Auger, Marianne Bitler, Thurston Domina, and Margaret Burchinal. 2017. "Boosting School Readiness: Should Preschools Target Skills or the Whole Child?" *Working Paperk.*

Efron, B, and R J Tibshirani. 1994. *An Introduction to the Bootstrap.* Vol. 57, CRC Press.

Fava, Giovanni A., Chiara Rafanelli, Grandi Silvana, Renzo Canestrari, and Murray A. Morphy. 1998. "Six-year outcome for cognitive behavioral treatment of residual symptoms in major depression." *American Journal of Psychiatry*, 155(10): 1443–1445.

Ferrari, Alize J., Fiona J. Charlson, Rosana E. Norman, Scott B. Patten, Greg Freedman, Christopher J.L. Murray, Theo Vos, and Harvey A. Whiteford. 2013. "Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010." *PLoS Medicine*, 10(11): e1001547.

Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore. 2016. "On Her Account: Can Strengthening Women's Financial Control Boost Female Labor Supply?" *Working Paper.*

Field, Thomas A, Eric T Beeson, and Laura K Jones. 2015. "The New ABCs: A Practitioner's Guide to Neuroscience-Informed Cognitive-Behavior Therapy." *Journal of Mental Health Counseling*, 37(3): 206–220.

Figlio, David N, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review*, 104(12): 3921–3955.

Fisher, Jane, Meena Cabral de Mello, Vikram Patel, Atif Rahman, Thach Tran, Sara Holton, and Wendy Holmesf. 2012. "Prevalence and determinants of common perinatal mental disorders in women in low-and lower-middle-income countries: A systematic review." *Bulletin of the World Health Organization*, 90(2): 139–149.

Frijters, Paul, David W Johnston, and Michael A Shields. 2014. "The effect of mental health on employment: evidence from Australian panel data." *Health Economics*, 23(9): 1058–1071.

Gao, Ling-ling, Sally Wai Chi Chan, Xiaomao Li, Shaoxian Chen, and Yuantao Hao. 2010. "Evaluation of an interpersonal-psychotherapy-oriented childbirth education programme for Chinese first-time childbearing women: A randomised controlled trial." *International Journal of Nursing Studies*, 47: 1208–1216.

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(10): 973–982.

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

Grossman, Michael. 1972. "On the concept of health capital and the demand for health." *Journal of Political Economy*, 80(2): 223–255.

Heckman, James J, and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics*, 6(1): 689–733.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, 24: 411–482.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes." *American Economic Review*, 103(6): 2052–2086.

Heller, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *The Quarterly Journal of Economics*, 132(1).

Imbens, Guido W, and Charles F Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica*, 72(6): 1845–1857.

Imhoff, Sarah, Millie Harris, Jason Weiser, and Brady Reynolds. 2014. "Delay discounting by depressed and non-depressed adolescent smokers and non-smokers." *Drug and Alcohol Dependence*, 135: 152–155.

Karlan, Dean, Beniamino Savonitto, Bram Thuysbaert, and Christopher Udry. 2017. "Impact of savings groups on the lives of the poor." *Proceedings of the National Academy of Sciences*, 114(12): 3079–3084.

Kessler, Ronald C. 2012. "The costs of depression." *The Psychiatric clinics of North America*, 35(1): 1–14.

Kessler, Ronald C, and Richard G Frank. 1997. "The impact of psychiatric disorders on work loss days." *Psychological medicine*.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Kung, Claryn S J, David W Johnston, and Michael A Shields. 2018. "Mental Health and the Response to Financial Incentives:Evidence from a Survey Incentives Experiment." *Journal of Health Economics*, 62: 84–94.

Lara, Ma A., Claudia Navarro, and Laura Navarrete. 2010. "Outcome results of a psycho-educational intervention in pregnancy to prevent PPD: A randomized control trial." *Journal of Affective Disorders*, 122: 109–117.

Lavy, Victor, Giulia Lotti, and Zizhong Yan. 2016. "Empowering Mothers and Enhancing Early Childhood Investment: Effect on Adults Outcomes and Children Cognitive and Non-cognitive Skills." *Working Paper*, NBER 22963.

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

Lempert, Karolina M., and Diego A. Pizzagalli. 2010. "Delay discounting and future-directed thinking in anhedonic individuals." *Journal of Behavior Therapy and Experimental Psychiatry*, 41(3): 258–264.

Lerner, Debra, and Rachel Mosher Henke. 2008. "What Does Research Tell Us About Depression, Job Performance, and Work Productivity?" *Journal of Occupational and Environmental Medicine*, 50(4): 401–410.

Lerner, Debra, David A. Adler, Hong Chang, Leueen Lapitsky, Maggie Y. Hood, Carla Perissinotto, John Reed, Thomas J. McLaughlin, Ernst R. Berndt, and William H. Rogers. 2004. "Unemployment, Job Retention, and Productivity Loss Among Employees With Depression." *Psychiatric Services*, 55(12): 1371–1378.

Lundberg, Shelly J, Robert A Pollak, and Terence J Wales. 1997. "Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit." *The Journal of Human Resources*, 32(3): 463.

MacLeod, Andrew K., and Eva Salaminiou. 2001. "Reduced positive future-thinking in depression: Cognitive and affective factors." *Cognition & Emotion*, 15(1): 99–107.

Maselko, Joanna, Siham Sikander, Sonia Bhalotra, Omer Bangash, Nima Ganga, Satadru Mukherjee, Helen Egger, Lauren Franz, Amina Bibi, Rakhshanda Liaqat, Misbah Kanwal, Tayyaba Abbasi, Maryam Noor, Nida Ameen, and Atif 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." *The Lancet Psychiatry*, 2(7): 609–17.

Milazzo, Annamaria. 2014. "Why are adult women missing? Son preference and maternal survival in India." *World Bank Working Paper*, 1–53.

Miller, Barbara D. 1981. *The Endangered Sex: Neglect of Female Children in Rural North India.* Ithaca, NY:Cornell University Press.

Mintz, Jim, Lois Imber Mintz, Mary Jane Arruda, and Sun Sook Hwang. 1992. "Treatments of Depression and the Functional Capacity to Work." *Archives of General Psychiatry*, 49(10): 761.

Mirowsky, John. 1985. "Depression and Marital Power: An Equity Model." *American Journal of Sociology*, 91(3): 557–592.

O'Brien, P C. 1984. "Procedures for comparing samples with multiple endpoints." *Biometrics*, 40(4): 1079–1087.

Patel, Vikram, Benedict Weobong, Helen A Weiss, Arpita Anand, Bhargav Bhat, Basavraj Katti, Sona Dimidjian, Ricardo Araya, Steve D Hollon, Michael King, Lakshmi Vijayakumar, A. La Park, David McDaid, Terry Wilson, Richard Velleman, Betty R Kirkwood, and Christopher G Fairburn. 2017. "The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial." *The Lancet*, 389(10065): 176–185.

Persson, Petra, and Maya Rossin-Slater. 2017. "Family Ruptures, Stress, and the Mental Health of the Next Generation." *American Economic Review.*

Rahman, Atif. 2007. "Challenges and opportunities in developing a psychological intervention for perinatal depression in rural Pakistan – a multi-method study." *Archives of Women's Mental Health*, 10(5): 211–219.

Rahman, Atif, Abid Malik, Siham Sikander, Christopher Roberts, and Francis Creed. 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, Atif, Siham Sikander, Abid Malik, Ikhlaque Ahmed, Barbara Tomenson, and Francis Creed. 2012. "Effective treatment of perinatal depression for women in debt and lacking financial empowerment in a low-income country." *The British Journal of Psychiatry*, 201(6): 451.

Rahman, A, Z Iqbal, and R Harrington. 2003. "Life events, social support and depression in childbirth: perspectives from a rural community in the developing world." *Psychological medicine*, 33(7): 1161–1167.

Ronda, Victor. 2016. "The Effect of Maternal Psychological Distress on Children's Cognitive Development." *Working Paper.*

Sathar, Zeba A, Gul Rashida, Sabahat Hussain, and Anushe Hassan. 2015. "Evidence of Son Preference and Resulting Demographic and Health Outcomes in Pakistan." *Islamabad Population Council.*

Shi, Riyi, and Lori L Altshuler. 2004. "Postpartum Depression: Risk Factors and Treatment Options." *Psychiatric Times*, October 1.

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

Takahashi, Taiki, Hidemi Oono, Takeshi Inoue, Shuken Boku, Yuki Kako, Yuji Kitaichi, Ichiro Kusumi, Takuya Masui, Shidn Nakagawa, Katsuji Suzuki, Teruaki Tanaka, Tsukasa Koyama, and Mark H B Radford. 2008. "Depressive patients are more impulsive and inconsistent in intertemporal choice behavior for monetary gain and loss than healthy subjects–an analysis based on Tsallis' statistics." *Neuro endocrinology letters*, 29(3): 351–8.

Tolin, David F. 2010. "Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review." *Clinical Psychology Review*, 30(6): 710–720.

Trivers, Robert L, and Dan E Willard. 1973. "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring." *Science*, 179(4068).

Vos, Theo, Abraham D Flaxman, Mohsen Naghavi, Rafael Lozano, Catherine Michaud, and Majid et al Ezzati. 2012. "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010." *The Lancet*, 380(9859): 2163–2196.

Walker, S, S Chang, C Powell, and S M Grantham-McGregor. 2005. "Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-retarded Jamaican children: a prospective cohort study." *The Lancet*, 366(9499): 1804–1807.

Webster-Stratton, Carolyn, M. Jamila Reid, and Mary Hammond. 2001. "Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start." *Journal of Clinical Child and Adolescent Psychology*, 30(3): 283–302.

Westfall, Peter H, and S Stanley Young. 1993. *Resampling-based multiple testing: examples and methods for P-value adjustment.* John Wiley & Sons.

Wiles, Nicola J, Laura Thomas, Nicholas Turner, Kirsty Garfield, Daphne Kounali, John Campbell, David Kessler, Willem Kuyken, Glyn Lewis, Jill Morrison, Chris Williams, Tim J Peters, and Sandra Hollinghurst. 2016. "Long-term effectiveness and cost-effectiveness of

cognitive behavioural therapy as an adjunct to pharmacotherapy for treatment-resistant depression in primary care: follow-up of the CoBalT randomised controlled trial." *The Lancet Psychiatry*, 3(2): 137–144.

World Health Organization. 2013. *Neonatal and child health country profile: Pakistan,.* World Health Organization.

Young, Alwyn. 2018. "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *The Quarterly Journal of Economics.*

Zaidi, Batool, and S Philip Morgan. 2016. "In the Pursuit of Sons: Additional Births or Sex-Selective Abortion in Pakistan?" *Population and Development Review*, 42(4): 693–710.

APPENDIX

TABLE A1—CHARACTERISTICS OF ATTRITORS AND DIFFERENCES BY INTERVENTION AND CONTROL CLUSTERS

	Characteristics of attritors				Attritor characteristics by treatment arm			
	(1) In sample Mean	(2) Attritor Mean	(3) Diff	(4) p-val	(5) T Mean	(6) C Mean	(7) Diff	(8) p-val
Mother's age	26.87	26.50	−0.37	0.30	26.14	26.94	−0.81	0.19
Mother's height (m)	1.56	1.56	−0.01	0.09	1.56	1.56	0.00	0.94
Mother's BMI	23.18	23.40	0.22	0.43	23.40	23.40	−0.00	1.00
Mother's education	4.06	4.06	0.01	0.98	3.94	4.21	−0.27	0.58
Empowered	0.56	0.48	−0.07	0.03	0.48	0.49	−0.01	0.87
Mother usually works	0.02	0.04	0.03	0.03	0.03	0.06	−0.02	0.36
Parity	2.25	2.22	−0.04	0.76	2.28	2.14	0.14	0.51
First child	0.17	0.22	0.05	0.08	0.21	0.23	−0.02	0.72
Index child is female	0.51	0.49	−0.01	0.69	0.48	0.51	−0.02	0.67
Share children female	0.53	0.52	−0.01	0.55	0.54	0.50	0.04	0.27
Depression score	14.49	14.88	0.39	0.17	15.11	14.61	0.50	0.29
Disability score	8.12	8.36	0.23	0.21	8.15	8.61	−0.46	0.11
General functioning	62.25	61.68	−0.57	0.12	61.21	62.24	−1.02	0.08
Perceived social support	46.01	43.26	−2.75	0.02	42.63	44.03	−1.40	0.45
Joint/extended family	0.59	0.58	−0.01	0.71	0.58	0.58	0.00	0.94
Mother-in-law present	0.43	0.44	0.00	0.93	0.45	0.42	0.02	0.66
Mother's mother present	0.06	0.07	0.01	0.73	0.08	0.06	0.02	0.39
Father's education	7.09	6.89	−0.20	0.48	6.91	6.86	0.05	0.91
Father employed	0.90	0.92	0.02	0.29	0.91	0.94	−0.03	0.31
Not in manual labor	0.29	0.28	−0.01	0.78	0.27	0.30	−0.03	0.60
SES (0=poor, 4=rich)	1.41	1.33	−0.08	0.22	1.24	1.43	−0.19	0.09
Wealth index	0.07	−0.13	−0.20	0.15	−0.20	−0.04	−0.17	0.47
Reason for LTFU:								
Abortion					0.01	0.02	−0.01	0.51
Stillbirth					0.07	0.12	−0.05	0.13
Child death/illness					0.20	0.14	0.06	0.18
Mother death/illness					0.07	0.10	−0.03	0.36
Refused					0.25	0.28	−0.03	0.52
Moved					0.36	0.33	0.03	0.59
Joint test (p-value)								0.59
N	585	318	903		174	144	318	

Notes: Table shows baseline characteristics and their differences for women who were lost to follow-up between baseline and year 7. Columns 1-4 compare the 7-year follow-up sample to attritors. Columns 5-8 compare baseline characteristics of attritors by treatment arm, including the reasons respondents were lost to follow-up. The last row reports the *p*-value of joint test that all attritors from the treatment arm were different to attritors from the control arm along baseline characteristics.

TABLE A2—BALANCE BY GENDER

	Baseline Sample - Girls N=384				Baseline Sample - Boys N=377			
	(1) Control Mean	(2) SD	(3) T-C Diff	(4) p-val	(5) Control Mean	(6) SD	(7) T-C Diff	(8) p-val
Mother's age	27.40	5.1	−0.83	0.10	26.62	5.0	−0.13	0.79
Mother's height (m)	1.56	0.1	0.02	0.01	1.56	0.1	−0.01	0.25
Mother's BMI	23.11	4.3	0.45	0.29	23.37	3.9	−0.20	0.62
Mother's education	3.62	3.8	0.68	0.09	3.92	3.9	0.57	0.17
Empowered	0.57	0.5	−0.04	0.41	0.52	0.5	−0.03	0.58
Mother usually works	0.03	0.2	−0.01	0.38	0.03	0.2	−0.00	0.81
Parity	2.45	1.7	−0.27	0.12	2.32	1.9	−0.29	0.12
First born	0.17	0.4	0.02	0.58	0.18	0.4	0.02	0.66
Share children female	0.55	0.3	−0.05	0.10	0.51	0.3	0.04	0.21
Depression score	14.85	4.0	−0.18	0.67	14.07	3.9	0.75	0.07
Disability score	8.25	2.8	−0.08	0.77	8.29	2.7	−0.25	0.37
General functioning	61.64	5.1	0.60	0.26	62.22	5.4	−0.13	0.82
Perceived social support	44.64	16.1	1.32	0.43	44.22	16.3	1.86	0.27
Joint/extended family	0.55	0.5	0.01	0.78	0.57	0.5	0.08	0.13
Mother-in-law present	0.40	0.5	0.07	0.20	0.39	0.5	0.09	0.08
Mother's mother present	0.06	0.2	−0.00	0.96	0.04	0.2	0.06	0.03
Father's education	7.36	3.8	−0.12	0.75	7.04	4.0	−0.17	0.68
Father employed	0.93	0.3	−0.04	0.18	0.88	0.3	0.02	0.63
Not in manual labor	0.29	0.5	0.05	0.37	0.30	0.5	−0.05	0.31
SES (0=poor, 4=rich)	1.35	1.0	0.07	0.47	1.34	1.0	0.09	0.38
Wealth index	0.03	1.9	0.12	0.56	−0.19	2.0	0.22	0.27
Joint test (p-value)				0.81				0.04

Notes: A total of 454 girls and 449 boys were born to trial women. We know the counts of all children born by gender and treatment status. However, child gender was not recorded in the dataset for a subset of attritors (45%). Thus, to evaluate baseline characteristics bz gender, the sample is reduced by 85 percent for each gender (384/454 girls and 377/449 boys).