I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11187

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

Victoria Baranov
Sonia Bhalotra
Pietro Biroli
Joanna Maselko

NOVEMBER 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11187

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

Victoria Baranov
University of Melbourne

Sonia Bhalotra
ISER, University of Essex and IZA

Pietro Biroli
University of Zurich and IZA

Joanna Maselko
University of North Carolina, Chapel Hill

NOVEMBER 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

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

We evaluate the long-term impact of treating maternal depression on women's financial empowerment and parenting decisions. We leverage experimental variation induced by a cluster-randomized control trial that provided psychotherapy to perinatally depressed mothers in rural Pakistan. It was one of the largest psychotherapy interventions in the world, and the treatment was highly successful at reducing depression. We locate mothers seven years after the end of the intervention to evaluate its long-run effects. We find that the intervention increased women's financial empowerment, increasing their control over household spending. Additionally, the intervention increased both time- and monetary-intensive parental investments, with increases in investments tending to favor girls.

JEL Classification: I15, I30, O15

Keywords: mental health, maternal depression, women's labor supply, empowerment, early life, parenting, child development, randomized controlled trial, Pakistan

Corresponding author:
Victoria Baranov
Department of Economics
University of Melbourne
111 Barry Street
Carlton VIC 3053
Australia

E-mail: victoria.baranov@unimelb.edu.au

* The authors thank Grand Challenges Canada for funding this research under their Saving Brains program Grant Number #0072-03 and the Centre for Microsocial Change at ISER, Essex for further support from the ESRC centre grant ES/L009153/1. We are grateful to Siham Sikander and his colleagues at the Human Development Research Foundation in Pakistan for implementing the original trial and the follow-up. Atif Rahman designed the intervention. We are grateful to Atif Rahman and to Siham Sikander for sharing the baseline and early follow-up data with us and for providing several insights. We received helpful feedback from Prashant Bharadwaj, David Byrne, Flavio Cunha, Alex Frankel, James Fenske, James Heckman, Anandi Mani, Satadru Mukherjee, Kjell Salvanes, Agne Suziedelyte, Marcos Vera-Hernandez, and seminar participants at Oxford, IFS, Essex, Alicante, Odense, HCEO Chicago, University of Michigan, University of Hawaii, University of Sydney, and conference participants at CSAE, PacDev, NEUDC, ASSA, and RES. All errors and opinions are our own.

NON-TECHNICAL SUMMARY

Major Depressive Disorder is the most common mental health problem and the single most important contributor to Years Lived With Disability, estimated to affect 13 percent of the global population. The burden of depression is about twice as high for women, for whom it is the leading source of disease burden. Women of child-bearing age are especially at risk due to higher rates of perinatal depression, a depressive episode around the time of childbirth. Perinatal depression rates are estimated at 12-20% in OECD countries and 20-35% in poorer countries, where the condition often goes undiagnosed and hence untreated.

Since the onset of maternal depression is often associated with economic and social disadvantage, it is hard to investigate its causal effects. We overcame this limitation by leveraging experimental variation in depression generated by a cluster randomized control trial providing cognitive behavioral therapy (CBT) treatment to a sample of women identified as clinically depressed in pregnancy in one of the largest psychotherapy trials to date. The intervention was called the Thinking Healthy Program and it was implemented in 40 communities in rural Pakistan in 2005/6, involving 903 women.

The intervention has been shown to have significantly reduced depression postpartum. In this paper we analyse follow-up data gathered seven years after the end of the therapy to investigate its impacts on economic decision making. We find that the intervention increased women's financial empowerment, increasing their control over household spending. Additionally, the intervention increased both time- and monetary-intensive parental investments, with increases in investments tending to favor girls. We provide the first causal estimates of impacts of treating depression on these outcomes, using what would appear to be the longest follow-up of a psychotherapy intervention.

Our findings suggest that maternal depression is potentially critical to women's empowerment and to the intergenerational transmission of inequality. Yet it is not systematically measured, and it is not one of the new Sustainable Development Goals. The results in this paper provide new evidence to motivate greater policy investment in recognizing and addressing maternal depression. Although these findings are of particular interest in the context of developing countries where fertility rates, poverty, and the incidence of depression are high, and where women's financial empowerment is more limited, the behavioral effects are likely to have wider relevance, illustrating how depression can bias decision-making.

1 Introduction

Major Depressive Disorder is the most common mental health problem and 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, for whom it is the leading source of disease burden (DeRubeis et al., 2008). Women of childbearing age are especially at risk due to higher rates of perinatal depression, a depressive episode around the time of childbirth. Perinatal depression rates are estimated at 12-20% in OECD countries and 20-35% in poorer countries, where the condition often goes undiagnosed and hence untreated (Rahman, 2005).

We leverage experimental variation from one of the largest psychotherapy trials in the world to date to investigate the impacts of treating depression on decision-making. The intervention provided cognitive behavioral therapy (CBT) to a sample of perinatally depressed women, and was shown to significantly reduce depression (Rahman et al., 2008; Baranov et al., 2017). Our analysis focuses upon follow-up data gathered 7 years after the end of the therapy to investigate its impacts on women's financial empowerment, their parenting choices, fertility, investments in children and, also, the children's developmental outcomes. In addition to providing the first causal estimates of economic impacts of treating depression, our study allows us to evaluate impacts over the long run.

Little is known about the causal effects of depression on economic decision-making. Depression is characterized by symptoms including sadness, pessimism, and loss of agency, and it has the potential to impair productivity and hamper economic decision-making (Kessler and Frank, 1997; Currie and Madrian, 1999). Yet, the hypothesis that maternal depression might impact women's economic and financial empowerment has received surprisingly little attention. Women tend to have low levels of control over household spending, especially but not only in low-income settings.[1] Given evidence that increasing women's share of household income results in increased expenditure on children (Lundberg et al., 1997), perinatal depression may also disadvantage children. Parenting and child rearing responsibilities fall largely in the domain of women, and while a growing literature highlights the importance of parental investments in early childhood for later life outcomes (Blimpo et al., 2016; Lavy et al., 2016; Macours et al., 2015; Almond and Mazumder, 2013), there is no causal evidence that perinatal depression can influence parenting decisions.

The psychotherapy intervention, called the Thinking Healthy Program (THP), was implemented in 40 communities in rural Pakistan. All pregnant women in these communities were clinically assessed for depression, and 903 women diagnosed as depressed were included in the trial. Women in the 20 communities in the control arm received routine maternal and health services from government-funded community health workers, and women in the 20 treated communities additionally received the psychotherapy intervention. Women in both arms received 16 home visits, starting in the third trimester of pregnancy and ending when the children were 10 months old. The CBT focused on identifying and modifying cognitive

[1] Several studies show that income shares impact the distribution of consumption in the household (Browning et al., 1994) and, in rich and poor countries alike, women earn a smaller share of income than men (Bertrand et al., 2015).

distortions common in depression, using techniques of active listening, collaboration with the family, and guided discovery of healthy thinking, aided by pictorial and verbal messages. The short-term follow-up of the THP showed a reduction in depression rates of 78% within the first year postpartum in the treatment group, a 33 percentage point difference with respect to the control group (Rahman et al., 2008).[2] The THP 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 WHO as a model for other countries.[3]

Seven years after the end of the intervention, we located 585 participants, 83% of the mother-child dyads who were last surveyed when the child was a year old. We gathered rich data with multiple indicators in each of the outcome domains related to the woman's decision-making and her child's development. In our 7-year follow-up, we also recruited 300 mother-child dyads from among the women who were excluded from the trial at baseline because they did not suffer perinatal depression. This allows us to compare the long-term outcomes of women who were depressed in pregnancy, but treated with CBT, against the long-term outcomes of women who were not depressed in pregnancy.

We find that 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 treated group 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 monetary-intensive parental investments. For example, we find that the intervention lead to sending the child to better quality schools and having more learning materials in the home, both objectively measured by the interviewer. We also find a 12 percentage point increase in the probably that the child was enrolled in private school.[4] Quantile treatment effect estimates indicate that these improvements are fairly homogeneous across the distribution of investments. However, we find no discernible average differences in parenting style or fertility between women in the treated and control communities.

For the long-run outcomes modified by the intervention, treatment effects are consistently larger for mothers who were pregnant with a girl at the time of the intervention: 0.47 SD for financial empowerment, 0.48 SD for monetary investments in children and 0.24 SD for parenting style. Control group means in the 7-year follow-up reveal that mothers who were pregnant with girls are disempowered relative to mothers of boys.[5] The estimates indicate that the intervention tends to close this gender gap.

The identified treatment effects on women's empowerment and on parental investments tend to close the descriptive gaps in these outcomes between women who were not depressed at baseline and women in the control arm who were depressed at baseline. In this sense, where

[2] In a companion paper, we show persistent impacts of the CBT intervention on mental health at the 7-year mark (Baranov et al., 2017). Although these impacts are much smaller than those found at the 1-year mark, this is nevertheless new evidence that CBT can have persistent impacts on mental health and important given the chronicity of depression (DeRubeis et al., 2008; Lund et al., 2010, 2011; Shi and Altshuler, 2004).

[3] See http://www.who.int/mental_health/maternal-child/thinking_healthy/en/.

[4] Mothers also report higher expected grade attainment. Responsive components of the time-intensive index are enrichment, family companionship, and family integration.

[5] Some recent studies document this more systematically in other contexts. Using data for India, Milazzo (2014) shows that survival rates of women without sons are lower.

it made a difference, the intervention was very effective. The findings that treatment improves women's control over household spending, and that it leads to greater investments in children line up given the evidence that additional resources in the hands of women (rather than men) tend to be directed toward children (Lundberg et al., 1997; Hoddinott and Haddad, 1995).[6] Our finding that, amongst children, girls benefit differentially, may be rationalized in terms of son preference being stronger amongst men than women (Thomas, 1994; Dahl and Moretti, 2008).

Since the intervention increased investments in children and gave mothers more financial empowerment, we may have expected to find improvements in indicators of child development in treated relative to control clusters. However, we find no discernible effect on children's cognitive or socio-emotional development at age 7, and only weak evidence of improvements in physical health and survival. We confirm that these averaged results are not concealing large effects in a particular part of the distribution, and that they are not the result of compensating investments in the control arm, nor do they appear to be the result of using measures of development that are inappropriate for the context. Comparing children of depressed women in the control arm to children of women who were not prenatally depressed reveals limited differences in outcomes for physical growth and cognitive function, indicating that these outcomes were unlikely to be affected at this age. It may be that the payoff to increased parental investments in the treated arm children will emerge at a later stage of the life cycle given that gaps in cognitive function and socio-emotional skills that emerge very early in life appear to persist and grow over time through dynamic complementarity (Heckman et al., 2006; Cunha and Heckman, 2007; Cunha et al., 2010; Conti and Heckman, 2014).

Our study contributes to four strands of the literature. First, we provide the first causal evidence that treating depression can have large impacts on economic decision-making and, rather remarkably, impacts that are evident seven years after the end of a one-off intervention. Previous studies show that poverty may exacerbate depression and have identified negative effects of unemployment, crime, price, and income shocks on mental health (Clark, 2003; Dustmann and Fasani, 2016; Farré et al., 2015; Adhvaryu et al., 2016; Haushofer and Shapiro, 2016). Recent research also documents how scarcity can alter cognitive function, leading individuals to make suboptimal decisions that reinforce their poverty (Shah et al., 2012; Mani et al., 2013; Mullainathan and Shafir, 2013; Dean et al., 2017; Schilbach et al., 2016).[7] For reasons set out in section (2), it is plausible that depression might also bias decision-making, and thus contribute to aggravation of poverty and low levels of human capital investment. We provide the first evidence for this direction of causation. This is important because, put together with the cited evidence, it suggests that depression has the potential to create a poverty trap (Haushofer and Fehr, 2014).[8]

[6] Similarly there is evidence that women legislators are more likely to favor pro-child policies; see the literature cited in Bhalotra and Clots-Figueras (2014).

[7] Scarcity may create tunneling, or excess focus and attention on the scarce resource at the expense of attention to other dimensions. Individuals tend to become more present-biased, and executive function with respect to tasks that are not immediately related to the scarce resource becomes hindered.

[8] There is considerable evidence linking depression and labor market outcomes (see Mintz et al. (1992) for an early meta-analysis of evidence based on associations). 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 for identification (see Frijters et al. (2014) for a review). For example, using survey data from Australia,

Second, this paper is related to recent studies in economics that analyze impacts of CBT-based interventions on skills (Heller et al., 2017; Blattman et al., 2017).[9] However, these studies targeted specific populations with impulse control problems, such as crime-prone youth in Chicago and Liberia, and did not aim at improving mental health. The CBT-based interventions in these studies effectively provided a form of skills training. As these interventions did not target mental health, they would not be informative about how depression impacts decision-making.

Third, we contribute to the literature on women's empowerment (e.g. Almas et al. (2015)) by providing the first evidence that depression may inhibit women's financial empowerment. Although economic development often brings empowerment of women, continuous policy action is likely necessary to achieve gender equality (Duflo, 2012). Most successful interventions aimed at improving women's empowerment (education, fertility planning, cash transfers, savings accounts) are targeted at adolescent girls (Bandiera et al., 2017).[10] Yet such interventions have generally failed to generate persistent effects (Baird and Özler, 2016) potentially because of social constraints or cultural norms (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.

Last, we contribute to the literature on parenting and early childhood development, providing causal evidence that maternal depression inhibits investments in children.[11] The early childhood environment has been shown to influence adult outcomes.[12] The first generation of empirical studies analyzed shocks to the physical health of pregnant women (Almond and Currie, 2011b,a; Currie, 2011; Almond et al., 2017), but there has been no direct analysis of shocks to their mental health, although recent studies analyze maternal stress (Aizer et al., 2016) or bereavement during pregnancy (Persson and Rossin-Slater, 2017; Black et al., 2016).[13]

Frijters et al. (2014) find that negative mental health shocks, instrumented using sudden death of a friend, negatively impact labor market participation. Other instruments that have been used, primarily in cross-sectional variation, are religiosity, whether the parent had mental disorders, perceived social support, and physical activity. However, these instruments are unlikely satisfy the exclusion restriction.

[9] In a large randomized field experiment in Chicago, Heller et al. (2017) find that in-school programming incorporating CBT reduced violent-crime arrests and generated sustained gains in schooling outcomes, and Blattman et al. (2017) find that a short term CBT program reduced criminal behavior and improved self-control and self-image in Liberia.

[10] Adolescence is a clear critical period for women's empowerment as interventions at this early stage have the capacity to impact education, timing of marriage, and age of first birth. While interventions targeting adult women have less scope for changing the woman's long-term wellbeing, they may have other important spillovers to the next generation (Lavy et al., 2016).

[11] Studies in the psychology literature have documented associations between mental health and negative outcomes for mothers and children, but these relationships are typically not causal (Lund et al., 2010; Murray et al., 1996, 1999). In a meta-analysis of interventions for common perinatal maternal depression administered by non-specialist community workers in low- and middle-income countries, Rahman et al. (2013) report benefits to the child which included improved mother-infant interaction, better cognitive development and growth, reduced diarrheal episodes, and increased immunization rates. However, to our knowledge no study examines whether psychotherapy influences decision-making, and in particular, several years after the end of the intervention.

[12] See Heckman et al. (2013); Heckman and Mosso (2014); Elango et al. (2015); Doyle et al. (2017); Bhalotra and Venkataramani (2011); Bhalotra et al. (2017c).

[13] For identification, these studies exploit the timing of a shock, occurring during gestation versus just before or just after. Persson and Rossin-Slater (2017) find that perinatal stress caused by the death of a close relative impacts later life mental health of the children *in utero*, while Black et al. (2016) find it has no impact on later life educational attainment and wages. Aizer et al. (2016) finds that maternal stress, measured using cortisol levels, is associated

The emphasis in this literature has shifted from focusing on children skills to parental investments (Almond and Mazumder, 2013). For example, Cunha (2017) and Bhalotra et al. (2017a) highlight the importance of subjective expectations of returns to investments in children, identifying a tendency for mothers of low socioeconomic status to under-estimate returns and hence, to under-invest. Since depression is positively associated with low socioeconomic status, depression may contribute to this tendency toward biased beliefs. Walker et al. (2005); Attanasio et al. (2014) underline the importance of early childhood stimulation. If depression diminishes the quality of the mother's time with children, this may also translate into lower stimulation.

Our findings suggest that maternal depression is potentially critical to women's empowerment and to the intergenerational transmission of inequality. Yet it is not systematically measured, and it is not one of the new Sustainable Development Goals. The results in this paper provide new evidence to motivate greater policy investment in recognizing and addressing maternal depression. Although these findings are of particular interest in the context of developing countries where fertility rates, poverty, and the incidence of depression are high, and where women's financial empowerment is more limited, the behavioral effects are likely to have wider relevance.

The rest of the paper is organized as follows. A conceptual framework that delineates how depression can shape economic decision-making is sketched in section 2; section 3 describes the intervention; section 4 describes the data, and section 5 outlines our empirical strategy which addresses potential threats to the validity of the experiment. Section 6 presents the results of the intervention. Finally, Section 7 discusses potential explanations of our findings.

2 Conceptual Framework

In order to better understand the mechanisms through which depression 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.

We describe a stylized optimization problem of the mother and consider how depression may affect her choices by impacting her economic primitives, i.e. her preferences, beliefs, or constraints. Depression may influence the mother's decisions over parenting effort and investments in children and, consequently, the child's human capital development. We start from a model of household production in which the mother makes decisions over consumption, leisure, and parenting inputs of time and monetary investment (Cunha and Heckman, 2007; Del Boca et al., 2014; Cobb-clark et al., 2016; Ronda, 2016). Specifically, consider a two-period set-up, where the mother cares about consumption and leisure today but only child

with worse cognitive function of children.

quality tomorrow:

$$\begin{aligned} \max_{\tau,x,a,h,c,l} U(Q,a,c,l;d) &= v(c,l) - e_d(a) + \beta_d E_d\left[u(Q)\right] \\ \text{s.t.} & \\ Q &= f_d(\tau,x,a;\varepsilon) \\ \Omega - s_d &= \tau + l + h \\ \alpha_d(Y + wh) &= px + c \end{aligned}$$

Mothers are characterized by their depression status $d \in [0,1]$ and maximize utility from current consumption and leisure $v(c,l)$, as well as expected utility from the future quality of the child $E_d\left[u(Q)\right]$, discounted at rate β_d. Future child quality Q is uncertain, and it is determined by a production function f_d that combines time investment τ, monetary investment x, parenting effort a, and a random shock $\varepsilon \sim \phi^*(\varepsilon)$ which is independent of the mother's choices. E_d is defined as the expectation under the distribution $\phi_d(\cdot)$, which captures the idea that subjective beliefs about the shock distribution $\phi_d(\cdot)$ might diverge from the true distribution $\phi^*(\cdot)$ in a way that depends upon depression.

Time investments τ come at the expense of hours worked h and leisure time l, summing up to the total effective time $(\Omega - s_d)$ at the disposal of the mother, which can be reduced by sick days s_d. Monetary expenditure x costs p and is rival to expenditure on other goods c, which both come out of her share α_d of household income (excluding her wages) Y and her wages wh. We do not explicitly model the bargaining process, however, α_d can be related to the woman's bargaining power in a collective model of household bargaining. Parenting effort has a utility cost captured by $e_d(a)$. We assume $v(\cdot)$, $u(\cdot)$, and $f_d(\cdot)$ to be increasing in their arguments and concave, and $e_d(\cdot)$ to be increasing and convex.

Based on the characteristic symptoms of the disease, we believe depression may impact decisions through a number of independent channels. The first are related to *preferences*, consisting of depression induced changes in time discounting and the disutility of effort in our example.[14] For example, depression may influence time discounting, β_d: having a bleak view of the future, depressed mothers could discount delayed rewards at a higher rate, putting more weight on the present. For instance, positive affect has been shown to reduce discounting of delayed rewards (Ifcher and Zarghamee, 2011; Lempert and Pizzagalli, 2010), and depression has been linked to impulsivity and inconsistent intertemporal choices (Takahashi et al., 2008), as well as present biased unhealthy behaviors such as drinking, smoking, or suicide (Dennhardt and Murphy, 2011; Imhoff et al., 2014). The second channel involves a higher cost of effort, $e_d(\cdot)$ (Cohen et al., 1982; Den Hartog et al., 2003). Depression has been related to stress and fatigue, and therefore can increase the mental cost of performing simple tasks, such as interacting with the child. In other words, depression can increase the disutility cost of effort. By making effort more costly, depression can also reduce a women's ability to advocate for herself, thus reducing her bargaining power (see for instance Mirowsky (1985)).

[14] It is also plausible that depression may impact the marginal utility of consumption or leisure as depressed individuals derive less pleasure from things they normally would have enjoyed. Indeed anhedonia is also a classic symptom of depression.

Second, depression may affect *expectations*, $E_d(\cdot)$, or subjective beliefs over the distribution of shocks, as posited by Beck's cognitive theory of anxiety and depression (Beck, 1967, 1979). For example, depression may induce a pessimistic view about the mother's ability to influence future outcomes such as her child's skills (De Quidt and Haushofer, 2016; MacLeod and Salaminiou, 2001).[15] More pessimistic beliefs can, in turn, change maternal aspirations for her child and thus inputs (Dalton et al., 2010; Genicot and Ray, 2009; Ray, 2006; Macours and Vakis, 2009; Bernard et al., 2011; Glewwe et al., 2015).[16]

Third, depression may impact the mother's decision-making through constraints, potentially tightening *constraints* on disposable time through increasing sick days s_d (Grossman, 1972), or technological constraints, for instance by lowering the productivity of maternal investments, $f_d(\cdot)$. Depression has been linked to lower productivity in the labor force (see for instance Lerner et al. (2004); Lerner and Henke (2008)). In the case of home production, depression could reduce maternal productivity in combining investments to improve child quality, holding fixed the level of inputs. For example, this might occur if depression affects cognitive function (Den Hartog et al., 2003), or the quality of parent-child interactions (Ronda, 2016). Additionally, depression may affect the mother's bargaining power (see for instance Mirowsky (1985)), reducing her ability to advocate for herself and resulting in a lower share, α_d, of the total family resources.

Equilibrium conditions and comparative statics, which are described in greater detail in Online Appendix A, imply that reducing depression would generally lead to higher child quality either by increasing investment in children, or by improving her productivity, or both.[17] For example, if reducing depression makes mothers more patient or reduces the disutility of effort, they will increase parenting inputs. Interestingly, however, the channels of expectations and productivity yield more nuanced results; the impact of depression on the optimal allocation of parenting inputs depends on the relative curvature of the utility and production functions. However, even in this parameterized model, the different pathways are not separately identified with our data.

3 The Intervention

The Thinking Healthy Program (THP) was a cluster randomized community trial of a perinatal depression intervention 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 DSM-IV criteria for diagnosis of perinatal depression are the same as those required to

[15] More specifically, depression could reduce maternal beliefs about the returns to her parenting investments: depressed mothers might believe that negative shocks $\varepsilon < 0$ are more likely (i.e., place greater mass on the lower tail of ϕ_d), and therefore believe that it is not worthwhile investing. See the work of Cunha et al. (2013, 2016); Boneva and Rau (2016) on parental beliefs. By distorting beliefs, depression might change aspirations such that depressed mothers might set a lower goal for their children.

[16] Expectations and aspirations are closely related to psychological concepts of locus of control and fatalism, which can be considered components of mental health. Expectations are different from aspirations, as expectations relate to beliefs over probabilities, while aspirations relate to setting goals and reflect optimal choices conditional upon a set of beliefs.

[17] The Online Appendix can be accessed at https://drive.google.com/open?id=1Zn6ySHlAXgG854Q2O_XwZV2mk2B-d_J1.

diagnose depression at any other stage of the life course.[18] They include at least five of the following 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).[19]

The trial was randomized across 40 clusters, the clusters being Union Councils, the smallest geopolitical unit in Pakistan. Twenty clusters were randomized into receiving the intervention and twenty clusters to the control arm. Women were enrolled between April 2005 and March 2006. All women in their third trimester of pregnancy (married, ages 16-45, no other significant illness) who met the Diagnostic and Statistical Manual of Mental Disorders, IV-TR (DSM-IV) criteria for Major Depressive Episode were invited to participate in the study.[20]

A total of 3518 women were assessed for clinical depression, with 903 (26%) identified as prenatally depressed, a prevalence consistent with a previous literature identifying the prevalence of prenatal depression in this region.[21] To our knowledge, this is the largest psychotherapy intervention to treat mild to moderate depression evaluated to date.[22]

There were 463 depressed mothers in the clusters randomized to the intervention, and 440 in the control arm clusters. Only women who were diagnosed as depressed completed the baseline survey. All women who were offered to participate in the study accepted the invitation, and those who were not offered participation were unable to receive the intervention treatment or other psychotherapies.[23]

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 et al., 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. A number of meta-analyses of CBT indicate that it often at least as effective as, if not more effective than other forms of therapy, including pharmacotherapy (Bolier et al., 2013; Tolin, 2010; Cuijpers et al., 2008a). Through extensive piloting (Rahman, 2007), the original study team further designed an intervention which could be delivered by primary health workers based in the

[18] Apart from the coincidental timing of the episode occurring around pregnancy, perinatal depression is not clinically different to a major depressive episode occurring at other points in the life course.

[19] Defined broadly, mental health goes beyond the absence of a mental disorder to include concepts such as subjective well-being, perceived self-efficacy, autonomy, competence, and the achievement of one's intellectual and emotional potential.

[20] 3898 women were identified, with 8% refusing before any assessment, and 2% were not found. Rates were not differential by treatment status, Table C.2 in the appendix shows the precise sample number by treatment cluster through time.

[21] Rahman et al. (2003) find antenatal depression rates of 25%, and that in more than 90% of women, postnatal depression was a continuation of a depressive episode during pregnancy.

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

[23] There are no psychologists in the public sector and only 3 psychiatrists (based in Rawalpindi city) for the whole district (Rahman, 2007).

village. They developed a manual with step-by-step instructions for each session to train the health workers and for them to keep for reference.[24]

The intervention was delivered by Lady Health Workers (LHWs), appointed by the federal government to deliver community health services. During the CBT-based sessions, the LHWs focused on identifying and modifying cognitive distortions common in depression, such as how the mother views 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 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., 2002, 2009; Gao et al., 2010; ling Gao et al., 2012; Lara et al., 2010; Mao et al., 2012), the component of the intervention that provided guided discovery of healthy behavior is unique to this study.

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 thereafter for the following 9 months. Mothers in the control arm received enhanced routine care with an equal number of visits, enhanced not because of content but because the frequency of visits was greater than the usual, which is once a month. The standard health visits included advice on infant health issues such as tetanus and immunizations, as well as advice about and encouragement of breastfeeding.[25] Each LHW is 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.

3.1 Follow-Up Surveys

The THP study conducted detailed follow-up surveys at 6 and 12 months postpartum to evaluate maternal mental health, infant outcomes, parenting behavior and other household characteristics. The timeline for the intervention and all follow-ups is summarized in figure 3.1.

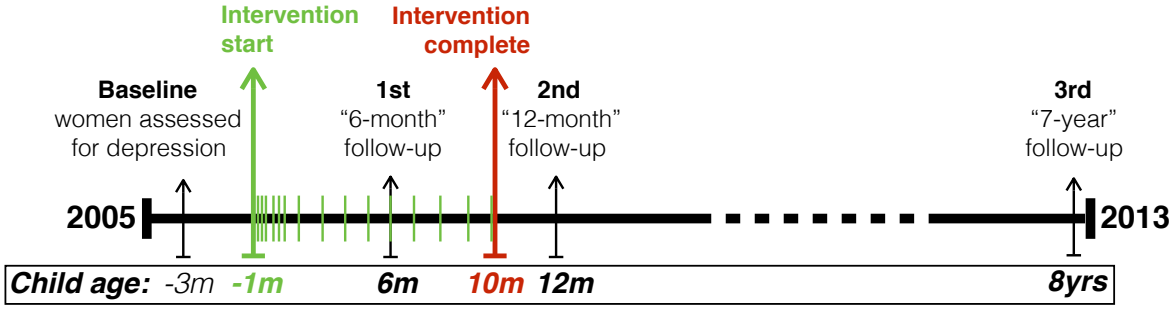
There had been no additional data collection, follow-up, or contact after the 12-month follow-up. In 2013, 7 years after the intervention concluded, a follow-up study was initiated in order to assess whether the reduction in postpartum depression noted in Rahman et al. (2008) had lasting effects.

As a first step we extracted a list of women with their contact information from the original trial and re-contacted those that completed the 12-month follow-up. Five field supervisors, who were blind to the woman's depression or trial status, worked directly with the LHWs to locate study participants. Additional queries with neighbors or relatives, as well as local hospital record checks assisted this. Fieldwork lasted between May 2013 and January 2014

[24]The manual can be download from the WHO website: http://www.who.int/mentalhealth/maternal-child/thinkinghealthy/en/.

[25]This enhanced frequency appears to be associated with better child outcomes for the control sample relative to the overall region. We find an infant mortality rate (of live births) of 4.6 percent, which is approximately two thirds of that of the Punjab region, which suffers an infant mortality rate of 77 per 1000 live births. Furthermore, rates of breastfeeding and completing immunization in our control sample were substantially higher than in the 2012-2013 Pakistan Demographic and Health Surveys (DHS)).

Figure 3.1 – Timeline of intervention and follow-ups



with a field team of 9 assessors. The assessors, who were also blind to treatment status, visited treated and control clusters at equal rates. Each dyad interview consisted of two parts: the first in the woman's home and the second either in the child's school or in the LHW's house, a commonly used meeting place. The purpose of the second session was to administer the cognitive function tests to the child in a quiet and more standardized environment than the home.

In the follow-up study, we also enrolled 300 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 women who were excluded from the original THP trial, the follow-up 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.

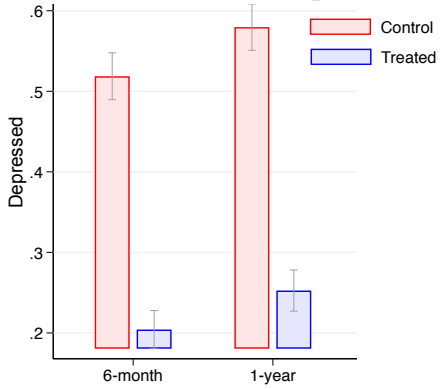
3.2 Impact of CBT Intervention on Depression Trajectories

The first-order aim of the THP intervention was to reduce the incidence of depression among prenatally depressed mothers. As such, the study was careful to measure clinical depression and mental health using the most rigorous methods. Maternal depression was assessed using the Structured Clinical Interview (SCID) for DSM-IV diagnosis. All mothers were evaluated by a psychiatrist at baseline, 6-months, and 1-year to determine if they were experiencing a major depressive episode (MDE). At the 7-year follow-up, maternal depression was also determined using the SCID interview, but administered by trained assessors. In addition to a binary indicator of clinical depression, the surveys at baseline, 6-months, and 1-year contained other measures related to mental health such as 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), and the Multidimensional Scale of Perceived Social Support (MSPSS). These are defined in Table D.4. The 7-year follow-up contained information on depression severity based on the SCID interview (number and severity of symptoms present) and the MSPSS.

The intervention was evaluated in the short run (at 6-months and 1-year) by the original study team and was shown to be extremely successful in reducing depression (Rahman et

al., 2008). Figure 3.2, adapted from Rahman et al. (2008), summarizes the main effects on depression rates of the women by intervention arm. CBT reduced depression rates by 31

Figure 3.2 –
Short-run effects on depression



percentage points in the treated group relative to the control group by the 1-year follow-up despite high rates of spontaneous recovery from depression among control arm women. Major depression episodes frequently subside over time, and the rate of recovery observed in the control group is consistent with findings from the broader literature. Although episodes often resolve spontaneously, they are also highly likely to recur. However, even after the major depressive episode is over, 20% to 30% of patients have residual symptoms, which can be distressing and associated with disability.

Incorporating the other measures of mental health into an index, we estimate that the intervention improved mental health among treated group women by 0.7 standard deviations in the 1-year follow-up, that the short-run effects were significant everywhere along the distribution of depression severity, and that the mental health benefits of the intervention persisted to the 7-year follow-up, although the difference between treated and control group women narrowed to 6 percentage points (Baranov et al., 2017).

4 Data

A detailed description of the data and outcomes as well as an analysis of sample flow, balance, and attrition are presented in appendix Section B.

Sample. Our analytical sample comprises of both the experimental group of mothers who were 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 and 440 control mothers. After 1 year, 412 treated mothers and 360 infants were compared to 386 control mothers and 345 infants in the control group. Attrition was mainly due to change of residence or stillbirths and infant deaths (see Table F.11). At the 7-year follow-up, the study team successfully re-enrolled 83% (n=585) of women and their children who completed the 1-year follow-up, with 85.5% (n=295) of the control group dyads and 80.3% (n=289) of the intervention arm dyads (see Figure C.2). The 7-year follow-up sample represents 64.8% of the baseline sample, where a third of the attrition was due to infant or child mortality (which was

not differential by treatment status). Finally, the non-experimental group also interviewed at the 7-year follow-up consists of 300 mother-child dyads, 150 from each cluster type, who were deemed not clinically depressed at baseline.

Baseline Balance. Balance of observable characteristics between the treated and the control group, shown in Table 1, is achieved at baseline. The only exception is household composition: treated women were more likely to have their mother-in-law or own mother present (each marginally significant). A joint test of balance of baseline covariates by intervention arm cannot be rejected (p-value of the joint test is 0.13). In the follow-up samples at 1 and 7 years, there is some imbalance. 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 is rejected for the 7-year follow-up sample. We discuss how we address this imbalance in the next section.

Attrition. Attrition was not significantly differential by treatment status, testament to the professional effort devoted to relocating mothers after 7 years. A two-sample Wilcoxon rank-sum test of difference in means cannot reject the null of no difference, with a p-value=0.12. Although fairly similar along most characteristics (see Table F.11), attritors (lost-to-follow-up, LTFU) were less empowered, perceived less social support, but 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). Estimated treatment effects on 6 and 12 months mental health outcomes are the same regardless of whether we use the full sample or the 7-year follow-up sample (see Table F.12), suggesting that attrition is unlikely to impact our findings. Nevertheless, we also test robustness of our results to accounting for attrition in two ways (detailed description and results described in Appendix B.2). First, we calculate treatment effects using Inverse Probability Weighting, where the weights were calculated 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 results are not sensitive to these corrections.

4.1 Outcomes

Extensive measures of mother and the child outcomes were collected in each survey. We first analyze how treatment affected women's decision-making and then analyse child outcomes. To organize the results, we group related measures of mother's decisions into five broad categories: mother's financial empowerment, their fertility trajectory, and parenting behaviors. Parenting behaviors are grouped in 3 domains: time-intensive investment, monetary-intensive investment, and parenting style. Parenting style captures parenting behavior that does not have explicit time or monetary costs. For example, how the mother speaks to the child, or the

style of discipline she uses. Differences in parenting style incur patience and effort more than time and money.

We define financial empowerment in this paper as access and control over resources (following Ahmad and Khan (2016); Karlan et al. (2017); Lavy et al. (2016) and others). In Pakistan, women's empowerment is exceptionally low, with strictly defined traditional gender norms. Most women in Pakistan are confined to their homes to do housework for the extended family and are excluded from main decision-making (Ahmad and Khan, 2016). For instance, in our sample only 3% of mothers reported usually working at baseline (recall that they are pregnant and baseline), and 52% reported having any control over spending money in the control group at 6 months.

We evaluate treatment effects on *trajectories* of empowerment and fertility. 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.[26]

To explore how treatment affected children, we report results for four broad groups of child outcomes at the 7-year follow-up: (1) physical development of the index child, (2) cognitive development of the index child, (3) socio-emotional development of the index child, and (4) survival of siblings.[27]

As there are numerous outcomes within each of these groupings, we generate summary indices to investigate effects across broad domains of decision-making before exploring effects on individual outcomes. We discuss the merits and assumptions of this approach below. We describe the outcome measures here and provide more precise definitions of all measures in Tables (D.4–D.6). Tables (D.1–D.3) show summary statistics for all outcomes used in the analysis. [28]

Mother's decision-making

1. **Financial empowerment** is measured using three outcomes: whether the woman was employed, her monthly earnings, and whether she has control over spending. At the 7-year follow-up, we have data on all three of these measures but for the earlier waves we have only control over spending.[29] From the earlier waves, only control over spending is avail-

[26] A limited number of self-reported measures were also collected from infancy, but they are not used in the main analysis since a subset of these outcomes were already analyzed in the original trial paper (Rahman, 2005). For completeness, we present the parenting results during infancy, all self-reported by the mother, in the appendix (see Table I.36).

[27] Developmental outcomes for the index child were also measured at infancy. Just as for parenting behaviors, our main analysis focuses on the richer outcomes collected during the 7-year follow-up, and since some of the early development outcomes have already been analyzed in the original trial paper (Rahman, 2005). At the 6-month and 12-month follow-ups, interviewers measured the length (height) and weight of the infants. They also asked the mother about recent diarrheal episodes and Acute Respiratory Infections (ARIs). However, for reference, results for these outcomes are presented in Table I.37).

[28] Although the women had other children that could have been affected by the intervention, the majority of our measures are based on extensive evaluation of the index child, i.e. the child who was in womb at the start of the intervention, at the 7-year follow-up.

[29] We use both measures of control over spending, even though they are highly correlated, because one phrasing was asked at the 6-month survey, while the other at the 12-month. Since the summary index places more weight on uncorrelated information, adding additional highly correlated variables does not affect the index. For example, in the limit, adding additional perfectly correlated variables does not affect the index at all.

able.[30] [31]

2. **Monetary-intensive investments** include the school quality index, whether the index child attends 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 subscales of the HOME inventory.[32] Mothers answered a detailed module on the home environment, which was a cultural adaptation of the short version of the HOME inventory (Caldwell and Bradley, 1984) similar to the HOME-SF used in a number of international studies, a prominent example being the National Longitudinal Survey of Youth, 1979. A number of questions within the HOME inventory are based on interviewer observations, and these are indicated in the detailed description of indices in D.4. Measures of school quality are based on assessor reports. The interviewer visited the child at school and recorded information about whether the school had certain amenities (playground, computers, library, etc), the number of teachers and classrooms, and classroom amenities.

3. **Time-intensive parental investments** include frequency of mother and father playing with the index child, whether someone helps with his or her studies, as well as the enrichment, family companionship, and family integration subscales of the HOME inventory (all measures are directed at the index child).

4. **Parenting style** is measured using the Parenting Practices Inventory (PPI) (Webster-Stratton et al., 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 subscales of the HOME score.

5. **Fertility trajectory** between the beginning of the intervention and the 7-year follow. 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 the child roster, we can learn how many surviving children the mother has had since the intervention began.[33] We analyze fertility in two ways. First, to summarize the results in a way that will be consistent with the other outcomes, we create an index with the total number of children born since treatment started, whether the mother was pregnant again at the 1-year follow-up (the only fertility outcome actually measured prior to the 7-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.

[30] Female labor force participation is extremely low in our setting. There is no variation in mother's employment in the first year, since none of the women surveyed reported working either at the 6 or the 12 month follow-ups, while only 12 percent of mothers work 7 years after the birth of the index child. Thus, at 6- and 12-month follow-ups, our empowerment measures only consist of control over spending.

[31] At 6 months, the question was phrased as "Does your husband/head of family give you spending money and if so do you decide how to use it", while at 12 months the question was phrased as "Do you get pocket money for spending on personal needs".

[32] Since expected grade attainment is related to schooling, we included it into the monetary index which contains all of the other questions pertaining to schooling

[33] At baseline and at the 7-year follow-up, mothers were asked about the number of her children that have died. From that we can infer the number of children that died between baseline and the 7-year follow-up and add that to the total number of surviving children to estimate total number of pregnancies. The results are essentially unchanged using this approach.

Child outcomes

1. **Physical development** was assessed with growth, fine motor skills, and illness. Interviewers measured height and weight, and motor skills were assessed using the Grooved Pegboard Test. The 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 between 2.5 and 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, 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)) which is designed to identify symptoms of various anxiety disorders. Both measures are based on sets of questions answered by the mother.
4. **Child 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 the surviving children at the 7-year follow-up to assess treatment effects.[34]

5 Econometric Specification

The identification strategy is straightforward since treatment assignment was random. Our principal estimating equation is:

$$Y_{ic} = \alpha + \beta T_c + \mathbf{\Gamma}'\mathbf{X_{ic}} + \varepsilon_{ic} \tag{5.1}$$

where Y_{ic} is the outcome for the mother or child i living in cluster c. We start by analyzing treatment effects on summary indices capturing measures related to women's decision-making (financial empowerment, her fertility trajectory, time-intensive parenting investment, monetary-intensive parenting investment, and parenting style) and child outcomes (cognitive, physical, and socio-emotional development of the index child, and child survival). All indices are standardized within the control group to have mean zero and standard deviation one, and so that 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. Standard errors are clustered at the Union

[34]Survival of the index child is an outcome in itself; 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. Similarly, while the gender composition of the index child is suggestive, it captures only the intervention effect on the index child.

Council level, the unit of randomization, using the sandwich estimator.[35] $\mathbf{X_{ic}}$ is a vector of controls. We first show a parsimonious model controlling only for interviewer fixed effects and date of interview. We then include the full set of baseline characteristics interacted with the treatment indicator, following the standard operating procedures laid out by Lin and Green (2016). This accounts for the fact that the follow-up sample is not balanced along all observable baseline characteristics.[36] 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 was empowered, wealth index, depression severity (Hamilton score), and perceived social support (MSPSS). Our results are not sensitive to the set of controls used.[37]

The parameter of interest is β, representing the Intention-To-Treat (ITT) effect. Since attrition was not a function of treatment, 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.[38] Therefore, we can only claim to estimate the causal effect of being offered treatment (ITT), encompassing treatment intensity.

Note that we identify the causal effect of the intervention (CBT 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 avoid using the randomized assignment as an instrument for depression, since it is possible that the intervention per se, through encouraging healthy thinking and bonding with the child, may have had direct impacts on some outcomes above and beyond affecting maternal depression, implying that the exclusion restriction might not be satisfied.

Heterogeneous Treatment Effects. We estimated heterogeneous treatment effects by gender, socio-economic status, demographic indicators, and baseline depression severity. For instance, for gender, the estimated equation is:

$$Y_{ic} = \alpha + \beta^g Girl_i \times T_c + \beta^b Boy_i \times T_c + \delta Girl_i + \mathbf{\Gamma'X_{ic}} + \varepsilon_{ic} \tag{5.2}$$

where $Girl_i$ is an indicator equal to one if the index child is female, and Boy_i is an indicator for male. Gender indicators are interacted with treatment, so that the coefficient β^g is interpreted

[35] We calculated p-values generated from the Wild-t bootstrap method to address small numbers of clusters following Cameron et al. (2008). However, there is little difference between the Wild-t bootstrapped p-values and those reported here.

[36] The baseline characteristics are centered by the sample mean and interacted with a treatment indicator such that coefficients on the baseline characteristics may vary by intervention arm. Thus we include $\mathbf{X_{ic}} + (\mathbf{X_{ic}} - \bar{\mathbf{X}}) \times T_c$.

[37] The children are of a single birth cohort born in 2005-2006. Child age is excluded from the controls as it is potentially endogenous. 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.

[38] Furthermore, in our data 12 women moved between clusters within the first year (6 treatment, 6 control). In these cases, women who were assigned to control and moved to treatment clusters did not receive treatment; however those in treatment clusters who moved into control clusters no longer received treatment. Thus there are 12 cases where treatment take-up differs from random assignment variable that we actually know about. We use the original treatment assignment in our analysis, yielding an estimate of the ITT. However, instrumenting actual treatment take-up with the original assignment yields similar results, albeit with slightly larger standard errors.

as the impact of treatment for dyads who had girls and β^b is the treatment effect for boy dyads.

Multiple Inference and Power. As there are many outcomes, following O'Brien (1984); Kling et al. (2007); Anderson (2008), results are reported using summary indices that are constructed as the most efficient weighted average of a set of outcomes. This approach addresses the problem of multiple inference by reducing the number of hypotheses tested but also improves the power of our statistical test for whether the intervention had effects within groups of related outcomes.[39]

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, following (Westfall and Young, 1993; Efron and Tibshirani, 1994). We also calculate the Family-Wise Error Rate (FWER)-adjusted p-values when we explore the effects of the intervention within the components of the indices.

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.

Quantile Treatment Effects. We further explore treatment effects across the distributions of parental inputs and child outcomes (there is not enough variation in other indices to perform such an analysis). We show quantile treatment effects (QTE) for the summary index, where the QTE is the horizontal distance between the treated and control group CDF at a given percentile, following Frölich and Melly (2013). Because treatment was randomized, the treatment effect at the quantiles is also identified.[40] Since the assumption of rank preservation for depression severity is not satisfied in our case, we cannot interpret the QTEs as the distribution of treatment effects, but only the effect of treatment across the distribution of the outcome.[41]

[39] 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 approach, we compute factor scores, which is a method more suited when the measures included in the factor score are proxies of an underlying one-dimensional latent factor, measured with noise. The results, presented in Table G.17, are qualitatively similar.

[40] We estimate the QTE for each quantile between 5 and 95. We use inverse propensity score weights to account for observables, controlling only for the interviewer and date of interview (though results are similar using the full set of controls). For inference, we construct point-wise confidence intervals at each quantile by bootstrapping using 1,000 replications, clustered at the Union Council level. We use a bootstrapping procedure to calculate the confidence intervals instead of the analytical calculations in order to account for the cluster-randomized design.

[41] Quantile treatment effects can be interpreted as the distribution of treatment effects under the assumption that treatment preserves the ranking of outcomes relative to the counter-factual ranking. Tests of this assumption can be made if the outcome is measured before the treatment. We reject that treatment preserved rank using depression severity.

6 Results

We identify meaningful and statistically significant impacts on mother's control over household resources and investments in children. Yet, in general, we find no detectable effects on fertility or on child development across multiple domains. Our main results are summarized in Tables 2 and 10. All coefficients can be interpreted as improvements in terms of percentages of one standard deviation of the outcome. The tables display results with parsimonious and enriched controls and report the FWER-adjusted p-values from the richer model. We first present the main tests of how treatment impacted the mother's decision-making; we then explore treatment effects on individual outcomes and heterogeneity by gender; we then analyze the effect of the intervention on potential mediators other than maternal depression; finally, we report the effect of the intervention on child outcomes, and discuss possible explanations for the lack of improvement in child development indicators.

6.1 Women's Decision-Making

Table 2 shows the treatment effects on indices of women's decision-making. The intervention significantly increased mother's financial empowerment by 0.29 of a standard deviation (SD), increased monetary-intensive parental investments by 0.28 SD, and time-intensive parental investments by 0.20 SD. This pattern across outcomes is consistent with previous research which suggests that investments in children are greater when the mother has greater control over household resources (Lundberg et al., 1997; Hoddinott and Haddad, 1995). On average, we find no effects on parenting style or the mother's fertility trajectory.

As detailed below, the financial power and parental investment results emerge primarily from mothers who were pregnant with girls at baseline, for whom we see a 0.47 SD increase in financial empowerment, a 0.48 SD increase for monetary investments in children and 0.24 for parenting style. These findings are consistent with Thomas (1994), whose findings suggest that girls benefit disproportionately from the mother's power within the household (while boys benefit more from father's control), and potentially with Dahl and Moretti (2008) and Blau et al. (2017) whose findings suggest that fathers prefer to have sons, from which it follows that girls will tend to receive a more equal share of resources when mothers' preferences matter more.

Quantile treatment effects show fairly homogeneous treatment effects over the distribution of parental inputs (Figure 8.1); we do not have sufficient continuity in the data to estimate QTE for financial empowerment. To put the magnitude of the estimated treatment effects into context, Table 3 reports the mean differences in outcomes in the 7-year follow-up between control arm women who were depressed at baseline and women who were excluded from the trial because they were not depressed. We see that perinatal depression is descriptively associated with significantly lower financial empowerment of the woman, and lower time and monetary investments in children 7 years later. There are no significant differences in parenting style on average, though there are for mothers of girls. It is striking that the estimated treatment effects more or less close these raw gaps, or that the treatment brings the long-term outcomes of perinatally depressed women close to those of their non-depressed counterparts.

Robustness. Our main findings survive both parametric and non-parametric corrections for attrition (Table F.13) and are not sensitive to the set of control variables used (Table G.16). Furthermore, a difference-in-differences specification provides an additional robustness check against cluster-level shocks. In particular, over the 7 years since the initial randomization, shocks potentially correlated with treatment assignment could undermine randomization. To address this concern, we leverage availability of data from both treated and control clusters on women who were excluded from the trial because they were not perinatally depressed at baseline. This approach assumes that any cluster-specific shocks similarly influence outcomes for the baseline depressed and the baseline non-depressed women in the cluster. Results using the difference-in-differences specification are almost identical (see B.4 for a description of the specification and Tables G.21 and G.22). These results allow us to reject the concern that our estimates of treatment effects are driven by favorable shocks to treatment clusters, the coefficients on the indicator for treatment clusters being insignificantly different from zero (and negative in many domains).

6.1.1 Exploring Treatment Effects on Individual Outcomes

Tables 4–8 provide results by sub-components of the indices on women's decision-making, and by the gender of the child in the womb when the intervention started.[42]

Empowerment. Table 4 shows treatment effects on the components of the financial empowerment index. Mother's control over spending shows a consistent improvement of about 10 percentage points in each of the 6-month, 12-month and 7-year follow-ups, and the 12-month and 7-year estimates are robust to adjusting for multiple inference. Interestingly, these results are corroborated by descriptive evidence showing that maternal depression over the years is correlated with lower maternal empowerment (see Table G.20).[43] In contrast, there is no robust evidence that the intervention affected the mother's labor supply or income. However, employment rates of women in this rural sample are low, and were lower at baseline because the baseline survey was conducted on pregnant women. We do identify a marginal effect of 1 percentage point on employment, which is a meaningful 10% increase, but this estimate is statistically indistinguishable from zero.

Monetary-intensive parental inputs. Components of the monetary investments index are shown in Table 5. The treatment effect is positive and sizeable for each of the six components, and statistically significant for learning materials, expectations of grade attainment, private school attendance, and school quality, but only learning materials is robust to controlling for multiple inference. It is also important that we see changes in both mother-reported and interviewer assessed measures of inputs (for example school quality and presence of learning

[42] Estimates of treatment effects on specific sub-scales are reported in Appendix (I).

[43] Although, where relevant, we use data from the short-run follow-ups in the indices, for consistency we restrict our main analysis to the long-run follow-up sample. Reported treatment effects using all available short-run data are very similar. If we repeat the analysis using only the long-run outcomes, the magnitude of the overall effect is smaller (0.18 of a standard deviation), but the results are qualitatively similar (see Tables G.18 and I.27).

materials in the home), suggesting that our finding are not driven by experimenter demand effects.

Time-intensive parental inputs. Table 6 shows that four of the five measures of time-intensive investments are positive and statistically significant. These are the subscales of the HOME inventory which indicate enrichment and family companionship, both of which improve by 31% of a standard deviation, and family integration, which improves by 17%. There is a 6% increase in the likelihood that someone in the family helps with studies, but this is not robust to adjusting for multiple inference. We find no effect on the frequency with which the mother plays with the child at 7-years, although this variable was highly significant at the 1-year follow-up.

Parenting Style. Table 7 shows more variation in treatment effects across items of the parenting style index than is the case for other inputs. There are large positive effects for subscales of HOME indicating higher responsivity (22%) and emotional climate (37%) in treated families compared to the control mean. There are small or even negative effects of treatment on encouragement of maturity and on elements of the Parenting Practices Inventory that indicate consistency and harshness of parenting styles. However as we note below, treatment effects on parenting styles differ so much between boys and girls that once they are separated, we see robust positive treatment effects on parenting style for girls.[44]

Fertility Trajectory. Table 8 shows that the intervention had no discernible effect on any of the components of the fertility trajectory index, both when the index child is a boy or a girl. All of the estimated treatment effects are small and indistinguishable from zero; the treatment effect on whether the index child is the last is marginally statistically significant only when additional controls are not added, and the result does not survive multiple hypothesis testing. Additionally, Figure C.5 shows the number of kids born after the index child, and C.4 plots the trajectory of fertility between baseline and the 7-year follow-up for the treatment and control groups, showing no differences in number or timing of subsequent births.

6.1.2 Heterogeneity of Treatment Effects

We first discuss heterogeneity of treatment effects by the gender of the index child, and then by other characteristics. There is a long-standing hypothesis that males are biologically more sensitive to stress *in utero* (Trivers and Willard, 1973; Giussani et al., 2011; Mueller and Bale, 2008; Almond and Edlund, 2007), and a large literature suggests that male and female children respond differently to early life shocks (see, among many others, Autor et al. (2016); Anderson (2008); Garcia et al. (2017); Bhalotra and Venkataramani (2011); Baird et al. (2016)). In South Asia, there is considerable evidence that gender differences in investments and outcomes arise from a preference for sons (Kristof and WuDunn, 2009; Atif et al., 2016; Sathar et al., 2015;

[44] Parts of these subscales are interviewer observed, which means they are less prone to reporting bias from the mother. Table J.38 shows an additional measure of positive parenting using the interviewer observed measures only.

Miller, 1981; Sen, 1992; Bhalotra et al., 2016). Not only are sons often favored, but mothers of sons are favored over women who have only daughters ((Milazzo, 2014).

Comparing control group means of outcomes in the 7-year follow-up for women who were pregnant with a girl versus a boy when the intervention started, we similarly find that women with daughters in our sample have more limited control over household spending and that their children receive lower monetary and time investments.

Importantly, we find that intervention effects tend to redress this imbalance. In Tables 4–8, columns 5-6 report outcome means in the control group by gender of the index child and columns 7-8 report the estimated treatment effects for mothers of girls and boys respectively. The difference between columns 5-6 and 7-8 tells us whether treatment is able to overcome the gender differences in outcomes that would arise in the absence of the intervention.[45]

We find overwhelming evidence that the average treatment effects are driven by mothers who were pregnant with a girl at the start of the intervention. The estimates for boys are consistently smaller and often indistinguishable from zero. With the exception of the index for parental time-intensive investment, the estimated gender differences in the outcomes are statistically different from each other. Overall, the intervention acted to narrow or close the existing gap between mothers of boys and mothers of girls. For example, among control group mothers, 59% of those with sons report some control over household spending, compared with 44% of those with daughters. The intervention was able to overcome this difference, empowering mothers of girls. The additional control over spending is plausibly one mechanism that led to higher investments in children, and in particular in girls. This is quite stark in the case of monetary investments. In the control group these are on average 34% of a standard deviation lower for girls, and the treatment effect for girls is 36 percentage points higher than for boys, perfectly closing the gap. The underlying gender differences in time-intensive investments are not so pronounced and the treatment effects for girls, although much bigger in magnitude, are on average not statistically different from treatment effects for boys. They are significantly larger for two of five components, namely enrichment and companionship subscales of the HOME index, and these components show a control-group mean difference that favors boys so, again, the treatment acts to redress the balance. A similar pattern is evident for parenting style. On average, the control group differences favor girls. The exceptions are responsivity and emotional climate which favor boys, and these are precisely the two scales that respond significantly to treatment, and only for girls.

Our evidence that treatment effects on investments favor girls and act to narrow the underlying advantage of boys lines up with previous evidence that the preference for sons is stronger amongst fathers (Thomas, 1994; Dahl and Moretti, 2008). An alternative interpretation of our findings may be that selective processes of gender-differentiated abortion, fertility or mortality result in the girls at the follow-up being more "wanted" and hence treated more equally relative to sons. To be relevant, any such processes would have to create a difference in wantedness between the treated and control group. In fact, we see an increase in the share of

[45]Rather than differentiate by whether the woman had a son at any time or not, we differentiate by the gender of the child that the woman was pregnant with when she was diagnosed as depressed and entered into the trial. This is directly relevant to intervention effects. It will also be positively correlated with the probability of having at least one son at the 7-year mark.

girls in treated relative to control clusters over time, suggesting less selective abortion or post-birth mortality of girls. Sex-selective abortion is very unlikely to play a role in this setting[46] but as we document below, we find some evidence of lower girl-mortality after birth in treated clusters.[47] In any case, if it did, this would be consistent with our maintained hypothesis that the empowerment of women in the treated group favored girls.

Heterogeneity by other characteristics. Estimates of heterogeneity in treatment effects by other baseline characteristics are in Table H.23. Treatment effects on women's financial empowerment are significantly larger among women with education above the median level, and significantly smaller among women who are younger (less than 27), and among women who had a grandmother in the household at baseline. There is no significant difference in treatment effects by wealth, or by whether the child in the womb at baseline was the first child. In the control group, the only significant associations are positive associations of financial empowerment with wealth and the presence of a grandmother of the child. Treatment effects on money and time investments, parenting style, and fertility do not vary significantly with any of the stated characteristics. However, in the control group we see the commonly documented positive associations of parental investments with wealth and the mother's education.

6.2 Potential Mediators

The CBT intervention was targeted at improving women's mental health postpartum and, as discussed, it succeeded in creating a large difference of 33 percentage points in depression rates of mothers in the treated and control arms in the 1-year follow-up. A hypothesis of interest is that postpartum improvements in depression drive the improvements in women's financial empowerment and parental investments at the 7-year mark. However, an alternative possibility is that the intervention influenced these outcomes through other intermediate outcomes between birth and the 7-year follow-up. 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 mother 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) and this may have been the pathway to higher financial empowerment;[48] or the

[46] Prenatal sex detection diagnostics were unavailable to this population (Zaidi and Morgan, 2016), and there is evidence that Muslims have a strong religious aversion to abortion, using other means to exercise their preference for sons (Bhalotra and Cochrane, 2010; Almond et al., 2013; Bhalotra et al., 2017b).

[47] The gender of the index child at birth was predetermined and not affected by the intervention. However, differences in gender ratios by treatment status in later follow-ups could have resulted from differential attrition due to child mortality. At birth, treated clusters had 3 percentage point more girls, and this difference grows to 5 percentage points in the 1-year follow-up sample (Table 1). In the 7-year follow-up, the treated arm had 6 percentage point more girls than the control arm (p=0.13). Thus the change in treatment-control difference in percent female between birth and 7-year follow-up is just 3 percentage points.

[48] In principle, we would also consider divorce but this is rare in this setting, with just 1% of women reporting to have divorced. Rates of divorce were not differential by treatment status, but it is of interest to note that divorced mothers were 34 percentage points ($p < 0.05$) more likely to be depressed at the 7-year follow-up. 97% of the women in our sample were still married, with 1.8% widowed (also not differential by treatment status, but unlike divorce, widowhood was not correlated with depression).

child's grandmother might have been more likely to have moved in to help care for the child in the treated group, and this might have contributed to the outcomes we document.

To test for potential mediating factors, we investigate whether husband's earnings, mother's physical health, relationship quality, or social support are directly affected by the intervention. As above, given the large number of outcomes, we generate indices for the trajectories of husband's labor earnings (at 6-month, 1-year, and 7-year follow-ups), mother's physical health (self reported measures at 7-year, and weight at 6-month follow-up), relationship quality (at 1-year and 7-year follow-ups), and social support (grandmothers at 6-month, 1-year, and 7-year follow-ups); again all indices are coded such that higher values of the index correspond to more favorable outcomes. Table 9 shows no evidence of effects on women's health or husband's earnings. Treatment led to improved relationship quality and an increased presence of grandmothers in the household.[49] Relationship quality improvements occur primarily at the 1-year follow-up and do not persist into the 7-year follow-up (Table I.35).There is no evidence of heterogeneity by gender of the index child. So, overall, we conclude that the identified impacts on women's decision-making at the 7-year mark associated with randomized exposure to CBT derive primarily from decreases in postpartum depression, which was the target of the intervention.

6.3 Child outcomes

Table 10 presents estimates of intervention effects on child development in the cognitive, socio-emotional, and physical domains, together with estimates for child survival, all measured in the 7-year follow-up.[50] We find no statistically significant effects of the CBT treatment on developmental indicators at age 8, irrespective of domain. Treatment effects on child development are, in general, not significantly different from zero in any part of the distribution or for a certain subpopulation.[51] The quantile treatment effects plotted in Figure 8.1 barely change over the distribution of children skills, and always gravitate around zero, though QTEs for physical development show the largest effects in the lower two-thirds of the distribution.

A detailed analysis of treatment effects on child development outcomes can be found in Appendix Section I. We refrain from interpreting treatment effects on specific components of indices for which we did not find statistically significant effects in our main tests in Table 10. Instead, we briefly summarize the patterns below, interpreting them as suggestive. Since intervention-led increases in parental investments favored girls, we allowed gender-specific coefficients but there is little evidence of improved developmental indicators for girls any more than for boys. The sporadic positive effects within components of indices, however, were con-

[49] Treatment effects within indices for intermediate outcomes that were not affected by treatment are presented in Tables I.32-I.34. No significant effects emerge for the components, nor were the null effects masked by heterogeneity by child gender.

[50] In previous work using this same data, Maselko et al. (2015) estimated the impact of the program on a subset of the child development outcomes that we consider in our analysis. Our primary focus in this paper is on the mother's financial empowerment and her decisions regarding investments in children.

[51] Heterogeneous treatment effects are absent also for mother's education, mother's age, whether she was pregnant with her first child at baseline, an index of baseline household wealth, and an indicator for whether the grandmother was present at baseline (Table H.24). Only two interaction terms are statistically significant, notably the treatment effect on child survival was larger in families in which the grandmother was not present at baseline. However these are two of twenty interaction coefficients estimated.

centrated among girls. For example, of the eight measures of physical development, the only one that is statistically significant at naive p-values is BMI-for-age which shows a positive treatment effect for girls and a slight negative treatment effect for boys.[52] Of the nine components of cognitive development, only processing speed and fluid reasoning improve in the treated group, and only for girls. Of the eleven components of socio-emotional development, only two respond to treatment, and only for girls. Treatment girls have fewer conduct problems but higher obsessive-compulsive subscale scores. The survival index is 0.17 standard deviations larger in the treated group, and this is driven by improved survival of girls. Survival of the index child does not seem to be impacted by the treatment, so the overall finding suggests a potential sibling spillover.[53]

So as to benchmark our findings, we compared child development indicators between women who were not prenatally depressed and prenatally depressed controls who did not receive treatment. This descriptive gap is shown in Table 11. It is striking to see no correlation between perinatal depression and the cognitive development of eight-year-old boys and girls. However, we do see the expected differences in physical development for both boys and girls (entirely driven by mother-reported illness and not growth), and in socio-emotional development especially for boys, with children of baseline depressed women exhibiting worse outcomes.[54] This descriptive benchmarking for the effectiveness of the intervention suggests that we should probably not have expected impacts on cognitive development at this age although, in principle, we may expect impacts in the other domains. We discuss alternative explanations for our null findings on child development further in Section 6.4.

6.4 Discussion: Exploring Results for Child Development

A priori we expected that lower rates of postpartum depression created by the intervention would influence child development indicators at later ages given growing evidence of the long-run benefits of an improved early life environment (Campbell et al., 2014; Almond et al., 2017). We even identify a potential mechanism for this improvement in skills, insofar as we find that parental investments (at age 1 and 7) improved as a result of the intervention. It may therefore be surprising that we see no improvements in cognitive or socio-emotional development at age 7 for girls or boys, and only small impacts on physical development.

We propose two possible explanations. One is that although additional time and money investments were made by parents in the treated group, these 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. 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

[52] In this context where nutritional resources are scarce, we may interpret this evidence as consistent with reallocation of nutrition from boys to girls.

[53] Table F.11 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 differential by treatment status.

[54] This is consistent with a prior anlaysis of correlates of child mental health in this sample (Maselko et al., 2016).

cognitive impacts of an intervention in Colombia that was designed to stimulate children and their estimates suggest that the parenting intervention was key. A second explanation is that differences between treated- and control-arm children will diverge over time, consistent with dynamic complementarity (Cunha and Heckman, 2007; Heckman and Mosso, 2014; Doyle et al., 2017; Almond et al., 2017), and so might become evident at later ages even if not yet evident at age 7. Verifying this explanation requires longer-term follow-up data.

Here we investigate (and undermine) some alternative hypothesis. These include fading of treatment effects, possibly through compensating investments in the control arm, or differential survival of weaker children.

A potential explanation is that there were treatment effects on child development but that these had faded by the 7-year follow-up. Fading has been noted, for instance in Chetty et al. (2011), and Heckman et al. (2013).[55] One mechanism underlying fading is catch up on the part of the control group. To investigate this, we sought evidence of compensating investments in the control arm, but could find none. In fact, as discussed, we find higher parental investments in treated arm children, consistent with reinforcing behavior (during infancy and at age 7). We also estimated grandmother co-residence in the 7-year follow-up as an outcome, to test whether support from the extended family increased more in the control arm to compensate the mother's depression. We again found reinforcing behavior, inasmuch as grandmothers were significantly more likely to live with the family at year 7 in the treated arm.

Alternatively, differential survival might bias the estimates toward zero if the marginal child surviving due to the treatment is negatively selected on birth endowments and hence 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 (see Table I.31 and C.3). Although we found some evidence of intervention effects on child survival, this did not stand up the multiple inference adjustment on average, and the evidence was restricted to girls (Table I.31). In contrast, our failure to find intervention effects on child development is similar for boys and girls.

Another possibility is that the intervention was initiated too late in pregnancy, or stopped too early, so that damage to the child might have already occurred (i.e. before the third trimester) and the postnatal investments we document may not have been sufficient to reverse it. However, if this were correct, we would expect to see systematically worse outcomes for children of mothers who were depressed vs non-depressed in pregnancy. Although we do see worse socio-economic and physical development among children of perinatally depressed women, we see no significant deficit in their cognitive development. This weakens the timing hypothesis.

We might imagine that the indices of child development we use are not sensible constructs for this population. To investigate this, we estimated regressions of each index on baseline family characteristics, and these reveal the expected associations with wealth and education, indicating that the indices are sensible constructs (see Tables D.7 and D.8). Additionally, previous analyses with this data utilized a narrower subset of measures traditionally used in the field

[55] The cognitive gains documented among toddlers in Attanasio et al. (2017) had also faded by the next follow-up; personal communication.

of child development and found similar results (Maselko et al., 2015). As another sensibility-check, we estimate each index of child development on parenting inputs (without accounting for their endogeneity), and the resulting coefficients are of the expected sign (see Table G.19).[56] Another possible concern is that not all of the relevant domains of child development were measured and so there were intervention effects but we do not capture them in our data. For instance, evaluation of a school intervention in the US, Project Star (Chetty et al. (2011)), shows early improvements in cognitive skills that subsequently fade but because individuals in the treated arm have higher adult income, the authors argue that there may have been (undocumented) positive effects on non-cognitive skills. However we measure non-cognitive skills using inventories with numerous questions, so this seems unlikely. A further measurement issue is that treated mothers might be more prone to report negative outcomes for the child if, for example, the intervention made them more attuned to the psychological disposition of their child. However, while this may explain the failure to identify intervention effects on socio-emotional development, which is more dependent on the mother's subjective responses, it is unlikely the results for cognitive and physical development, the indices of which include several objective measurements.

7 Conclusion

Since the onset of maternal depression is often associated with economic and social disadvantage, it is hard to investigate its causal effects. We overcame this limitation by leveraging exogenous variation in depression generated by a cluster randomized control trial providing cognitive behavioral therapy (CBT) treatment to a sample of depressed pregnant women in one of the largest psychotherapy trials to date. The intervention cost $10 per woman per year and was provided by training public sector health workers in administration of CBT.

Perinatal depression is a widespread mental health problem, estimated to affect about 20-35% of women and their children in poor countries (Rahman, 2005) and 10-20% worldwide (WHO). Maternal depression is often undiagnosed and hence untreated, and in many cases is incorrectly perceived as a transient condition. The standard approach to treating depression in general, at least in richer countries, is to prescribe antidepressants. However, the evidence is that antidepressants only provide symptom relief and that even when adherence to the drugs leads to recovery, patients who stop taking the drugs have a significantly higher risk of relapse than the average member of the population (DeRubeis et al., 2008). In contrast, a one-off exposure to CBT has the potential to have lasting effects to the extent that it addresses the source of the problem, changing mindsets and encouraging learning of positive ways of thinking. In any case it has no known adverse side-effects and no recurrent costs. The existing body of

[56] Parenting measures at age 7 are associated with child development. Monetary investment is associated with positive physical and cognitive development, but not with socio-emotional development; time investment is not significantly associated with any domain, although the biggest coefficients can be found for socio-emotional development; parenting style is positively associated with both cognitive and socio-emotional development. We also see that maternal financial empowerment is positively related to physical and cognitive development, while fertility is not correlated with child developmental outcomes. Including the sample of prenatally non-depressed women does not affect these gradients. We also see again that prenatal depression is associated with worse physical and socio-emotional development but the coefficient in this associational estimate is in fact positive for cognitive development.

evidence of long-run effects of CBT is small and the results are mixed (Cuijpers et al., 2008a,b).

We contribute to this literature, first, with the longest follow-up of a stand-alone CBT intervention; second, documenting impacts of CBT on women's financial empowerment and their investment in children; third, providing evidence for a low-income country; and fourth, focusing on a CBT initiated during pregnancy. This is particularly important as many women are worried about potential impacts on their infant when taking antidepressants during pregnancy or while breastfeeding. The finding that CBT withdrawal does not encourage relapse but has persistent impacts is also of considerable importance given that women who have an episode of maternal depression experience a 50% greater risk of a subsequent episode (Shi and Altshuler, 2004).

Seven years after the CBT intervention, we identify fairly large impacts on women's financial empowerment, and on the monetary and time investments made in their children, and also improvements in parenting style for girls. All of the improvements we identify emerge from the sample of women who were pregnant with a girl at the beginning of the intervention. Data from the control group confirm that mothers of girls are less financially empowered and invest less money in their children, in line with the sway of patriarchal norms in Pakistan. The intervention was able to empower mothers of girls, and to bring about convergence in investments (especially monetary investments) in girls versus boys. Among potential mediators, there are no significant impacts on trajectories of fertility, maternal physical health, or husband's labor supply by the seven-year mark. We find evidence of reinforcing investments in that the grandmother of the index child was more likely to move in with the family in the treated group, and relationship quality within the household improved. However, overall the evidence suggests that the identified long-run impacts of the intervention derive from the intervention's success in reducing postnatal depression.

References

Adhvaryu, Achyuta R, James Fenske, Namrata Kala, and Anant Nyshadham, "Fetal Origins of Mental Health: Evidence from Africa," *Working Paper*, 2016.

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

Aizer, Anna, Laura Stroud, and Stephen Buka, "Maternal Stress and Child Outcomes: Evidence from Siblings," *Journal of Human Resources*, aug 2016, *51* (3), 523–555.

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

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

__ **and Janet Currie**, "Human capital development before age five," *Handbook of Labor Economics*, 2011, *4*, 1315–1486.

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

__ **and Lena Edlund**, "Trivers-Willard at birth and one year: evidence from US natality data 1983-2001," *Proceedings of the Royal Society of London B: Biological Sciences*, 2007, *274* (1624).

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

__ **, Lena Edlund, and Kevin Milligan**, "Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada," *Population and Development Review*, 2013, *39* (March), 75–95.

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

Anderson, Michael L., "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*, dec 2008, *103* (484), 1481–1495.

Atif, Khaula, Muhammad Zia Ullah, Afeera Afsheen, Syed Abid Hassan Naqvi, Zulqarnain Ashraf Raja, and Saleem Asif Niazi, "Son Preference in Pakistan; A Myth or Reality.," *Pakistan journal of medical sciences*, 2016, *32* (4), 994–8.

Attanasio, O. P., C. Fernandez, E. O. a. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina, "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*, 2014, *349* (sep29 5), g5785–g5785.

Attanasio, Orazio P, "The determinants of human capital formation during the early years of life: theory, measurement, and policies," *Journal of the European Economic Association*, dec 2015, *13* (6), 949–997.

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

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

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

__ **, Joan Hamory Hicks, Michael Kremer, and Edward Miguel**, "Worms at Work: Long-run Impacts of a Child Health Investment," *The Quarterly Journal of Economics*, nov 2016, *131* (4), 1637–1680.

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

Baranov, Victoria, Sonia Bhalotra, Joanna Maselko, Siham Sikander, and Atif Rahman, "The Long-term Impact of Treating Maternal Depression: Evidence from a Randomized Controlled Trial in Pakistan," *Working Paper*, 2017.

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

__ , *Cognitive therapy and the emotional disorders*, Penguin, 1979.
Bernard, Tanguy, Stefan Dercon, and Alemayehu Seyoum Taffesse, "Beyond fatalism-an empirical exploration of self-efficacy and aspirations failure in Ethiopia.," *Working Paper*, 2011.
Bertrand, Marianne, Emir Kamenica, and Jessica Pan, "Gender Identity and Relative Income within Households," *The Quarterly Journal of Economics*, may 2015, *130* (2), 571–614.
Bhalotra, Sonia, Adeline Delavande, Paulino Font, and Joanna Maselko, "The impact of early cognitive and non-cognitive skills on later outcomes," *Mimeo*, 2017.
__ **and Irma Clots-Figueras**, "Health and the Political Agency of Women," *American Economic Journal: Economic Policy*, may 2014, *6* (2), 164–197.
__ **and T Cochrane**, "Where Have All the Young Girls Gone? Identifying Sex-Selective Abortion in India," *IZA Discussion Paper 5381*, 2010.
__ **, I Clots-Figueras, and L Iyer**, "Religion and Abortion: The Role of Politician Identity," *Mimeo*, 2017.
__ **, Martin Karlsson, and Therese Nilsson**, "Infant Health and Longevity: Evidence from A Historical Intervention in Sweden," *Journal of the European Economic Association*, mar 2017, *29* (5), 208–230.
Bhalotra, Sonia R. and Atheendar Venkataramani, "The Captain of the Men of Death and His Shadow: Long-Run Impacts of Early Life Pneumonia Exposure," *IZA Discussion Papers*, 2011.
Bhalotra, Sonia Radhika, Abhishek Chakravarty, Selim Gulesci, Sonia Bhalotra, Abhishek Chakravarty, and Selim Gulesci, "The Price of Gold: Dowry and Death in India," *IZA Discussion Paper*, 2016.
Black, Sandra E, Paul J Devereux, and Kjell G Salvanes, "Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes," *American Economic Journal: Applied Economics*, jan 2016, *8* (1), 193–223.
Blattman, Christopher, Julian C Jamison, and Margaret Sheridan, "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia," *American Economic Review*, apr 2017, *107* (4), 1165–1206.
Blau, Francine, Lawrence Kahn, Peter Brummund, Jason Cook, and Miriam Larson-Koester, "Is There Still Son Preference in the United States?," *NBER working paper*, sep 2017, *23816*.
Blimpo, Moussa P, Pedro Carneiro, and Nathalie Lahire, "Improving Parental Investments in Children: Experimental Evidence from the Gambia," *Working paper*, 2016.
Bolier, Linda, Merel Haverman, Gerben Westerhof, Heleen Riper, Filip Smit, and Ernst Bohlmeijer, "Positive psychology interventions: a meta-analysis of randomized controlled studies.," *BMC public health*, 2013, *13*, 119.
Boneva, Teodora and Christopher Rau, "Parental Beliefs about Returns to Educational Investments - The Later the Better?," *Working Paper*, 2016.
Browning, Martin, François Bourguignon, Pierre-André Chiappori, and Valérie Lechene, "Income and Outcomes: A Structural Model of Intrahousehold Allocation," *Journal of Political Economy*, dec 1994, *102* (6), 1067–1096.
Buvinic, Mayra and Rebecca Furst-nichols, "Promoting Women's Economic Empowerment," *World Bank Policy Research Working Paper*, 2016, *7087*.
Caldwell, B. and R. Bradley, *Home Observation for Measurement of the Environment (HOME)-revised edition.*, Little Rock, AR: University of Arkansas, Little Rock., 1984.
Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," 2008.
Campbell, Frances A, Gabriella Conti, James J Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth P Pungello, and Yi Pan, "Early Childhood Investments Substantially Boost Adult Health," *Science*, mar 2014, *343* (6178), 1478–1485.
Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How does your kindergarten classroom affect your earnings? Evidence from project star," *Quarterly Journal of Economics*, 2011, *126* (November), 1593–1660.
Clark, Andrew E, "Unemployment as a Social Norm: Psychological Evidence from Panel Data," *Journal of Labor Economics*, apr 2003, *21* (2), 323–351.
Cobb-clark, Deborah A, Nicolas Salamanca, and Anna Zhu, "Parenting Style as an Investment in Human Development," *Working Paper*, 2016.
Cohen, Robert M, Herbert Weingartner, Sheila A Smallberg, David Pickar, and Dennis L Murphy, "Effort and Cognition in Depression," *Archives of General Psychiatry*, may 1982, *39* (5), 593.
Conti, Gabriella and James J Heckman, *Economics of Child Well-Being*, Springer, 2014.

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

Cooper, Peter J., Mireille Landman, Mark Tomlinson, Christopher Molteno, Leslie Swartz, and Lynne Murray, "Impact of a mother-infant intervention in an indigent peri-urban South African context. Pilot study," *British Journal of Psychiatry*, 2002, *180*, 76–81.

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

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

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

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

_ , Irma Elo, and Jennifer Culhane, "Eliciting maternal expectations about the technology of skill formation," *NBER Working Paper Series*, 2013, *19144*.

_ , James J Heckman, and Susanne Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation.," *Econometrica*, may 2010, *78* (3), 883–931.

_ , Orazio P Attanasio, and Pamela Jervis, "Parental Beliefs and Investments in Human Capital," *Working Paper*, 2016.

Currie, Janet, "Inequality at birth: Some causes and consequences," *American Economic Review*, 2011, *101*, 1–22.

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

Dahl, Gordon B and Enrico Moretti, "The Demand for Sons," *Review of Economic Studies*, oct 2008, *75* (4), 1085–1120.

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

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

Dean, Emma Boswell, Frank Schilbach, and Heather Schofield, "Poverty and Cognitive Function," in "The Economics of Asset Accumulation and Poverty Traps," University of Chicago Press, jun 2017.

Del Boca, Daniela, Christopher J Flinn, and Matthew Wiswall, "Household Choices and Child Development," *The Review of Economic Studies*, jul 2014, *81* (1), 137–185.

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

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

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

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

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

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

Dustmann, Christian and Francesco Fasani, "The Effect of Local Area Crime on Mental Health," *The Economic Journal*, jun 2016, *126* (593), 978–1017.

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

Elango, Sneha, Jorge Luis Garcia, James J Heckman, and Andres Hojman, "Early Childhood Education," in Robert Moffitt, ed., *Economics of Means-Tested Transfer Programs in the United States, volume 2*, 2015, chapter 4.

Farré, Lídia, Francesco Fasani, and Hannes Felix Mueller, "Feeling Useless: The Effect of Unemployment on Mental Health in the Great Recession," *Working Paper*, 2015.

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

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

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

First, Michael B, Rrobert L Spitzer, Miriam Gibbon, and Janet B W Williams, *Structured Clinical Interview for DSM-IV-TR Axis I Disorders Research Version, Patient Edition with Psychotic Screen (SCID-I/P)*, New York: Biometrics Research, New York State Psychiatric Institute, 2002.

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

Frölich, Markus and Blaise Melly, "Unconditional Quantile Treatment Effects Under Endogeneity," *Journal of Business & Economic Statistics*, 2013, *31* (3), 346–357.

Garcia, Jorge Luis, James J Heckman, and Anna Ziff, "Gender Differences in the Benefits of an Influential Early Childhood Program," *NBER working paper*, may 2017.

Genicot, Garance and Debraj Ray, "Aspirations, inequality, investment and mobility," *Georgetown University and New York University, mimeo*, 2009.

Giussani, Dino A, Andrew J W Fletcher, and David S Gardner, "Sex Differences in the Ovine Fetal Cortisol Response to Stress," *Pediatric Research*, feb 2011, *69* (2), 118–122.

Glewwe, Paul, Phillip H Ross, and Bruce Wydick, "Developing Aspirations: The Impact of Child Sponsorship on Self-Esteem and Life Expectations," *Working Paper*, 2015.

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

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

Hamilton, Max, "A rating scale for depression.," *Journal of neurology, neurosurgery, and psychiatry*, feb 1960, *23*, 56–62.

Haushofer, Johannes and Ernst Fehr, "On the Psychology of Poverty," *Science*, 2014, *344* (6186).

__ **and Jeremy Shapiro**, "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *The Quarterly Journal of Economics*, nov 2016, *131* (4), 1973–2042.

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

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

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

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

Hoddinott, John and Lawrence Haddad, "Does female income share influence household expenditures? Evidence from Cote d'Ivoire," *Oxford Bulletin of Economics and Statistics*, feb 1995, *57* (1), 77–96.

Ifcher, John and Homa Zarghamee, "Happiness and Time Preference: The Effect of Positive Affect in a Random-Assignment Exper," *The American Economic Review*, dec 2011, *101* (7), 3109–3129.

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

Kaplow, Julie B, Patrick J Curran, Adrian Angold, and E Jane Costello, "The Prospective Relation Between Dimensions of Anxiety and the Initiation of Adolescent Alcohol Use," *Journal of Clinical Child Psychology*, 2001, *30* (3), 316–326.

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

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

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

Kristof, Nicholas D and Sheryl WuDunn, *Half the Sky: Turning Oppression into Opportunity for Women Worldwide*, Alfred A. Knopf, 2009.

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

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

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

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

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

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

Lin, Winston and Donald P Green, "Standard Operating Procedures: A Safety Net for Pre-Analysis Plans," *PS: Political Science & Politics*, 2016, *49* (03), 495–500.

ling Gao, Ling, Sally Wai chi Chan, and Ke Sun, "Effects of an interpersonal-psychotherapy-oriented childbirth education programme for Chinese first-time childbearing women at 3-month follow up: Randomised controlled trial," *International Journal of Nursing Studies*, 2012, *49*, 274–281.

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

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

_ , Mary De Silva, Sophie Plagerson, Sara Cooper, Dan Chisholm, Jishnu Das, Martin Knapp, and Vikram Patel, "Poverty and Mental Disorders: Breaking the Cycle in Low-Income and Middle-Income Countries," *Lancet*, oct 2011, *378*, 1502–1514.

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

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

Macours, Karen and Renos Vakis, "Changing Households' Investments and Aspirations through Social Interactions," *World Bank Research Working papers*, 2009, pp. 1–45.

_ , Patrick Premand, Norbert Schady, and Renos Vakis, "Experimental Evidence from an Early Childhood Parenting Intervention in Nicaragua.," *Working Paper*, 2015.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *science*, 2013, *341* (6149), 976–980.

Mao, Hong Jing, He Jiang Li, Helen Chiu, Wai Chi Chan, and Shu Ling Chen, "Effectiveness of Antenatal Emotional Self-Management Training Program in Prevention of Postnatal Depression in Chinese Women," *Perspectives in Psychiatric Care*, 2012, *48*, 218–224.

Maselko, Joanna, Siham Sikander, Omer Bangash, Sonia Bhalotra, Lauren Franz, Nima Ganga, Divya Guru Rajan, Karen O'Donnell, and Atif Rahman, "Child mental health and maternal depression history in Pakistan," *Social Psychiatry and Psychiatric Epidemiology*, 2016, *51* (1), 49–62.

_ , _ , 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, "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*, jul 2015, *2* (7), 609–17.

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

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

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

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

Mueller, Bridget R and Tracy L Bale, "Sex-Specific Programming of Offspring Emotionality after Stress Early in Pregnancy," *Journal of Neuroscience*, 2008, *28* (36).

Mullainathan, Sendhil and Eldar Shafir, *Scarcity: Why Having Too Little Means So Much*, Macmillan, 2013.

Murray, Lynne, Agnese Fiori-Cowley, Richard Hooper, and Peter Cooper, "The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome," *Child development*, 1996, *67* (5), 2512–2526.

_ , Dana Sinclair, Peter Cooper, Pierre Ducournau, Patricia Turner, and Alan Stein, "The socioemotional development of 5-year-old children of postnatally depressed mothers," *Journal of Child Psychology and Psychiatry*, 1999, *40* (08), 1259–1271.

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

Organization, World Health, *Mental health and the workplace*, World Health Organization, 2005.

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

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

Rahman, Atif, "Maternal depression and child health: the need for holistic health policies in developing countries," *Harv Health Pol Rev*, 2005, *6*, 70–80.

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

_ , Abid Malik, Siham Sikander, Christopher Roberts, and Francis Creed, "Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised controlled trial," *The Lancet*, 2008, *372* (9642), 902–909.

_ , Jane Fisher, Peter Bower, Stanley Luchters, Thach Tran, M Taghi Yasamy, Shekhar Saxena, and Waquas Waheed, "Interventions for common perinatal mental disorders in women in low- and middle-income countries: a systematic review and meta-analysis.," *Bulletin of the World Health Organization*, 2013, *91* (8), 593–601I.

Ray, Debraj, "Aspirations, poverty, and economic change," *Understanding poverty*, 2006, pp. 409–421.

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

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

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan, "The Psychological Lives of the Poor," *American Economic Review: Papers and Proceedings*, 2016, *106* (5), 435–440.

Sen, Amartya K, "Missing women," *British Medical Journal*, 1992, *304* (6827), 587–8.

Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir, "Some consequences of having too little," *Science*, 2012, *338* (6107), 682–685.

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

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

Syed, Ehsan Ullah, Sajida Abdul Hussein, Syed Iqbal Azam, and Abdul Ghani Khan, "Comparison of Urdu version of Strengths and Difficulties Questionnaire (SDQ) and the Child Behaviour Check List (CBCL) amongst primary school children in Karachi," *Journal of the College of Physicians and Surgeons Pakistan*, 2009, *19* (6), 375–379.

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, "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*, jun 2008, *29* (3), 351–8.

Thomas, Duncan, "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height," *The Journal of Human Resources*, 1994, *29* (4), 950.

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

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

Vos, Theo, Abraham D Flaxman, Mohsen Naghavi, Rafael Lozano, Catherine Michaud, and Majid et al Ezzati, "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*, dec 2012, *380* (9859), 2163–2196.

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

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

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

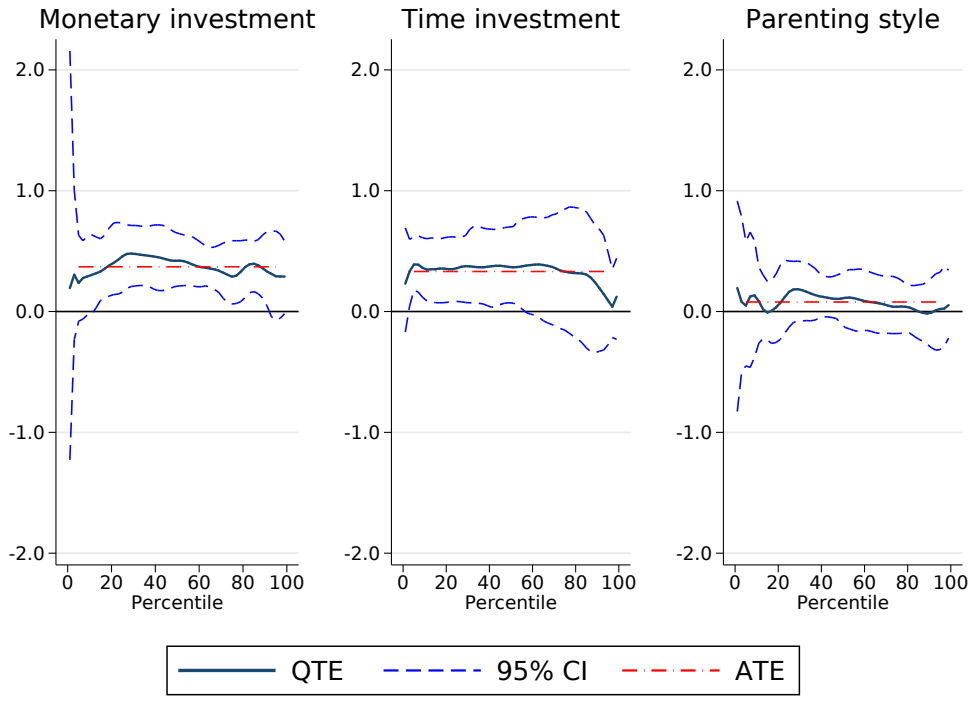
World Health Organization, *Neonatal and child health country profile: Pakistan,*, World Health Organization, 2013.

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

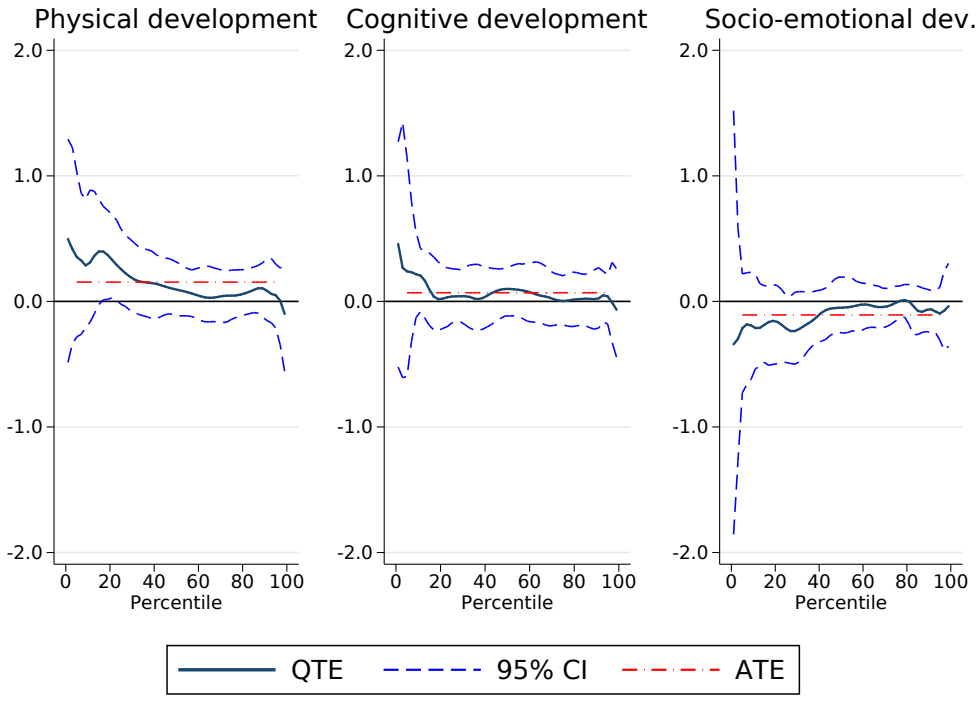
Zimet, Gregory D., Nancy W. Dahlem, Sara G. Zimet, and Gordon K. Farley, "The Multidimensional Scale of Perceived Social Support," *Journal of Personality Assessment*, mar 1988, *52* (1), 30–41.

8 Figures and tables

Figure 8.1 – Quantile Treatment Effects on parenting and child development



(a) Parenting



(b) Child development

Notes: Quantile Treatment Effects (QTE) of THP Intervention on parenting and child development outcomes measured at the 7-year follow-up. 95% confidence intervals for the QTE were calculated by bootstrapping using 1,000 replications, clustering at the UC level.

Table 1 – Balance

	Baseline Sample N=903				1-year Sample N=704		7-year Sample N=595	
	(1) Control Mean	(2) (st.dev.)	(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**
Index child is first born	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
% children female	0.52	(0.3)	0.02	0.44	0.00	0.93	0.01	0.84
Hamilton depression score	14.43	(4.0)	0.40	0.15	0.25	0.41	0.31	0.35
Baseline BDQ score	8.30	(2.7)	−0.17	0.34	−0.16	0.45	−0.03	0.88
General functioning (GAF)	62.05	(5.3)	−0.01	0.97	0.35	0.37	0.58	0.18
Perceived social support score	44.49	(16.3)	1.08	0.32	1.82	0.13	2.63	0.05*
Joint/extended family structure	0.57	(0.5)	0.04	0.27	0.05	0.17	0.06	0.18
Mother-in-law lives with	0.40	(0.5)	0.06	0.05*	0.08	0.02**	0.09	0.04**
Mother's mother lives with	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
Father not manual worker	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 (N=903)	0.33	(0.5)	0.05	0.14				
LTFU from 1y (N=704)[b]	0.15	(0.4)			0.04	0.12		
Joint test (p-value)				0.13		0.12		0.01

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: This table tests for balance along a number of baseline characteristics among the 1-year follow-up sample (Rahman et al., 2008), and in the 7-year follow-up sample. See section B.1 of the Appendix. Columns show the means and standard deviations (in parentheses) as noted, by intervention arm for the 1-year follow-up and 7-year follow-up samples. The p-value of the difference between intervention and control for each sample is also reported.

[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).

[b] Only those mother-child dyads that were interviewed at the THP 1-year follow-up were considered for the 7-year follow-up. The number of mothers in the treatment group at baseline was 463, and 440 in the control group. Between baseline and 1-year, 22% of the sample was LTFU, but not differential by treatment status. Attrition between baseline and 7-year follow-up was 35%.

Table 2 – Mother's decision-making: financial empowerment, investment in children and fertility

	Coefficient on Treat (β / (s.e.))		FWER-adj. test
	(1) No controls	(2) All controls	(3) FWER p-val (all controls)
Mother's financial empowerment	0.29**	0.29***	0.01**
	(0.11)	(0.09)	
Parental investment (monetary)	0.35***	0.28***	0.00***
	(0.07)	(0.06)	
Parental investment (time-intensive)	0.20***	0.20***	0.01**
	(0.07)	(0.06)	
Parenting style	0.04	0.05	0.74
	(0.08)	(0.08)	
Fertility trajectory	0.01	−0.00	0.99
	(0.10)	(0.09)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993).

Table 3 – Differences in outcomes between baseline non-depressed mothers and depressed controls

	Dep-NonDep (No controls)	Dep-NonDep (With controls)		
	(1) β (s.e.)	(2) β (s.e.)	(3) Girls	(4) Boys
Mother's financial empowerment (7y)	−0.26***	−0.22**	−0.37***	−0.12
	(0.09)	(0.09)	(0.13)	(0.11)
Parental investment (monetary)	−0.31***	−0.11*	−0.11	−0.11
	(0.09)	(0.06)	(0.09)	(0.09)
Parental investment (time-intensive)	−0.28***	−0.33***	−0.16	−0.46***
	(0.06)	(0.08)	(0.13)	(0.10)
Parenting style	−0.09	−0.04	−0.18*	0.07
	(0.07)	(0.09)	(0.11)	(0.14)
# kids born past 7yrs	−0.24***	−0.10	−0.00	−0.16**
	(0.07)	(0.06)	(0.10)	(0.08)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Table reports difference in means between prenatally non-depressed mothers and their children (N=300) and prenatally depressed controls and their children (N=293) 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. Column 1 shows the differences only controlling for interviewer FEs and date of interview. Columns 2-4 report adjusted differences using all baseline controls that are available for both groups of women, namely mother's age and its square, mother and father education, parity at baseline, interview date, and interviewer fixed effects. Columns 3 and 4 report the depression gap by child gender.

Table 4 – Mother's financial empowerment

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Mother's financial empowerment	0.00	0.29**	0.29***	0.00***	−0.08	0.07	0.45***	0.12	0.06
	(1.00)	(0.11)	(0.09)				(0.11)	(0.14)	
Mother controls spending (6m)	0.53	0.07	0.09**	0.12	0.52	0.53	0.15***	0.02	0.11
	(0.50)	(0.06)	(0.04)				(0.05)	(0.07)	
Mother gets pocket money (1y)	0.68	0.12**	0.11**	0.08*	0.67	0.69	0.14**	0.08	0.46
	(0.47)	(0.05)	(0.05)				(0.06)	(0.06)	
Mother controls spending (7y)	0.52	0.08**	0.09***	0.03**	0.44	0.59	0.20***	−0.03	0.01
	(0.50)	(0.03)	(0.03)				(0.05)	(0.05)	
Mother gets pocket money (7y)	0.57	0.08***	0.09***	0.01**	0.51	0.63	0.18***	−0.00	0.03
	(0.50)	(0.03)	(0.03)				(0.05)	(0.05)	
Mother employed (7y)	0.10	0.02	0.01	0.75	0.11	0.09	−0.00	0.03	0.57
	(0.30)	(0.03)	(0.03)				(0.03)	(0.04)	
Mother's income (100s PKR)	3.09	1.41	0.85	0.75	3.09	3.08	0.49	1.27	0.69
	(12.07)	(1.70)	(1.39)				(1.57)	(1.82)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table 5 – Parental investment: monetary-intensive investment

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Parental investment (monetary)	−0.00	0.35***	0.28***	0.00***	−0.18	0.16	0.47***	0.11	0.02
	(1.00)	(0.07)	(0.06)				(0.11)	(0.08)	
HOME: Learning materials	2.64	0.32***	0.31**	0.08*	2.56	2.70	0.40***	0.23	0.37
	(1.48)	(0.11)	(0.12)				(0.14)	(0.15)	
HOME: Physical environment	4.65	0.27	0.20	0.30	4.64	4.66	0.39*	0.00	0.31
	(2.37)	(0.17)	(0.15)				(0.20)	(0.28)	
Monthly expend. on educ (ln)	7.31	0.20*	0.13	0.30	7.22	7.38	0.21	0.06	0.39
	(1.18)	(0.11)	(0.10)				(0.15)	(0.12)	
Expected grade attainment	14.06	0.46**	0.33*	0.25	13.56	14.53	0.77***	−0.06	0.03
	(2.69)	(0.20)	(0.19)				(0.28)	(0.24)	
Private school	0.38	0.15***	0.12**	0.14	0.29	0.46	0.16**	0.10	0.48
	(0.49)	(0.06)	(0.05)				(0.07)	(0.07)	
School quality	−0.42	0.64***	0.51**	0.14	−0.65	−0.21	0.80**	0.19	0.14
	(2.21)	(0.21)	(0.23)				(0.36)	(0.24)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table 6 – Parental investment: time-intensive investment

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Parental investment (time-intensive)	0.00	0.20***	0.20***	0.00***	0.02	−0.02	0.28***	0.10	0.23
	(1.00)	(0.07)	(0.06)				(0.10)	(0.09)	
HOME: Enrichment	2.65	0.31**	0.31**	0.08*	2.50	2.79	0.51***	0.14	0.07
	(1.39)	(0.13)	(0.13)				(0.15)	(0.16)	
HOME: Family companionship	2.95	0.38***	0.31**	0.08*	2.77	3.11	0.64***	−0.02	0.01
	(1.77)	(0.12)	(0.12)				(0.16)	(0.19)	
HOME: Family integration	2.61	0.20**	0.17**	0.08*	2.61	2.62	0.21*	0.11	0.52
	(0.92)	(0.07)	(0.06)				(0.11)	(0.10)	
Frequency of mother play	0.17	0.00	−0.00	0.81	0.18	0.16	−0.00	−0.01	0.94
	(0.31)	(0.02)	(0.02)				(0.03)	(0.02)	
Someone helps with studies	0.54	0.04	0.06*	0.18	0.60	0.50	0.06	0.04	0.79
	(0.50)	(0.03)	(0.03)				(0.04)	(0.05)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table 7 – Parental investment: parenting style

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Parenting style	−0.00	0.04	0.05	0.52	0.04	−0.03	0.24***	−0.20	0.00
	(1.00)	(0.08)	(0.08)				(0.08)	(0.12)	
PPI: Not harsh	13.46	0.81	0.76	0.49	13.83	13.12	1.53*	−0.35	0.05
	(8.22)	(0.50)	(0.54)				(0.80)	(0.59)	
PPI: Not harsh for age	8.71	−0.17*	−0.12	0.49	8.77	8.66	−0.05	−0.23	0.41
	(1.04)	(0.10)	(0.09)				(0.11)	(0.16)	
PPI: Consistent	9.57	0.40	0.40	0.49	9.74	9.41	0.55	0.17	0.47
	(3.62)	(0.30)	(0.30)				(0.35)	(0.46)	
HOME: Responsivity	8.75	0.21	0.22*	0.25	8.58	8.91	0.63***	−0.22	0.00
	(1.75)	(0.13)	(0.12)				(0.15)	(0.19)	
HOME: Encouragement of maturity	5.25	−0.11	−0.16	0.49	5.36	5.15	−0.03	−0.36*	0.20
	(1.56)	(0.14)	(0.13)				(0.16)	(0.20)	
HOME: Emotional climate	4.50	0.35**	0.37**	0.16	4.43	4.57	0.90***	−0.24	0.00
	(1.93)	(0.16)	(0.17)				(0.21)	(0.25)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table 8 – Mother's fertility trajectory

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Fertility trajectory	0.00	0.01	−0.00	0.99	−0.15	0.13	0.13	−0.10	0.12
	(1.00)	(0.10)	(0.09)				(0.10)	(0.13)	
Ideal # kids (7y)	3.36	−0.14	−0.04	0.91	3.50	3.23	−0.14	0.05	0.31
	(1.23)	(0.11)	(0.11)				(0.13)	(0.16)	
# kids born past 7yrs	1.01	0.09	0.03	0.91	1.08	0.95	−0.05	0.08	0.19
	(1.00)	(0.09)	(0.06)				(0.08)	(0.08)	
Pregnant at 1y	0.08	−0.01	−0.01	0.91	0.11	0.05	−0.04	0.01	0.34
	(0.27)	(0.02)	(0.02)				(0.04)	(0.03)	
Index not last child	0.60	0.08*	0.04	0.59	0.61	0.60	0.03	0.05	0.62
	(0.49)	(0.04)	(0.04)				(0.04)	(0.05)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table 9 – Potential mediators

	Coefficient on Treat (β / (s.e.))		FWER-adj. test
	(1) No controls	(2) All controls	(3) FWER p-val (all controls)
Mother's physical health	0.07	0.07	0.63
	(0.07)	(0.08)	
Husband's income trajectory	−0.02	−0.04	0.69
	(0.10)	(0.10)	
Relationships	0.14*	0.16*	0.21
	(0.09)	(0.09)	
Grandmother trajectory index	0.34***	0.16**	0.07*
	(0.08)	(0.07)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993).

Table 10 – Child outcomes: child survival and physical, cognitive, and socio-emotional development by age 7

	Coefficient on Treat (β / (s.e.))		FWER-adj. test
	(1) No controls	(2) All controls	(3) FWER p-val (all controls)
Physical development index	0.15	0.14	0.35
	(0.09)	(0.09)	
Cognitive development index	0.09	0.04	0.64
	(0.08)	(0.08)	
Socio-emotional development index	−0.11	−0.08	0.41
	(0.07)	(0.07)	
Child survival index	0.19**	0.17**	0.18
	(0.08)	(0.08)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993).

Table 11 – Differences in outcomes between baseline non-depressed mothers and depressed controls

	Dep-NonDep (No controls)	Dep-NonDep (With controls)		
	(1)	(2)	(3)	(4)
	β (s.e.)	β (s.e.)	Girls	Boys
Physical development index	-0.22^{**}	-0.24^{**}	-0.24^{*}	-0.25^{*}
	(0.08)	(0.09)	(0.12)	(0.12)
Cognitive development index	-0.11	0.04	-0.04	0.14
	(0.10)	(0.08)	(0.13)	(0.10)
Socio-emotional development index	-0.16^{*}	-0.15^{*}	-0.04	-0.21^{*}
	(0.08)	(0.08)	(0.11)	(0.12)
Child survival index	-0.15^{**}	-0.12^{*}	-0.18^{*}	-0.05
	(0.07)	(0.06)	(0.09)	(0.08)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Table reports difference in means between prenatally non-depressed mothers and their children (N=300) and prenatally depressed controls and their children (N=293) for key outcome indices and variables at the 7-year follow-up. Column 1 shows the differences only controlling for interviewer FEs and date of interview. Columns 2-4 report adjusted differences using all baseline controls that are available for both groups of women, namely mother's age and its square, mother and father education, parity at baseline, interview date, and interviewer fixed effects. Columns 3 and 4 report the depression gap by child gender.

Appendix: For Online Publication

for paper "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Large Randomized Control Trial" by Baranov, Bhalotra, Biroli, & Maselko

List of Figures

3.1 Timeline of intervention and follow-ups . . . 10
3.2 Short-run effects on depression . . . 11
8.1 Quantile Treatment Effects on parenting and child development . . . 35
C.1 Map of treatment and control clusters . . . 55
C.2 Sample sizes . . . 56
C.3 Distributions of selected components of outcomes at the 7-year follow-up . . . 57
C.4 Effects on fertility and birth-spacing . . . 58
C.5 Effects on fertility: Number of children born since treatment . . . 58

List of Tables

1 Balance . . . 36
2 Mother's decision-making: financial empowerment, investment in children and fertility . . . 37
3 Differences in outcomes between baseline non-depressed mothers and depressed controls . . . 37
4 Mother's financial empowerment . . . 38
5 Parental investment: monetary-intensive investment . . . 39
6 Parental investment: time-intensive investment . . . 40
7 Parental investment: parenting style . . . 41
8 Mother's fertility trajectory . . . 42
9 Potential mediators . . . 43
10 Child outcomes: child survival and physical, cognitive, and socio-emotional development by age 7 . . . 43
11 Differences in outcomes between baseline non-depressed mothers and depressed controls . . . 44
D.1 Summary Statistics for Women's Decision Making . . . 59
D.2 Summary Statistics for Children's Outcomes . . . 60
D.3 Summary Statistics for Mediators . . . 61
D.4 Descriptions of mental health scales . . . 62
D.5 Descriptions of Outcome Variable . . . 63
D.6 Descriptions of Scales and Inventories . . . 65
D.7 Correlates of Parental Investment Behavior at Age 7 . . . 67
D.8 Correlates of Child Development at Age 7 . . . 68
E.9 Balance in non-depressed sample: Characteristics by cluster assignment at 7-yr follow-up . . . 69
E.10 Characteristics in 2013 by Baseline Depression Status . . . 70
F.11 Characteristics of attritors and differences by intervention and control clusters . . . 71
F.12 Differences in short-run treatment effects by sample . . . 72
F.13 Attrition corrected treatment effects: Inverse Probability Weights and Bounds . . . 73
F.14 Sample flow and attrition by gender . . . 74
F.15 Attrition corrections by child gender . . . 74
G.16 Control sensitivity . . . 75
G.17 Control sensitivity (indices by factor score) . . . 76
G.18 Mother's decision-making (only year 7 outcomes): financial empowerment, investment in children and fertility . . . 76
G.19 Do measures of parental investment predict child development? . . . 77
G.20 Depression and empowerment: associations . . . 77
G.21 Difference-in-differences with baseline non-depressed, and health worker fixed effects . . . 78
G.22 Difference-in-differences with baseline non-depressed . . . 79
H.23 Heterogeneous treatment effects for mother's decision-making . . . 81
H.24 Heterogeneous treatment effects for child outcomes . . . 82
H.25 Heterogeneous treatment effects for specific outcomes . . . 83
I.26 Mother's depression trajectory . . . 84
I.27 Mother's financial empowerment (at 7-year follow-up) . . . 85
I.28 Child physical development and health at age 7 . . . 85
I.29 Child cognitive development at age 7 . . . 86
I.30 Child's socio-emotional development at age 7 . . . 87
I.31 Child survival . . . 88
I.32 Husband's income trajectory . . . 88
I.33 Trajectory of mother's physical health . . . 89
I.34 Trajectory of social support: presence of grandmothers in the household . . . 89
I.35 Trajectory of relationship quality with husband and mother-in-law . . . 90
I.36 Parenting inputs during infancy . . . 91
I.37 Infant development . . . 91
J.38 HOME score components . . . 92
J.39 School quality components . . . 93

J.40 WPSSI components 94
J.41 Spence Anxiety Scale components 94
J.42 SDQ components 95

A Conceptual Framework Derivations

In this section we show the equilibrium conditions and some simple comparative statics for changes in depression in the conceptual framework presented in section 2, and connect these theoretical predictions to the empirical findings.

We consider a simple two-period model of household production, where the mother maximizes an additively separable utility function:

$$\begin{aligned} \max_{\tau,x,a,h,c,l} U(Q,a,c,l;d) &= v(c,l) - e_d(a) + \beta_d E_d\left[u(Q)\right] \\ \text{s.t.} & \\ Q &= f_d(\tau,x,a;\varepsilon) \\ \Omega - s_d &= \tau + l + h \\ \alpha_d(Y + wh) &= px + c \end{aligned}$$

Mothers are characterized by their depression status $d \in [0,1]$ and maximize utility from current consumption and leisure $v(c,l)$, as well as expected utility from the future quality of the child $E_d\left[u(Q)\right]$, discounted at rate β_d. Future child quality Q is uncertain, and it is determined by a production function f_d that combines time investment τ, monetary investment x, parenting effort a, and a random shock $\varepsilon \sim \phi^*(\varepsilon)$ which is independent of the mother's choices. E_d is defined as the expectation under the distribution $\phi_d(\cdot)$, which captures the idea that subjective beliefs about the shock distribution $\phi_d(\cdot)$ might diverge from the true distribution $\phi^*(\cdot)$ in a way that depends upon depression.

Time investments τ come at the expense of hours worked h and leisure time l, summing up to the total effective time $(\Omega - s_d)$ at the disposal of the mother, which can be reduced by sick days s_d. Monetary expenditure x costs p and is rival to expenditure on other goods c, which both come out of her share α_d of household income (excluding her wages) Y and her wages wh. We do not explicitly model the bargaining process, however, α_d can be related to the woman's bargaining power in a collective model of household bargaining. Parenting effort has a utility cost captured by $e_d(a)$.

Assuming decreasing marginal returns for the main functions in the model,[57] an interior solution is described by the following first order conditions equating the expected marginal benefits and the current marginal costs of time investments τ, monetary investments x, and effort a:

$$\beta_d E_d\left[\frac{\partial u}{\partial Q}\frac{\partial f_d}{\partial \tau}\right] = \frac{\partial v}{\partial l} \tag{A.1}$$

$$\beta_d E_d\left[\frac{\partial u}{\partial Q}\frac{\partial f_d}{\partial x}\right] = p\frac{\partial v}{\partial c} \tag{A.2}$$

$$\beta_d E_d\left[\frac{\partial u}{\partial Q}\frac{\partial f_d}{\partial a}\right] = \frac{\partial e_d}{\partial a} \tag{A.3}$$

Furthermore, the mother will be working ($h^* > 0$) if the market wage is high enough to equal the marginal rate of substitution between (lost) leisure and (additional) consumption, but otherwise she will not work:

$$\frac{\partial v/\partial l}{\partial v/\partial c} \geq \alpha_d w \tag{A.4}$$

[57] More specifically, we assume differentiability for the functions v, e, u and f; positive partial first derivatives for every argument of the functions $v', e', u', f' \geq 0$; concavity for the utility and production functions $v'', u'', f'' \leq 0$, and convexity of the effort-cost function $e'' \geq 0$. The partial second derivatives determine the degree of complementarity or substitutability of the different arguments.

Let us now consider the potential effects of depression (and thus, implicitly, treatment for depression) on outcomes, encompassing the mother and child outcomes we analyze in this paper. For simplicity of exposition, we shut down all the channels but one at a time, and evaluate the expected change in outcomes that would arise if depression had an effect only through this one channel.[58]

Depression effects on time discounting β_d influence how mothers currently value the future expected quality of the child.[59] All else equal, more future-oriented mothers (higher β_d) are more willing to sacrifice leisure l for more time investments τ (higher marginal utility $\frac{\partial v}{\partial l}$ in equation A.1), to forgo consumption c in favor of monetary investments x (higher marginal utility $p\frac{\partial v}{\partial c}$ in equation A.2), and to exert more effort a today (lower disutility of effort $\frac{\partial e_d}{\partial a}$ in equation A.3).

All else equal, lower disutility of effort $e_d(\cdot)$ induced by improvement in depression would directly translate into a higher provision of effort a. Although effort (and cognitive load) could influence virtually every decision that the mother makes, a natural proxy for effort in our dataset is the parenting style index. In other words, having appropriate discipline, being consistent, responsive, encouraging and not harsh when dealing with the child does not require more time or more money, but simply more cognitive effort. Finding no direct effect of treatment on parenting style would indicate the absence of a direct effect of depression on the disutility of parenting effort, so that $e_d = e \quad \forall d$, consistent with the findings of Den Hartog et al. (2003). Results on the other investment choices would depend on the degree of complementarity or substitutability between these inputs; in the case of low levels of complementarity or substitutability $\left(\frac{\partial f}{\partial a \partial x} \approx \frac{\partial f}{\partial a \partial \tau} \approx 0\right)$ we would expect only small decreases in the other inputs. We find that treatment-induced reductions in depression affect parenting style for girls but not boys. Since the treatment effect on maternal depression at the 7-year follow-up (when parenting style was measured) also favored girls, our results appear to be consistent with depression affecting the disutility of effort. The other results on parenting inputs, however, suggest that disutility of effort is not the only channel.

If lower levels of depression decrease sick days s_d leaving more time at the disposal of the mother, we would expect an increase in the probability of working, and in time devoted to both leisure l and time investments τ. The magnitude of the increase would depend on the relative benefits of these two choices: the marginal utility of leisure $\left(\frac{\partial v}{\partial l}\right)$ and the expected future benefit of more time investments $\left(\beta_d E_d\left[\frac{\partial u}{\partial Q}\frac{\partial f_d}{\partial \tau}\right]\right)$, as guided by equation A.1. Changes in the other investment choices would depend on the degree of complementarity or substitutability between these inputs; in the case of low levels of complementarity or substitutability $\left(\frac{\partial f}{\partial \tau \partial x} \approx \frac{\partial f}{\partial \tau \partial a} \approx 0\right)$ we might find small but not significant decreases in other investments. Finding a positive effect on the probability of mother working is consistent with this channel being at work. In fact our findings are somewhat ambiguous as we find a meaningful 10 percent increase in employment, but this is not statistically significant.

An increase in bargaining power α_d would be related to an increase in mother's disposable income in two ways: outside of the home, it might increase her hours worked h and total wage bill wh; inside of the home, it might increase the share household income from other sources Y that is now at her disposal. Both of these mechanisms would lead to an increase in her consumption c and monetary investments x, but possibly a slight decrease in time investments τ if the mother works. Just as for the case of improved time availability, the magnitude of this increase would depend on the marginal rate of substitution between consumption and the return to monetary investments. Again, assuming low levels of complementarity or substitutability $\left(\frac{\partial f}{\partial \tau \partial x} \approx \frac{\partial f}{\partial \tau \partial a} \approx 0\right)$ would mean a small decrease in either time investment τ or effort a.

[58] Empirically we don't have a way of holding all other channels constants, so the results presented in section 6 cannot tell us which channel is at play, but only which channels are compatible with the empirical results.

[59] However, Lempert and Pizzagalli (2010) find that anhedonic individuals (a trait highly correlated with depression) had a higher time discount factor, "possibly due to their decreased responsiveness to immediate rewards."

Recent work has highlighted the importance of maternal beliefs regarding the productivity of investment in children by Cunha et al. (2013, 2016); Boneva and Rau (2016). However, the results for depression-induced changes in expectations and the productivity of parenting inputs are more nuanced, and depend on the relative curvature of the utility function and the production function of human capital.[60] If treating depression improves the productivity of parenting inputs, this would shift the production function f_d upward, increasing the marginal returns to investments, and eventually making the mother better off. All else equal, we would expect an increase in child quality Q. Results for consumption c, work or leisure h, l, and parenting inputs x, τ, a depend on the relative curvature of the utility function $\left(\frac{\partial v/\partial \cdot}{\partial u/\partial Q}\right)$ and the production function of human capital $\left(\frac{\partial f_d}{\partial \cdot}\right)$. If the curvature of the production function is small enough, then an increase in the productivity of inputs would induce mothers to substitute away from consumption and leisure, and into more parenting, according to the marginal productivity of each of these inputs $\left(\frac{\partial f_d}{\partial \cdot}\right)$ as compared to the rate of substitution between future utility from child quality $\left(\beta_d E_d\left[\frac{\partial u}{\partial Q}\right]\right)$ and current increase in the utility of consumption $\left(\frac{\partial v}{\partial c}\right)$ or leisure $\left(\frac{\partial v}{\partial l}\right)$. On the other hand, a high curvature of the production function would induce mothers to target a similar level of child quality, which now could be achieved with fewer inputs. Mothers would then substitute away from investments and into either leisure or consumption. Therefore, an increase in parental investment due to the treatment does not necessarily imply that depression reduces the productivity of maternal inputs.[61]

A similar result is induced by changes in maternal expectations and beliefs $E_d(\cdot)$. However, while improvement in productivity leads to improvements in child quality (even without increases in inputs), that is not necessarily true for the beliefs channel. It is also theoretically possible that, if depression only affected beliefs by making mothers overly pessimistic, this would induce them to over-invest in parenting inputs in order to hedge against negative shocks. In this case, reducing depression could actually lead to a *decrease* in investments and ultimately in child quality.

Summing up, our empirical findings are consistent with the following theoretical formulations of depression: increasing effects of depression on the disutility of parental effort; negative effects of depression on maternal productivity, in the case of a small curvature of the production function; negative effects of depression on time discounting (less patient and future-oriented); negative effects of depression on both effective time (less slack in the time constraint) and bargaining power (less slack in the budget constraint).

[60] We are grateful to Flavio Cunha and Marcos Vera-Hernandez for pointing this out.

[61] A more direct test of this hypothesis would come from the estimation of the production function f_d, and comparing the shape of this production function across treatment groups. However, identification of such a production function would require plausibly exogenous variation in the inputs in both treatment and control groups, see for example the work of Attanasio (2015).

B Data appendix

Section D reports summary statistics and descriptive correlations of the rich measures that we use throughout the paper. We report summary statistics for maternal decision-making outcomes (Table D.1), children's outcomes (Table D.2), and potential mediators (Table D.3). We then provide a description of the mental health scales available in the data (Table D.4), of the six outcome indices used (Table D.5), and of the parental investments and child development inventories (Table D.6). Finally, we show the conditional correlations of observable characteristics at baseline, 6, and 12 months with parental investment at age 8 (Table D.7), and with child developmental outcomes at age 8 (Table D.8).

In section E we report further information on the mothers who were excluded from the intervention because they were not clinically depressed at baseline. Table (E.9) checks for balance of baseline observable characteristics between the treatment and control clusters for mothers who were non-depressed at baseline. This is an additional check of the validity of the randomization that leverages information from the non-experimental sample. Table (E.10) reports the average outcomes for the baseline non-depressed sample, and the baseline depressed mother-child dyads (both treated and control).

B.1 Baseline Balance

Table 1 shows baseline characteristics for the sample of women who were interviewed at baseline, the 1-year follow-up (the target sample for the 7-year follow-up), and the 7-year follow-up.[62] At baseline, the samples were balanced along observable characteristics (p-value of the joint test is 0.13), with the exception of the composition of household: treated mothers were 10 percentage points ($p < 0.01$) more likely to have a grandmother of the index child (henceforth, just grandmother, though 90% of the cases this is the mother-in-law) living with them. This pattern remains in the 1-year follow-up sample, treated women are 12 percentage points more likely to have a grandmother of the index child, and the 7-year follow-up sample (difference of 11pp).

Additionally, there are several notable differences in characteristics between treated and control groups in both follow-up samples. Treated women reported 0.75 more years of education, and 0.29 fewer children. The 7-year follow-up sample displays similar patterns: perceived social support and presence of grandmothers were still greater in the intervention arm, and treatment women had more education and fewer children. Jointly testing all variables, we fail to reject the null hypothesis that treatment and control clusters were balanced in the 1-year follow-up sample ($p = 0.12$). While the magnitudes of the differences between treatment and control were similar using the 1-year sample, we reject the null that the 7-year follow-up sample is balanced by intervention arm. The differences between treatment and control clusters in mother's education, parity, and presence of a grandmother are also evident in the 300 women who were excluded at baseline (ie, perinatally non-depressed), and although the differences are not statistically significant, they are very similar in magnitude. See Table E.9. Differences that are observed between treatment and controls clusters of the baseline non-depressed sample would be subtracted away in a difference-in-differences analysis, providing an alternative method to "controlling" for baseline imbalance between treatment and controls clusters.

[62] The wealth index, also used as a control, is the first principal component of the following measures of house quality and asset ownership: brick walls, electricity, piped water, flush toilet, water pump, washing machine, air conditioning, refrigerator, TV, radio, bicycle, and car. Additionally, it includes if the mother reports having enough money for food, and the assessor-rated SES measure (5-point Likert scale from poorest to richest).

B.2 Attrition

Although attrition was not differential by treatment status, the changes in balance from the baseline to 7-year follow-up samples are due to attrition. LTFU (attritors) and mothers that were re-enrolled at year 7 were fairly similar along most characteristics (Table F.11); however, LTFU mothers were less empowered, perceived less social support, and were more likely to normally work. Table F.11 also shows baseline characteristics of the LTFU women by treatment group. Consistent with the similar balance between the original baseline sample and the 7-year follow-up sample reported in Table 1, there were no differences between treated and control LTFU mothers at the 5% significance level, and we fail to reject the joint test that characteristics of attritors in treatment were different to controls (p=0.59).[63]

Another way to investigate whether differential attrition between the 1-year follow-up and 7-year follow-up affects our estimates is to compare *short-run* treatment effects on maternal depression outcomes calculated using the full samples at 6-month and 1-year follow-ups (N= 818 and 791 respectively) to treatment effects calculated using the 7-year follow-up sample (N=585). Table F.12 shows treatment effects at 6 and 12 months for five mental health outcomes (depressed, depression severity, disability score, functioning, and perceived social support) using the full and 7-year follow-up samples. The comparison reveals very little difference in treatment effects between the two samples: differences range between 2 and 5 percent of a standard deviation of that outcome, with an average different of 3 percent of a standard deviation across the five outcomes. Furthermore, the differences in treatment effects between the two samples are not statistically significant for any of the mental health outcomes (individually or jointly). Estimating a Seeming Unrealated Regression (SUR) model of the mental health outcomes at 6 and 12 months as a function of treatment and (future) attrition status and their interaction, the joint test of the coefficient on the interaction term yields a p-value=0.60 for the 6-month outcomes and 0.95 for 12-month outcomes.

We take two approaches to account for attrition: one parametric and one non-parametric. First, we present estimates of the main results using Inverse Probability Weighting, where the weights were calculated 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. Because neither test of whether rates of attrition differ by treatment status nor whether characteristics of attritors differ by treatment status was rejected, we do not adjust the treatment effects presented in the main tables. The attrition-corrected results are presented in Table F.13.

Finally, we investigate whether attrition was differential by gender. Table F.14 shows the sample flow by gender of the index child and treatment status. Attrition was not differential by treatment status among women who had girls, however it was differential by treatment status among women who had boys. For example, 60% of treatment dyads were located at 7-year follow-up compared to 69% of control. Furthermore, the differential attrition in boy dyads was driven entirely by attrition between the 1-year and 7-year follow-up. Since our results are mainly driven by girls, the finding that attrition was greater among treated boys does not affect the interpretation of our results. Nevertheless, we present attrition bounds of treatment effects by gender in Table F.15. We still find that treatment impacted empowerment, and parental investment of time and monetary resources in mothers with girls. Meanwhile, the bounds of treatment effects for boys include zero and are generally fairly wide.

[63] Reasons for attrition were extensively monitored during the first year, but not thereafter, although the re-enrollment study at the 7-year follow-up did record reason for LTFU in some cases. Reasons such as death of a child or death of the mother are also not differential by treatment status. Additionally, attrition due to moving or refusing were also not differential by treatment status.

B.3 Robustness

Section G shows several robustness checks in terms of controls used, construction of indices, and observational associations between baseline covariates and outcomes of interest. The results are not sensitive to inclusion of different sets of controls (Table G.16), or the construction of summary indices using factor scores (Table G.17), or the fact of using only outcomes measured in year 7 instead of the full trajectory of mother financial empowerment (Table G.18).

We then show that the observed conditional correlations between post-treatment variables are reasonable and of the expected sign: higher investment in children is usually associated with better child development measures (Table G.19), and maternal depression over the years is correlated with lower maternal empowerment (Table G.20).

B.4 Difference-in-differences with prenatally non-depressed mothers.

Since the treatment was randomized, a single difference is sufficient to estimate the causal effect of the intervention. However, a difference-in-differences specification provides an additional robustness check against cluster-level shocks. In particular, over the 7 years since the initial randomization, shocks potentially correlated with treatment assignment could undermine randomization. To address this concern, we estimate a difference-in-differences specification, exploiting availability of data from both treated and control clusters on women who were excluded from the trial because they were not perinatally depressed at baseline. This approach assumes that any cluster-specific shocks similarly influence outcomes for the baseline depressed and the baseline non-depressed women in the cluster. The estimating equation is

$$Y_{ic} = \alpha + \beta T_c \times Depressed_{ic} + \delta Depressed_{ic} + \eta T_c + \mathbf{\Gamma}'\tilde{\mathbf{X}}_{\mathbf{ic}} + \lambda_{LHW} + \varepsilon_{ic} \tag{B.1}$$

where $Depressed_{ic}$ is a dummy that equals one if the mother was prenatally depressed at baseline. The coefficient β on the interaction $T_c \times Depressed_{ic}$ will pick up the intervention effect. The parameter η is of interest as it indicates the average difference between treated and control clusters for mothers who were not depressed at baseline.[64] If η were positive and significant, this would suggest that treatment clusters experienced favorable shocks relative to control. Alternatively, it could signify positive spillovers of the intervention to women in treated clusters who were not offered the treatment (because they were not depressed at baseline).[65] The parameter δ provides an estimate of the difference in outcomes between control mothers who were and were not prenatally depressed. This descriptive difference gives an idea of the magnitude of the correlation between perinatal depression and long-term outcomes, providing a useful benchmark for the potential effect of treatment.[66]

The vector of controls in $\mathbf{\Gamma}'\tilde{\mathbf{X}}_{\mathbf{ic}}$ is different to that in equation 5.1 because we do not have baseline characteristics for prenatally non-depressed mothers. Instead, we include time-invariant individual specific demographic characteristics.[67] This specification affords a further sensitivity check, allowing us to control

[64] We test if there is balance along fixed demographic characteristics among prenatally non-depressed women along the dimension of randomization and we cannot reject that the two samples are different (with p-value=0.38, Table E.9).

[65] We could alternatively include Union Council (UC) fixed effects, γ_c, which absorb the indicator for T_c, that is, being assigned to a treatment cluster. Since the parameter η is of interest, we present the results using the more parsimonious specification. However, the results are similar when including UC fixed effects.

[66] Given negative selection into depression, and under the assumption that women who were not prenatally depressed remain less likely to be depressed through the next seven years than women who were prenatally depressed, the parameter δ could provide an upper bound on treatment effects. However, this assumption may not hold, so we cast this comparison in descriptive terms.

[67] The individual specific controls are mother's age and its square, mother's and father's education, parity at baseline (estimated based on parity in 2013 and the reported number of children born since the index child), date of interview and interviewer fixed

for Lady Health Worker fixed effects (λ_{LHW}) which would be collinear with the treatment variable in the main specification.[68] Tables G.21 and G.22 show the results from a difference-in-differences estimation discussed in the main text.

B.5 Heterogeneous treatment effects

We test for heterogeneity of treatment effects by baseline depression severity, education, wealth, family structure, mother's age, and whether the index child is the first child. We present heterogeneous treatment effects by estimating an extension of the main equation:

$$Y_{ic} = \alpha + \beta_1 Het_i + \beta_2 T_c + \beta_3 Het_i \times T_c + \mathbf{\Gamma' X_{ic}} + \varepsilon_{ic} \tag{B.2}$$

where Het_i is the dimension of heterogeneity we are exploring, measured at baseline. The coefficient on the interaction term, β_3, estimates the differential effect of the intervention along that specified dimension. Results are in Section H.

B.6 Treatment effects on individual outcomes

In section I, we show in detail all of the treatment effects for the individual outcomes we measure in the dataset. We group outcomes based on the index they belong to, and show the treatment effect on the index as a whole as well as for each of the individual outcomes, controlling for multiple hypothesis testing within each table. Table I.27 reports the results for mother financial empowerment using only measures available at the 7-year follow-up; tables (I.28-I.31) report the results for the child development indices; tables (I.32-I.35) show the treatment effects for all of the mediators; finally tables I.36 and I.37 respectively report the results for parental investments and child development measures available in the 6 and 12 month follow-up.

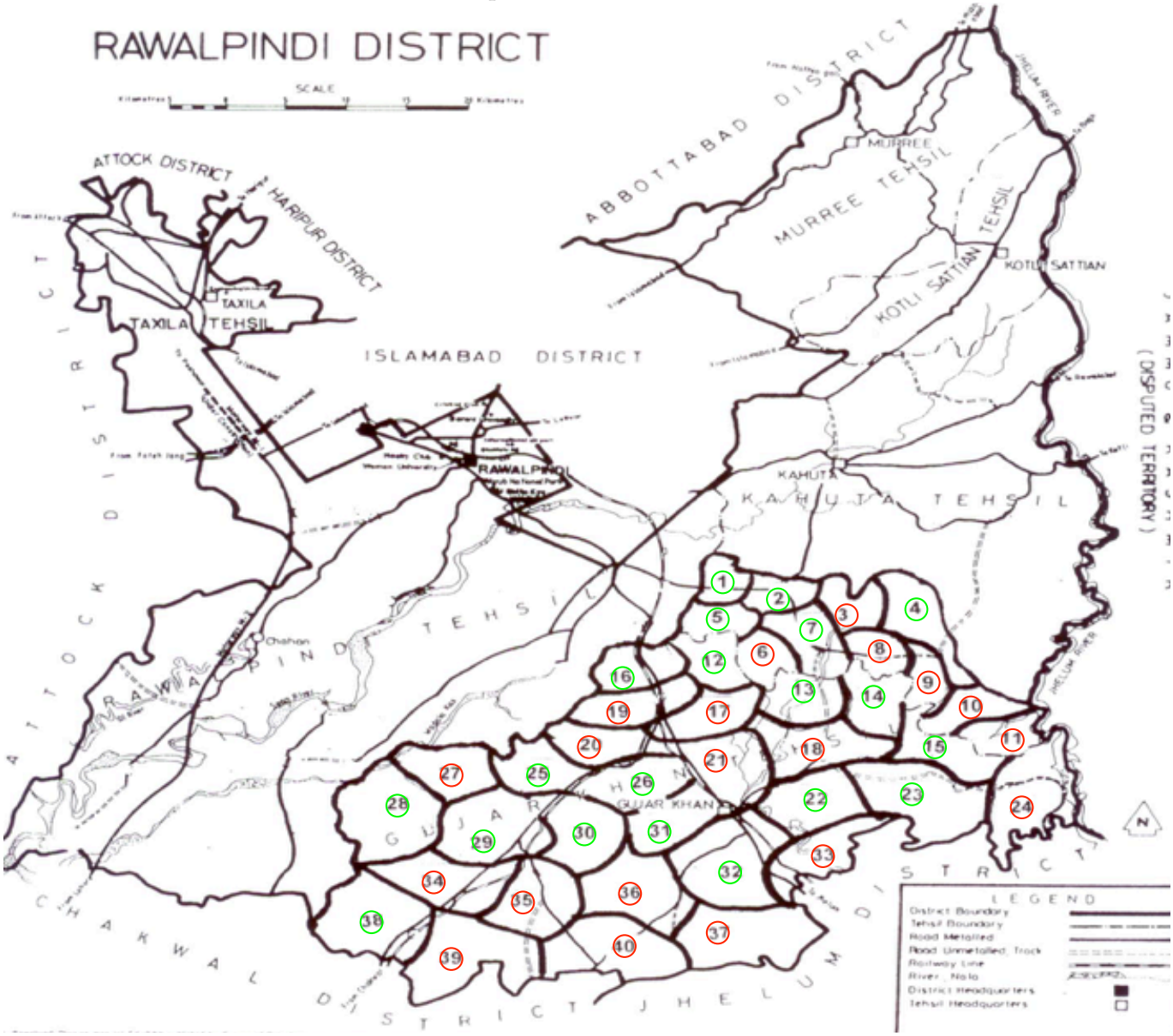
In section J, we report treatment effects for each of the components that make up the scales used to measure parenting (HOME score, Table (J.38); school quality, Table (J.39)) or scales used to measure child development (WPSSI, Table (J.40); Spence Anxiety Scale, Table (J.41); Strength and Difficulties Questionnaire, Table (J.42)). As above, we control for multiple hypothesis testing within each table.

effects.

[68] LHWs delivered the intervention. In the 7-year follow-up data, we only identify the LHWs who are currently serving the families and this is not an identical set to the original 40 LHWs from the intervention since some LHWs moved, retired, or stopped work for other reasons. At the 7-year follow-up, there were a total of 65 LHWs. We present the results without LHW fixed effects because shifting and reallocation of LHWs may be endogenous to treatment; however, the results are similar with LHW FEs (Table G.21).

C Appendix Figures

Figure C.1 – Map of treatment and control clusters



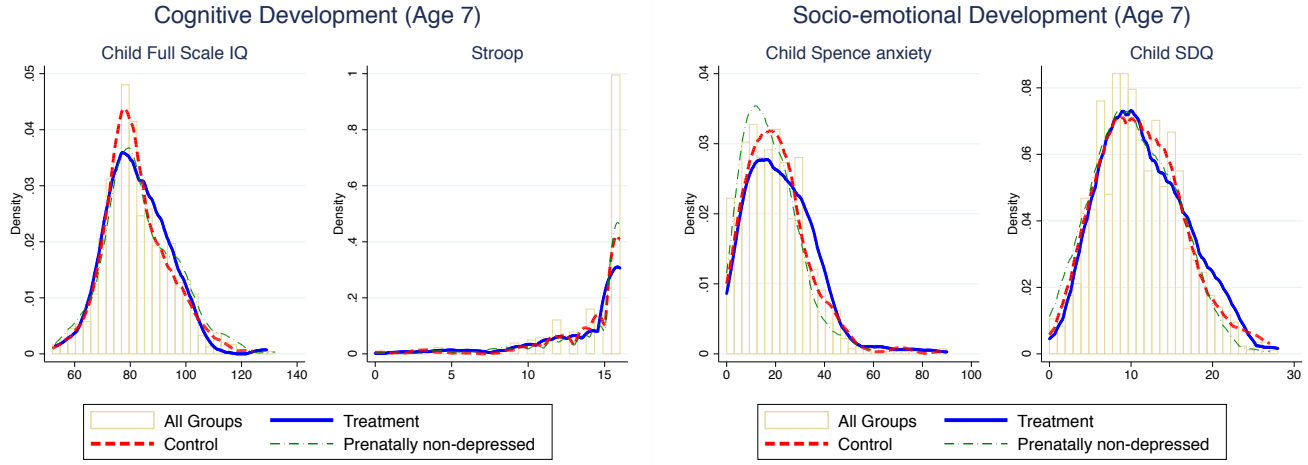
Notes: Treatment clusters are indicated by green circle, and control are indicated by red.

Figure C.2 – Sample sizes

	Treatment		Control		Total		T-C p-value
Pregnant women identified	**1967**		**1931**		**3898**		
refusals	140	7%	159	8%	299	8%	0.19
not found	40	2%	40	2%	80	2%	
Screened at baseline	**1787**	91%	**1731**	90%	**3518**	90%	0.20
excluded	138	8%	138	8%	276		
Depressed (completed survey)	**463**	26%	**440**	25%	**903**	26%	0.74
boys at birth	223	48%	226	51%	449	50%	0.95
Attrited btw baseline & 1yr	**103**	22%	**95**	22%	**198**	22%	
total child mortality/illness	52	11%	41	9%			0.34
stillbirths/abortions	15	3%	21	5%			0.24
infant mortality (of live births)	31	7%	18	4%			0.10
mother mortality	2	0%	3	1%			0.99
refused	11	2%	11	3%			0.90
moved	38	8%	40	9%			0.64
Complete dyads at 1yr	**360**		**345**		**705**		
Attrited btw 1yr & 7yr	**72**	20%	**51**	15%	**123**	17%	0.07
LTFU	62	13%	44	10%	106	12%	0.10
child mortality	4		3				
mother mortality	3		1				
child disabled/not eligible	2		2				
Attrited btw baseline & 7yr	**174**	38%	**145**	33%	**319**	35%	
child death/illness	55	32%	44	30%	99	31%	0.37
child death (of live births)	35	8%	21	5%	56	6%	0.09
mother death	5	3%	4	3%	9	3%	0.80
refused/moved/LTFU/not eligible	112	64%	96	66%	208	65%	0.39
Complete dyads at 7yr	**289**	62%	**295**	67%	**584**	65%	0.15
dyads at 7yr who completed 1yr	289	80%	295	86%		83%	0.07
in our data	289		296				0.13

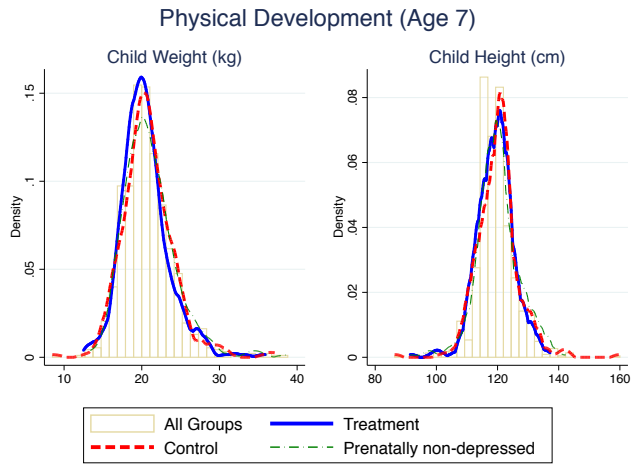
Notes: Table shows the sample flow from the start of the intervention when pregnant women were identified to the 7-year follow-up. Percentages are not defined in the same way from row to row. P-values of simple χ^2 tests of differences in rates across treatment and control groups are in the last column.

Figure C.3 – Distributions of selected components of outcomes at the 7-year follow-up

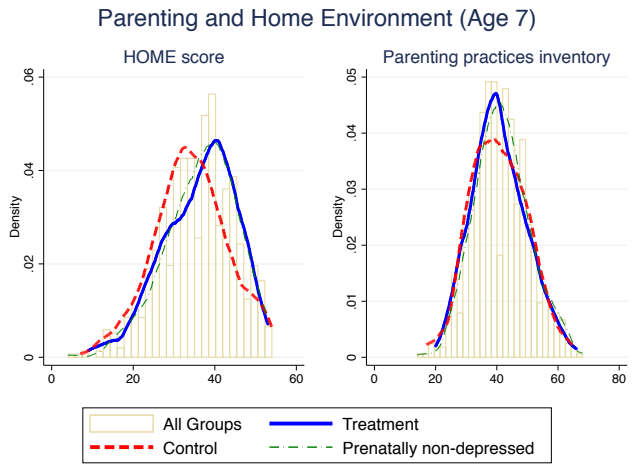


(a) Cognitive development

(b) Socio-emotional development



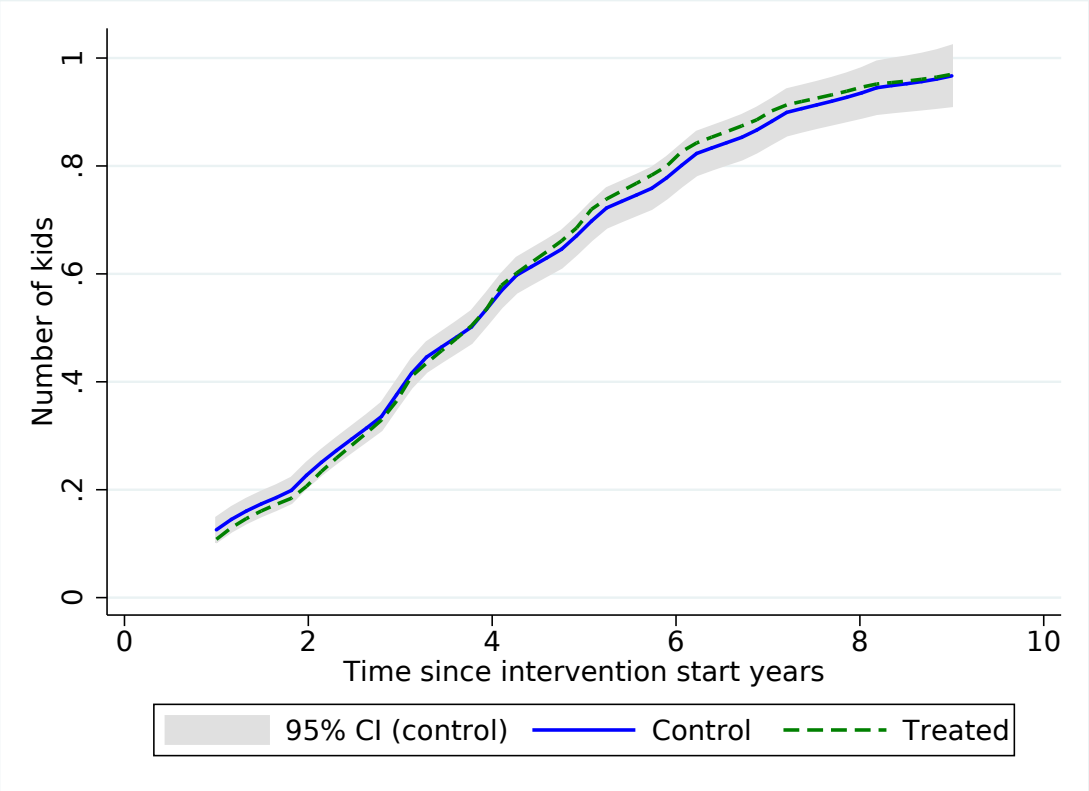
(c) Child height and weight (7 years)



(d) Parenting and home environment

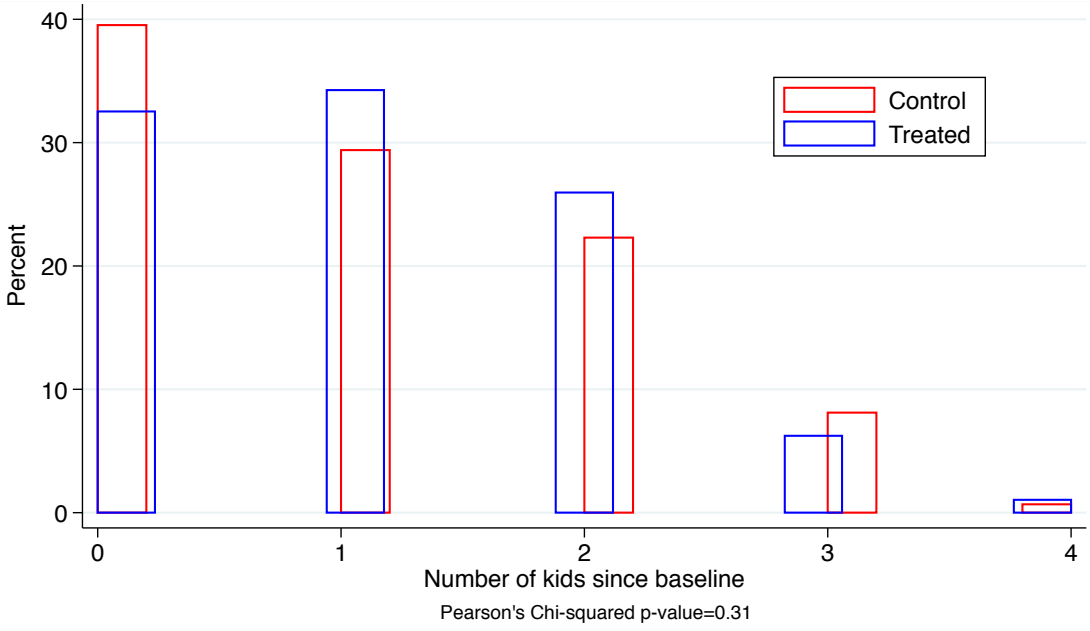
Notes: Distributions of child outcomes at the 7-year follow-up for main outcome variables, by treatment arm. Distributions for prenatally non-depressed mothers are also plotted for comparison. Histograms of the data for all groups combined (treatment, control, and non-depressed) are plotted in the background.

Figure C.4 – Effects on fertility and birth-spacing



Notes: This figure show the average number of births women reported since the start of the intervention until the 7-year follow-up. Birth histories were constructed from the listing of children and their ages at the 7-year follow-up. 95% confidence interval, not adjusted for clustered errors or autocorrelation, is presented (and is thus tighter than the true CI).

Figure C.5 – Effects on fertility: Number of children born since treatment



Notes: This figure plots the histograms of total number of births women reported since the start of the intervention until the 7-year follow-up. Birth histories were constructed from the listing of children and their ages at the 7-year follow-up.

D Description of indices and measures

Table D.1 – Summary Statistics for Women's Decision Making

	Mean	SD	Median	Min.	Max.	Total Obs
Mother's financial empowerment[a]	0.21	1.2	0.4	-1.8	6.0	885
Mother controls spending (6m)	0.57	0.5	1.0	0.0	1.0	584
Mother gets pocket money (1y)	0.74	0.4	1.0	0.0	1.0	582
Mother controls spending (7y)	0.60	0.5	1.0	0.0	1.0	885
Mother gets pocket money (7y)	0.65	0.5	1.0	0.0	1.0	859
Mother employed (7y)	0.12	0.3	0.0	0.0	1.0	885
Mother's income (100s PKR) (7y)	4.33	15.5	0.0	0.0	100.0	876
Parental investment (monetary)[a]	0.23	1.0	0.3	-5.6	2.5	885
HOME: Learning materials	2.86	1.5	3.0	0.0	6.0	885
HOME: Physical environment	4.86	2.4	5.0	0.0	8.0	885
Monthly expend. on educ (ln) (7y)	7.37	1.3	7.6	2.3	10.6	748
Expected grade attainment	14.45	2.5	16.0	0.0	21.0	881
Private school	0.47	0.5	0.0	0.0	1.0	878
School quality	0.00	2.3	-0.1	-5.3	4.4	850
Parental investment (time-intensive)[a]	0.23	1.0	0.2	-2.3	3.6	885
HOME: Enrichment	2.94	1.4	3.0	0.0	5.0	885
HOME: Family companionship	3.36	1.7	3.0	0.0	6.0	885
HOME: Family integration	2.82	1.0	3.0	0.0	12.0	885
Frequency of mother play	0.19	0.3	0.0	0.0	1.0	885
Someone helps with studies	0.58	0.5	1.0	0.0	1.0	885
Parenting style[a]	0.05	1.0	0.2	-4.3	2.2	885
PPI: Not harsh	13.97	8.1	14.0	0.0	33.0	885
PPI: Not harsh for age	8.61	1.4	9.0	0.0	9.0	885
PPI: Consistent	9.88	3.5	10.0	0.0	18.0	885
HOME: Responsivity	8.96	1.6	10.0	1.0	10.0	885
HOME: Encouragement of maturity	5.23	1.6	5.0	0.0	7.0	885
HOME: Emotional climate	4.74	1.9	5.0	0.0	8.0	885
Fertility trajectory[a]	-0.08	1.1	-0.0	-4.3	2.7	885
Ideal # kids (7y)	3.29	1.2	3.0	0.0	9.0	874
# kids born past 7yrs	1.12	1.0	1.0	0.0	5.0	885
Pregnant at 1y	0.07	0.3	0.0	0.0	1.0	583
Index not last child	0.67	0.5	1.0	0.0	1.0	885

Notes: Index variables, created such that the control group has mean 0, standard deviation 1, are in bold. The individual variables that make up each index are listed below. The sample includes the intervention (baseline depressed mothers in treatment and control groups) and non-intervention (baseline non-depressed mothers) groups.

[a] Index variables were created following Anderson (2008), with positive values always associated with more favorable outcomes for all indices.

Table D.2 – Summary Statistics for Children's Outcomes

	Mean	SD	Median	Min.	Max.	Total Obs
Physical development index[a]	0.12	0.9	0.3	-4.3	2.3	885
Height-for-age (z)	-0.82	1.1	-0.8	-4.9	3.2	879
BMI-for-age (z)	-0.96	1.2	-1.0	-4.7	4.6	879
Not stunted (height >-2SD)	0.86	0.3	1.0	0.0	1.0	885
Motor function	-0.00	0.7	0.2	-3.4	0.8	885
No hospitalization	0.85	0.4	1.0	0.0	1.0	885
No severe illness	0.73	0.4	1.0	0.0	1.0	885
No eyesight problems	0.96	0.2	1.0	0.0	1.0	885
No hearing problems	0.98	0.1	1.0	0.0	1.0	885
Cognitive development index[a]	0.06	1.0	0.1	-3.7	2.7	885
WPPSI: Verbal comprehension	86.41	14.5	85.0	45.0	146.0	882
WPPSI: Visual spatial	86.94	14.6	86.0	45.0	148.0	883
WPPSI: Fluid reasoning	78.55	12.5	77.0	45.0	133.0	884
WPPSI: Working memory	99.62	15.9	100.0	58.0	146.0	884
WPPSI: Processing speed	77.74	10.1	77.0	45.0	112.0	877
Urdu score	6.73	3.7	6.0	0.0	12.0	877
Math score	9.35	3.5	11.0	0.0	16.0	876
Executive function (Stroop)	14.15	3.1	16.0	0.0	16.0	885
Grade-for-age	0.75	0.3	0.8	0.0	1.5	872
Socio-emotional development index[a]	0.02	1.0	0.1	-4.3	2.2	885
SDQ: Emotional	2.20	2.0	2.0	0.0	10.0	885
SDQ: Conduct problems	3.28	2.1	3.0	0.0	10.0	885
SDQ: Hyperactivity	3.55	2.6	3.0	0.0	10.0	885
SDQ: Peer problems	1.98	1.6	2.0	0.0	8.0	885
SDQ: Prosocial	2.40	2.5	2.0	0.0	10.0	885
SCAS: Panic and agoraphobia	1.51	2.8	0.0	0.0	25.0	885
SCAS: Separation	5.75	4.1	6.0	0.0	17.0	885
SCAS: Injury fear	5.89	3.7	6.0	0.0	15.0	885
SCAS: Social phobia	2.15	2.7	1.0	0.0	17.0	885
SCAS: Obsessive-compulsive	1.33	2.2	0.0	0.0	15.0	885
SCAS: General anxiety	3.42	3.1	3.0	0.0	18.0	885
Child survival index[a]	0.12	0.9	0.2	-7.4	1.5	885
Share of boys	0.51	0.2	0.5	0.0	1.0	885
# died <1 year of age	0.25	0.6	0.0	0.0	6.0	881
# died btw 1 & 5 years old	0.04	0.2	0.0	0.0	3.0	881
# died > 5 years old	0.02	0.1	0.0	0.0	1.0	882

Notes: Index variables, created such that the control group has mean 0, standard deviation 1, are in bold. The individual variables that make up each index are listed below. The sample includes the intervention (baseline depressed mothers in treatment and control groups) and non-intervention (baseline non-depressed mothers) groups.

[a] Index variables were created following Anderson (2008), with positive values always associated with more favorable outcomes for all indices.

Table D.3 – Summary Statistics for Mediators

	Mean	SD	Median	Min.	Max.	Total Obs
Mother's physical health[a]	0.17	1.1	0.3	-3.6	3.6	885
Mother never been unwell (7y)	0.67	0.5	1.0	0.0	1.0	885
Overall health (0-4) (7y)	1.95	0.9	2.0	0.0	4.0	885
Healthy days in past 30 (7y)	26.52	7.1	30.0	0.0	30.0	621
Weight (kg) (6m)	54.18	11.5	52.0	30.0	116.0	584
Weight (kg) (1y)	52.99	11.6	50.0	30.0	115.0	585
Husband's income trajectory[a]	0.04	1.1	0.3	-5.7	4.1	824
Monthly income (ln) (7y)	9.01	1.4	9.2	2.3	13.8	719
Monthly income (ln) (1y)	7.28	2.3	8.0	2.3	10.6	554
Monthly income (ln) (6m)	7.30	2.2	8.0	2.3	10.5	554
Relationships[a]	0.17	1.0	0.4	-3.7	2.7	880
Marital quality scale (7y)	5.31	1.4	6.0	0.0	6.0	859
Relationship husband (7y)	4.06	0.9	4.0	1.0	5.0	859
Husband nonviolent (7y)	0.76	0.4	1.0	0.0	1.0	859
Relationship m-in-law (7y)	3.46	1.0	4.0	1.0	5.0	566
Marital quality scale (1y)	3.56	1.0	4.0	0.0	4.0	582
Relationship husband (1y)	4.04	0.9	4.0	1.0	5.0	582
Husband nonviolent (1y)	0.72	0.4	1.0	0.0	1.0	582
Relationship m-in-law (1y)	4.81	2.5	4.0	1.0	9.0	585
Grandmother trajectory index[a]	0.17	1.1	0.1	-1.0	1.6	885
Grandmother present (7y)	0.39	0.5	0.0	0.0	1.0	885
Grandmother present (1y)	0.48	0.5	0.0	0.0	1.0	584
Grandmother present (6m)	0.54	0.5	1.0	0.0	1.0	584

Notes: Index variables, created such that the control group has mean 0, standard deviation 1, are in bold. The individual variables that make up each index are listed below. The sample includes the intervention (baseline depressed mothers in treatment and control groups).
[a] Index variables were created following Anderson (2008), with positive values always associated with more favorable outcomes for all indices.

Table D.4 – Descriptions of mental health scales

Maternal Mental Health	
Depressed (Baseline,6m,1y,7y)	Binary indicator for clinically diagnosed major depressive episode at the time of interview^f. Diagnosis of major depression was made using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), a semi-structured interview for making the major DSM-IV Axis I diagnoses (First et al., 2002). At the 7y follow-up, the SCID interview was administered by trained assessors, whereas the baseline through 1y diagnoses were done by clinical psychiatrists. The SCID contains 10 questions about depressive symptoms, which were individually recorded on a 3-point scale (only at the 7y follow-up). We use the SCID interview to construct a measure of depression severity (number of depressive symptoms) at the 7y follow-up^f.
Hamilton (Baseline,6m,1y)	Hamilton depression severity scale^f is a clinician-evaluated depression severity measure. Scale based on a questionnaire, ranging from 0-23, higher values indicate more severe depression. It is one of the most widely used and accepted outcome measures for evaluating the severity of depression symptoms (Hamilton, 1960).
BDQ (Baseline,6m,1y)	Brief Disability Questionnaire (BDQ)^f. A 8-item questionnaire gaging the extent to which the mother's health condition disables her from doing physical activities, participating in hobbies, and taking part in family activities. Each item is recorded on a 3-point scale.
GAF (6m,1y)	Generalized Assessment of Functioning (GAF). Scale between 0-100, where larger values indicate better functioning. GAF is assigned by the clinician, based on criteria and scaling set out by the DSM-IV-TR (p. 34): Consider psychological, social, and occupational functioning on a hypothetical continuum of mental health-illness. Do not include impairment in functioning due to physical (or environmental) limitations.
MSPSS (Baseline,6m,1y,7y)	Multi-dimension Scale of Perceived Social Support (MSPSS) is designed to measure perceptions of support from 3 sources: family, friends, and spouse (Zimet et al., 1988). The questionnaire is comprised of a total of 12 items, each measured on a scale of 1-7. Higher values indicate more perceived social support.

Notes: The follow-up waves of when each outcome was measured is listed in parentheses under the name in the first column. Mental health index variables are generated following Anderson (2008), a GLS-weighted average of outcomes within the index group. More positive values of the index indicate more favorable outcomes (thus certain outcomes, as indicated above with ^{f}, are "flipped", i.e., such that more positive values are associated with favorable outcomes).

Table D.5 – Descriptions of Outcome Variable

Parental Investment	
Parenting Style Index	A standardized weighted average of the 4 subscales of the Parenting Practices Inventory (harsh, harsh for age, consistent, and appropriate disciplining) and the following 3 subscales of the HOME inventory: responsivity^{o}, encouragement of maturity^{o}, and emotional climate^{o}.
Time-intensive Investment Index	A standardized weighted average of frequency of mother play (to help learn new things) with index child, frequency of father play with index child, if anyone else in the family helps child with studies, and the following 3 subscales of the HOME inventory: enrichment, family companionship, and family integration.
Monetary-intensive Investment Index	A standardized weighted average of family education expenditure in past month, mother's expected grade attainment for index child, whether index child attends a private school, school quality* (index constructed as the primary component of class size*, number of teachers in the school*, number of rooms in the school*, classroom amenities* [3 items: backboard, backboard functional, other materials], school amenities* [10 items: school has office, playground, computers, library, clean drinking water, toilets for girls]), and 2 subscales of the HOME inventory (learning materials and physical environment*).
Child Development	
Cognitive Development Index	A standardized weighted average of the 5 subscales of the WPPSI IQ*, Urdu score*, Math score*, Stroop executive function test*, and current grade (teacher report)*. The Stroop-like Day/Night test gages inhibition and working memory. Basic literacy and numeracy tests were administered, providing math and Urdu scores based on the number correct out of 16 and 12 respectively.
Physical Development Index	A standardized weighted average of (flipped) binary indicators of whether the child had been hospitalized, has had a severe illness, has eyesight or hearing problems, motor function (assessed using the Grooved Pegboard Test, which asks the child to place pegs in a correct orientation on a board and records the amount of time the child took to complete the task, also flipped), and measured height-for-age* and weight-for-age* (Z scores calculated according to WHO criteria) and if the child was not stunted* or thin* (according to WHO criteria).

Continued on next page

Table D.5 – continued from previous page

Socio-emotional Development Index	A standardized weighted average of the 5 subscales of the Strengths and Difficulties Questionnaire (4 subscales that make up the total SDQ score + the prosocial component) and 6 subscales of the Spence child anxiety scale (SCAS).

Notes: Items indicated with an asterisk (*) indicates outcomes measured by observation (ie, assessor observed or administered test), and ^{o} indicates the subscale comprises of some (but not all) direct observation. All other outcomes are self-reported by the mother. Index variables are generated following Anderson (2008), a GLS-weighted average of outcomes within the index group. More positive values of the index indicate more favorable outcomes (thus certain outcomes, as indicated above, are "flipped", i.e., redefined as the value subtracted from the maximum). An alternative construction of indices using the factor scores was also used, and the results, which are robust to the alternative method of defining the indices, are presented in the Appendix (see Table G.17).

Table D.6 – Descriptions of Scales and Inventories

Parental Investments	
HOME	Home Observation for Measurement of the Environment (HOME) Inventory (Caldwell and Bradley, 1984). The HOME assessment used in our experiment contains 54 items in total and 8 subscales: (1) Learning Materials; (2) Encouragement of Maturity; (3) Physical Environment; (4) Responsivity of Parent to Child; (5) Family Companionship; (6) Family Integration; (7) Variety in Daily Stimulation and Enrichment; and (8) Emotional Climate. 19 items are based on observation. The HOME assessment is one of the most used child assessments. For example, the short form of the HOME is the primary measure of the quality of a child's home environment included in the NLSY79 child survey.
PPI	The Parent Practices Interview (PPI) is a modified (shortened) version of a 72-item questionnaire adapted from the Oregon Social Learning Center's Discipline Questionnaire and revised for young children. It is composed of four subscales –Harsh Discipline, Harsh for Age, Inconsistent Discipline, and Appropriate Discipline – rated on a 4-point scale ranging from 0 (never) to 3 (always). The PPI has bee used in several studies, including a study of the effectiveness of parent and teacher training for Head Start mothers and their 4 year old children and Head Start teachers (Kaplow et al., 2001).
School Quality	An index constructed as the primary component of class size, number of teachers in the school, number of rooms in the school, classroom amenities [3 items: backboard, backboard functional, other materials], and school amenities [10 items: school has an office, a playground, a library, a water source, clean drinking water, toilet for girls, fencing, computers, and if books and computers were visibly in use]. All items within the index were reported by the interviewer, who visited the school.
Child Development	
WPPSI	The Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV), fourth edition, is an intelligence test designed for children ages 2.5 years to 7.5 years. WPPSI-IV provides primary index scales for verbal comprehension (VCI), visual-spatial (VSI), fluid reasoning (FRI), working memory (WMI), and processing speed (PSI).
SCAS	The Spence Children's Anxiety Scale (SCAS) to assess anxiety (Spence, 1998). The SCAS is also parent administered and consists of six different subscales in addition to an overall anxiety score: panic and agoraphobia, separation anxiety, physical injury fears, social phobia, obsessive-compulsive problems, and generalized anxiety. The sum of items ranges from 0 to 114. Higher values indicate more anxious behavior.

Continued on next page

Table D.6 – continued from previous page

SDQ	The Strengths and Difficulties Questionnaire (SDQ) is used to measure behavioral and emotional problems. The SDQ is parent administered and has been validated in Pakistan (Syed et al., 2009). The questionnaire consists of 20 questions about the child's difficulties in four areas (emotional, conduct problems, hyperactivity, and peer problems) and a positive prosocial domain. The SDQ total difficulties score is generated by addition of the problem scale scores and ranges from 0 to 40. Higher values indicate more behavioral and emotional problems.
Stroop	The Stroop test is considered to measure selective attention, cognitive flexibility and processing speed, and it is used as a tool in the evaluation of executive functions. An increased interference effect is found in disorders such as attention-deficit hyperactivity disorder, or a variety of mental disorders such as schizophrenia, addictions, and depression. A Stroop-like Day/Night test was administered and is measured as the number correct out of 16.
Grooved Pegboard	The Grooved Pegboard is a manipulative dexterity test. This unit consists of 25 holes with randomly positioned slots. Pegs, which have a key along one side, must be rotated to match the hole before the can be inserted. This test requires more complex visual-motor coordination than most pegboards. The measure used in this paper is a factor score of the time (in minutes) to complete the task using both the dominant and nondominant hands, the number of pegs dropped, and the number of peg not placed. Results are similar using the time to complete task using the dominant hand (or nondominant hand), however more meaningful variation, assessed by how the measure varying with key covariates, was generated using the factor score.

Table D.7 – Correlates of Parental Investment Behavior at Age 7

	Time investment index			Monetary investment index			Parenting style index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Girl	0.01	0.01	0.04	−0.38***	−0.35***	−0.33***	0.04	0.05	0.06
	(0.11)	(0.11)	(0.09)	(0.08)	(0.08)	(0.07)	(0.11)	(0.11)	(0.09)
Age of index child	0.14	0.16	0.28	−0.13	−0.07	−0.08	0.52*	0.55*	0.57
	(0.27)	(0.27)	(0.25)	(0.37)	(0.38)	(0.41)	(0.30)	(0.29)	(0.34)
Wealth score (at baseline)	0.06**	0.06**	0.08***	0.08**	0.08**	0.10***	0.00	0.00	0.02
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Mother's years of education	0.05***	0.05***	0.03**	0.07***	0.07***	0.06***	0.05***	0.05***	0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Father's years of education	0.04**	0.03**	0.01	0.07***	0.06***	0.06***	0.01	0.01	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Mother's age	0.06	0.06	0.01	0.03	0.03	−0.01	0.18**	0.18**	0.12*
	(0.05)	(0.05)	(0.05)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Mother's age^2	−0.00	−0.00	0.00	−0.00	−0.00	0.00	−0.00**	−0.00**	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No. kids (at baseline)	−0.00	−0.00	0.01	−0.02	−0.01	0.00	0.01	0.01	0.05
	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
Grandmother at baseline	0.10	0.08	0.03	−0.04	−0.08	−0.13	0.02	−0.01	−0.08
	(0.10)	(0.10)	(0.09)	(0.12)	(0.11)	(0.10)	(0.10)	(0.10)	(0.12)
Mother depressed (at 7-year followup)		−0.16	0.03		−0.42***	−0.37***		−0.26**	−0.11
		(0.11)	(0.07)		(0.12)	(0.11)		(0.12)	(0.14)
Baseline depression severity		0.03	0.06		0.03	0.06*		0.07	0.09
		(0.06)	(0.04)		(0.04)	(0.03)		(0.05)	(0.06)
Mother play (at 1-year followup)			0.74***			0.23*			0.30**
			(0.11)			(0.13)			(0.13)
Father play (at 1-year followup)			0.59***			0.30*			0.51***
			(0.11)			(0.16)			(0.11)
Diarrhea (at 1-year followup)			−0.10			−0.04			0.00
			(0.10)			(0.10)			(0.09)
Breastfeeding (at 6-month followup)			0.03			0.18			0.32***
			(0.15)			(0.22)			(0.11)
ARI (at 1-year followup)			0.09			−0.08			−0.00
			(0.10)			(0.10)			(0.07)
Observations	295	295	276	295	295	276	295	295	276
R^2	0.44	0.44	0.65	0.31	0.35	0.40	0.23	0.24	0.33

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: This table shows associations of parenting behavior with potential mediating infant inputs and key demographic and socioeconomic characteristics (which were used as controls in the main analysis). The sample consists only of mothers in the control group. Column 1 shows the associations by regressing the child development outcome on baseline demographic/socioeconomic characteristics. Column 2 adds mother's depressed status at the 7-year follow-up and her baseline depression severity. Column 3 adds mediating infant inputs and infant health. The parental behavior indicators are measured using three broad domains and calculated as a summary index following Anderson (2008). All regressions control for interviewer fixed effects. Heteroskedasticity robust standard errors are clustered at the Union Council level.

Table D.8 – Correlates of Child Development at Age 7

	Cognitive development index			Physical development index			Socio-emotional development index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Girl	0.10	0.10	0.09	−0.03	−0.02	−0.06	−0.25**	−0.21*	−0.24*
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.11)	(0.11)	(0.12)
Age of index child	0.31	0.33	0.22	0.45	0.46	0.54	0.46	0.50	0.47
	(0.39)	(0.39)	(0.42)	(0.53)	(0.53)	(0.54)	(0.57)	(0.58)	(0.53)
Wealth score (at baseline)	0.06*	0.06*	0.06	0.02	0.02	−0.00	−0.01	−0.03	−0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mother's years of education	0.05***	0.05***	0.04***	−0.00	−0.00	−0.01	0.02	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Father's years of education	0.05***	0.05***	0.05***	−0.00	−0.00	−0.00	0.03*	0.03*	0.04**
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Mother's age	0.16***	0.16**	0.14**	0.23**	0.23**	0.22**	0.02	0.02	0.03
	(0.06)	(0.06)	(0.05)	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)	(0.06)
Mother's age^2	−0.00**	−0.00**	−0.00**	−0.00**	−0.00**	−0.00**	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No. kids (at baseline)	−0.09**	−0.08**	−0.09**	−0.02	−0.02	−0.06	−0.00	−0.00	−0.02
	(0.04)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)
Grandmother at baseline	0.04	0.02	−0.01	0.09	0.08	0.09	−0.06	−0.09	−0.14
	(0.09)	(0.08)	(0.08)	(0.12)	(0.12)	(0.12)	(0.14)	(0.14)	(0.17)
Mother depressed (at 7-year followup)		−0.23**	−0.17		−0.06	−0.01		−0.19	−0.19
		(0.10)	(0.12)		(0.16)	(0.17)		(0.13)	(0.13)
Baseline depression severity		0.06	0.06		−0.05	−0.03		−0.16**	−0.17**
		(0.05)	(0.05)		(0.07)	(0.06)		(0.06)	(0.07)
Mother play (at 1-year followup)			−0.03			0.04			−0.08
			(0.13)			(0.17)			(0.19)
Father play (at 1-year followup)			0.35***			0.15			−0.14
			(0.07)			(0.14)			(0.17)
Diarrhea (at 1-year followup)			−0.31***			−0.08			−0.12
			(0.09)			(0.17)			(0.14)
Breastfeeding (at 6-month followup)			−0.01			−0.11			−0.16
			(0.11)			(0.18)			(0.17)
ARI (at 1-year followup)			−0.07			−0.19			0.06
			(0.09)			(0.12)			(0.09)
Observations	295	295	276	295	295	276	295	295	276
R^2	0.23	0.24	0.29	0.09	0.09	0.12	0.12	0.16	0.17

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: This table shows associations of child development with potential mediating infant inputs and key demographic and socioeconomic characteristics (which were used as controls in the main analysis). The sample consists only of mothers in the control group. Column 1 shows the associations by regressing the child development outcome on baseline demographic/socioeconomic characteristics. Column 2 adds mother's depressed status at the 7-year follow-up and her baseline depression severity. Column 3 adds mediating infant inputs and infant health. The child development indicators are measured using three broad domains and calculated as a summary index following Anderson (2008). All regressions control for interviewer fixed effects. Heteroskedasticity robust standard errors are clustered at the Union Council level.

E Baseline non-depressed

Table E.9 – Balance in non-depressed sample: Characteristics by cluster assignment at 7-yr follow-up

	Non-experimental Sample at 7-year followup					
	Control Mean	(s.d.)	T-C Diff	(s.e.)	p-val	N
Age	33.86	(5.2)	0.42	(0.71)	0.56	300
Parity	4.65	(3.0)	−0.23	(0.30)	0.46	300
Mother's education	4.85	(4.3)	1.39	(0.73)	0.07*	300
Father's education	7.89	(3.3)	0.24	(0.47)	0.61	300
Grandmother lives with	0.40	(0.5)	0.09	(0.07)	0.20	300
Adults in house	4.01	(2.6)	0.27	(0.30)	0.38	299
Index child is girl	0.48	(0.5)	−0.03	(0.07)	0.69	300
Age of index child	7.57	(0.1)	0.00	(0.01)	0.80	300
Mother's Financial Autonomy Index	0.27	(1.1)	0.08	(0.14)	0.55	300
Father's Employment Index	0.03	(0.7)	0.04	(0.08)	0.60	299
Household Wealth Index	0.13	(0.8)	0.22	(0.14)	0.12	300
Relationship Quality Index	0.37	(0.8)	0.03	(0.09)	0.73	295
Mother's Health Index	0.14	(1.3)	0.19	(0.16)	0.24	300
Mental health index (7y)	0.52	(0.5)	0.03	(0.06)	0.63	300
Joint test (p-value)					0.38	
Observations		150				

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: This table tests for balance in characteristics at the 7-year follow-up for women excluded (non-depressed) at baseline, by treatment and control clusters.

[a] The wealth index is a PCA-weighted index of household income, health worker SES rating, house materials, water and waste infrastructure, and a number of other assets.

Table E.10 – Characteristics in 2013 by Baseline Depression Status

Sample Characteristics:	(1) Non-depressed	(2) Depressed	(3) p-value
Mother's characteristics			
Mother's age	34.06	34.73	0.10 *
Mother's education	5.54	4.02	0.00 ***
Number of kids	4.00	4.31	0.00 ***
Number of kids born to mother in last 7 years	1.24	0.87	0.00 ***
Avg age if kids born to mother in last 7 yrs	3.68	3.71	0.81
Mother's general health (1=vgood 5=vbad)	2.87	3.14	0.00 ***
Mother's Mental Health			
Currently depressed (MDE)	0.11	0.27	0.00 ***
Perceived social support score (MSPSS)	41.69	37.94	0.00 ***
Recovered permanently	0.00	0.39	0.00 ***
Never recovered	0.00	0.13	0.00 ***
Depressed ever between 2008-2013	0.13	0.31	0.00 ***
Depressed between 2008-2013 (recall only)	0.03	0.14	0.00 ***
Number of recalled depressive episodes	0.03	0.15	0.00 ***
Number of depressive episodes since 2007	0.12	0.33	0.00 ***
Duration of recalled depressive episodes (yrs)	0.03	0.11	0.00 ***
Family characteristics			
Joint/extended family structure	0.60	0.60	0.93
Grandmother lives with	0.44	0.37	0.03 **
Number of adults living with	4.14	3.72	0.01 ***
Father's characteristics			
Father's education	8.01	6.96	0.00 ***
Father employed	0.90	0.87	0.25
Father's occupation non-manual worker	0.09	0.05	0.01 ***
Household income and SES			
SES (1=Rich, 5=Poor)	3.34	3.48	0.01 ***
Has debt	0.56	0.63	0.05 **
Piped drinking water	0.06	0.08	0.28
Flush toilet	0.65	0.57	0.03 **
Sample size	300	585	885

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The table shows sample means for characteristics for perinatally depressed and perinatally non-depressed mother measure at the time of the 7-year follow-up. Column 3 shows the p-value of the difference in means between the depressed and non-depressed groups.

F Attrition

Table F.11 – Characteristics of attritors and differences by intervention and control clusters

	Characteristics of attritors N=903				Attritor characteristics by treatment arm N=318			
	(1) In sample Mean N=585	(2) Attritor Mean N=318	(3) Diff	(4) *p*-val	(5) T Mean N=174	(6) C mean N=144	(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
Index child is first born	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
% children female	0.53	0.52	−0.01	0.55	0.54	0.50	0.04	0.27
Hamilton depression score	14.49	14.88	0.39	0.17	15.11	14.61	0.50	0.29
Baseline BDQ score	8.12	8.36	0.23	0.21	8.15	8.61	−0.46	0.11
General functioning (GAF)	62.25	61.68	−0.57	0.12	61.21	62.24	−1.02	0.08*
Perceived social support score	46.01	43.26	−2.75	0.02**	42.63	44.03	−1.40	0.45
Joint/extended family structure	0.59	0.58	−0.01	0.71	0.58	0.58	0.00	0.94
Mother-in-law lives with	0.43	0.44	0.00	0.93	0.45	0.42	0.02	0.66
Mother's mother lives with	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
Father not manual worker	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[a]	0.07	−0.13	−0.20	0.15	−0.20	−0.04	−0.17	0.47
LTFU: abortion					0.01	0.02	−0.01	0.51
LTFU: stillbirth					0.07	0.12	−0.05	0.13
LTFU: child death/illness					0.20	0.14	0.06	0.18
LTFU: mother death/illness					0.07	0.10	−0.03	0.36
LTFU: refused					0.25	0.28	−0.03	0.52
LTFU: moved					0.36	0.33	0.03	0.59
Joint test (*p*-value)								0.59

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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.

[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).

Table F.12 – Differences in short-run treatment effects by sample

	Coefficient on Treat (β / (s.e.))		Difference between samples		
	(1)	(2)	(3)	(4)	(5)
	Full sample	7-yr followup sample	Raw Diff.	Diff. in st.devs.	p-value
Depressed (6m)	-0.30^{***}	-0.32^{***}	0.03	5%	0.26
Depressed (1y)	-0.32^{***}	-0.33^{***}	0.02	4%	0.45
Depression severity (6m)	-4.20^{***}	-4.34^{***}	0.14	2%	0.76
Depression severity (1y)	-5.29^{***}	-5.60^{***}	0.31	4%	0.43
BDQ disability score (6m)	-1.91^{***}	-1.97^{***}	0.06	2%	0.81
BDQ disability score (1y)	-2.96^{***}	-3.07^{***}	0.11	3%	0.60
GAF general functioning (6m)	7.14^{***}	7.54^{***}	-0.41	4%	0.49
GAF general functioning (1y)	8.68^{***}	9.11^{***}	-0.43	4%	0.48
Perceived social support (6m)	7.06^{***}	7.72^{***}	-0.66	4%	0.40
Perceived social support (1y)	8.44^{***}	8.76^{***}	-0.32	2%	0.67
Joint test at 6m (p-value)					0.60
Joint test at 1y (p-value)					0.98

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Table shows the treatment effects (as a simple difference in means between treatment and control, T-C) on short-run depression and mental health outcomes using the full samples at 6 and 12 months (N=818 and 791 respectively) and compares the estimated treatment effects to those using the 7-year follow-up sample (N=584). The third columns reports the differences in the estimated treatment effects between the two samples, and the fourth column reports the difference in estimate effects as a percentage of a standard deviation in the outcome. Negative treatment effects in the first 3 outcomes (depressed, depression severity score, and disability score) correspond to more favorable outcomes, while positive treatment effects for the last 2 outcomes (functioning and social support scores) correspond to better outcomes. Across all mental health outcomes, the differences in treatment effects range between 2 and 5% of a standard deviation of the outcome, averaging at 3% of a standard deviation. The difference between the two samples always favors the non-attriting sample; however, no individual treatment effect differed statistically between the two samples nor jointly. The joint test was conducted by running a Seemingly Unrelated Regression (SUR) model on all 5 outcomes, with the GAF and social support outcomes flipped (so that higher values indicate worse mental health outcomes like the rest of the measures), regressing outcomes on treat, an indicator for attritor, and the interaction ($y^k = \beta_0^k + \beta_1^k Treat + \beta_2^k LTFU + \beta_3^k Treat \times LTFU + \varepsilon^k$ for $k \in (1, \ldots, 5)$ where each variable represents a vector and k denotes the outcome). The p-value reported in the last row is the joint test of the hypothesis that the interaction ($Treat \times LTFU$) is different from zero. The joint test of whether mental health outcomes favored nonattritors also suggests no differences (p=0.96).

Table F.13 – Attrition corrected treatment effects: Inverse Probability Weights and Bounds

	Inverse Prob. Weighted		Lee Bounds CI 1-year Followup		Lee Bounds CI THP Sample	
	(1) No controls β/(s.e.)	(2) All controls β/(s.e.)	(3) Lower	(4) Upper	(5) Lower	(6) Upper
Mother's financial empowerment (7y)	0.17*	0.18**	0.07	0.53	0.06	0.56
	(0.09)	(0.07)				
Mother's financial empowerment	0.27**	0.29***	0.13	0.62	0.14	0.67
	(0.11)	(0.09)				
Parental investment (monetary)	0.29***	0.27***	0.09	0.62	0.04	0.68
	(0.07)	(0.06)				
Parental investment (time-intensive)	0.18***	0.20***	0.05	0.59	0.02	0.67
	(0.06)	(0.06)				
Parenting style	0.05	0.06	−0.26	0.30	−0.28	0.37
	(0.08)	(0.08)				
Fertility trajectory	0.01	0.00	−0.28	0.35	−0.34	0.39
	(0.10)	(0.09)				
Physical development index	0.16*	0.14	−0.19	0.35	−0.26	0.39
	(0.09)	(0.09)				
Cognitive development index	0.07	0.04	−0.27	0.27	−0.32	0.33
	(0.08)	(0.08)				
Socio-emotional development index	−0.09	−0.08	−0.44	0.14	−2.35	0.84
	(0.07)	(0.06)				
Child survival index	0.21**	0.16*	−0.21	0.37	−0.63	1.28
	(0.09)	(0.08)				
Mother's physical health	0.05	0.07	−1.51	3.98	−2.57	1.39
	(0.07)	(0.08)				
Husband's income trajectory	−0.05	−0.05	−0.36	0.27	−0.38	0.32
	(0.11)	(0.10)				
Relationships	0.12	0.15*	−0.15	0.38	−9.30	8.11
	(0.09)	(0.09)				
Grandmother trajectory index	0.28***	0.17**	0.10	0.91	0.07	0.95
	(0.08)	(0.07)				

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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).Columns 1-2 replicate the main results using IPW (Inverse Probability Weighting) to account for attrition. Columns 3 and 4 show 95% confidence intervals for the treatment effect using attrition bounds based on Lee (2009), using the starting sample of $N = 704$, and columns 5 and 6 show attrition bounds using the baseline starting sample with $N = 903$.

Table F.14 – Sample flow and attrition by gender

	Boy		Girl	
	(1) Treatment	(2) Control	(3) Treatment	(4) Control
Baseline	223	226	240	214
6-month	185	190	202	179
1-year	179	180	198	174
7-year	133	155	156	141
Follow-up rates at 7-year follow-up:				
from baseline (%)	0.60**	0.69**	0.65	0.66
from 1-year (%)	0.72***	0.86**	0.77	0.85

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The upper panel reports the total number of mother-child dyads in our sample from baseline to the 7-year follow-up, split by gender and treatment status. The lower panel reports the ratio of dyads present at the 7-year follow-up over the total number of dyads present at baseline or at the 1-year follow-up. Stars indicate significant differences in follow-up rates between columns (1) and (2), or (3) and (4).

Table F.15 – Attrition corrections by child gender

	Girls			Boys		
	(1) No controls β/(s.e.)	(2) Lower	(3) Upper	(4) No controls β/(s.e.)	(5) Lower	(6) Upper
Mother's financial empowerment	0.46***	0.16	0.94	0.13	−0.28	0.65
	(0.14)			(0.15)		
Parental investment (monetary)	0.55***	0.10	0.92	0.16*	−0.24	0.67
	(0.10)			(0.09)		
Parental investment (time-intensive)	0.25**	−0.01	0.66	0.15	−0.16	0.83
	(0.10)			(0.09)		
Parenting style	0.22**	−0.26	0.52	−0.20	−0.69	0.33
	(0.08)			(0.13)		
Fertility trajectory	0.11	−0.37	0.58	−0.08	−0.52	0.54
	(0.11)			(0.13)		
Physical development index	0.20	−0.31	0.44	0.08	−0.49	0.52
	(0.13)			(0.11)		
Cognitive development index	0.15	−0.49	0.41	0.03	−0.47	0.40
	(0.13)			(0.09)		
Socio-emotional development index	−0.06	−0.53	0.27	−0.16	−0.70	0.30
	(0.11)			(0.12)		
Child survival index	0.32***	−0.24	0.59	0.01	−0.46	0.34
	(0.10)			(0.10)		
Mother's physical health	0.16*	−0.21	0.55	−0.06	−0.52	0.48
	(0.09)			(0.11)		
Husband's income trajectory	−0.09	−0.46	0.33	0.06	−0.51	0.47
	(0.13)			(0.13)		
Relationships	0.10	−0.30	0.37	0.18	−0.36	0.69
	(0.11)			(0.11)		
Grandmother trajectory index	0.26**	0.00	1.02	0.44***	0.01	1.07
	(0.11)			(0.10)		

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.Columns 1 and 4 show the treatment effects on indices split by gender. Columns 2-3 and 5-6 show 95% confidence intervals for the treatment effect using attrition bounds based on Lee (2009), split by child gender. The starting sample for the bound is 454 girls and 449 boys.

G Robustness checks

Table G.16 – Control sensitivity

	Coefficient on Treat (β / (s.e.))			
	(1) No controls	(2) Interviewer FEs	(3) + Individual controls	(4) + Ind. × T controls
Mother's financial empowerment (7y)	0.23**	0.18*	0.18**	0.18**
	(0.11)	(0.09)	(0.08)	(0.07)
Mother's financial empowerment	0.34**	0.29**	0.29***	0.29***
	(0.13)	(0.11)	(0.09)	(0.09)
Parental investment (monetary)	0.36***	0.35***	0.28***	0.28***
	(0.09)	(0.07)	(0.06)	(0.06)
Parental investment (time-intensive)	0.32**	0.20***	0.19***	0.20***
	(0.16)	(0.07)	(0.06)	(0.06)
Parenting style	0.06	0.04	0.06	0.05
	(0.09)	(0.08)	(0.08)	(0.08)
Fertility trajectory	0.02	0.01	0.00	−0.00
	(0.11)	(0.10)	(0.09)	(0.09)
Physical development index	0.12	0.15	0.15	0.14
	(0.12)	(0.09)	(0.09)	(0.09)
Cognitive development index	0.03	0.09	0.06	0.04
	(0.10)	(0.08)	(0.08)	(0.08)
Socio-emotional development index	−0.12	−0.11	−0.07	−0.08
	(0.08)	(0.07)	(0.07)	(0.07)
Child survival index	0.17**	0.19**	0.17**	0.17**
	(0.08)	(0.08)	(0.08)	(0.08)
Mother's physical health	0.08	0.07	0.08	0.07
	(0.09)	(0.07)	(0.08)	(0.08)
Husband's income trajectory	0.04	−0.02	−0.03	−0.04
	(0.10)	(0.10)	(0.09)	(0.10)
Relationships	0.16*	0.14*	0.16*	0.16*
	(0.09)	(0.09)	(0.09)	(0.09)
Grandmother trajectory index	0.31***	0.34***	0.16**	0.16**
	(0.09)	(0.08)	(0.06)	(0.07)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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).

Table G.17 – Control sensitivity (indices by factor score)

	Coefficient on Treat (β / (s.e.))			
	(1) No controls	(2) Interviewer FEs	(3) + Individual controls	(4) + Ind. × T controls
Mother's financial autonomy (7y)	0.16	0.14	0.10	0.09
	(0.12)	(0.13)	(0.11)	(0.11)
Parental investment (monetary)	0.25***	0.27***	0.19***	0.18***
	(0.07)	(0.07)	(0.06)	(0.05)
Parental investment (time-intensive)	0.28**	0.19***	0.17***	0.17***
	(0.13)	(0.06)	(0.06)	(0.06)
Parenting style	0.14	0.11	0.11*	0.11
	(0.09)	(0.07)	(0.07)	(0.07)
Physical development index	−0.01	−0.02	−0.03	−0.03
	(0.08)	(0.07)	(0.07)	(0.07)
Cognitive development index	0.06	0.11	0.07	0.06
	(0.10)	(0.08)	(0.07)	(0.07)
Socio-emotional development index	−0.11	−0.08	−0.06	−0.07
	(0.08)	(0.06)	(0.07)	(0.06)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary variables generated by factor score. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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).

Table G.18 – Mother's decision-making (only year 7 outcomes): financial empowerment, investment in children and fertility

	Coefficient on Treat (β / (s.e.))		FWER-adj. test
	(1) No controls	(2) All controls	(3) FWER p-val (all controls)
Mother's financial empowerment (7y)	0.18*	0.18**	0.07*
	(0.09)	(0.07)	
Parental investment (monetary)	0.35***	0.28***	0.00***
	(0.07)	(0.06)	
Parental investment (time-intensive)	0.20***	0.20***	0.01**
	(0.07)	(0.06)	
Parenting style	0.04	0.05	0.78
	(0.08)	(0.08)	
Fertility trajectory	0.01	−0.00	0.99
	(0.10)	(0.09)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993).

Table G.19 – Do measures of parental investment predict child development?

	Physical development		Cognitive development		Socioemotional development	
	(1)	(2)	(3)	(4)	(5)	(6)
Monetary investment	0.11*	0.12**	0.16***	0.15***	−0.04	−0.01
	(0.06)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Time investment	−0.05	−0.03	−0.00	0.03	0.08	0.07
	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)
Parenting style	−0.01	0.03	0.07*	0.07**	0.30***	0.32***
	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)
Mother's financial empowerment	0.07**	0.04**	0.06*	0.02	−0.03	−0.04
	(0.03)	(0.02)	(0.04)	(0.03)	(0.05)	(0.03)
Fertility trajectory	−0.02	−0.03	0.06*	0.03	0.02	−0.01
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
Baseline depressed		−0.14**		0.03		−0.18***
		(0.06)		(0.07)		(0.06)
Observations	584	884	584	884	584	884
R^2	0.13	0.12	0.26	0.21	0.19	0.19

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Sample includes children of mothers who were depressed at baseline as well those who were not depressed at baseline, in both treatment and control clusters. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.All regressions control for interviewer fixed effects, age of mother and its square, father's and mother's education, parity, the date of interview, and child gender and age at interview.

Table G.20 – Depression and empowerment: associations

	Empowered before 7y			Empowered at 7y followup					
	(1) at baseline	(2) at 6m	(3) at 1y	(4) All	(5) Girls	(6) Boys	(7) All	(8) Girls	(9) Boys
Depressed (7y)				−0.16**	−0.13*	−0.18**	−0.13	−0.05	−0.22*
				(0.06)	(0.07)	(0.08)	(0.08)	(0.07)	(0.12)
Depressed (1y)			−0.24***				−0.17**	−0.19**	−0.15
			(0.05)				(0.06)	(0.08)	(0.10)
Depressed (6m)		−0.36***							
		(0.06)							
Depressed (baseline)				−0.08**	−0.18**	0.02			
				(0.04)	(0.07)	(0.04)			
Dep. severity (baseline)	−0.07***	−0.05*	−0.02						
	(0.02)	(0.03)	(0.03)						
Index child is girl		−0.02	−0.01	−0.06			−0.16**		
		(0.06)	(0.05)	(0.04)			(0.07)		
Observations	903	369	351	596	279	317	295	139	156

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Sample includes children of mothers who were depressed at baseline as well those who were not depressed at baseline, in both treatment and control clusters. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.

Table G.21 – Difference-in-differences with baseline non-depressed, and health worker fixed effects

	Coefficient on		
	(1) Treat	(2) Treat × Prenatally Depressed	(3) Prenatally Depressed
Mother's financial empowerment (7y)	−0.04	0.30**	−0.19*
	(0.13)	(0.14)	(0.10)
Mother's financial empowerment	−0.08	0.51***	−0.25*
	(0.16)	(0.17)	(0.14)
Parental investment (monetary)	0.01	0.32**	−0.15
	(0.13)	(0.12)	(0.09)
Parental investment (time-intensive)	−0.08	0.20	−0.20*
	(0.11)	(0.13)	(0.10)
Parenting style	−0.27*	0.32**	−0.27**
	(0.14)	(0.15)	(0.12)
Fertility trajectory	−0.06	0.18	0.20
	(0.15)	(0.18)	(0.13)
Physical development index	0.20*	0.01	−0.21**
	(0.12)	(0.15)	(0.10)
Cognitive development index	0.16	−0.11	0.10
	(0.13)	(0.16)	(0.11)
Socio-emotional development index	−0.03	−0.16	−0.12
	(0.13)	(0.15)	(0.10)
Child survival index	0.06	0.15	−0.12
	(0.11)	(0.13)	(0.09)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=885. Sample includes children of mothers who were depressed at baseline as well those who were not depressed at baseline, in both treatment and control clusters. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.All regressions control for interviewer fixed effects, age of mother and its square, father's and mother's education, parity, the date of interview, and health worker fixed effects.

Table G.22 – Difference-in-differences with baseline non-depressed

	Coefficient on		
	(1) Treat	(2) Treat × Prenatally Depressed	(3) Prenatally Depressed
Mother's financial empowerment (7y)	−0.08	0.22	−0.20*
	(0.12)	(0.14)	(0.11)
Mother's financial empowerment	−0.07	0.33	−0.16
	(0.17)	(0.20)	(0.16)
Parental investment (monetary)	0.06	0.22*	−0.07
	(0.12)	(0.12)	(0.09)
Parental investment (time-intensive)	0.03	0.15	−0.17**
	(0.10)	(0.11)	(0.07)
Parenting style	−0.19	0.22*	−0.13
	(0.13)	(0.12)	(0.08)
Fertility trajectory	0.01	0.02	0.21***
	(0.13)	(0.11)	(0.08)
Physical development index	0.15	0.01	−0.15*
	(0.11)	(0.13)	(0.08)
Cognitive development index	0.13	−0.11	0.10
	(0.13)	(0.15)	(0.10)
Socio-emotional development index	0.05	−0.18	−0.12
	(0.13)	(0.13)	(0.10)
Child survival index	0.05	0.11	−0.11
	(0.10)	(0.12)	(0.07)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=885. Sample includes children of mothers who were depressed at baseline as well those who were not depressed at baseline, in both treatment and control clusters. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.All regressions control for interviewer fixed effects, age of mother and its square, father's and mother's education, parity, and the date of interview.

H Heterogeneous treatment effects

Table H.23 – Heterogeneous treatment effects for mother's decision-making

	Coefficient on:		
	(1) Treat	(2) Treat × Baseline characteristic	(3) Baseline characteristic
	Baseline characteristic: Mother's education		
Mother's financial empowerment	0.05	0.05**	0.01
	(0.13)	(0.02)	(0.02)
Parental investment (monetary)	0.20*	0.02	0.07***
	(0.11)	(0.02)	(0.01)
Parental investment (time-intensive)	0.16*	0.00	0.04***
	(0.09)	(0.02)	(0.01)
Parenting style	0.01	0.00	0.03*
	(0.10)	(0.02)	(0.02)
Fertility trajectory	−0.02	0.01	−0.01
	(0.14)	(0.02)	(0.01)
	Baseline characteristic: Younger mother (age < 27)		
Mother's financial empowerment	0.48***	−0.34**	0.09
	(0.13)	(0.14)	(0.08)
Parental investment (monetary)	0.49***	−0.25	0.10
	(0.15)	(0.21)	(0.18)
Parental investment (time-intensive)	0.29***	−0.16	0.07
	(0.10)	(0.14)	(0.11)
Parenting style	0.06	−0.02	−0.19
	(0.12)	(0.17)	(0.13)
Fertility trajectory	0.14	−0.23	−0.24**
	(0.13)	(0.14)	(0.11)
	Baseline characteristic: First child		
Mother's financial empowerment	0.27**	0.10	−0.16
	(0.12)	(0.18)	(0.15)
Parental investment (monetary)	0.35***	−0.04	0.18
	(0.08)	(0.21)	(0.15)
Parental investment (time-intensive)	0.24***	−0.20	0.10
	(0.07)	(0.16)	(0.11)
Parenting style	−0.01	0.34	−0.33*
	(0.09)	(0.26)	(0.19)
Fertility trajectory	0.01	0.03	−0.48**
	(0.10)	(0.23)	(0.19)
	Baseline characteristic: Wealth index		
Mother's financial empowerment	0.26**	−0.00	0.09***
	(0.11)	(0.04)	(0.03)
Parental investment (monetary)	0.30***	0.05	0.15***
	(0.07)	(0.04)	(0.03)
Parental investment (time-intensive)	0.18**	−0.00	0.09***
	(0.07)	(0.03)	(0.02)
Parenting style	0.02	0.07**	0.04
	(0.08)	(0.03)	(0.02)
Fertility trajectory	0.01	0.01	−0.01
	(0.10)	(0.04)	(0.03)
	Baseline characteristic: Grandmother present		
Mother's financial empowerment	0.53***	−0.53**	0.42***
	(0.13)	(0.20)	(0.13)
Parental investment (monetary)	0.35***	−0.04	0.22**
	(0.10)	(0.14)	(0.11)
Parental investment (time-intensive)	0.23**	−0.08	0.14
	(0.10)	(0.14)	(0.12)
Parenting style	0.01	0.08	−0.04
	(0.10)	(0.14)	(0.10)
Fertility trajectory	−0.00	0.07	−0.22***
	(0.12)	(0.15)	(0.08)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.Controls include interviewer FEs and date of interviewer.

Table H.24 – Heterogeneous treatment effects for child outcomes

	Coefficient on:		
	(1) Treat	(2) Treat × Baseline characteristic	(3) Baseline characteristic
	Baseline characteristic: Mother's education		
Physical development index	0.24*	−0.02	−0.00
	(0.12)	(0.02)	(0.01)
Cognitive development index	0.06	−0.01	0.07***
	(0.11)	(0.02)	(0.01)
Socio-emotional development index	−0.18	0.01	0.02
	(0.11)	(0.02)	(0.02)
Child survival index	0.09	0.02	0.01
	(0.13)	(0.02)	(0.02)
	Baseline characteristic: Younger mother (age < 27)		
Physical development index	0.17	−0.04	−0.02
	(0.12)	(0.15)	(0.12)
Cognitive development index	0.22*	−0.25	0.07
	(0.13)	(0.16)	(0.11)
Socio-emotional development index	−0.23*	0.24	−0.34***
	(0.12)	(0.16)	(0.11)
Child survival index	0.24	−0.09	0.10
	(0.15)	(0.17)	(0.13)
	Baseline characteristic: First child		
Physical development index	0.15	0.00	0.08
	(0.10)	(0.23)	(0.18)
Cognitive development index	0.05	0.13	0.23
	(0.10)	(0.21)	(0.17)
Socio-emotional development index	−0.09	−0.07	−0.02
	(0.08)	(0.23)	(0.13)
Child survival index	0.18	0.08	−0.07
	(0.11)	(0.22)	(0.17)
	Baseline characteristic: Wealth index		
Physical development index	0.14	0.01	0.01
	(0.09)	(0.03)	(0.02)
Cognitive development index	0.04	0.03	0.11***
	(0.09)	(0.04)	(0.03)
Socio-emotional development index	−0.13*	0.11***	0.00
	(0.07)	(0.04)	(0.02)
Child survival index	0.18**	−0.03	0.08***
	(0.08)	(0.04)	(0.02)
	Baseline characteristic: Grandmother present		
Physical development index	0.11	0.05	0.07
	(0.11)	(0.14)	(0.10)
Cognitive development index	0.09	−0.06	0.25***
	(0.14)	(0.16)	(0.08)
Socio-emotional development index	−0.17	0.12	0.04
	(0.14)	(0.20)	(0.11)
Child survival index	0.36***	−0.36*	0.25
	(0.13)	(0.19)	(0.15)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.Controls include interviewer FEs and date of interviewer.

Table H.25 – Heterogeneous treatment effects for specific outcomes

	Coefficient on:		
	(1) Treat	(2) Treat × Baseline characteristic	(3) Baseline characteristic
	Baseline characteristic: Mother's education		
Mother controls spending (7y)	0.03	0.01	0.01
	(0.05)	(0.01)	(0.01)
HOME inventory	0.45	0.25*	0.51***
	(0.94)	(0.15)	(0.12)
Private school	0.11	0.01	0.02***
	(0.07)	(0.01)	(0.01)
BMI-for-age (z)	−0.08	0.04	−0.04**
	(0.16)	(0.03)	(0.02)
WPPSI Full Scale IQ	0.17	0.11	0.57***
	(1.16)	(0.21)	(0.18)
	Baseline characteristic: Younger mother (age < 27)		
Mother controls spending (7y)	0.10**	−0.05	−0.00
	(0.05)	(0.07)	(0.05)
HOME inventory	2.69**	−1.25	−0.33
	(1.14)	(1.49)	(1.02)
Private school	0.17*	−0.02	0.03
	(0.08)	(0.09)	(0.08)
BMI-for-age (z)	−0.01	0.15	−0.05
	(0.16)	(0.20)	(0.13)
WPPSI Full Scale IQ	2.49	−2.45	0.75
	(1.70)	(1.94)	(1.41)
	Baseline characteristic: First child		
Mother controls spending (7y)	0.08**	−0.03	−0.01
	(0.04)	(0.11)	(0.07)
HOME inventory	2.02***	−0.28	1.02
	(0.71)	(1.41)	(0.97)
Private school	0.16**	−0.07	0.18*
	(0.06)	(0.12)	(0.10)
BMI-for-age (z)	0.01	0.34	−0.15
	(0.12)	(0.27)	(0.21)
WPPSI Full Scale IQ	0.62	2.56	2.02
	(1.26)	(2.25)	(1.41)
	Baseline characteristic: Wealth index		
Mother controls spending (7y)	0.07**	0.01	0.01
	(0.03)	(0.02)	(0.01)
HOME inventory	1.60**	0.26	1.24***
	(0.63)	(0.29)	(0.19)
Private school	0.13**	0.02	0.05***
	(0.05)	(0.02)	(0.01)
BMI-for-age (z)	0.08	0.11**	−0.12***
	(0.11)	(0.05)	(0.03)
WPPSI Full Scale IQ	0.75	0.45	1.01**
	(1.06)	(0.56)	(0.40)
	Baseline characteristic: Grandmother present		
Mother controls spending (7y)	0.14***	−0.16*	0.12**
	(0.05)	(0.08)	(0.06)
HOME inventory	2.25**	−0.81	1.54
	(0.84)	(1.17)	(0.92)
Private school	0.10	0.08	0.08
	(0.07)	(0.07)	(0.05)
BMI-for-age (z)	−0.03	0.23	−0.22*
	(0.15)	(0.20)	(0.12)
WPPSI Full Scale IQ	1.22	−0.68	2.48**
	(1.64)	(1.82)	(1.06)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses.Controls include interviewer FEs and date of interviewer.

I Treatment effects for individual outcomes in other indices

Table I.26 – Mother's depression trajectory

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Depression trajectory	0.00	−0.59***	−0.58***	0.00***	0.01	−0.01	−0.66***	−0.49***	0.18
	(1.00)	(0.09)	(0.07)				(0.09)	(0.10)	
Depressed (6m)	0.52	−0.32***	−0.33***	0.00***	0.48	0.56	−0.31***	−0.35***	0.56
	(0.50)	(0.05)	(0.04)				(0.05)	(0.06)	
Depressed (1y)	0.58	−0.33***	−0.32***	0.00***	0.58	0.58	−0.36***	−0.27***	0.20
	(0.49)	(0.05)	(0.04)				(0.06)	(0.05)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.27 – Mother's financial empowerment (at 7-year follow-up)

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Mother's financial autonomy (7y)	−0.00	0.18*	0.18**	0.02**	−0.10	0.09	0.32***	0.03	0.11
	(1.00)	(0.09)	(0.07)				(0.10)	(0.14)	
Mother controls spending (7y)	0.52	0.08**	0.09***	0.02**	0.44	0.59	0.20***	−0.03	0.01
	(0.50)	(0.03)	(0.03)				(0.05)	(0.05)	
Mother gets pocket money (7y)	0.57	0.08***	0.09***	0.01**	0.51	0.63	0.18***	−0.00	0.03
	(0.50)	(0.03)	(0.03)				(0.05)	(0.05)	
Mother employed (7y)	0.10	0.02	0.01	0.76	0.11	0.09	−0.00	0.03	0.57
	(0.30)	(0.03)	(0.03)				(0.03)	(0.04)	
Mother's income (100s PKR)	3.09	1.41	0.85	0.76	3.09	3.08	0.49	1.27	0.69
	(12.07)	(1.70)	(1.39)				(1.57)	(1.82)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 *p*-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.28 – Child physical development and health at age 7

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Physical development index	−0.00	0.15	0.14	0.13	0.01	−0.01	0.19	0.08	0.49
	(1.00)	(0.09)	(0.09)				(0.13)	(0.12)	
Height-for-age (z)	−0.83	−0.02	−0.05	0.98	−0.76	−0.90	−0.13	0.03	0.47
	(1.11)	(0.09)	(0.08)				(0.13)	(0.14)	
BMI-for-age (z)	−0.98	0.06	0.05	0.98	−1.08	−0.88	0.34***	−0.28*	0.00
	(1.18)	(0.11)	(0.11)				(0.12)	(0.15)	
Not stunted (height > 2SD)	0.84	0.03	0.02	0.98	0.87	0.83	−0.01	0.05	0.32
	(0.36)	(0.03)	(0.03)				(0.04)	(0.04)	
Motor function	−0.02	0.03	0.05	0.96	−0.15	0.10	0.14*	−0.01	0.12
	(0.68)	(0.05)	(0.05)				(0.08)	(0.06)	
No hospitalization	0.81	0.05	0.06	0.59	0.84	0.79	0.05	0.07	0.70
	(0.39)	(0.03)	(0.04)				(0.05)	(0.05)	
No severe illness	0.70	0.03	0.02	0.98	0.72	0.68	−0.02	0.06	0.15
	(0.46)	(0.03)	(0.03)				(0.05)	(0.04)	
No eyesight problems	0.94	0.01	0.01	0.97	0.97	0.92	−0.01	0.03	0.28
	(0.23)	(0.02)	(0.02)				(0.02)	(0.03)	
No hearing problems	0.98	0.01	0.01	0.98	0.98	0.98	0.02	−0.00	0.44
	(0.14)	(0.01)	(0.01)				(0.02)	(0.02)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 *p*-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.29 – Child cognitive development at age 7

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Cognitive development index	−0.00	0.09	0.04	0.62	0.03	−0.02	0.11	−0.04	0.30
	(1.00)	(0.08)	(0.08)				(0.11)	(0.11)	
WPPSI: Verbal comprehension	85.15	1.54	0.67	0.98	85.31	84.99	1.75	−0.43	0.27
	(13.59)	(1.37)	(1.30)				(1.65)	(1.56)	
WPPSI: Visual spatial	87.34	−0.41	−0.83	0.96	86.74	87.88	0.41	−2.03	0.29
	(15.06)	(1.07)	(1.05)				(1.60)	(1.49)	
WPPSI: Fluid reasoning	77.26	1.76**	1.43*	0.39	76.65	77.82	3.19***	−0.37	0.06
	(11.42)	(0.80)	(0.76)				(1.05)	(1.32)	
WPPSI: Working memory	99.57	0.88	0.42	0.98	99.26	99.85	1.37	−0.30	0.42
	(15.69)	(1.10)	(1.02)				(1.59)	(1.35)	
WPPSI: Processing speed	76.32	2.44***	2.52***	0.02**	77.28	75.43	3.29***	1.30	0.22
	(9.58)	(0.74)	(0.77)				(1.15)	(1.05)	
Urdu score	6.33	0.37	0.04	0.98	6.69	6.01	0.16	−0.20	0.49
	(3.50)	(0.31)	(0.25)				(0.32)	(0.40)	
Math score	8.98	0.54*	0.25	0.96	8.90	9.06	0.73**	−0.27	0.03
	(3.63)	(0.32)	(0.30)				(0.35)	(0.39)	
Executive function (Stroop)	14.11	0.08	−0.09	0.98	13.98	14.24	0.05	−0.19	0.61
	(3.20)	(0.24)	(0.24)				(0.34)	(0.31)	
Grade-for-age	0.75	−0.02	−0.02	0.98	0.78	0.73	−0.03	−0.02	0.82
	(0.33)	(0.03)	(0.03)				(0.04)	(0.04)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.30 – Child's socio-emotional development at age 7

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Socio-emotional development index	0.00	−0.11	−0.08	0.22	−0.02	0.01	−0.02	−0.14	0.51
	(1.00)	(0.07)	(0.07)				(0.09)	(0.13)	
SDQ: Emotional	2.34	0.14	0.14	0.95	2.50	2.19	−0.07	0.35*	0.19
	(2.06)	(0.14)	(0.14)				(0.23)	(0.20)	
SDQ: Conduct problems	3.33	−0.05	−0.14	0.95	3.18	3.46	−0.50***	0.36	0.01
	(2.04)	(0.16)	(0.17)				(0.18)	(0.26)	
SDQ: Hyperactivity	3.50	0.13	0.02	0.95	2.99	3.96	0.09	0.09	0.99
	(2.56)	(0.18)	(0.17)				(0.22)	(0.30)	
SDQ: Peer problems	1.93	0.04	0.08	0.95	1.99	1.88	0.12	0.03	0.70
	(1.57)	(0.10)	(0.10)				(0.16)	(0.16)	
SDQ: Prosocial	2.49	−0.09	−0.09	0.95	2.06	2.87	−0.19	0.17	0.27
	(2.53)	(0.18)	(0.17)				(0.20)	(0.27)	
SCAS: Panic and agoraphobia	1.54	0.31	0.33	0.57	1.64	1.46	0.23	0.42	0.73
	(2.71)	(0.22)	(0.20)				(0.34)	(0.33)	
SCAS: Separation	5.96	0.24	0.16	0.95	6.50	5.48	−0.21	0.45	0.26
	(4.05)	(0.28)	(0.27)				(0.43)	(0.34)	
SCAS: Injury fear	6.07	−0.05	−0.08	0.95	7.15	5.10	−0.26	−0.25	0.99
	(3.65)	(0.27)	(0.27)				(0.38)	(0.39)	
SCAS: Social phobia	2.40	−0.11	−0.19	0.95	2.71	2.12	−0.63	0.25	0.17
	(2.93)	(0.21)	(0.21)				(0.41)	(0.35)	
SCAS: Obsessive-compulsive	1.18	0.56***	0.50***	0.08*	1.16	1.20	0.65***	0.31	0.32
	(1.93)	(0.17)	(0.18)				(0.20)	(0.28)	
SCAS: General anxiety	3.40	0.23	0.28	0.95	3.77	3.06	−0.06	0.56	0.25
	(3.28)	(0.26)	(0.27)				(0.39)	(0.37)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 *p*-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.31 – Child survival

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Child survival index	0.00	0.19**	0.17**	0.04**	0.15	−0.14	0.27***	−0.02	0.02
	(1.00)	(0.08)	(0.08)				(0.09)	(0.10)	
Share of boys	0.52	−0.03	−0.04*	0.21	0.40	0.63	−0.03	0.00	0.23
	(0.25)	(0.02)	(0.02)				(0.02)	(0.02)	
# died <1 year of age	0.29	−0.02	0.00	1.00	0.33	0.25	−0.05	0.05	0.18
	(0.62)	(0.05)	(0.05)				(0.07)	(0.06)	
# died btw 1 & 5 years old	0.05	−0.02	−0.01	0.64	0.05	0.05	−0.02	−0.00	0.67
	(0.23)	(0.01)	(0.01)				(0.02)	(0.02)	
# died > 5 years old	0.04	−0.03*	−0.02	0.30	0.06	0.01	−0.05*	−0.01	0.10
	(0.19)	(0.02)	(0.02)				(0.02)	(0.02)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.32 – Husband's income trajectory

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Husband's income trajectory	0.00	−0.02	−0.04	0.70	0.05	−0.05	−0.05	−0.03	0.93
	(1.00)	(0.10)	(0.10)				(0.13)	(0.12)	
Monthly income (ln) (7y)	8.94	0.11	0.03	0.99	9.00	8.89	0.06	−0.01	0.76
	(1.40)	(0.18)	(0.17)				(0.21)	(0.18)	
Monthly income (ln) (1y)	7.16	0.05	0.06	0.99	7.12	7.20	0.36	−0.30	0.16
	(2.33)	(0.20)	(0.21)				(0.30)	(0.33)	
Monthly income (ln) (6m)	7.24	0.04	0.04	0.99	7.40	7.09	−0.03	0.11	0.73
	(2.19)	(0.21)	(0.22)				(0.29)	(0.31)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.33 – Trajectory of mother's physical health

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Mother's physical health	−0.00	0.07	0.07	0.42	0.05	−0.05	0.14	−0.04	0.25
	(1.00)	(0.07)	(0.08)				(0.10)	(0.13)	
Mother never been unwell	0.63	−0.01	−0.01	0.98	0.67	0.59	−0.06	0.05	0.17
	(0.48)	(0.03)	(0.03)				(0.04)	(0.06)	
Overall health (0-4)	1.83	0.06	0.05	0.94	1.84	1.83	0.12	−0.04	0.32
	(0.97)	(0.07)	(0.08)				(0.11)	(0.11)	
Healthy days in past 30	26.16	0.20	0.05	0.98	26.22	26.10	0.32	−0.36	0.63
	(7.66)	(0.57)	(0.62)				(0.91)	(0.99)	
Weight (kg) (6m)	53.87	0.39	0.46	0.94	53.93	53.82	2.41**	−1.90	0.02
	(11.08)	(0.80)	(0.81)				(1.15)	(1.30)	
Weight (kg) (1y)	52.41	1.15	1.14	0.61	52.60	52.23	2.98**	−1.07	0.04
	(10.93)	(0.84)	(0.87)				(1.29)	(1.22)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 *p*-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.34 – Trajectory of social support: presence of grandmothers in the household

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Grandmother trajectory index	−0.00	0.34***	0.16**	0.02**	0.01	−0.01	0.13	0.20**	0.56
	(1.00)	(0.08)	(0.07)				(0.10)	(0.08)	
Grandmother present (7y)	0.31	0.11***	0.06	0.16	0.33	0.30	0.06	0.06	0.98
	(0.46)	(0.04)	(0.04)				(0.06)	(0.05)	
Grandmother present (1y)	0.42	0.15***	0.06**	0.08*	0.43	0.41	0.04	0.08**	0.55
	(0.49)	(0.04)	(0.03)				(0.05)	(0.03)	
Grandmother present (6m)	0.47	0.16***	0.07**	0.08*	0.45	0.48	0.05	0.11**	0.30
	(0.50)	(0.04)	(0.03)				(0.04)	(0.05)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 *p*-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.35 – Trajectory of relationship quality with husband and mother-in-law

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Relationships	−0.00	0.14*	0.16*	0.08*	0.05	−0.04	0.14	0.19	0.76
	(1.00)	(0.09)	(0.09)				(0.12)	(0.12)	
Marital quality scale (7y)	5.10	0.11	0.17	0.66	5.18	5.03	0.11	0.22	0.64
	(1.59)	(0.13)	(0.13)				(0.16)	(0.20)	
Relationship husband (7y)	3.91	0.03	0.05	0.74	3.93	3.90	0.01	0.12	0.49
	(0.98)	(0.08)	(0.08)				(0.11)	(0.11)	
Husband nonviolent (7y)	0.74	−0.01	−0.02	0.74	0.75	0.73	−0.04	0.01	0.52
	(0.44)	(0.04)	(0.04)				(0.05)	(0.07)	
Relationship m-in-law (7y)	3.27	0.15	0.16	0.66	3.30	3.25	0.15	0.17	0.90
	(1.10)	(0.11)	(0.13)				(0.18)	(0.15)	
Marital quality scale (1y)	3.49	0.13	0.16	0.48	3.59	3.40	0.10	0.20	0.54
	(1.07)	(0.10)	(0.10)				(0.12)	(0.13)	
Relationship husband (1y)	3.91	0.26***	0.25**	0.10*	3.97	3.85	0.22*	0.28**	0.67
	(0.89)	(0.09)	(0.10)				(0.12)	(0.13)	
Husband nonviolent (1y)	0.70	0.06	0.04	0.66	0.71	0.69	0.04	0.04	0.96
	(0.46)	(0.04)	(0.03)				(0.04)	(0.05)	
Relationship m-in-law (1y)	4.72	0.17	0.40**	0.26	4.80	4.65	0.31	0.51*	0.62
	(2.68)	(0.19)	(0.20)				(0.29)	(0.27)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.36 – Parenting inputs during infancy

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Parenting inputs by 1y	−0.00	0.63***	0.59***	0.00***	0.08	−0.07	0.55***	0.62***	0.73
	(1.00)	(0.11)	(0.11)				(0.15)	(0.15)	
Exclusive breastfeeding (6mo)	0.11	0.09**	0.09**	0.06*	0.11	0.11	0.09*	0.10*	0.89
	(0.32)	(0.04)	(0.04)				(0.05)	(0.06)	
Breastfeeding (6mo)	0.91	0.00	−0.00	0.91	0.93	0.90	−0.01	0.00	0.83
	(0.28)	(0.02)	(0.02)				(0.03)	(0.03)	
Mother play frequency with infant (12mo)	2.37	0.39***	0.35***	0.00***	2.35	2.39	0.38***	0.32***	0.69
	(0.77)	(0.07)	(0.07)				(0.11)	(0.09)	
Father play frequency with infant (12mo)	2.28	0.28***	0.24***	0.04**	2.35	2.21	0.20	0.27**	0.69
	(0.91)	(0.09)	(0.09)				(0.14)	(0.11)	
Completed immunization	0.84	0.10***	0.11***	0.01**	0.84	0.85	0.10**	0.11**	0.90
	(0.36)	(0.03)	(0.03)				(0.04)	(0.04)	
Discussed child's development with family (12mo)	0.14	0.07*	0.05	0.21	0.19	0.10	0.02	0.08**	0.21
	(0.35)	(0.04)	(0.03)				(0.05)	(0.04)	
Selected appropriate place for delivery	0.75	0.18***	0.16***	0.00***	0.74	0.75	0.19***	0.15***	0.58
	(0.44)	(0.04)	(0.04)				(0.05)	(0.05)	
Arranged transport for delivery	0.70	0.22***	0.19***	0.00***	0.69	0.71	0.20***	0.19***	0.78
	(0.46)	(0.05)	(0.05)				(0.06)	(0.06)	
Arranged finances for delivery	0.75	0.17***	0.16***	0.01**	0.74	0.75	0.17***	0.14**	0.69
	(0.44)	(0.05)	(0.05)				(0.06)	(0.06)	
Practicing birth spacing (12mo)	0.55	0.10**	0.11***	0.02**	0.55	0.55	0.13*	0.10**	0.76
	(0.50)	(0.04)	(0.04)				(0.07)	(0.04)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table I.37 – Infant development

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Infant development index	−0.00	0.38***	0.32***	0.00***	0.10	−0.09	0.34***	0.28**	0.70
	(1.00)	(0.09)	(0.08)				(0.11)	(0.12)	
Height-for-age (z) 6m	−0.70	0.07	0.01	0.96	−0.58	−0.82	0.03	−0.04	0.63
	(0.98)	(0.08)	(0.07)				(0.09)	(0.12)	
Weight-for-age (z) 6m	−0.82	0.02	−0.03	0.96	−0.74	−0.90	0.00	−0.09	0.60
	(0.99)	(0.08)	(0.08)				(0.11)	(0.12)	
No Diarrhea episodes (6mo)	0.55	0.09**	0.06	0.47	0.57	0.53	0.04	0.07	0.68
	(0.50)	(0.04)	(0.04)				(0.05)	(0.06)	
No Acute Respiratory Infection (6mo)	0.55	0.06	0.06	0.54	0.55	0.55	0.09	0.03	0.54
	(0.50)	(0.04)	(0.04)				(0.06)	(0.06)	
Height-for-age (z) 1y	−1.34	0.23**	0.16**	0.21	−1.11	−1.56	0.09	0.19	0.61
	(1.18)	(0.09)	(0.08)				(0.13)	(0.13)	
Weight-for-age (z) 1y	−2.12	0.06	0.02	0.96	−1.95	−2.28	0.06	−0.07	0.49
	(1.04)	(0.08)	(0.07)				(0.11)	(0.12)	
No Diarrhea episodes (12mo)	0.58	0.08*	0.06	0.54	0.60	0.57	0.05	0.08	0.77
	(0.49)	(0.05)	(0.05)				(0.06)	(0.06)	
No Acute Respiratory Infection (12mo)	0.47	0.24***	0.24***	0.00***	0.46	0.48	0.26***	0.22***	0.60
	(0.50)	(0.04)	(0.04)				(0.05)	(0.05)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: N=585. Summary indices were created (following Kling et al. (2007); Anderson (2008)) such that they are mean 0 and standard deviation of 1 in the control group, with positive values always associated with more favorable outcomes. Description of the construction of the indices can be found in Appendix Section 4.1. Tables D.1 and D.2 show the summary statistics for the variables included in each index. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

J Treatment effects for specific scales and their subcomponents

Table J.38 – HOME score components

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
HOME inventory	34.01	1.92***	1.72***	0.01**	33.45	34.52	3.65***	−0.35	0.00
	(8.98)	(0.67)	(0.63)				(0.81)	(0.93)	
HOME: Responsivity	8.75	0.21	0.22*	0.19	8.58	8.91	0.63***	−0.22	0.00
	(1.75)	(0.13)	(0.12)				(0.15)	(0.19)	
HOME: Encouragement of maturity	5.25	−0.11	−0.16	0.34	5.36	5.15	−0.03	−0.36*	0.20
	(1.56)	(0.14)	(0.13)				(0.16)	(0.20)	
HOME: Emotional climate	4.50	0.35**	0.37**	0.15	4.43	4.57	0.90***	−0.24	0.00
	(1.93)	(0.16)	(0.17)				(0.21)	(0.25)	
HOME: Learning materials	2.64	0.32***	0.31**	0.10*	2.56	2.70	0.40***	0.23	0.37
	(1.48)	(0.11)	(0.12)				(0.14)	(0.15)	
HOME: Enrichment	2.65	0.31**	0.31**	0.10	2.50	2.79	0.51***	0.14	0.07
	(1.39)	(0.13)	(0.13)				(0.15)	(0.16)	
HOME: Family companionship	2.95	0.38***	0.31**	0.10	2.77	3.11	0.64***	−0.02	0.01
	(1.77)	(0.12)	(0.12)				(0.16)	(0.19)	
HOME: Family integration	2.61	0.20**	0.17**	0.10*	2.61	2.62	0.21*	0.11	0.52
	(0.92)	(0.07)	(0.06)				(0.11)	(0.10)	
HOME: Physical environment	4.65	0.27	0.20	0.34	4.64	4.66	0.39*	0.00	0.31
	(2.37)	(0.17)	(0.15)				(0.20)	(0.28)	
Positive parenting (interviewer obs.)	9.18	0.30	0.33*	0.19	8.92	9.41	0.84***	−0.19	0.00
	(2.33)	(0.18)	(0.17)				(0.21)	(0.27)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 *p*-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table J.39 – School quality components

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
School quality	−0.42	0.64***	0.51**	0.04**	−0.65	−0.21	0.80**	0.19	0.14
	(2.21)	(0.21)	(0.23)				(0.36)	(0.24)	
Classroom amenities	2.94	−0.02	−0.02	0.96	2.96	2.93	−0.05*	0.02	0.04
	(0.26)	(0.03)	(0.03)				(0.03)	(0.04)	
School has office	0.76	0.08*	0.06	0.73	0.68	0.84	0.17**	−0.05	0.01
	(0.43)	(0.04)	(0.04)				(0.07)	(0.04)	
School has playground	0.81	0.06	0.06	0.67	0.79	0.83	0.04	0.09	0.53
	(0.39)	(0.04)	(0.04)				(0.04)	(0.06)	
School has library	0.38	0.10*	0.08	0.67	0.34	0.42	0.15*	0.00	0.16
	(0.49)	(0.05)	(0.05)				(0.08)	(0.05)	
Library books visably in use	0.35	0.10**	0.09*	0.55	0.32	0.39	0.15*	0.03	0.18
	(0.48)	(0.05)	(0.05)				(0.08)	(0.05)	
School has water source	0.89	0.03	0.02	0.96	0.87	0.90	0.05	−0.01	0.15
	(0.32)	(0.04)	(0.03)				(0.04)	(0.04)	
School has clean drinking water	0.92	−0.01	−0.00	0.96	0.89	0.96	0.04	−0.05	0.13
	(0.26)	(0.03)	(0.03)				(0.05)	(0.03)	
School has fencing	0.97	0.02	0.01	0.83	0.96	0.98	0.03	−0.01	0.19
	(0.16)	(0.01)	(0.01)				(0.02)	(0.02)	
School has computers	0.26	0.12***	0.09*	0.46	0.24	0.28	0.10	0.08	0.86
	(0.44)	(0.04)	(0.04)				(0.07)	(0.06)	
Computers visably in use	0.24	0.13***	0.09*	0.46	0.21	0.25	0.10	0.09	0.90
	(0.43)	(0.04)	(0.05)				(0.07)	(0.06)	
Has toilets for girls	0.46	0.07	0.06	0.67	0.51	0.41	0.08	0.03	0.51
	(0.50)	(0.04)	(0.04)				(0.06)	(0.06)	
Total teachers (ln)	2.10	0.19***	0.14*	0.47	2.01	2.18	0.20**	0.09	0.29
	(0.70)	(0.07)	(0.07)				(0.10)	(0.08)	
Total rooms (ln)	1.99	0.17**	0.12	0.73	1.88	2.09	0.23*	0.02	0.12
	(0.78)	(0.08)	(0.09)				(0.13)	(0.10)	
Class size	20.21	0.85	0.86	0.96	19.48	20.88	1.54	0.24	0.50
	(10.74)	(1.26)	(1.38)				(1.59)	(1.79)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: The school quality index is computed as the principal component of individual measures. Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table J.40 – WPSSI components

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
WPPSI Full Scale IQ	81.95	1.27	0.85	0.38	81.99	81.92	2.48*	−1.01	0.04
	(11.43)	(1.02)	(0.96)				(1.31)	(1.17)	
WPPSI: Verbal comprehension	85.15	1.54	0.67	0.85	85.31	84.99	1.75	−0.43	0.27
	(13.59)	(1.37)	(1.30)				(1.65)	(1.56)	
WPPSI: Visual spatial	87.34	−0.41	−0.83	0.80	86.74	87.88	0.41	−2.03	0.29
	(15.06)	(1.07)	(1.05)				(1.60)	(1.49)	
WPPSI: Fluid reasoning	77.26	1.76**	1.43*	0.22	76.65	77.82	3.19***	−0.37	0.06
	(11.42)	(0.80)	(0.76)				(1.05)	(1.32)	
WPPSI: Working memory	99.57	0.88	0.42	0.85	99.26	99.85	1.37	−0.30	0.42
	(15.69)	(1.10)	(1.02)				(1.59)	(1.35)	
WPPSI: Processing speed	76.32	2.44***	2.52***	0.01***	77.28	75.43	3.29***	1.30	0.22
	(9.58)	(0.74)	(0.77)				(1.15)	(1.05)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table J.41 – Spence Anxiety Scale components

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
Spence Child Anxiety Scale	20.56	1.18	1.00	0.25	22.93	18.41	−0.28	1.74	0.37
	(13.32)	(0.91)	(0.85)				(1.47)	(1.28)	
SCAS: Panic and agoraphobia	1.54	0.31	0.33	0.34	1.64	1.46	0.23	0.42	0.73
	(2.71)	(0.22)	(0.20)				(0.34)	(0.33)	
SCAS: Separation	5.96	0.24	0.16	0.75	6.50	5.48	−0.21	0.45	0.26
	(4.05)	(0.28)	(0.27)				(0.43)	(0.34)	
SCAS: Injury fear	6.07	−0.05	−0.08	0.76	7.15	5.10	−0.26	−0.25	0.99
	(3.65)	(0.27)	(0.27)				(0.38)	(0.39)	
SCAS: Social phobia	2.40	−0.11	−0.19	0.72	2.71	2.12	−0.63	0.25	0.17
	(2.93)	(0.21)	(0.21)				(0.41)	(0.35)	
SCAS: Obsessive-compulsive	1.18	0.56***	0.50***	0.02**	1.16	1.20	0.65***	0.31	0.32
	(1.93)	(0.17)	(0.18)				(0.20)	(0.28)	
SCAS: General anxiety	3.40	0.23	0.28	0.72	3.77	3.06	−0.06	0.56	0.25
	(3.28)	(0.26)	(0.27)				(0.39)	(0.37)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Spence total score is the sum of all subcomponents. Higher values of all measures indicate less favorable outcomes (more problems). Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.

Table J.42 – SDQ components

	Full sample				By child gender				
	(1) Control mean	(2) No controls β (s.e.)	(3) All controls β (s.e.)	(4) FWER p-value	(5) Girl control mean	(6) Boy control mean	(7) β^{Girl} (s.e.)	(8) β^{Boy} (s.e.)	(9) $\beta^{Girl} = \beta^{Boy}$ p-value
SDQ Total Score	11.10	0.26	0.11	0.77	10.67	11.50	−0.36	0.83	0.15
	(5.24)	(0.37)	(0.37)				(0.46)	(0.64)	
SDQ: Emotional	2.34	0.14	0.14	0.85	2.50	2.19	−0.07	0.35*	0.19
	(2.06)	(0.14)	(0.14)				(0.23)	(0.20)	
SDQ: Conduct problems	3.33	−0.05	−0.14	0.89	3.18	3.46	−0.50***	0.36	0.01
	(2.04)	(0.16)	(0.17)				(0.18)	(0.26)	
SDQ: Hyperactivity	3.50	0.13	0.02	0.89	2.99	3.96	0.09	0.09	0.99
	(2.56)	(0.18)	(0.17)				(0.22)	(0.30)	
SDQ: Peer problems	1.93	0.04	0.08	0.89	1.99	1.88	0.12	0.03	0.70
	(1.57)	(0.10)	(0.10)				(0.16)	(0.16)	
SDQ: Prosocial	2.49	−0.09	−0.09	0.89	2.06	2.87	−0.19	0.17	0.27
	(2.53)	(0.18)	(0.17)				(0.20)	(0.27)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: SDQ total score is the sum of all subcomponents excluding prosocial. Higher values of all measures indicate less favorable outcomes (more problems). Heteroskedasticity robust standard errors, clustered by Union Council, in parentheses. All models control for interview date and interviewer fixed effects. 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 p-values which control for the family-wise error rate (FWER), calculated using a free step-down resampling method (Westfall and Young, 1993). Columns 7 and 8 report treatment effects by gender of the index child (controlling for all baseline characteristics) and Column 9 reports the test of equality in treatment effects between the two samples.