Early Childhood Development in the Slums of Cuttack , Odisha, India
Baseline report*

28th April 2015

Alison Andrew[⋆], Orazio Attanasio[⋆,†], Britta Augsburg[⋆], Sally Grantham-McGreggor[⋆], Smriti Pahwa[§], Costas Meghir[⋆,ϒ], and Marta Rubio-Codina[⋆]

[⋆]The Institute for Fiscal Studies, London, UK
[†]Univeristy College London, UK
[§]Pratham Education Foundation- ASER Centre, India
[ϒ]Yale University, USA

1 Overview

1.1 Introduction

The very earliest years of life are key to fulfilled, productive and meaningful lives. Children's brain and physical development is at its most rapid during these first years as they develop skills and capabilities that affect lifetime outcomes as diverse as lifetime earnings, wellbeing and criminality. Gaps that open up between children, often along familiar lines of wealth and income, during this stage typically persist and are exacerbated over time. Thus, these years are key to understanding the transmission of poverty across generations. For many children growing up in poorer countries, these earliest years don't offer conditions that are always sufficient to reach their developmental potential. Poverty, malnutrition and disease-ridden and unstimulating environments can all contribute to children falling short developmentally of what they otherwise would have been capable of. Excitingly, however, a vibrant research agenda demonstrates that a child's development is not predetermined but highly malleable: it is heavily affected by environmental factors which can be altered by policy or behaviour change. This creates a clear rationale for intervening early in life, especially for the most disadvantaged children.

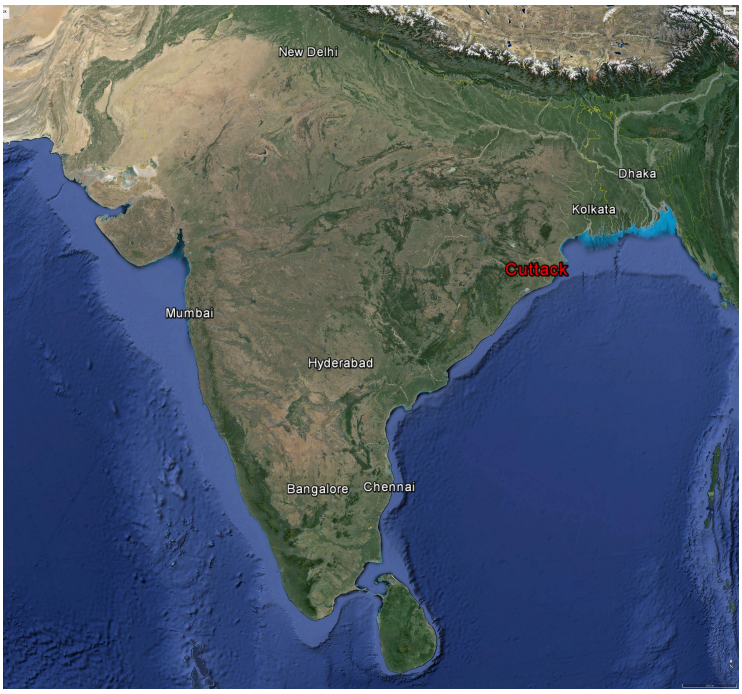
*This research was made possible through the generous gift of Mr Martin Rushton-Turner and the support of the Waterloo Foundation; Orazio Attanasio thanks the ESRC for funding under the Professorial Fellowship RES-051-27-0135 and we also thank the ESRC Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies. Responsibility for any errors is ours. The research team would also like to thank JPal for their invaluable work on this project and the ASER Centre social sector unit team provided inputs on the health and nutrition inquiry segments.

In this study researchers at the IFS, UCL and Yale, in collaboration with Pratham, JPal-SA, will evaluate an intensive Early Childhood Development (ECD) intervention, in Cuttack, India (see figure 1 for location). The home visiting programme aims to increase the development of cognitive and non-cognitive skills, through increasing the level of psychosocial stimulation children are exposed to. It consists of 18 months of weekly home visits, from trained local women following a structured curriculum, and activities that mothers are encouraged to do between visits. The aim is to improve levels of interaction and attachment between mothers and their infants, creating a more stimulating environment for the child and increasing his or her expected level of development.

The impact evaluation of this intervention, which is based on a randomised control trial design, will directly and rigorously study the effectiveness of the home visiting intervention over a broad range of child development indicators. Further work will attempt to analyse the mechanisms through which the programme impacted (if at all) these measures. The evaluation design includes a baseline survey before the intervention as well as an endline survey after the implementation of the program. This reports focuses on the baseline survey.

This report offers a detailed description of sampling methodology, the practicalities of baseline data collection and descriptives of the data itself. Baseline was collected between 3 October 2013 and 24 February 2014, just before the intervention began. We provide descriptive statistics of our sample over a many facets of dimensions such as household structure and characteristics, economic indicators of income and expenditure, education, health and indicators of child development. These descriptive statistics provide an interesting snapshot of our target population – households with young children living in sahis around the medium sized city of Cuttack, Odisha. In terms of checking the validity of our ultimate evaluation we test for any systematic differences between treatment and control groups which could undermine the argument that our randomisation led to two treatment groups balanced on observable and unobservable characteristics.

Figure 1: Location of Cuttack within India



1.2 Project Background

In 2004 India was home to more disadvantaged children under the age of five than any other country: 65 million under fives met the criteria of disadvantage, defined as either being stunted, living in poverty or both [1]. These children face huge barriers, from a lack of resources and unhealthy and unstimulating home environments, to fulfilling their developmental potential in the early years. This often later results in poor performance within the formal education system and later in the labour market. Partly driven by their own poverty they are often unable to provide high quality care for their own children who will subsequently also have poor life chances. Not only do these 65 million children represent a huge loss of potential healthy and productive lives they also represent an enormous social and economic opportunity. If policy interventions are found which can both mitigate some of the early disadvantages these children face, and that could be rolled out at a large scale sustainably and at a suitable cost, the potential gains are huge. We would hope to see improvements in young children's developmental levels, school performance, labour market success - factors that would ultimately weaken the bind of poverty and its intergenerational transmission.

India created formal institutions responsible for early childhood development early by international standards. In 1975 the Integrated Child Development Services (ICDS), now the world's largest integrated early childhood programme, was created with the following stated objectives[2]:

1. to improve the nutritional and health status of children in the age-group 0-6 years
2. to lay the foundation for proper psychological, physical and social development of the child
3. to reduce the incidence of mortality, morbidity, malnutrition and school drop-out

4. to achieve effective co-ordination of policy and implementation amongst the various departments to promote child development

5. to enhance the capability of the mother to look after the normal health and nutritional needs of the child through proper nutrition and health education

The plan was to achieve these objectives through providing the following services[2]:

1. Supplementary Nutrition for all children below six years old and pregnant and lactating mothers

2. Immunisation for all children below six years old and pregnant and lactating mothers

3. Health check-ups for all children below six years old and pregnant and lactating mothers

4. Referral services for all children below six years old and pregnant and lactating mothers

5. Pre-school education for all children between three and six years old

6. Nutrition and Health education for all women between 15 and 45 years old

The backbone of the ICDS was the creation of Anganwadi Centres (AWCs) - each with an Anganwadi Worker (AWW) and an Anganwadi Helper (AWH) - in every village, and subsequently in every settlement. This has created a huge network of institutions formally dedicated to improving outcomes in early childhood and has contributed to significant gains being made to children's services on a very large scale. However, it is evident, and particularly so with ever increasing evidence on the importance of the earliest years in laying the foundations for lifetime achievement and wellbeing, that there could be potential gains from filling gaps in the current role and functionality of the ICDS and Anganwadi services. Particularly relevant to the context of the present study is the current scope and performance of the pre-school education service. The pre-school education role of ICDS aims to provide "a natural, joyful and stimulating environment, with emphasis on necessary inputs for optimal growth and development" [2]. However, this function of the ICDS often falls short of its aims - an appraisal in 2006 found that almost half of AWCs lacked space to conduct outdoor and indoor activities which was inhibiting the functioning of the pre-school activities. Likewise, 44% of AWCs were found to be lacking pre-school education kits which are an essential part of planned pre-school activities. Furthermore, the service only covers children between the ages of three and six despite increasing evidence that programmes encouraging stimulating play in younger children can be very effective at improving lifetime outcomes. Clearly an extension/restructuring of the ICDS programmes to increase stimulation to younger ages would require a different model. A first step in this direction was undertaken through the ICDS restructuring which was issued in 2012, through the National Early Childhood Care and Education (ECCE) Policy.This present study evaluates a home-visiting programme which goes beyond the changes proposed in that restructuring process, particularly focusing on a closer engagement with mothers and following a very structured curricula. Whilst more time intensive this model would lead to increased opportunities to induce behaviour change and increase the stimulation young children face within the home as well as at the AWC and hence could, possibly, with further adaptations, prove a valuable model in attempting to extend further the role of ICDS in this area.

Home visiting schemes for disadvantaged mothers with young children have been very effective in various parts of the world at mitigating some of the unfavourable influences of poverty and improving

the developmental levels of young children. In Jamaica a home visiting programme centred on increasing levels of psychosocial stimulation of children and strengthening the mother (or main caregiver)-child bond had significant positive impacts on children's levels of cognitive functioning, non-cognitive skills, educational and labour market outcomes both in the short, medium and long term [3, 4, 5, 6]. A larger scale programme in Colombia demonstrated that an adapted version of the Jamaican curriculum could significantly improve cognition and language development when delivered in a scalable manner through the existing institutional infrastructure of a conditional cash transfer scheme [7]. This finding is important since it showed an ambitious aim to improve child development through altering complex caregiver-child relationships and interactions, could be achieved using local resources and through pre-existing institutions. This suggests that such a programme could be implemented at a much larger scale. This evaluation furthers this research agenda by asking whether an adaptation of the Jamaican curriculum could be effective at increasing levels of child development when delivered through local women in poor Indian urban neighbourhoods. Like the Colombian study this programme also makes use of existing institutional infrastructure although this time leveraging the infrastructure of one of the country's largest non-governmental organisations, Pratham.

1.3 Home Visiting Programme

The home visiting programme evaluated in this study is based on the model and curriculum designed by Sally Grantham-McGregor for use in Jamaica. As described in section 1.2 the Jamaican home visiting programme has had very impressive impacts on cognitive and non-cognitive child development, as well as on much longer term outcomes like educational attainment and labour market success[3, 4, 5, 6]. Positive impacts have also been found when the programme was adapted for other countries and contexts[7].

The basic idea of the programme is weekly home visits for a sustained period of time during which the home visitors follow a structured curriculum. Below we provide more details on the curriculum, the home visitors and their training and mentoring.

1.3.1 Curriculum

The core of the programme is structured curriculum of play and other developmental activities that the home visitor follows every week when she visits the target child and his or her main caregiver (usually his or her mother). Such developmental activities could include pushing different shaped blocks through the correct holes or playing with a rubber toy in a bowl of water. The home visitor does these activities with the mother and child during the visit but also encourages the mother to continue with the activities in the coming week, before the next visit. The home visitor leaves any materials required with the mother for that week. The activities are all designed to be stimulating for the child and to enhance one or more domains of child development. In parallel to increasing levels of stimulation through performing these activities the curriculum is designed to increase the interaction between mother and child even when not doing these specific activities. It encourages mothers to be creative in finding fun and stimulating opportunities for play and learning that could be build into an everyday routine, without specific need for extra bought toys or materials. For example, mothers are given ideas about games they could play with their child during washing or preparing food. In addition, most of the toys that are used during home visits are made with locally available materials,

often discarded objects such as empty plastic bottles. The idea is that integrating psychosocially stimulating activity into everyday routine is the best way of increasing the overall level of stimulation children are exposed to and doing so in such a way that does not put unrealistic demands on the household in terms of material or time resources.

Clearly, the original curriculum developed for Jamaica was not directly suitable for use in Odisha. In addition to translation it needed very significant adaptations to make it as relevant as possible for poor urban households in Cuttack. The International Centre for Diarrhoeal Disease Research, Bangladesh - an international health research institution located in Dhaka - took the lead in adapting the curriculum. In this were supported by Sally Grantham-McGregor, the creator of the original Jamaican curriculum, and other researchers.

1.3.2 Home visitors

The home visitors, who deliver the curriculum, are local women without any specific experience in child development, who are trained specifically by experts to deliver the home visiting intervention. Using paraprofessionals is an important design element of the home visiting programme for both theoretical and practical reasons. Firstly, there is a clear cost and scalability imperative to use paraprofessionals rather than professionals in child development. It would be near impossible to find sufficient numbers of professionals in Cuttack and this would be extremely expensive. Since paraprofessionals have much lower formal qualifications they can be employed at much lower rates. Furthermore, paraprofessionals could be more effective than child development professionals since they may be well known in the community and there is likely a lesser gap in socio-economic status between them and the targeted families. This would suggest paraprofessionals might be better at encouraging households to take up the intervention and could be better at relating to the target mothers and making the target mother more at ease.

Because the curriculum is delivered by women who do not have an explicit background in child development it is carefully crafted to cater for their needs. The result is that it is far more structured, in terms of instructing the home visitor what activity leads into another and what advice to give to mothers, than it would be if the intervention was delivered by professionals. This is because we cannot rely on the home visitors having the background knowledge and experience needed for creating a successful visit without guidance.

Pratham, the implementation partner, was responsible for recruiting the 27 home visitors. Many of them were recruited through Pratham's extensive network in the sahis (slums) - many had previously been involved in other programmes Pratham had run in the sahi. The criteria for being eligible for selection as a home visitor was as follows:

1. **Education**: 12th class completion preferable, 10th class completion a must
2. **Age**: Minimum 18 years
3. **Experience** of working with children preferable
4. **Residence location:** either in the same sahi or close by

The characteristics of the home visitors and their initial knowledge of child development are described in section 7.

1.3.3 Training and monitoring

Home visitors were trained by three Pratham mentors (who were themselves trained by experts in the home visiting curriculum from ICDDR,B in Bangladesh). In total the initial home visitors had three weeks of training in Cuttack (24 September - 9 October 2013 and 21 October - 28 October 2013) with an additional week of toy making workshops (29 October - 4 November 2013). Home visitors that joined the program later were trained by the Pratham mentors at their convenience. All home visitors had two days of refresher training between 14 July - 15 July 2014.

Throughout the intervention a proportion of home visits (approximately one per home visitor per week) are observed by and discussed with Pratham mentors. This provides a ongoing opportunity for training. In addition, each home visitor meets with her mentor every Saturday morning to discuss the progress of each child, again providing an opportunity for continuous learning.

1.4 Our Evaluation Study

Our evaluation study will test whether a programme of weekly home visits, delivered over 18 months, can improve various indicators of early childhood development. We also aim to study how the intervention affects child development by looking how it affects investments made by families into young children. The full list of outcomes we intend to study are detailed in section 2.4. We hope that the evidence that we generate from this study will help donors and policy makers target spending on early childhood development policies that are most effective at boosting key outcomes and working to mitigate the intergenerational transmission of poverty.

Our study is complex and diverse in its components - from recruiting and implementing the home visiting programme through local networks, to adapting and training intricate assessments of child development, to large scale data collection, to data analysis. Below is a list of the partner institutions involved in the intervention and evaluation:

1. The academic lead on the project lies with researchers at the **Institute for Fiscal Studies** (IFS), the **University College London** (UCL) and **Yale University**, working closely with researchers at the implementing agency.

2. The implementing agency is **Pratham**, one of the largest non-governmental organisation in India, which works towards the provision of quality education to the underprivileged children in India.

3. The adaptation of the intervention to the local context was lead by the **International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B)**, an international health research institution located in Dhaka.

4. Data collection was managed by the **Abdul Latif Jameel Poverty Action Lab**, South Asia (JPAL-SA).

2 Methodology

2.1 Terminology

Here we briefly spell out the terminology we use to describe the subjects of our study:

1. **Target children** are the key subjects of interest in the study. To have been selected into the study sample target children must meet to full eligibility criteria: (1) being between 10 and 20 months at the time of baseline, (2) residing in one of the 54 study sahis, (3) not being a twin and (4) not having a physical or mental disability. In the intervention sahis target children are all eligible for the home visiting intervention. The evaluation design will directly compare the outcomes of the target children in the intervention sahis with those in the control after the end of the intervention.

2. **Biological mothers** of the target children are of key interest in our study since the maternal-child bond is so crucial in child development. They are typically also the main caregiver to the target child and thus the agent through which we hope the intervention will induce behaviour change.

3. **Main caregivers** are only identified for children where their biological mothers are not the person who predominantly cares for the child and his or her wellbeing. In our sample few children have main caregivers other than their biological mothers (29 out of 421).

4. **Households** of target children are defined for the purpose of the study as a group of people who typically share a cooking pot.

5. **Sahis** are the 54 study sites (or slums), in and around Cuttack.

6. **Home visitors** are only defined in the 27 treatment sites. They are the women identified and employed by Pratham, the implementing partner, to deliver the home visiting intervention.

2.2 Evaluation design

We are using a cluster randomised controlled trial (RCT) to evaluate the effects of the home visiting programme, with the unit of randomisation being the sahi. Thus, half the sahis were randomly allocated to receive the home visiting intervention whilst half were randomly allocated to be part of the control group. At the end of the programme the average difference in the outcomes between the two groups can be interpreted as the average effect of being offered the programme.

2.2.1 Randomisation

The 54 sahis chosen to be part of the study (see section 2.2 for details of sample selection) were randomly allocated to one of two possible treatment[1] states: (1) treatment - that is receiving the ECD home-visiting intervention and (2) control - receiving no additional service. See table 53 in the Appendix for the full randomisation outcome.

[1] The terms 'treatment' and 'control' come from the medical literature where individuals in the treatment group are given some treatment (or intervention) and those in the control are do not receive any active treatment.

Randomising is vital for our evaluation approach since it ensures that the children and households in the treatment and control groups will be, in expectation, identical in terms of all observed *and unobserved* characteristics. This means that any (statistically significant) difference in the outcomes of interest that we observe between children and households in the two groups after the intervention can be attributed to the effects of the intervention. If we were to select which sahis were to receive the intervention by some other means, for example selecting the poorest sahis, then the two groups would look different in terms of observable characteristics. More importantly they would also look different in terms of characteristics that we cannot directly observe but that are likely correlated with income. Since we cannot observe these characteristics we cannot control for them in our analysis and there is no way, without making strong assumptions, to know whether any difference in outcomes between the two groups, at the end of the intervention, can be attributed to the effects of the intervention or whether the difference arose from pre-existing unobservable differences between the groups. This is what we refer to as *sample selection bias* and is what we solve through randomisation.

When designing a randomised controlled trial there are various options for the unit, or level, of randomisation: we could have randomised the intervention across households, or streets, or sahis. We chose sahis as an appropriate geographical level of randomisation because:

1. randomising at the level of this larger geographical unit, which encompasses much social interaction, diminishes the likelihood of treatment contamination where the controls are indirectly exposed to the programme through information or resources diffusing through social networks.

2. randomising at a larger level means that the intervention takes place at the level of the whole sahi. This makes the organisation of the intervention simpler and also more similar to what might occur in a non-experimental setting.

We stratified our randomisation of treatment status on the number of children in the target age range in the sahi. There were two stratas: (1) sahis that had more than nine children in the target age range (10-20 months), a total of 36 sahis, and (2) sahis that had fewer children in the target age range, a total of 18 sahis. This is because in the larger sahis we were only offering the intervention to a random nine children whereas in the smaller sahis we were offering the intervention to all children. The reason for this is that our budget allowed us to employ only one home visitor per sahi and it was considered that nine children is the maximum number of children a home visitor could work with without being stretched. Thus, if we anticipate that this might affect the effectiveness of the intervention then stratifying the randomisation on this dimension will increase balance.

2.2.2 Evaluation specification

After 18 months of the intervention we will return to the same households who were included in the baseline sample and collect measures of child development and outcomes of interest (to be discussed in section 2.3), as well as supporting data on the household characteristics. For each outcome of interest we will estimate the impact of eligibility for the home visiting programme by running a regression of the following form:

$$y_{ij} = \alpha + \beta T_j + \gamma X_{ij} + \nu_j + \epsilon_{ij} \tag{1}$$

where y_{ij} is the outcome of interest for household (or child) i, in sahi j , T is a dummy variable equal

to one if sahi j was allocated to the treatment group receiving the home visiting intervention and equal to zero otherwise, X_{ij} is a vector of observed household, and sahi level, characteristics measured at baseline (including the baseline measure of the outcome of interest), υ_j is a cluster-specific unobserved effect and $\epsilon_{ij}$is a random error term. Note that the error term, ϵ_{ij}, cannot be assumed to be independent between households (or children) since households living in the same sahis may be subject to correlated unobserved shocks or their unobserved characteristics may be correlated. Therefore for our inference we will cluster errors at the level of the sahi, allowing for arbitrary correlation between error terms of households in the same sahi. In this regression framework the most interesting parameter is β, our estimate of the impact of being eligible for the home visiting programme. It is the size and significance of this parameter that will tell us the impact of the intervention on the outcome of interest and the degree of uncertainty associated with that estimate.[2]

It is important to note that this set-up is an *Intention to Treat* analysis - we are estimating the impact of being in a sahi that was allocated to the treatment group, and thus that the child was eligible for the home visiting programme. This may be different from the impact of actually receiving the intervention if some households decide not to participate in the programme even though they were eligible, for example if they perceived the programme would be of no benefit to their child. Using an Intention to Treat framework is optimal in our case for two reasons. Firstly, if the households that chose not to participate were different, on important dimensions, from those that did then this would introduce selection bias as we would be comparing only the treatment group who had *chosen* to take part with *all* of the control group. These groups may look different in terms of the underlying distribution of observable and unobservable characteristics which would undermine the randomisation. Secondly, the *unconditional* effect of a household being offered the home visiting programme is arguably more useful from a policy perspective than the effect *conditional* on choosing to participate. The unconditional effect of being offered the programme is our best estimate on the effect on the 'average' child of expanding the policy on a larger scale.

This specification controls for baseline values of the outcome of interest and other characteristics, measured at baseline. This will not affect the expected value of our estimator of the treatment effect β, which will be unbiased regardless of whether we control for these variables or not. However, it will increase the precision of our estimate, i.e. it will reduce the standard errors associated with our estimates, which will increase the power of our evaluation to detect small effects of the intervention. Collecting a rich set of characteristics at baseline is also important as a check that randomisation was successful. In this report we check that the two treatment groups do indeed look similar in terms of observed characteristics. From this we have to infer that they are likely similar in terms of unobservable characteristics. Another important use of baseline data is to analyse any attrition that occurs at follow-up due to households who refuse to participate in the follow-up survey or cannot be tracked. We can use the baseline data to check whether there are systematic differences between households that are lost during the follow-up survey, and particularly whether we see different attrition patterns in the treatment group than in the control.

[2] An alternative to using post-treatment data only or conduct a difference-in-differences analysis. The decision which approach is appropriate boils down to whether the variance of time-invariant individual effects is greater or smaller than the variance of transitory shocks. If the former is smaller, using post-treatment data only is the appropriate strategy. If it is greater, we should use difference-in-differences. McKenzie (2012) shows that difference-in-differences may limit statistical power if autocorrelation in the outcomes is limited. Ex-ante we do not have information on the relative size of these variances and will hence take this up once we have endline data available.

2.3 Sampling selection strategy

2.3.1 Selecting sahis

The study comprises of 54 sahis across Cuttack (see figure 2). An initial list of 100 sahis was drawn up by Pratham, the implementation partner, and from this a first 54 were selected by J-PAL SA following the criteria that these sahis should have between 100 and 350 households. J-PAL SA then conducted mapping and census (see below for more detail) in these sahis and found that 7 of these 54 did not have at least seven children within the 10 to 20 month age range needed for the intervention. Therefore another seven sahis, from the initial list of 100 sahis were selected to replace these sahis leaving a total of 54 sahis which did have enough children in the target age range.

Figure 2: Location of the study's sahis

2.3.2 Selecting target children

We aimed to have 9 target children in each of the 54 sahis. We identified and selected target children who met the inclusion criteria (aged 10-20 months at baseline, living in the 54 study sahis, excluding twins and children with physical or mental disabilities) through the following procedure summarised in figure 3:

1. First, each sahi was mapped using pencil and paper. This process defined the boundaries of the study area and all streets within this area. This stage was crucial to ensure that no household was missed during the census.

2. Next, census teams carried out a digital door-to-door survey with every household in the sahi. If the household contained any children under the age of two the census team asked a set of more detailed questions to confirm the date of birth, the identities of the biological mother and main

caregiver and the households intention to remain in the sahi over the duration of the study. In total we identified 778 children in the target age range living in the 54 sahis through the census

3. If there was nine or fewer children in the target age range (and meeting the full inclusion criteria listed above) identified in the census above then we aimed included all of them in our study (this was the case in 18 sahis). If there were more than nine (as was the case in 36 sahis) then we randomly chose nine children within the age range who we aimed to include. This resulted in an initial list of 459 children who we aimed to include in the study.

4. We attempted to collect baseline data from all of these children. However, in 105 cases baseline could not be completed due to either the date of birth initially having been recorded wrongly so the child was out of age range, the household having relocated or planning to relocate, the household refusing the baseline questionnaires and measurements either during the questionnaire, the Ages and Stages Questionnaire (ASQ) or the anthropometric measurements (see table 1 for breakdown). These children had to be dropped. To replace them, in sahis where there were more children identified during census than had been initially selected, we drew randomly from these extra children until we had nine children who had completed baseline or there were no more extra children. In total we attempted to do baseline with 100 such replacement children of which baseline was completed with 67.

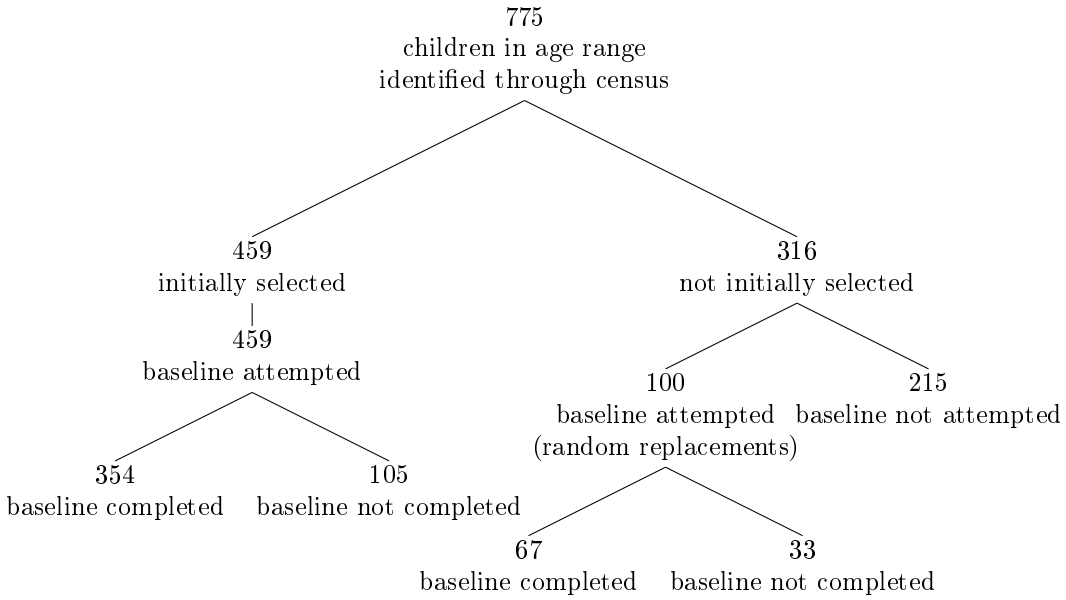
5. This gave us a total of 421 target children for which we have complete baseline data

It is important to note that while the randomisation happened simultaneous to the sample selection all field staff working on the census and baseline did not know the result of the randomisation. Therefore no selection bias could have been introduced through this.

Table 1: Reasons for baseline incomplete

	Number of children
Immediately refused to take part in baseline	20
Refused during household questionnaire/ASQ or between the household questionnaire and ASQ	14
Refused during anthropometric measurements	2
Total refusals	**36**
Household relocated between census and baseline	61
Household relocated after household questionnaire and before anthropometric measurements	11
Total relocation	**72**
Total errors (largely misrecorded ages)	**30**
Total children for whom baseline was attempted but was incomplete	**138**

Figure 3: Process of selecting target children



2.4 Outcome indicators

We are interested in how the psychosocial home visiting programme affects outcomes that fall into two categories: (1) indicators of child development (cognitive, non-cognitive and health) and (2) indicators of inputs to the process of child development, such as the quality of the home environment. It is important to distinguish these two categories. The first category of outcome indicators evaluate *if* the programme affected levels of child development in deprived children, what it was designed to achieve. The second category of indicators allow us to analyse *how* (if at all) the programme affected child development. We call them indicators of child development inputs. They allow us to look at the mechanisms through which the programme works. For example, they might suggest that the programme had a big impact on the amount of time that mothers spent playing with their children suggesting that this was an important mechanism in the programme being effective. Further work, following on the impact evaluation, will involve using this data and econometric methods to estimate structural models of the processes through which child development occurs. Such work is important for extrapolating findings from particular programmes in particular populations to other populations and differently designed programmes.

In the list below we also highlight outcome indicators that we label as primary and secondary (in accordance with the trial registry).

2.4.1 Indicators of child development

2.4.1.1 Cognitive development (Primary outcome) A key aim of the psycho-social home visiting programme is to promote cognitive development. Cognitive development is an umbrella term for different, but interlinked, domains of child development that including sensorimotor development, exploration and manipulation, object relatedness, concept formation, memory as well as language development [8]. Cognition is key in determining success in education and the labour market. It is

during the earliest years of life that areas of the brain that are important for cognition, especially the pre-frontal cortex, are developing at the fastest pace. Thus cognitive development that occurs in these years has big impacts on cognitive functioning for the rest of life and, therefore, on a broad range of adult indicators. The process of cognitive development during these early years is malleable - it is hugely influenced by a child's environment - and can, therefore, be altered by policy interventions, such as the home visiting programme studied in this evaluation.

At follow-up, to evaluate the overall impact of the programme on cognitive development, we will use the Bayley Scales of Infant and Toddler Development, Third Edition [9] which are widely taken to be the gold standard of cognitive development for children under 42 months. However, at baseline, we measured cognitive development using an adapted (see section 4.1 for details of adaptation) version of the the Ages and Stages Questionnaires, Third Edition [10], hereafter referred to as the ASQ-3. The ASQ-3 are a series of questionnaires to be completed by parents of young children about their child. The questionnaires can be answered in the home and no specialist materials are required. This marks a key difference from the Bayley which must be administered by a professional in a test centre with a large number of different aides and materials. The result is that the ASQ-3 is much quicker, cheaper and easier to administer than the Bayley. It also has good psychometric properties. The original ASQ [11] had a concurrent validity with a range of standardised measures was 85%[12]. It also displayed high test-retest reliability, inter-observer reliability, and internal consistency[12]. The test has subsequently been used in many contexts and languages across the world.

2.4.1.2 Non-cognitive development (Primary outcome) Non-cognitive skills are increasingly being seen on parity to cognitive skills when it comes to affecting educational attainment and labour market success [13, 14, 15, 16, 17]. Qualities such as motivation, resilience and social competencies correlate closely with a whole range of desirable outcomes although proving a causal relationship, as with cognitive skills, is more difficult (see Gutman and Schoon [18] for a recent literature review). For non-cognitive skills there is less evidence than for cognitive skills on which periods in a child's development are most critical for the development of these skills and periods where skill development can be moulded by policy.

At baseline we measured one dimension of non-cognitive skills, using the Personal Social scale on the ASQ-3 [10] which we selected as a convenient and reliable measure for the same reasons listed in section 2.4.1.1. Personal social skills are skills involved in caring for oneself (for example, dressing or washing) and skills involved in interacting with others (for example, sharing toys with other children). This is just one of many types of non-cognitive skills (see, again, the recent literature review by Gutman and Schoon [18]).

2.4.1.3 Motor development (Primary outcome) Motor development combines gross motor development - larger body movements, such as crawling or walking, - and fine motor - more intricate movement, such as picking up objects. This domain of child development influences motor skills in later life, skills that may be crucial for later work. Motor skills also impact heavily on levels of physical activity and thus later fitness and health since children with lower levels of motor development are less likely to engage in physical play and exercise[19]. Furthermore, recent research has suggested links between early motor development and cognitive and other forms of development and wellbeing. See Adolph for a recent overview of this area [20].

In this baseline data collection we measured motor development through two subscales of n the ASQ-3[10] - the Fine Motor subscale and the Gross Motor subscale.

2.4.1.4 Health and morbidity (Secondary outcome) Another vital measure of child development whose effects extend into adult life is health. In developing countries poor health and high morbidity constrains all aspects of child development in a very real way as well as having devastating impacts on child survival and quality of life through into adulthood. Malnutrition and infectious disease, for example, inhibit brain and motor development. This relationship in so strong that stunting (height less than two standard deviations below the mean of a standardised distribution), an excellent indicator of long term malnutrition and poor health, is very strongly correlated to a huge range of poor developmental outcomes[1].

At baseline we measured health and morbidity in two ways. Firstly, we took children's anthropometric measurements (height and weight). Child height and weight (and combinations of these two measures with each other and age) are the most usual measure of a child's nutritional status over the medium and longer terms since poor nutrition, or poor absorption due to disease, persistently inhibits children's growth. Secondly, to look at shorter term health and morbidity we collect reports from the child's mother or main caregiver on whether the child has suffered from diarrhoea and/or symptoms of malaria in the past two weeks.

2.4.2 Indicators of child development inputs

2.4.2.1 Quality of the home environment (Secondary outcome) Poor stimulation within the home environment impacts negatively on many developmental domains [21]. At baseline of this study we used an adapted version of the Family Care Indicators (FCI) questionnaire developed by UNICEF to assess levels of stimulation within the home environment. We measure the quality of the home stimulation environment on five subscales: (1) play activities, (2) variety of play materials, (3) sources of play materials, (4) household books and (5) household magazines. The FCI was derived from the much longer and complex tool to measure the quality of the home stimulation environment - Home Observations for Measurement of the Environment - and was designed to be quicker, cheaper and easier to administer in large survey settings. The tool and these subscales have been shown to have good reliability qualities as well as good predictive power over child developmental outcomes (cognitive, language and motor) as measured by the Bayley Scales of Infant Development, in Bangladesh [22].

We also measure the quality of the home environment in promoting good health, in terms of access to sanitation and a smoke-free living environment.

2.4.2.2 Child nutrition, healthcare and birth Good quality nutrition and access to healthcare services, both preventative and when sick, is crucial for children to grow up healthy. To measure nutrition ask mothers to report whether children ate foods belonging to a variety of categories (e.g. pulses) during the past 24 hours. From this we can get some idea of children's intakes of different food groups and macro- and micro-nutrients, as well as measures of dietary diversity. We also measure the breastfeeding history of the target children. In particular we collect information that allows us to construct an outcome variable for a child having been exclusively breastfed for the first six months of life. Exclusive breastfeeding during this period is recommended by the World Health Organisation and has been shown to have many advantages for child growth, development and protection from disease.

In terms of child healthcare we measure whether the child was born in a hospital or clinic. From sahi level data we also know the type of healthcare services available in the sahi and ward. As an important starting condition for all further inputs to child health we measure outcomes around the child's birth - the period of gestation and birthweight.

2.4.2.3 Knowledge of child development Knowledge about children's developmental needs and how to best fulfil them is key to children being raised in a healthy and stimulating environment. Indeed, increasing mother's and caregiver's levels of knowledge and understanding about child development, and thus inducing behaviour change, is a crucial mechanism through which we hypothesise the home visiting intervention may work.

At baseline, as we will also do at follow-up, we measured maternal and caregiver knowledge of key principals of child development using an adapted and shortened version of the Knowledge of Infant Development Inventory [23]. This tool attempts to measure knowledge on parental practices, child development processes and infant norms of behaviour by giving various statements to mothers and caregivers and asking whether they the statement "is true", "is partly true" or "is not true". From these answers we construct aggregate scores which measure knowledge under the following domains: (1) praising/paying attention to child, (2) punishing child, (3) school readiness and expectations, (4) importance of maternal interactions and play and, (5) age appropriate expectations.

2.4.2.4 Quality and quantity of maternal time The quality and quantity of time that young children spend playing and interacting with their mothers and other caregivers directly impacts upon the level of stimulation children experience and thus their development. Good quality time also creates strong attachments between child and mother which further impacts on quality of care and stimulation. We attempt to measure some aspects of the quantity and quality of time that mothers spend with the target children through the FCI (see section 2.4.2.1). However, we also specifically ask the mother to estimate the length of time she spent doing various activities on the previous week day, in order to gain additional measures of the proportion of her time that was dedicated to the care of the target child and, out of this time, how much involved play and stimulating interaction. This data will also give us an idea of the constraints that mothers face on their time which will be important in understanding how the home visiting programme can increase stimulation within these constraints.

2.4.2.5 Maternal and main caregiver wellbeing and education (Secondary outcome) Mothers and main caregivers are crucial in shaping the home environment her child grows up in. Therefore her own health, wellbeing and education are important determinants of a child's environment in the very early years. Mothers in better health (mental and physical) are more able to engage in energetic, active play with their children. Mothers who report higher levels of wellbeing and fewer depressive symptoms may be more able to form secure attachments with their children. More educated and empowered mothers are more likely and able to understand a child's developmental needs and thus be able to cater for them.

At baseline we measured maternal depression symptoms through a shortened and adapted version of the Center for Epidemiological Studies Depression Scale [24], a short self-report scale that is useful in study settings when full clinical assessments would be infeasible. We measured maternal levels of education and basic literacy. We asked a series of questions to measure how empowered the mother is

to make decisions related to her own and her child's wellbeing.

2.4.2.6 Economic resources (Secondary outcome) The economic resources at the disposal of a household affect many factors we believe are important in child development - from good nutrition and access to healthcare to time spent interacting with children. In this baseline household survey we collected many measures relating to the economic resources households have and the economic decisions they make. We collected information on household assets, labour market outcomes for all household members, savings, debts and loans, income and transfers and expenses.

2.5 Instruments for data collection

2.5.1 Household questionnaires

The majority of our data was collected through a series of household questionnaires, divided into eight modules which addressed different members of the household on different topics. Table 2 provides a brief summary of the information collected by each module as well as the average time taken to complete the module. Some of the material used in the household questionnaires was taken from other research projects looking at Early Childhood Development interventions and adapted and translated for use in Odisha. Other material was created specifically for these questionnaires. All the questionnaires were piloted for two weeks.

These household questionnaires were done on paper by a team of 15 surveyors and 3 supervisors over a period of 44 days between 5 November 2013 and 22 December 2013 (although 95% were completed by 5 December) . There was lower productivity than anticipated, in terms of completed households, with a two surveyor team completing, on average, just one household (household questionnaire plus ASQ) per day. This was partly due to many revisits being needed due to mothers and children being absent at the time of the first visit or because respondent fatigue or other commitments meant the questionnaires had to be left incomplete and the survey team had to revisit. In addition, refusals, relocation or errors in recording date of birth meant that baseline was attempted, and often started, for more children than it was finally completed. In addition to the 421 target children for whom baseline was finally completed baseline was attempted with an additional 138 (for details see figure 3).

Table 2: Structure of baseline household questionnaires

Module	Title	Content	Avg time (min)
1	Dwelling and the household	Characteristics of dwelling; education of all household members; property; assets; savings and debts; expenses; income and transfers; shocks; frequency of food intake for children younger than six years	27
2	Workforce	Workforce participation of household members above 12 years	15
2a	Roster	Name gender, relationship , marital status of every household member	14
3	Children younger than six years	Institutional and non-institutional care; nutritional status, growth and development; habitat conditions; morbidity	11
3a	Target child	Interaction with mother and biological parents; target child care information; household environment quality scale (based on FCI questionnaire developed by UNICEF)	12
4	Biological mother	Use of time; empowerment; family information; reproductive and contraceptive history; birth and lactation of biological children; physical development of biological children; food and feeding; child feaces; knowledge on child development (based on KIDI)	26
4a	Mother depression	Depression scale - CESD-6	6
5	Main carer	Identification and address; education; workforce; use of time; food and feeding; child feaces; knowledge on child development (based on KIDI)	20

2.5.2 Ages and Stages Questionnaires

In addition to carrying out the household questionnaires, the same 15 interviewers and 3 supervisors collected ASQ data for every target child. Interviewers and supervisors received specialist training on administering the ASQ by staff from the Centre for Early Childhood Education and Development (CECED) at Ambedkar University, under the guidance of the ICDDR,B. This was usually done on the day after the household questionnaires because of respondent fatigue or other commitments. On average the ASQ took 20 minutes to administer.

2.5.3 Anthropometrics

Two surveyors separately collected anthropometric measurements from all target children and their mothers between the 13 November 2013 and the 11 January 2014. They collected the weight and height of both child and mother following a strict protocol. The scales used were Seca 874.

2.5.4 Home visitor questionnaire

A short questionnaire was administered to home visitors before they began their training. It gathered details of their background characteristics, educational attainment, experience and initial knowledge on key aspects of child development.

2.5.5 Sahi questionnaire

Another two surveyors collected data at the level of the sahi. The sahi site questionnaire was administered to someone who was knowledgeable about the sahi, such as the Anganwadi worker, the sahi head or a teacher. The questionnaire lasted for 27 minutes on average and asked about the location, population and infrastructure of the sahi; public services; transport and distances; health institutions located in the sahi; childcare institutions located in the sahi; educational institutions located in the sahi; commercial activity in the sahi and; social programmes in the sahi.

Two surveyors collected data on 54 sahis between the 11 February 2014 and 24 February 2014.

2.6 Problems in data collection

The biggest problem faced in terms of data collection was caused by two cyclones during which many households left their communities and only returned some time after the end of Cyclone Phalin (the most disruptive). These cyclones occurred after the surveyors had already been trained but before they had begun the data collection. Therefore, an additional refresher training was needed before they could begin their work once the largest cyclone was over. Overall this delayed the baseline survey by a total of four weeks.

2.7 Purpose of this baseline report

The data analysis presented in the remaining sections of this report has two main purposes. First, it provides an interesting snapshot child development and its determinants in a population where this type of study has never before been carried out. It will hopefully serve as a useful tool in thinking about child development interventions that could be effective here though helping to understand the developmental level of children in this population along with the most important constraints and challenges to them reaching their developmental potential.

The second purpose of this report is to formally test whether we see any systematic differences between the treatment and control group prior to the intervention starting. As discussed in section 2.2 our evaluation methodology is based on the comparison of outcomes for children and households in sahis allocated to receive the home visiting intervention and those in the control group, at the end of the intervention period. In order for this methodology to be valid it is very important that the two groups are similar in all respects, other than treatment. Since we randomised which sahis were allocated to the intervention group and which to the control group we expect that this will be the case - we know that it will be *in expectation*. The randomisation removes sample selection bias so that, in theory, the only difference between the intervention and the control group is eligibility for the home visiting programme thus meaning any differences we observe between the two groups can be attributed to the programme. This means we can estimate an unbiased effect of the programme on all outcomes of interest.

At this stage in the evaluation, we check that the randomisation did, indeed, give us a balanced allocation of treatment and control - i.e. treatment and control samples that appear similar in terms of observable characteristics (which would also suggest they are similar in terms of unobservable characteristics). Because of the randomisation we know that, in expectation, this will be the case. However, in finite samples it is always possible that, by chance, there are systematic differences between the intervention and control groups. This is what we formally test in this report. For all variables we report

in this report we compare the mean values for the intervention and the control units. We conduct two tailed hypothesis tests to see if any differences in mean values we observe are statistically significant at conventional levels[3]. In all our analysis we allow for arbitrary correlation in unobservables for all units within the sahi (cluster) by using cluster robust standard errors.

The rest of the report proceeds as follows. In section 3 we present a summary snapshot of who are sample of households and children are and the structure of the data we have on them. We present here general characteristics of our target children, such as the age and sex distribution. Then we present general characteristics of the households they come from, in terms of the structure of those households, the religion and the dwelling they abide in. We then proceed to analyse the baseline results for all baseline values of outcome indicators and other factors we believe may be important in determining child development. We follow the structure set out for our outcomes indicators in section 2.4. We divide indicators into indicators of child development and indicators of child development inputs. Section 4 presents our indicators of child development - indicators of cognitive, non-cognitive and motor development and or health and morbidity. Section 5 then goes onto present indicators of child development inputs across the following domains: (1) quality of the home environment, (2) child nutrition, healthcare and birth, (3) knowledge of child development, (4) quality and quantity of maternal time, (5) maternal and main caregiver wellbeing and education and (6) economic resources. After analysing indicators of child development and inputs to child development we provide some brief description and analysis of the 54 study sahis (in section 6) and of the 27 home visitors employed in the treatment sahis (in section 7).

[3]By a 'statistically significant difference' we means that we can be confident, at a given probability, that the difference in the sample means represents a difference in the expected value of the underlying distribution, rather than just having occurred by chance. If we test a null hypothesis that the two population means are equal at a significance level of 0.05, this corresponds to a 5% chance of falsely rejecting the null when infact the population means were equal. The p-values which we report in the tables correspond to the marginal probability at which we are indifferent between rejecting or not the null hypothesis. The smaller the p-value, therefore, the more likely it is that the true population means between treatment and control, for this variable, are different. The stars on the tables represent whether we reject the null at conventional significance levels (* for rejecting at 0.05, ** for rejecting at 0.01 and *** for rejecting at 0.001).

3 Baseline data – general characteristics

3.1 Summary of data structure

Table 3 summarises the the structure of the baseline sample: 421 target children, living in 418 households, in 54 sahis. Three pairs of target children live in the same household although none of these pairs were twins (twins were excluded by the inclusion criteria).

Number of target children	421
Number of households	418
Number of sahis	54

Table 3: Baseline sample structure

Table 4 summarises the structure of missing data for the 421 target children. There are only four cases where a key measurement is missing - either due to the data being lost or due to miscommunication in the field. y There were 29 cases when interviewers completed the module for the main carer in addition to the module for the biological mother.

✓ Module 1	Module 2	Module 2a	Module 3	Module 3a	Module 4	Module 4a	Module 5	ASQ	Anthropometrics	Number of cases	Notes
✓	✓	✓	✓	✓	✓	✓		✓	✓	338	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	29	29 (+1 below) children had main caregivers who were different from their mother. For all of these children the mother was also present in the household
✓	✓	✓	✓	✓	✓	✓	✓	✓		1	Anthropometric data for one child was either lost or never collected
✓	✓	✓	✓	✓	✓			✓	✓	1	Mother depression for one mother was either lost of never recorded
✓	✓		✓	✓	✓	✓		✓	✓	2	Two rosters were lost/not entered

Table 4: Baseline missing data structure

3.2 General target child characteristics

Table 5 shows that, as expected, our sample was split pretty evenly between boys and girls with 52.1% of our target children being male. In terms of ages report the age at the time that the Ages and Stages Questionnaire was administered (typically the day after the households questionnaire). We would have expected our sample selection strategy (see section 2.3.2) would have resulted in a pretty uniform distribution between the ages of 10 and 20 months. Overall, we do see this, except for a small dips and the upper and lower age limits which are perhaps due to interviewers having problems determining exactly whether the child should be in the sample and therefore dropping them. We do have four

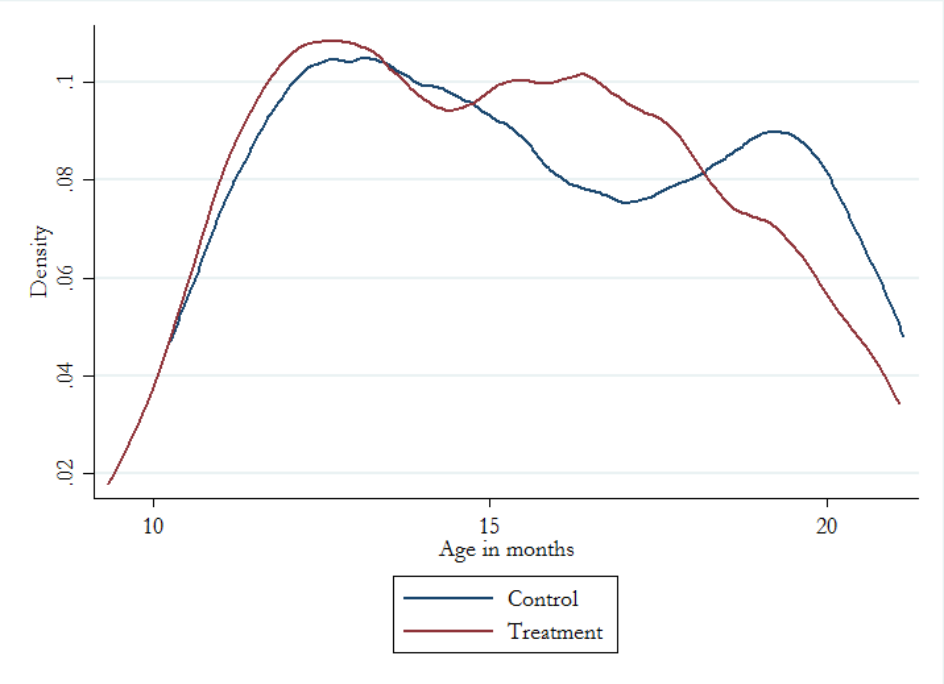
children in our sample who lie outside the age range (10 to 20 months) specified in the inclusion criteria - one child who appear to be only 9 months and three who appear to be 21 months. The latter could well have been caused by delays in the baseline team reaching the household meaning the child had just turned 21 months when baseline was completed. Figure 4 shows that we do see a slightly different distribution of children's ages across treatment and control, with fewer control children falling in the middle of the age distribution. However, these differences are not statistically significant (nor is there any statistically significant difference in the mean age between groups) and by controlling for age in our evaluation they should not cause a problem.

Table 5: Age and sex of target children

	Whole Sample	N	Control	Treatment	P-value
Male (%)	52.1	420	47.9	56.5	0.126
9 months (%)	0.2	421	0.0	0.5	0.319
10 or 11 months (%)	16.9	421	17.0	16.7	0.944
12 or 13 months (%)	20.4	421	18.9	22.0	0.427
14 or 15 months (%)	18.5	421	18.9	18.2	0.851
16 or 17 months (%)	17.1	421	14.2	20.1	0.074
18 or 19 months (%)	17.1	421	20.3	13.9	0.081
20 months (%)	8.3	421	9.0	7.7	0.674
21 months (%)	0.7	421	0.9	0.5	0.568

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Figure 4: Distribution of children's ages in treatment and control groups



3.3 General household characteristics

3.3.1 Household size and structure

We now move on to look at the structure of the households (defined as those who regularly cook from the same pot) in our sample. Table 6 shows that, on average, our sample households contained almost six members, although there is a relatively high amount of variation here. On average, there were almost four adults, defined as being at least 16 years of age, with equal numbers of men and women. There were, on average, almost two children under the age of 16 in the household with the large majority of these being under the age of six.

Table 7 shows all of the biological mothers of the sample target children live in the same household as their target child. This holds true for 97.6% of biological fathers. Just over half of households also had a biological brother or sister of the target child. 45.0% of sample households contained a biological grandparent of the target child, the vast majority of these being paternal grandparents. Likewise, in many (22.4%) sample households the target child's paternal uncle(s) lived in the household. The structure of households appears balanced between treatment and control.

Table7 shows which household member (in relation to the target child) is named as the household head in the survey. In just over half of sample households the household head is listed as the father of the target child. This is balanced between treatment and control.

3.3.2 Religion and caste

Data on household caste and religion, specifically on the caste and religion of the mother, was collected during the census and therefore not collected again during baseline. However, this led to some problems

Table 6: Household size

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Number of people	5.803 (3.112)	416	5.713 (2.836)	5.894 (3.372)	0.627
Number of adults (>16 years)	3.767 (2.437)	416	3.684 (2.176)	3.850 (2.677)	0.571
Number of adult women	1.930 (1.313)	416	1.866 (1.131)	1.995 (1.473)	0.391
Number of adult men	1.837 (1.281)	416	1.818 (1.187)	1.855 (1.372)	0.813
Number of children (16 or younger)	1.957 (1.081)	416	1.986 (0.988)	1.928 (1.170)	0.569
Number of young children (6 or younger)	1.433 (0.670)	416	1.459 (0.665)	1.406 (0.675)	0.473

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 7: Household structure

	Whole Sample	N	Control	Treatment	P-value
TC's biological mother lives in household (%)	100.0	416	100.0	100.0	.
TC's biological father lives in household (%)	97.8	416	98.1	97.6	0.704
TC's sister(s) and/or brother(s) lives in household (%)	52.2	416	53.1	51.2	0.662
TC's paternal grandmother lives in household (%)	40.4	416	38.3	42.5	0.446
TC's paternal grandfather lives in household (%)	27.9	416	28.2	27.5	0.886
TC's maternal grandmother lives in household (%)	4.6	416	5.3	3.9	0.513
TC's maternal grandfather lives in household (%)	3.1	416	3.8	2.4	0.428
TC's paternal aunt lives in household (%)	8.4	416	8.6	8.2	0.894
TC's paternal uncle lives in household (%)	22.1	416	21.5	22.7	0.790
TC's maternal aunt lives in household (%)	1.9	416	2.4	1.4	0.510
TC's maternal uncle lives in household (%)	3.1	416	3.8	2.4	0.391
TC's cousin(s) lives in household (%)	11.5	416	12.0	11.1	0.805

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 8: Relationship of household head to target child

	Whole Sample	N	Control	Treatment	P-value
Household head is TC's biological mother (%)	1.2	418	2.4	0.0	0.018*
Household head is TC's biological father (%)	53.8	418	55.0	52.6	0.667
Household head is TC's paternal grandmother (%)	12.4	418	12.9	12.0	0.778
Household head is TC's paternal grandfather (%)	23.9	418	22.0	25.8	0.397
Household head is TC's paternal uncle (%)	0.5	418	0.5	0.5	1.000
Household head is other household member (%)	8.1	418	7.2	9.1	0.474

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

since there were a lot of missing values, especially for religion where this data was missing for 131 out of 418 households. More problematic still was that this missing data was concentrated in certain sites so for 18 sites (exactly one third of the sample) the religion of the majority of households is missing.

For those households where the religion data is available we find that religious groups are clustered, in different sahis. On average, 92% of sample households within a sahi belonged to the same religion. Table 9 shows that roughly two-thirds of the 54 sample sahis are more than 60% Hindu, just two sahis are more than 60% Muslim and three are mixed (Hindu and Muslim). For 29.6% of the sahis, more than 60% of religion data is missing.

Table 9: Most common religion in sahi

	Whole Sample	N	Control	Treatment	P-value
More than 60% Hindu (%)	61.1	54	63.0	59.3	0.783
More than 60% Muslim (%)	3.7	54	3.7	3.7	1.000
More than 60% Missing (%)	29.6	54	25.9	33.3	0.556
Mixed (%)	5.6	54	7.4	3.7	0.557

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 10 shows the proportion of different households in the sample who belong to different religions. Looking at the p-values listed in the table shows us that there is no statistical significance between these proportions between treatment and control areas.

3.3.3 Dwelling status

Moving on to the characteristics of the dwellings in which our sample households live, table 11 shows that 39.0% of our sample households lived in a dwelling that only housed their household (a household

Table 10: Household's religion

	Whole Sample	N	Control	Treatment	P-value
Hindu (%)	61.0	418	64.9	57.0	0.482
Muslim (%)	6.7	418	7.6	5.8	0.707
Christian (%)	0.7	418	0.9	0.5	0.565
Missing (%)	31.6	418	26.5	36.7	0.357

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

being defined as those who regularly cook from the same pot). 23.4% lived in dwellings that they shared with three or more other households. More than 61.5% of households owned the dwelling (or the part of the dwelling) where they resided. These characteristics were balanced across treatment and control. We discuss additional information on the household structure below.

Table 11: Basic dwelling characteristics

	Whole Sample	N	Control	Treatment	P-value
Single household in dwelling (%)	39.0	418	39.8	38.2	0.798
Two or three households in dwelling (%)	36.6	418	35.1	38.2	0.583
Four or more households in dwelling (%)	23.4	418	24.2	22.7	0.774
Household owns dwelling (%)	61.5	418	60.2	62.8	0.645

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

4 Baseline data - indicators of child development

4.1 Cognitive, non-cognitive and motor development

Our main measure of current levels of child development at baseline was the Ages and Stages Questionnaires (third edition). As discussed in sections 2.4.1.1, 2.4.1.2 and 2.4.1.3, the ASQ-3 has been shown to be a reliable and practical measure of cognitive, non-cognitive and motor development. The ASQ-3 consists of five subscales, each measuring a different developmental domain - problem solving, communication, personal-social, fine motor and gross motor. Each subscale on the original test is made up of six questions, such as "Does your baby walk beside furniture while holding on with only one hand?", ordered by developmental stage. For each question the respondent, the mother or the person who knows most about the target child's development, answers "yes", "sometimes" or "not yet". A "yes" is scored as 10, "sometimes" as 5 and "not yet" as 0. The maximum possible score for each subscale is therefore 60 and the maximum possible score for the test as a whole is 300. The test is divided into age-specific questionnaires so a 11 month old does the 11-12 month questionnaire whilst an 18 month old does the 17-18 month questionnaire. The different questionnaires were originally calibrated so each *should* have the same distribution of scores, however it will increase precision to control for age when analysing the results[4].

The ASQ-3 was originally designed as a screener to screen children for developmental delays. It was also originally designed for populations of children in developed countries. For these two reasons we were concerned that the range of difficulty of questions might not be sufficient to adequately measure children with particularly high or low levels of development. Therefore we extended each subscale, in both directions, by adding the following non-overlapping (non-matching) questions from the previous and next questionnaire. This meant our adapted test was scored out of a maximum of 120 points for each subscale and 600 for the whole test.

The ASQ-3 was carried out by the household interviewers (all of whom were female) after they received two weeks of specialist training by child development professionals. Generally our adapted version of the ASQ-3 was found to be acceptable to the mothers/ other respondents. However, from looking at our data we do have some concerns over the administration and how this impacted on data quality. Respondents were given three options for responding to questions - "yes" (scoring 10), "not yet" (scoring 0) and "sometimes" (scoring 5). We would expect a reasonable proportion of questions to be answered with "sometimes" since the test is targeted at the expected developmental stage of a child of that age and therefore we would expect many children would learning that particular skill and therefore exhibiting it "sometimes". However, Table 12 shows that only around one percent of answers were "sometimes" which seems very low (although this is balanced across treatment and control). In future work using the ASQ-3 extra attention must be given to the importance of this middle option.

Table 13 presents the raw scores from this adapted test, with 12 question subscales, for all ages combined. Each subscale is scored out of a maximum of 120 points. We see that the target children's scores are balanced over treatment status across four domains (problem solving, communication, fine motor, gross motor). For the fifth (personal social) we observe a slight discrepancy in scores, where the control group's mean score is higher than the treatment group. Whilst this difference is not significant at conventional levels ($p<0.05$) the p-value of 0.0717 on this domain does raise some concerns for the

[4]Because of the randomisation age should be orthogonal to treatment allocation so it is also viable to analyse the scores without controlling for age.

Table 12: Percentage of questions scored 0 (not yet), 5 (sometimes) and 10 (yes) by subscale.

	% 0s	% 5s	% 10s
Problem solving	28.2	1.0	70.8
Communication	30.1	1.0	68.9
Personal social	28.6	1.9	69.5
Fine motor	31.3	1.4	67.2
Gross motor	16.0	0.5	83.5

balance of the sample of target children on this domain. We will therefore account for this slight imbalance at follow-up. Table 14 shows the mean and standard deviation of each subscale, broken down by each age-specific test. We do see a fair amount of variation in scores between the different age-specific tests. Regression analysis (not shown here) confirms that for each subscale we can reject the hypothesis that the mean score is equal across each age specific test, using the Wald test. However, given a limited sample size it is impossible to disentangle whether the different age specific tests are of different difficulty (relative to the age they are targeting) or whether this is showing a true effect of age on various domains of child development. The relative ordering of scores on different subscales (e.g. problem solving vs. communication) also changes with the age specific test, although, by and large means were highest for the gross motor subscale and lowest for the fine motor and language subscales.

Since these scores are not standardised to any particular population the scale is not particularly meaningful in itself. Likewise, it is difficult to make comparisons across subscales based on these scores since the scores are not tied to any common metric. The real value of this data will be to compare children within the sample which will become very useful when we analyse follow-up data. We will be able to control for pre-existing differences in child developmental levels by using variation we observe in these ASQ-3 scores.

For comparison to other work using the ASQ-3 we also include, in table 15, the means and standard deviations of each subscale, broken down by age-specific test, for the middle six questions (i.e. the original questions) on each subscale.

One final set of analysis we present in relation to the ASQ-3 data is a measure of internal reliability. In table 16 we present Cronbach's alpha for each subscale and each age-specific test. Cronbach's alpha measures the correlation between items (in our case questions) making up one scale. It always takes a value between 0 and 1, with higher values representing higher degrees of correlations between items. Values closer to one are therefore suggestive that the twelve items in the subscale are indeed measuring the same underlying construct. A common view is that a test must have a Cronbach alpha of 0.7 for it to have good internal validity [25]. Looking at table 16 we see that half of our alphas (for each subscale-age-specific test combination) fall short of that criterion. This does leave us some cause for concern on the internal reliability of the ASQ-3 in our setting. Partly on the basis of these concerns our research team is currently involved in investigating alternative adaptations of the ASQ-3 for use in future projects.

Table 13: Ages and stages questionnaire - raw scores from 12 question subscales

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Problem solving domain	85.52	421	85.09	85.96	0.733
	(24.70)		(24.68)	(24.77)	
Communication domain	83.28	421	83.35	83.21	0.958
	(24.66)		(23.71)	(25.65)	
Personal social domain	84.54	421	86.46	82.58	0.0717
	(22.35)		(20.57)	(23.90)	
Fine motor domain	81.56	421	82.12	80.99	0.614
	(22.72)		(21.89)	(23.58)	
Gross motor domain	100.5	421	100.7	100.3	0.877
	(24.45)		(24.40)	(24.57)	

Each subscale is scored out of a maximum of 120 points.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 14: Mean and standard deviation for each subscale, by age specific test

	10 months	12 months	14 months	16 months	18 months	20 months
Problem solving	87.9	78.6	79.7	90.5	88.6	92.2
	(22.3)	(28.0)	(27.7)	(21.2)	(21.6)	(20.2)
Communication	88.3	93.6	79.1	75.9	81.0	81.7
	(20.1)	(22.1)	(22.7)	(21.7)	(26.6)	(28.3)
Personal social	79.3	76.3	84.3	87.6	90.2	89.7
	(24.7)	(23.6)	(24.2)	(20.0)	(20.1)	(18.5)
Fine motor	97.4	83.4	74.6	80.6	86.3	77.9
	(16.1)	(22.8)	(22.8)	(22.0)	(20.1)	(23.9)
Gross motor	98.4	91.9	105.3	105.7	102.4	100.8
	(22.2)	(30.1)	(25.6)	(24.6)	(18.1)	(16.4)
N	29.0	98.0	81.0	76.0	59.0	78.0

Each subscale is scored out of a maximum of 120 points. Notation is: Mean(SD).

Table 15: Mean and standard deviation for each subscale (middle six questions), by age specific test

	10 months	12 months	14 months	16 months	18 months	20 months
Problem solving	45.3	42.3	38.0	43.7	40.3	43.3
	(13.0)	(15.6)	(18.9)	(13.3)	(13.5)	(10.3)
Communication	45.9	48.6	44.1	39.5	37.6	35.6
	(13.2)	(12.8)	(14.9)	(12.9)	(15.4)	(18.5)
Personal social	37.8	42.6	40.1	41.5	52.7	43.6
	(15.3)	(15.1)	(13.8)	(15.1)	(10.5)	(12.8)
Fine motor	49.1	47.1	39.1	40.2	43.5	41.1
	(8.9)	(12.9)	(13.3)	(14.3)	(12.8)	(13.1)
Gross motor	52.2	49.0	53.0	53.3	56.0	55.8
	(13.2)	(15.9)	(14.5)	(14.8)	(9.9)	(8.7)
N	29.0	98.0	81.0	76.0	59.0	78.0

Each subscale is scored out of a maximum of 60 points. Notation is: Mean(SD).

Table 16: Cronbach's alpha for each subscale (out of 120) and age group

	All	10 months	12 months	14 months	16 months	18 months	20 months
Problem solving	0.694	0.710	0.787	0.786	0.645	0.692	0.662
Communication	0.699	0.573	0.685	0.720	0.692	0.782	0.822
Personal social	0.580	0.678	0.663	0.689	0.602	0.653	0.427
Fine motor	0.641	0.545	0.743	0.709	0.691	0.664	0.758
Gross motor	0.841	0.795	0.885	0.897	0.883	0.786	0.723
N	421.000	29.000	98.000	81.000	76.000	59.000	78.000

4.2 Health and morbidity

4.2.1 Anthropometrics

As discussed in section 2.4.1.4, height and weight (and measures of these in relation to age and one another) are important measures of medium and long term nutritional status and health. In table 17 we present the raw height and weight of the target children in our sample. We also report:

1. **Weight for age (Underweight)**. This is probably the most common assessment of child nutrition status. It represents a suitable combination of both linear growth and body proportion and thus can be used for the diagnosis of underweight children.

2. **Weight for Height (Wasting)**. This is a measure of current body mass. It is generally seen as a measure of acute or short-term inadequate nutrition and/or poor health status. It is the best index to use to reflect wasting malnutrition, when it is difficult to determine the exact ages of the children being measured.

3. **Height for age (Stunting)**. This is a measure of linear growth. Stunting refers to shortness. A deficit in height for age is generally assumed to indicate exposure to an unhealthy environment, such as poor nutrition, unhygienic environment or disease in the past and hence captures long-term, cumulative effects.

More specifically we report z-scores of these measures. A z-score describes a point (in our case the height-for-age or weight-for-age of a specific child) in relation to the distribution of that measure in some reference population. In this case the reference population is a population of infants that the WHO deemed to be healthy and raised in environments that do not constrain growth, as documented in the WHO Child Growth Standards. Details on these standards and how they were constructed can be found in publications by the WHO Multicentre Growth Reference Study Group[26]. More specifically a z-score tells us how many standard deviations (measured in standard deviations of the reference population) the point is away from the mean of the reference population. Therefore, a z-score of -1 on height for age would tell us that that child is one standard deviation smaller than the mean child in the reference population of healthy children.

We see that our target children are significantly shorter and lighter than the WHO's reference population of healthy children. On average, our children are roughly one standard deviation (of the reference population) shorter and lighter than children of their age from the WHO reference population. They are roughly, on average, half a standard deviation lighter than children of their height in the WHO reference population. Table 18 shows that 16.2% of our sample are classified as underweight, 22.2% are classified as stunted and 9.4% are classified as wasted. Figures 6, 5 and 7 visually present the distribution of these measures in relation to the WHO reference population. For each measure we see our distribution is shifted to the left and show a higher degree of dispersion. This is immediately suggestive of poor nutritional status amongst our study population.

Table 17: Anthropometric measurements

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Weight of child (kg)	9.252	420	9.166	9.340	0.285
	(1.535)		(1.574)	(1.493)	
Height of child (cm)	76.40	418	76.27	76.54	0.559
	(4.357)		(4.327)	(4.392)	
Weight-for-age z-score	-0.892	418	-1.001	-0.781	0.115
	(1.287)		(1.351)	(1.212)	
Height-for-age z-score	-1.038	416	-1.142	-0.931	0.123
	(1.280)		(1.257)	(1.297)	
Weight-for-height z-score	-0.523	416	-0.574	-0.472	0.396
	(1.172)		(1.204)	(1.140)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 18: Proportion of children who are classified as underweight, stunted or wasted following WHO criterion

	Whole Sample	N	Control	Treatment	P-value
Underweight (weight-for-age z-score < -2) (%)	15.8	418	18.0	13.5	0.255
Stunted (length-for-age z-score < -2) (%)	22.1	416	23.3	20.9	0.564
Wasted (weight-for-height z-score < -2) (%)	9.4	416	9.6	9.1	0.850

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Figure 5: Distribution of height-for-age of target children compared with WHO reference population

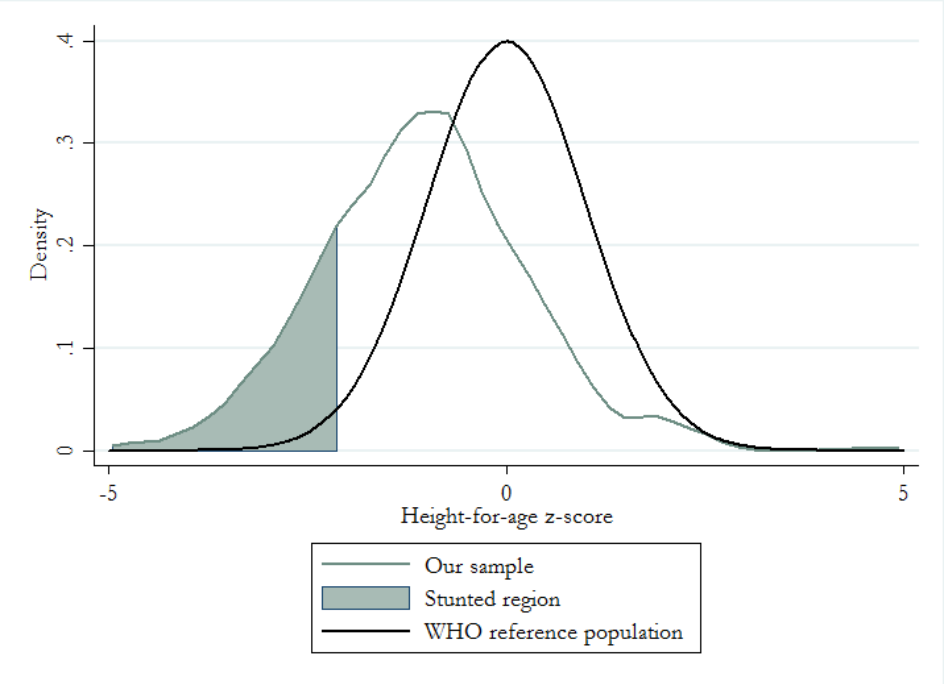


Figure 6: Distribution of weight-for-age of target children compared with WHO reference population

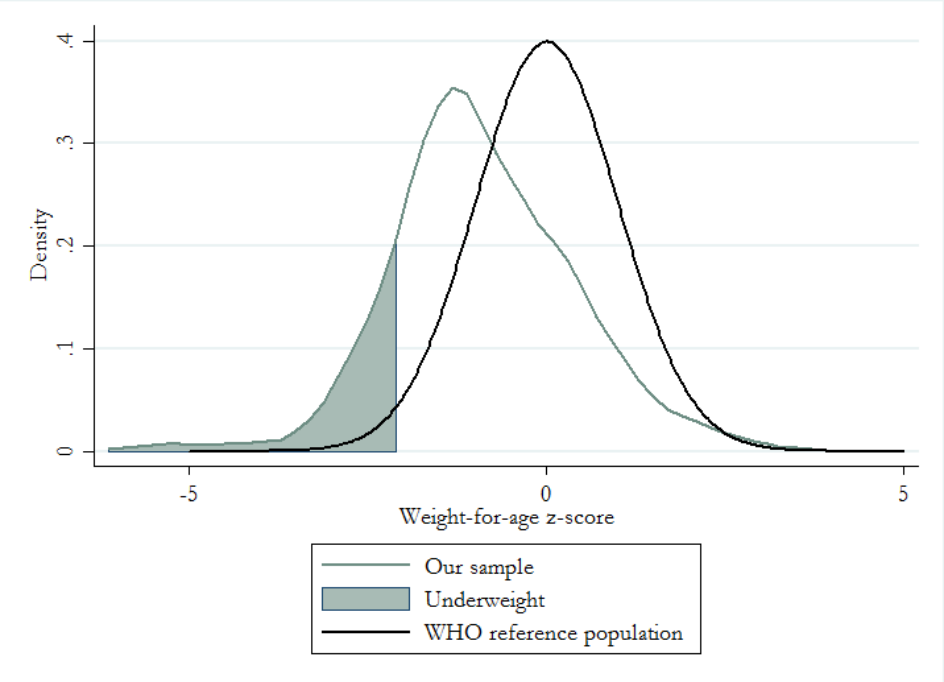
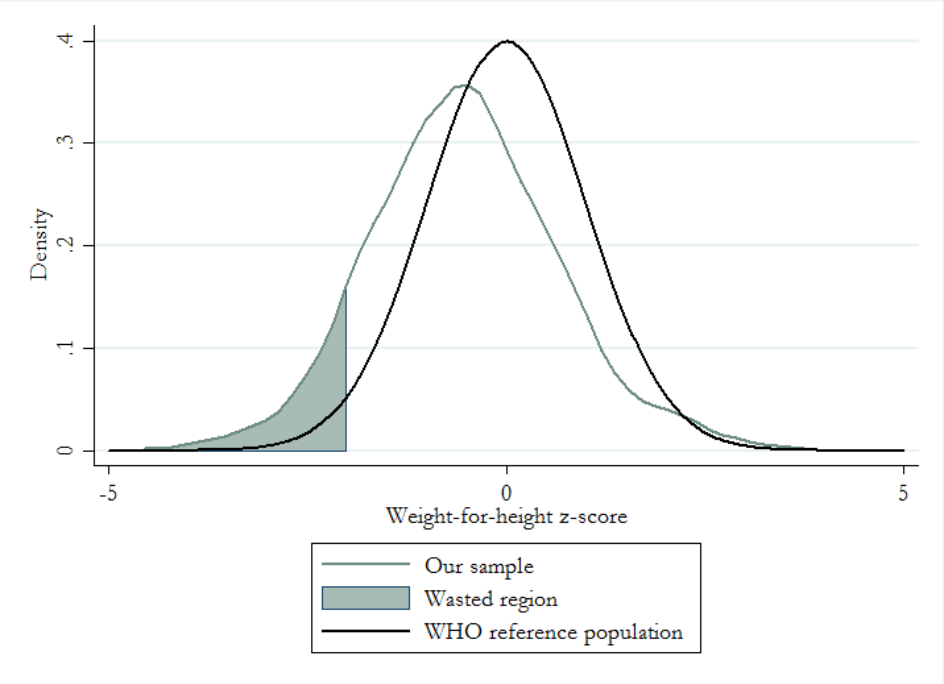


Figure 7: Distribution of weight-for-height of target children compared with WHO reference population



4.2.2 Morbidity

We only collected limited morbidity data for our target children. 15.7% of respondents reported that the target child had had diarrhoea in the past seven days, 81.0% reported that the target child had had a cold, cough or flu with fever in the past 15 days (which could be symptomatic of malaria). There were no significant differences in these measures of morbidity between treatment and control.

Table 19: Morbidity

	Whole Sample	N	Control	Treatment	P-value
Diarrhoea in past seven days (%)	15.7	420	18.4	13.0	0.147
Cold, cough or flu with fever in past 15 days (%)	81.0	420	79.2	82.7	0.375

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5 Baseline data - indicators of child development inputs

5.1 Quality of the home environment

5.1.1 Quality of the home stimulation environment

The quality of the home environment in terms of the amount of stimulation it provides is crucial for understanding the driving forces behind child development, as discussed in section 2.4.2.1. We measured the quality of the home environment using an adapted version of the Family Care Indicators (FCI) questionnaire developed by UNICEF. From the FCI we construct the five different subscales discussed by Hamadani *et al.*[22]:

1. **Sources of play materials:** constructed by adding indicators for whether child has played with at least one homemade toy and four or more bought toys in the past 30 days. (Maximum score of 2.)

2. **Variety of play materials:** the number of different types of play materials (types listed in table 22) the child has played with in the past 30 days. (Maximum score of 7.)

3. **Play activities:** the the number of die rent play activities (listed in table 23) the child has done with a household member over the age of 15 in the past 3 days. (Maximum score of 7.)

4. **Household books:** number of books for adults in household (not including school books). Top censored at 6.

5. **Household newspapers and magazines:** number of newspapers and magazines in household. Top censored at 6.

Hamadani *et al.* [22] show that, in their sample of 801 Bangladeshi children the subscales of the FCI that were most correlated with levels of child development (as measured through the Bayley Scales of Infant and Child Development-III) were the 'play activities' and 'variety of play materials' subscales.

Table 20: Family Care Indicator subscales

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Sources of play materials (/2)	1.197 (0.741)	421	1.222 (0.711)	1.172 (0.771)	0.473
Variety of play materials (/7)	2.713 (1.772)	421	2.660 (1.628)	2.766 (1.908)	0.494
Play activities (/7)	2.563 (1.585)	421	2.453 (1.518)	2.675 (1.646)	0.108
Household books	0.806 (1.613)	421	0.583 (1.351)	1.033 (1.816)	0.0130*
Household newspapers and magazines	0.337 (0.770)	419	0.245 (0.636)	0.428 (0.877)	0.0364*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 20 shows the scores for each subscale, broken down by treatment and control. We see that the sample appears balanced over treatment and control on measures of sources of play materials, variety of play materials and play activities. We do see a slight imbalance on the number of books in the household and the number of newspapers and magazines in the household - the treatment group appear to have slightly more of each. A rough comparison of these results with summary statistics presented for a rural Bangladeshi population[22] shows that our population appears to own significantly fewer books (although this may be because we specifically excluded school books whereas this was not the case in the Bangladeshi instrument) but similar scores for the number of newspapers and magazines in the household, variety of play materials and play activities. Our measure of sources of play materials is not scored comparably.

Tables 21 to 24 show the component parts of each subscale. In table 21 we see that 83.8% of children played with at least one bought toy in the past 30 days, 51.1% played with at least four while 68.8% played with at least one homemade toy. The sample appears balanced across treatment and control in these measures. In table 22 , detailing the variety of play materials played with in the past 30 days, we see that the most common type of play material, from the given list, is 'toys that induce constant physical movement'. 'Toys to play music' and 'dolls and other objects that aid role play and fantasy games' were also common but very few children had played with 'toys to learn shapes and/or colours' or 'picture books for children (not school books)'. The proportion of children who had played with the latter was much lower than reported in the Bangladeshi study. We see a slight imbalance across treatment and control in this last measure but the sample appeared reasonably balanced over the other play materials. In terms of play activities, table 23 shows that the vast majority of children had left the house to go to 'the market, park or other place' or had 'played together with child's toys' with a household member during the past three days. Around half of children had sang with a household member. Far fewer had 'made drawings, paintings or writing' or had 'played naming objects or colours, or counting'. The proportion of children engaging in all activities was balanced across treatment and control. Table 24 shows that most households had no books and most had no magazines belonging to the household. The proportion of households having books and magazines were slightly higher in the treatment group.

Table 21: Sources of play materials in past 30 days

	Whole Sample	N	Control	Treatment	P-value
Some homemade toys (%)	68.6	421	72.2	65.1	0.213
Some bought toys (%)	83.8	421	81.6	86.1	0.247
Four or more bought toys (%)	51.1	421	50.0	52.2	0.693

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5.1.2 Quality of the home health environment

Table 25 presents various measures related to sanitation and cooking facilities that impact on health. In the whole sample 72.2% of households live in a dwelling which has some kind of sanitation facility,

Table 22: Variety of play materials played with in past 30 days

	Whole Sample	N	Control	Treatment	P-value
Toys to play music (%)	25.2	421	22.6	27.8	0.207
Toys to ensemble or build things (%)	7.1	421	5.7	8.6	0.258
Things for drawing, painting and/or writing (%)	6.2	421	4.7	7.7	0.254
Toys that induce constant physical movement (%)	55.8	421	59.4	52.2	0.097
Dolls and other objects that aid role play and fantasy games (%)	18.3	421	15.6	21.1	0.197
Picture books for children (not school books) (%)	3.6	421	1.9	5.3	0.045*
Toys to learn shapes and/or colours (%)	2.6	421	2.4	2.9	0.768

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 23: Different play activities performed with household members in past 3 days

	Whole Sample	N	Control	Treatment	P-value
Read or looked at picture books (%)	20.0	421	17.9	22.0	0.349
Told stories (%)	21.9	421	19.3	24.4	0.133
Sang (%)	48.5	421	45.8	51.2	0.147
Went out to market, park or other place (%)	76.0	421	74.1	78.0	0.397
Played together with child's toys (%)	65.6	421	66.5	64.6	0.711
Made drawings, paintings or writing (%)	8.1	421	7.1	9.1	0.414
Played naming objects or colours, or counting (%)	16.4	421	14.6	18.2	0.365

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 24: Books, newspapers and magazines in household

	Whole Sample	N	Control	Treatment	P-value
No books (%)	72.7	421	78.3	67.0	0.022*
Between 1 and 2 books (%)	15.4	421	13.7	17.2	0.337
Between 3 and 5 books (%)	6.9	421	5.2	8.6	0.178
6 more books (%)	5.0	421	2.8	7.2	0.058
No newspapers and magazines (%)	79.8	421	84.4	75.1	0.056
Between 1 and 2 newspapers and magazines (%)	18.5	421	13.7	23.4	0.051
Between 3 and 5 newspapers and magazines (%)	0.7	421	0.9	0.5	0.564
6 more newspapers and magazines (%)	0.5	421	0.0	1.0	0.149

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

the most common being a septic tank. Of those whose dwelling had no facility a small minority used a community or neighbour's toilet but the vast majority typically practised open defecation - 26.5% of control households and 18.4% of treatment households. In terms of the food preparation environment, 37.6% of households had no window in the room where food preparation was done which can lead to breathing in unhealthy levels of smoke.

Table 25: Quality of the home health environment

	Whole Sample	N	Control	Treatment	P-value
Sanitation - dwelling has toilet connected to septic tank (%)	61.7	418	53.1	70.5	0.053
Sanitation - dwelling has toilet connected to drain (%)	9.3	418	10.4	8.2	0.555
Sanitation - dwelling has other toilet/latrine (%)	1.2	418	0.5	1.9	0.153
Sanitation - household members typically use other toilet (%)	5.3	418	9.5	1.0	0.025*
Sanitation - household members typically openly defecate (%)	22.5	418	26.5	18.4	0.344
Food preparation - done in room with windows (%)	62.4	418	57.8	67.1	0.088
Food preparation - done in room with chimney/opening (%)	5.7	418	7.1	4.3	0.227
Food preparation - cook mainly over wood or charcoal (%)	26.1	418	26.5	25.6	0.892

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5.2 Child nutrition, healthcare and birth

5.2.1 Birth and breastfeeding

Table 26: Birth and breastfeeding

	Whole Sample	N	Control	Treatment	P-value
Child born in hospital or clinic (%)	92.8	417	92.9	92.7	0.952
Child was born preterm (<37 weeks) (%)	46.3	417	46.9	45.6	0.775
Child was born very preterm (<32 weeks) (%)	0.7	417	0.9	0.5	0.567
Child born with low birthweight (<2500g) (%)	16.8	417	18.5	15.0	0.332
Child was breastfed (%)	99.0	417	99.1	99.0	0.980
Child was exclusively breastfed for first six months (%)	54.0	417	53.1	54.9	0.738

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

In table 26 we look at the birth and breastfeeding history of our target children. 92.8% of our target children were born in a hospital or clinic which is a key aim of the Indian health system. 46.3% of our sample were born preterm, that is before 46.3 weeks of gestation. Whilst this high proportion may be partly due to many respondents not knowing their exact period of gestation, India does have a high rate of prematurity amongst newborns, partly due to underlying infection and chronic conditions of mothers. Mothers could, in general, recall the birthweight of the target child. Indeed, only three said they did not know the birthweight. From the mother's recall we found that 16.8% of our target children had a low birthweight. A low birthweight is associated with high levels of child mortality, morbidity and disability and the effects can also reach into adulthood. A low birthweight is often indicative or poor nutritional status of the mother during pregnancy.

The World Health Organisation recommend that all children, except under certain medical conditions, are breastfed and are exclusively breastfed, meaning they do not consume any other substance (including water), for the first six months of life [27]. 99.0% of our target children were breastfed. From maternal reports of when they first gave their child any other substance than breastmilk we inferred that 54.0% had been exclusively breastfed for the first six months, in accordance with WHO guidelines.

All measures relating to birth and breastfeeding appear balanced across treatment and control.

5.2.2 Nutrition

Good nutrition is vital for child development. Whilst anthropometric measurements are a good indicator of long term nutritional status it is useful to also directly capture information of children's diets. We collect this information in a simple 24 hour recall period where respondents, usually the child's mother or whoever knew most about the care of the target child, were asked to indicate whether the target children had consumed food from each of a list of categories over the past 24 hours. Table 27 shows the raw results of this exercise, broken down into treatment and control. We see that our target

children's diets were high in cereals and whole grains (this includes rice) and pulses yet relatively low in foods high in proteins and haem iron.

Dietary diversity, defined as the number of foodgroups consumed in a given period of time, is an important indicator of the quality of a child's diet since diverse diets are more likely to contain sufficient quantities of the wide range of nutrients essential for healthy development. We construct a measure of dietary diversity based upon one proposed by Arimond and Ruel[28] which was shown to correlate well with broad measures of nutritional status. We place these above foodgroups into the seven larger foodgroups listed below, scoring each child as a 1 if they consumed some food in this foodgroup in the past 24 hours and as a 0 if they did not:

> *1) starchy staples (foods made from grain, roots, or tubers); 2) legumes; 3) dairy (milk other than breast milk, cheese, or yogurt); 4) meat, poultry, fish, or eggs; 5) vitamin A-rich fruits and vegetables (pumpkin; red or yellow yams or squash; carrots or red sweet potatoes; green leafy vegetables; fruits such as mango, papaya, or other local vitamin A-rich fruits); 6) other fruits and vegetables (or fruit juices); and 7) foods made with oil, fat, or butter.[28]*

We then construct dietary diversity scores by simply to summing the total number of these seven foodgroups consumed by the child in the past 24 hours. Arimond and Ruel*[28]* use indicators of whether a child has consumed this food three or more times in the past seven days, however we do not have seven day recall data so we adopt the method they use for Haiti in their study and use an indicator over the past 24 hours. We also use Arimond and Ruel's cut off's for dietary diversity, dividing children into those who consumed 0 to 2 foodgroups, those who consumed 3 to 4 groups and those who consumed 5 to 7 groups.

Table 28 presents our results. 95.5% of target children had consumed 'starchy staples' in the past 24 hours, 86.1% legumes yet only 19.9% had consumed meat, fish or egg and only 13.4% foods made with oil, fats and butter. In terms of the aggregate scores of diet diversity, 19.7% of target children had consumed two or less distinct foodgroups in the past 24 hours, which Arimond and Ruel term as low diet diversity. 60.1% of target children had a middle diet diversity score of 3 to 4 whilst 19.5% had a high diversity score of 5 or more. These measures were balanced across treatment and control.

5.3 Knowledge of child development

Knowledge about children's developmental needs and how to best fulfil them is key to children being raised in a healthy and stimulating environment. Indeed, increasing levels of knowledge and understanding about child development to mothers and caregivers, and thus inducing behaviour change, is a crucial mechanism through which we hypothesise the home-visiting intervention may work. Therefore, it will be important to assess how the programme has affected mothers' knowledge of this area. Controlling for maternal knowledge at baseline will be important in estimating this effect precisely.

We measure maternal and caregiver knowledge of key principals of child development using an adapted and shortened version of the Knowledge of Infant Development Inventory [23]. This tool attempts to measure knowledge on parental practices, child development processes and infant norms of behaviour. Mothers are read various statements and asked to give their opinion on whether the statement "is true", "is partly true" or "is not true". From these answers we construct aggregate scores which measure knowledge under the following domains: (1) praising/paying attention to child, (2)

Table 27: Child food consumption (last 24 hours)

	Whole Sample	N	Control	Treatment	P-value
Cereals and whole grains (%)	90.7	420	91.5	89.9	0.613
Roots and tubers (%)	81.0	420	78.8	83.2	0.280
Meat, fish or eggs (%)	19.8	419	22.7	16.8	0.220
Pulses (%)	86.2	420	84.9	87.5	0.489
Fats and oils (%)	13.8	420	14.2	13.5	0.832
Mother's milk (%)	85.5	420	82.5	88.5	0.087
Milk and milk products (%)	37.1	420	39.2	35.1	0.439
Dark yellow and orange fruits (%)	16.2	420	17.0	15.4	0.703
Citrus or juicy fruits (%)	17.6	420	17.5	17.8	0.938
Other fruits (%)	29.5	420	30.7	28.4	0.650
Green leafy vegetables (%)	17.4	420	20.3	14.4	0.089
Other vegetables (%)	51.9	420	46.2	57.7	0.068
Sugar (%)	68.3	420	70.3	66.3	0.427
Nuts (%)	29.5	420	27.8	31.3	0.407

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 28: Dietary diversity measure

	Whole Sample	N	Control	Treatment	P-value
Starchy staples (%)	95.5	418	95.2	95.7	0.842
Legumes (%)	86.1	418	84.8	87.5	0.469
Dairy (excluding breast milk) (%)	36.8	418	38.6	35.1	0.506
Meat, fish, egg (%)	19.9	418	22.9	16.8	0.216
Vitamin A rich fruit or vegetables (%)	29.9	418	33.3	26.4	0.107
Other fruit and vegetables (%)	67.0	418	64.3	69.7	0.331
Foods made with oil, fats or butter (%)	13.4	418	13.3	13.5	0.968
Diet diversity score of 0 to 2 (%)	19.7	421	19.8	19.6	0.964
Diet diversity score of 3 to 4 (%)	60.1	421	60.4	59.8	0.900
Diet diversity score of 5 to 7 (%)	19.5	421	18.9	20.1	0.698

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

punishing child, (3) school readiness and expectations, (4) importance of maternal interactions and play and, (5) age appropriate expectations.

Table 29 presents the percentage scores of the mothers on this instrument for each domain and the the instrument as a whole. We see that there is no significant difference between treatment groups over this measure of maternal knowledge about child development. In terms of the different domains of knowledge mothers typically scored higher on 'school readiness and expectations', which contained statements like "children who know lots of words learn to read earlier", and lower on 'praising/paying attention to child', which contained statements like "too much love and attention will spoil a child".

5.4 Quantity and quality of maternal time

In section 2.4.2.4 we highlighted how the quantity and quality of maternal time spent with and caring for children affects how stimulating a child's everyday environment is. Good quality time spent together also promotes strong attachment between mother and child, which has further beneficial effects on child development. In the household survey we asked all mothers about their time use during the previous working day (Monday to Friday). The aim was to capture how much time mothers spend each day primarily interacting with their child(ren) (i.e. their child being the sole object of their attention rather than just being present) and to capture how much of this time was engaged in play and games with their child(ren). For each category (e.g. cleaning house) we asked the mother to estimate how much time, in minutes, she had spent doing that activity.

From table 30 we can see that, of the categories of potential time use we asked about, the mothers in our sample spent little time working for a wage, or on their own land or business. Instead, a large proportion of their time was spend in housework activities. Mothers spent, on average, 70 minutes of the day playing with small children in the household and 15 minutes reading or telling stories to them.

Table 29: Knowledge of child development % scores

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Praising/paying attention to child	57.38 (20.36)	421	59.38 (20.24)	55.34 (20.32)	0.0385*
Punishing child	78.90 (22.57)	421	78.69 (23.53)	79.11 (21.61)	0.889
School readiness and expectations	81.24 (18.13)	421	81.21 (18.50)	81.26 (17.78)	0.982
Importance of maternal interactions and play	74.06 (16.41)	421	74.16 (17.25)	73.95 (15.55)	0.896
Age appropriate expectations	71.52 (14.85)	421	71.86 (15.37)	71.19 (14.34)	0.721
Total child development knowledge score	69.90 (8.379)	421	70.24 (9.060)	69.56 (7.633)	0.462

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

This is generally balanced between treatment and control sahis.

In table 31 we aggregate these categories into four larger categories. We see that housework, work and travel accounts for, on average, around five and a half hours of the mothers' time. Looking after (not including playing with or reading to) young children accounted for just under two hours whilst playing, reading and storytelling with young children accounted for around an hour and a half. Mothers had, on average, one and a half hours doing activities one might count as recreation. There are some obvious problems with these statistics - we didn't ask mothers about their time use in a way that was mutually exclusive so it is very possible that some tasks were done simultaneously and we would not capture that. Furthermore, when condensing the categories into the four broad categories the distinction between 'looking after (not playing with) children' and 'playing/reading/storytelling with children' is clearly not a tight one and, indeed, the home visiting programme advocates play activities that can be done during mothers' standard routines.

5.5 Maternal mental health, empowerment and education

5.5.1 Mental health

Maternal mental health and wellbeing is a crucial determinant of a mother's relationship and interaction with her child. There is substantial evidence that maternal depression affects parenting behaviours and child outcomes[29][30]. Evidence suggests that the most significant negative impacts of depression are for mothers of infants, the age of our target children[29].

We measured depressive symptoms in our sample of biological mothers using a shortened and adapted version of the Center for Epidemiologic Studies Depression Scale [24], a short self-report scale that is useful in study settings when full clinical assessments would be infeasible. Mothers were asked six questions on whether they experienced different symptoms of depression over the last seven days. For each question mothers were asked to respond with one of four options: (1) 'almost never or never

Table 30: Time use in minutes during last working day

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Working for wage or for own land or business	18.24	421	23.07	13.35	0.248
	(73.75)		(81.56)	(64.71)	
Travelling to work, market etc.	24.95	421	30.92	18.90	0.0601
	(71.59)		(79.11)	(62.67)	
Cooking for household	157.9	421	147.2	168.7	0.0245*
	(93.50)		(91.18)	(94.79)	
Washing/cleaning/ironing clothes	47.79	419	48.88	46.69	0.470
	(37.68)		(41.67)	(33.25)	
Cleaning house	50.17	419	49.83	50.50	0.880
	(43.83)		(44.38)	(43.37)	
Collecting and carrying water	12.69	420	12.40	12.99	0.832
	(25.51)		(22.10)	(28.58)	
Other household activities	10.28	420	9.336	11.23	0.471
	(26.23)		(20.05)	(31.27)	
Taking children to school or Anganwadi	3.755	421	4.514	2.986	0.493
	(23.00)		(30.49)	(11.13)	
Looking after small children in household	105.2	421	100.9	109.5	0.235
	(77.99)		(71.32)	(84.17)	
Playing with small children in household	70.36	420	69.29	71.44	0.779
	(83.42)		(83.40)	(83.62)	
Reading or telling stories with small children	15.24	421	15.71	14.76	0.774
	(37.04)		(39.82)	(34.08)	
Looking after ill member of household	5.178	421	5.731	4.617	0.781
	(43.16)		(42.85)	(43.57)	
Watching TV	77.39	421	71.82	83.04	0.150
	(85.27)		(82.71)	(87.62)	
Going to place of worship	6.912	421	6.816	7.010	0.945
	(31.35)		(34.19)	(28.27)	
Other activities	24.96	421	24.25	25.69	0.801
	(66.11)		(55.72)	(75.33)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 31: Time use in minutes during last working day

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Housework, work or travel	326.0 (170.0)	419	325.1 (168.1)	326.9 (172.4)	0.911
Looking after (not playing with) children	109.0 (80.68)	421	105.5 (76.63)	112.5 (84.62)	0.354
Playing/reading/storytelling with children	85.61 (92.01)	420	85.02 (93.07)	86.20 (91.14)	0.889
Recreation	84.30 (90.81)	421	78.63 (89.62)	90.05 (91.85)	0.167

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

(less than one day)', (2) 'a few times (between one and two days', (3) 'many times (between three and four times)' or (4) 'almost all the time (between five and seven days)'. These answers were scored 0, 1, 2 and 3 respectively and the total depression symptom score was created by summing these scores, so the total score was out of a maximum of 18. A mother was classified as 'depressed' if she had a total score of greater than 6. The short version of the CES-D is typically 10 questions and the binary cut-off for depression is taken as 10. Here we used only 6 questions and scaled the cut-off accordingly, to 6. Note that this is a short screener and the results should be interpreted as indicating symptoms consistent with depression rather than a diagnosis of clinical depression.

This instrument was a difficult one to implement and we did have concerns about the quality of the translation. Many of the words and concepts expressed in the questions were difficult to find simple equivalents to in the Oriya language. Even when translation was linguistically possible we ran into problems of mothers being very unfamiliar with evaluating their feelings or symptoms of depression in the way we were asking them to do. For this reason we are investigating alternative adaptations and measures of maternal mental health for use in similar projects.

Table 32 shows that just over a quarter of mothers scored six or more on the CESD scale which classified them as having substantial symptoms of depression. This was balanced between treatment and control.

Table 32: CESD depression scale

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Did you feel sad?	0.650 (0.898)	420	0.660 (0.902)	0.639 (0.895)	0.800
Did you feel lonely?	0.414 (0.772)	420	0.406 (0.758)	0.423 (0.789)	0.822
Did you have crying spells?	0.374 (0.660)	420	0.415 (0.687)	0.332 (0.630)	0.194
Did you enjoy life? (reverse scored)	1.736 (1.141)	420	1.797 (1.093)	1.673 (1.187)	0.248
Did you feel depressed?	0.411 (0.835)	419	0.398 (0.795)	0.423 (0.876)	0.788
Did you feel little interest or pleasure in doing things?	0.562 (0.756)	420	0.604 (0.775)	0.519 (0.735)	0.241
Total CESD score (/18)	4.145 (3.255)	420	4.278 (3.110)	4.010 (3.398)	0.405
Depressed (CESD score greater or equal to 6)	0.252 (0.435)	420	0.274 (0.447)	0.231 (0.422)	0.331

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5.5.2 Education and empowerment

As discussed in section 2.4.2.5 maternal education and empowerment are important factors in determining how household decisions that impact on child development are made. For example, there has long been observed a strong correlation between maternal education and child health, despite causal relationships being difficult to establish due to potential confounding[31]. In relation to empowerment, a recent literature review concluded that women's empowerment was associated with increases in child height and weight[32], key measures of child health and nutritional status.

Maternal education is presented, along with educational levels of the father and the household head, in table 33. We see that mothers had slightly less education than the biological fathers. 85.9% had completed at least one year of formal schooling however, only 26.8% had finished secondary school (tenth standard). As seen in table 34 the majority of mothers are functionally literate in the sense that they can read a newspaper and write a letter. Interestingly, while the proportion of mothers who can read and write is very similar to the proportion of fathers the proportion of mothers who can write a text message is substantially smaller than the proportion of fathers - 51.0% compared to 70.6%. We do observe an imbalance in maternal functional literacy - significantly more mothers in the treatment group can read and write than in the control, however the magnitude of this imbalance is relatively small and we will account for it in our endline analysis.

Table 33: Highest level of education completed

	Whole Sample	N	Control	Treatment	P-value
Biological mother - first standard or higher (%)	85.9	418	83.9	87.9	0.294
Biological mother - fifth standard or higher (%)	77.8	418	72.0	83.6	0.014*
Biological mother - tenth standard or higher (%)	26.8	418	23.2	30.4	0.209
Biological mother - higher secondary or higher (%)	15.8	418	14.7	16.9	0.645
Biological mother - graduate or post-graduate or higher (%)	7.2	418	7.1	7.2	0.963
Biological father - first standard or higher (%)	81.8	418	81.5	82.1	0.904
Biological father - fifth standard or higher (%)	73.9	418	71.1	76.8	0.298
Biological father - tenth standard or higher (%)	33.0	418	30.3	35.7	0.374
Biological father - higher secondary or higher (%)	22.0	418	18.0	26.1	0.143
Biological father - graduate or post-graduate or higher (%)	12.9	418	10.9	15.0	0.286
Household head - first standard or higher (%)	69.9	418	71.6	68.1	0.540
Household head - fifth standard or higher (%)	60.3	418	57.8	62.8	0.388
Household head - tenth standard or higher (%)	22.7	418	20.9	24.6	0.462
Household head - higher secondary or higher (%)	14.1	418	10.0	18.4	0.033*
Household head - graduate or post-graduate or higher (%)	7.7	418	6.2	9.2	0.284

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 34: Functional literacy

	Whole Sample	N	Control	Treatment	P-value
Biological mother - Can read newspaper (%)	81.8	418	76.7	87.0	0.036*
Biological mother - Can write a letter (%)	76.3	418	69.5	83.2	0.015*
Biological mother - Can write a text message (%)	51.0	418	45.7	56.3	0.077
Biological father - Can read newspaper (%)	82.4	409	78.2	86.7	0.051
Biological father - Can write a letter (%)	79.0	409	74.3	83.7	0.059
Biological father - Can write a text message (%)	70.7	409	68.0	73.4	0.336
Household head - Can read newspaper (%)	72.2	417	71.3	73.1	0.775
Household head - Can write a letter (%)	66.2	417	64.1	68.3	0.505
Household head - Can write a text message (%)	43.4	417	40.7	46.2	0.374

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

In terms of empowerment, we asked a series of questions to the biological mother about her role in key decisions affecting her and her child's wellbeing, such as what to do when a child is ill or how much money to spend on food. The results are presented in table 35. Roughly half of the mothers in our sample are allowed to go to the market alone. 25.4% said they would decide alone whether their child should see a doctor if he/she was ill whilst another 30.2% would make this decision with the biological father. In terms of decisions over nutrition and budgeting within the household mother seem to have comparatively little power - just 12.8% of mothers decide alone how much is spend on food and just 7.4% make this decision with their husband. Membership of Self-Help Groups which promote women's empowerment is relatively low amongst our sample - just 11.9% of mothers are members.

5.6 Economic resources

5.6.1 Work

Tables 36 to 39 describe the work behaviour of the biological mother of the target child, the biological father and the highest earner. A comparison between tables 36 and 37 will highlight that are far greater proportion of sample fathers, 94.6%, described work as their main activity in the previous week (Monday to Sunday) compared to just 10.4% of mothers. 87.9% of mothers stated that their main activity in the previous week was housework. For the subsample of household members whose main activity during the previous week was work we recorded the total hours worked last week and in a typical week. We see that, even of those whose main activity was work, biological fathers typically worked more hours (56.9 hours) per week than biological mothers (32.4 hours). We also have a very noisy estimate of earnings, from which we estimate yearly earnings. From also estimating hours worked per year we also create an estimate of an average hourly wage. For biological father's we estimate the average hourly wage to be Rs.72 per hour (USD 1.15). However, this is highly sensitive to outliers that are very likely measurement error. If we trim the top and bottom 5% of hourly wages we obtain a mean hourly wage, for biological fathers, of Rs. 36 per hour (USD 0.57). We also trim the data in

Table 35: Empowerment of biological mothers

	Whole Sample	N	Control	Treatment	P-value
Mother decides alone if ill child goes to doctor (%)	25.4	421	24.1	26.8	0.525
Mother and father decide together if ill child goes to doctor (%)	30.2	421	31.6	28.7	0.506
Mother decides alone if child goes to school (%)	26.4	421	24.5	28.2	0.425
Mother and father decide together if child goes to school (%)	7.6	421	9.0	6.2	0.249
Mother decides alone how much to spend on food (%)	12.8	421	13.7	12.0	0.622
Mother and father decide together how much to spend on food (%)	7.4	421	8.5	6.2	0.363
Mother is allowed to go to market alone (%)	49.4	421	51.9	46.9	0.296
Mother has asset she could sell without consent of husband (%)	3.6	421	2.4	4.8	0.183
Mother went to gathering to discuss children in past 6 months (%)	2.4	421	2.8	1.9	0.506
Mother is member of Self-Help Group (%)	11.9	421	10.8	12.9	0.627

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

a similar way for estimated yearly earnings. Once we trim the estimated work earnings data we do find that our mean estimate for the treatment is significantly greater than the control, for both the biological father and the main earner (who are more often than not the same person). However, given the noisiness of our estimates we don't perceive this as too much of a worry. For all other work related measures there is a good balance between treatment and control.

Table 37 presents the occupation of the highest earner in each household. We see that roughly one-fifth of household's highest earners are, respectively, private employees with a fixed income, daily labourers and self employed on own farm or business. Occupation of the highest earner is balanced across treatment and control.

5.6.2 Non-work income and transfers

In addition to income from work some sample households also receive income and transfers from other sources. 17.9% of the sample received some benefits from government schemes during the last year, 10.0% received some money from a pension scheme whilst it was less common for households to receive money from dowry [5], transfers/remittances or income from rent. Table 40 shows that the receipt of income from these sources was balanced with respect to treatment status.

5.6.3 Assets

Measuring assets that households own is a a good measure of household wealth and may be a better measure of economic wellbeing that studying income alone. In the household questionnaire we asked respondents whether their household owned each of 16 items of value, ranging from a chair to a computer. In the entire sample the most common assets that households owned were fan(s) (96.1%),

[5] Households may have been wary reporting this income source since it is illegal.

Table 36: Work and work earnings of biological mother

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Proportion who mainly worked last week	0.103 (0.304)	419	0.129 (0.336)	0.0766 (0.267)	0.100
Proportion who mainly did housework last week	0.883 (0.322)	419	0.857 (0.351)	0.909 (0.288)	0.121
Total hours worked last week	31.30 (17.60)	43	31.26 (18.05)	31.38 (17.40)	0.986
Total hours worked in typical week	32.42 (16.56)	45	33 (16.37)	31.38 (17.40)	0.794
Estimated work earnings last year (Rs.)	104347.5 (321097.9)	45	64397.7 (75345.0)	176756.6 (532130.6)	0.335
Estimated work earnings last year (Rs.) - trimmed	53792.5 (57975.9)	41	54968.3 (56684.4)	51259.9 (62962.8)	0.865
Estimated hourly wage (Rs.)	61.89 (179.8)	43	40.58 (35.11)	97.86 (293.6)	0.366
Estimated hourly wage (Rs.) - trimmed	33.93 (28.06)	39	39.73 (30.22)	24.65 (22.07)	0.0796

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Total hours worked last week reported for subsample who mainly worked last week. Total hours worked in a typical week reported for subsample who typically work. Trimming defined as dropping the top and bottom 5% of the distribution.

Table 37: Work and work earnings of biological father

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Proportion who mainly worked last week	0.944	409	0.942	0.946	0.854
	(0.231)		(0.235)	(0.227)	
Total hours worked last week	56.29	386	56.02	56.57	0.809
	(19.96)		(19.41)	(20.54)	
Total hours worked in typical week	56.96	399	55.85	58.07	0.329
	(19.13)		(19.00)	(19.24)	
Estimated work earnings last year (Rs.)	193998.3	350	161409.7	227723.7	0.305
	(553353.8)		(517384.1)	(587883.5)	
Estimated work earnings last year (Rs.) - trimmed	93009.9	316	76475.9	111076.8	0.00310**
	(83749.2)		(52719.3)	(105172.4)	
Estimated hourly wage (Rs.)	72.78	343	65.38	80.68	0.570
	(227.7)		(189.2)	(262.9)	
Estimated hourly wage (Rs.) - trimmed	36.42	309	32.18	41.21	0.0516
	(34.55)		(29.13)	(39.36)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Total hours worked last week reported for subsample who mainly worked last week. Total hours worked in a typical week reported for subsample who typically work. Trimming defined as dropping the top and bottom 5% of the distribution.

Table 38: Work and work earnings of highest earner

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Total hours worked last week	54.00	366	52.69	55.42	0.209
	(19.11)		(19.78)	(18.29)	
Total hours worked in typical week	54.58	377	52.66	56.60	0.0587
	(18.68)		(19.28)	(17.86)	
Estimated work earnings last year (Rs.)	218134.3	377	182146.7	256285.1	0.270
	(617949.3)		(542612.7)	(688355.1)	
Estimated work earnings last year (Rs.) - trimmed	105874.3	341	89239.0	124255.2	0.00399**
	(94555.8)		(64512.0)	(116757.6)	
Estimated hourly wage (Rs.)	82.66	368	76.37	89.45	0.620
	(234.7)		(191.1)	(274.5)	
Estimated hourly wage (Rs.) - trimmed	44.47	332	40.52	48.92	0.113
	(44.46)		(40.62)	(48.18)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Total hours worked last week reported for subsample who mainly worked last week. Total hours worked in a typical week reported for subsample who typically work. Trimming defined as dropping the top and bottom 5% of the distribution.

Table 39: Occupation of highest earner

	Whole Sample	N	Control	Treatment	P-value
Highest earner is a private employee (fixed income) (%)	20.6	374	21.4	19.8	0.750
Highest earner is a government employee (%)	9.1	374	7.8	10.4	0.537
Highest earner is a daily labourer (%)	21.1	374	24.5	17.6	0.235
Highest earner is a rickshaw/trolley puller (%)	3.5	374	1.6	5.5	0.081
Highest earner is a auto/taxi/car driver (%)	15.0	374	12.5	17.6	0.160
Highest earner is a petty trader vendor or hawker (%)	3.5	374	2.6	4.4	0.510
Highest earner is a boss, owner or partner on a farm or in a business (%)	21.4	374	24.5	18.1	0.258
Highest earner has other occupation (%)	5.9	374	5.2	6.6	0.607

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 40: Proportion of households who receive non-work income and transfers in the last year

	Whole Sample	N	Control	Treatment	P-value
Pensions (%)	10.0	418	9.0	11.1	0.535
Benefits from government schemes (%)	17.9	418	18.0	17.9	0.976
Dowry income (%)	1.4	418	0.9	1.9	0.371
Transfers/Remittances (%)	3.3	418	3.8	2.9	0.629
Income from rent (%)	8.6	418	6.2	11.1	0.088

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

mobile phone(s) (93.1%) and televisions (81.3%). Fewer households had items of furniture - 74.4% had a cot or bed, 72.0% had a chair, 57.7% a mattress and 55.5% a table. In terms of transportation assets, 69.9% of households owned a bicycle whilst 44.7% owned a motorcycle and just 10.8% a car. By means of household appliances 76.8% of households owned a pressure cooker, 35.5% a fridge , 18.1% a sewing machine and 14.4% a washing machine. Table 41 shows asset ownership by treatment status. We do see some imbalance in asset ownership by treatment status. For four asset types (car/auto rickshaw, sewing machine, mobile phone and sound equipment) the treatment group's ownership of these assets is significantly higher than the controls. On other asset categories there is no significant imbalance. Because we observe this imbalance here we will take care to control for it in any work assessing how the intervention affects economic decisions and/or measures of wealth and income.

Table 41: Proportion of households who own various assets

	Whole Sample	N	Control	Treatment	P-value
Bicycle (%)	69.9	418	67.8	72.0	0.407
Motorbike (%)	44.7	418	39.3	50.2	0.097
Car/autorickshaw (%)	10.8	418	6.6	15.0	0.005**
Fridge/freezer (%)	35.5	417	30.5	40.6	0.126
Fan (%)	96.2	418	95.3	97.1	0.381
Washing machine (%)	14.4	418	10.9	17.9	0.118
Pressure cooker (%)	76.8	418	75.4	78.3	0.629
Sewing machine (%)	18.2	418	13.7	22.7	0.031*
Table (%)	55.5	418	50.7	60.4	0.075
Chairs (%)	72.0	418	68.7	75.4	0.209
Cot/bed (%)	74.4	418	70.1	78.7	0.116
Mattress (%)	57.7	418	52.1	63.3	0.114
TV (%)	81.3	418	83.4	79.2	0.364
Mobile phone (%)	93.1	418	89.6	96.6	0.006**
Computer/laptop (%)	9.8	418	7.1	12.6	0.087
Sound equipment/radio/vcd/woofer (%)	24.2	418	19.4	29.0	0.042*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5.6.4 Expenditure

Household expenditures are a good, albeit noisy, measure of material standard of living along multiple dimensions. Household expenditure on food, for example, is a noisy measure of food quality and quantity. Expenditure data is also a good measure of total economic resources available to the household as in some cases this data may be less prone to measurement error than data specifically

on income and pay. In the household survey we asked respondents how much their household spent on goods in a variety of different categories during the past week, month, six months or year. We didn't include expenditure on rent, so these expenditure figures are all net of housing costs. Table 42 shows the results of these questions broken down by treatment group. We also calculated a rough measure of total expenditure over the past year by summing over these different categories. Over the whole sample, households had a mean expenditure of Rs.141,000 ($2,270) per year. This was balanced across treatment and control as was each individual component except water, electricity and mobile telephone costs.

We investigate further what these measures of household expenditure across different categories tell us about underlying economic resources and an economic decision over consumption. We do this using factor analysis. Factor analysis analyses whether variation amongst a set of correlated variables can be explained in terms of fewer variables - factors - by modelling the variation in the set of variables in terms of variation in underlying factors plus an error term. We use the well known Kaiser criterion to determine the number of factors we use. Table 43 presents our results. We find that 63.4% of the variation in the expenditure variables can be explained through one underlying factor, presumably resulting from an underlying consumption decision and determined by economic resources and constraints facing the household. It is interesting to see which categories of expenditure have the highest loading on, or the highest correlation with, the underlying factor. We find the loading is the highest for expenditures on on clothes for children under 6, on clothes and shoes for children under 6, water, electricity and mobile telephones, on education, and on household repairs/maintenance. These are all categories of expenditure that could be seen as particularly indicative of a household's underlying economic condition.

In addition to looking at the absolute value of household expenditure and the correlations between different types of expenditure it is also important to consider its breakdown. Table 44 shows expenditure on each category as a proportion of estimated total expenditure over the last year. Our estimates suggest that just under half of household's expenditure, on average, is on food. Fuel, hygiene products, transport and water, electricity and mobile telephone expenses together account for, on average, just under 30% of total expenditure. On average 8.11% is spent on health but as we would expect this has a very large variance due to the unpredictability of health shocks and little access to insurance mechanisms. This leaves little room for other expenditures and, on average, just 5.20% of household expenditure is on specialist products for young children under six years - hygiene products, toys, clothes or books.

5.6.5 Savings, debts and loans

We asked all respondents about the finances of the biological mother and father of the target child and in particular their savings, debts and loans. Table 45 shows the responses broken down by treatment group. In the sample as a whole 51.2% of couples had no savings. Those who did have savings had a median of Rs.10,000 (the mean was Rs. 43,600 although this is not representative due to extreme skewness and outliers). 59.8% of couples had no debts although 16.5% of the couples had over Rs.10,000 of debt. This last measure was not balanced between treatment and control - a significantly greater proportion of the treatment group had more than Rs.10,000 in debt than the control.

Table 42: Household expenditures (Rs.)

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Food (7d)	1228.2 (954.5)	384	1244.0 (995.0)	1212.6 (914.8)	0.761
Cigarettes, tobacco and alcohol (7d)	160.6 (471.7)	374	138.4 (249.9)	183.0 (620.1)	0.389
Hygiene products (exc. children <6) (7d)	213.6 (365.0)	389	214.6 (400.2)	212.5 (326.9)	0.961
Transport (1m)	1037.7 (2075.1)	390	1066.5 (2520.3)	1007.4 (1474.2)	0.776
Water, electricity, mobile etc (1m)	824.9 (1260.9)	378	658.5 (831.8)	986.2 (1554.0)	0.0416*
Fuel (1m)	647.5 (600.8)	401	696.5 (581.0)	597.3 (617.9)	0.131
Salaries for workers (1m)	130.7 (1010.9)	407	79.47 (727.1)	183.8 (1237.8)	0.317
Education (1m)	315.4 (1120.8)	410	378.5 (1344.4)	251.8 (835.4)	0.248
Health expenses (1m)	1029.1 (2630.0)	410	1130.3 (3248.4)	926.9 (1805.2)	0.452
Services (1m)	90.66 (137.0)	381	90.78 (130.3)	90.53 (143.9)	0.986
Hygiene products for children under 6 (1m)	292.2 (538.9)	401	281.7 (424.2)	302.9 (635.5)	0.684
Toys for children under 6 (6m)	325.3 (487.9)	402	339.4 (483.6)	310.7 (493.1)	0.619
Books for children under 6 (6m)	83.68 (394.7)	410	75.36 (436.6)	92.42 (346.2)	0.699
Clothes for children under 6 (6m)	928.3 (971.7)	398	955.9 (1074.6)	900.0 (856.0)	0.641
Clothes and shoes for members over 6 (6m)	2774.6 (4936.8)	382	2594.8 (4494.5)	2958.1 (5356.7)	0.530
Household repairs/ maintenance (12m)	1848.9 (6482.4)	383	1317.9 (4352.8)	2405.5 (8113.6)	0.115
Weddings including dowry (12m)	2769.6 (15632.0)	410	2275.5 (11075.2)	3278.5 (19245.7)	0.526
Festivals/ entertainment/ recreation (12m)	3607.4 (5174.1)	372	3738.9 (4651.1)	3470.1 (5678.7)	0.625
Estimated total expenditure (12 months)	141464.4 (108871.5)	418	139469.0 (112561.7)	143498.4 (105210.7)	0.731

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

7d=past week, 1m=past month, 6m=past 6 months, 12m=past year

Table 43: Factor analysis of household expenditure

	Factor Loading	Uniqueness
Food	0.273	0.926
Cigarettes, tobacco and alcohol	0.076	0.994
Hygiene products (exc. children <6)	0.206	0.957
Transport	0.261	0.932
Water, electricity, mobile etc	0.706	0.501
Fuel	0.343	0.882
Salaries for workers	0.159	0.975
Education	0.631	0.602
Health expenses	0.233	0.946
Services	0.403	0.837
Hygiene products for children under 6	0.227	0.949
Toys for children under 6	0.417	0.826
Books for children under 6	0.126	0.984
Clothes for children under 6	0.533	0.716
Clothes and shoes for members over 6	0.811	0.343
Household repairs/ maintenance	0.577	0.667
Weddings including dowry	0.157	0.975
Festivals/ entertainment/ recreation	0.217	0.953
Eigenvalue	3.034	
Proportion of variance	0.636	

Factor analysis performed using principal factor method. Only factor(s) with eigenvalue of greater than one kept, as per Kaiser criterion.

Table 44: Estimated proportion of total HH expenditure (past year) (% of total recorded expenditure)

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Food	45.79	384	46.55	45.05	0.498
	(18.99)		(17.68)	(20.22)	
Cigarettes, tobacco and alcohol	6.204	374	6.206	6.201	0.997
	(10.40)		(11.06)	(9.721)	
Hygiene products (exc. children <6)	7.719	388	8.539	6.898	0.118
	(8.394)		(9.687)	(6.788)	
Transport	8.016	389	7.038	9.039	0.0920
	(11.62)		(11.08)	(12.11)	
Water, electricity, mobile etc	6.783	377	5.357	8.156	0.00363**
	(7.090)		(4.482)	(8.703)	
Fuel	7.000	401	8.032	5.942	0.0322*
	(8.316)		(10.34)	(5.347)	
Salaries for workers	0.702	406	0.480	0.930	0.416
	(4.920)		(4.368)	(5.432)	
Education	2.211	409	2.667	1.753	0.0643
	(4.663)		(5.442)	(3.677)	
Health expenses	8.112	409	8.452	7.770	0.593
	(12.10)		(13.12)	(11.01)	
Services	1.078	381	0.823	1.336	0.337
	(5.227)		(1.134)	(7.334)	
Hygiene products for children under 6	2.641	400	2.684	2.597	0.792
	(4.111)		(3.869)	(4.351)	
Toys for children under 6	0.623	401	0.686	0.558	0.383
	(1.511)		(1.833)	(1.085)	
Books for children under 6	0.113	409	0.0771	0.150	0.169
	(0.458)		(0.269)	(0.594)	
Clothes for children under 6	1.819	397	2.147	1.486	0.118
	(4.556)		(6.209)	(1.603)	
Clothes and shoes for members over 6	3.694	381	3.653	3.735	0.876
	(5.266)		(5.722)	(4.774)	
Household repairs/ maintenance	1.040	382	0.794	1.298	0.0631
	(2.598)		(1.964)	(3.111)	
Weddings including dowry	1.261	409	1.071	1.456	0.543
	(6.264)		(4.993)	(7.350)	
Festivals/ entertainment/ recreation	2.709	371	2.776	2.638	0.690
	(3.369)		(3.192)	(3.550)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 45: Savings, debts and loans

	Whole Sample	N	Control	Treatment	P-value
Wife and husband have no savings (%)	51.2	418	54.0	48.3	0.265
Wife and husband have between zero and Rs.5000 in savings (%)	6.2	418	6.2	6.3	0.959
Wife and husband have between Rs.5000 and Rs.10000 in savings (%)	5.0	418	5.2	4.8	0.866
Wife and husband have more than Rs.10000 in savings (%)	6.0	418	7.1	4.8	0.322
Wife and husband have no debts (%)	59.8	418	61.6	58.0	0.459
Wife and husband have between zero and Rs.5000 in debt (%)	8.6	418	9.0	8.2	0.783
Wife and husband have between Rs.5000 and Rs.10000 in debt (%)	7.2	418	8.5	5.8	0.262
Wife and husband have more than Rs.10000 in debt (%)	16.5	418	11.4	21.7	0.009**
Wife and husband have no money lent to third parties (%)	90.4	418	90.5	90.3	0.950

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5.6.6 Ration card

In theory the type of ration card a household has (Below Poverty Line, Above Poverty Line, Antyodaya/Annapurna or other) is a useful benchmark of a household's economic position and poverty level. Currently, households qualify for a Below Poverty Line status if they score a certain number of points on an assessment of overall deprivation. However, there are many claims of significant corruption and distortions in the distribution of Below Poverty Line cards resulting in the cards not being allocated to the households for which they are intended. Indeed a recent study estimated that, in Odisha, as many as 30% of non-poor households had a BPL card whereas only 58.8% of poor households did (Ram et al., 2009). In general, a lower proportion of urban households (as in our sample) posses BPL cards than rural households. In table 46 we see that 10.8% of households in our sample possessed a BPL ration card. This was balanced between treatment and control.

Table 46: Household's ration card status

	Whole Sample	N	Control	Treatment	P-value
Below Poverty Line card (%)	10.8	418	11.8	9.7	0.525
Above Poverty Line card (%)	3.8	418	3.8	3.9	0.975
Antyodaya/Annapurna card (%)	1.7	418	1.9	1.4	0.790
Other ration card (%)	15.6	418	12.3	18.8	0.088

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

5.6.7 Quality of dwelling

Table 47 shows that more than three-quarters of sample households were classified (by the interviewer) as living in a pucca house, meaning a house with a strong structure. 90.2% of households had cement or brick walls and 85.4% had cement floors. Almost all households (all but six) lived in dwellings that were connected to electricity. Most (58.4%) had a piped water connection but none were connected to the sewage system. All of these physical characteristics of the dwellings, where our sample lived, were balanced across treatment and control.

Table 47: Physical dwelling characteristics

	Whole Sample	N	Control	Treatment	P-value
Pucca house (strong structure) (%)	77.0	418	76.3	77.8	0.833
Semi-pucca house (%)	15.6	418	17.1	14.0	0.487
Kutcha house (weak structure) (%)	7.4	418	6.6	8.2	0.687
Floor material - cement (%)	85.4	418	84.4	86.5	0.661
Floor material - tiles (%)	7.9	418	8.5	7.2	0.669
Floor material - mud/earth/soil/cow-dung (%)	5.5	418	5.2	5.8	0.873
Wall material - concrete/brick (%)	90.2	418	91.9	88.4	0.468
Wall material - mud/brick/stone (%)	5.0	418	3.8	6.3	0.395
Wall material - mud/wooden plank (%)	3.1	418	2.8	3.4	0.804
Roof material - cement/RCC (%)	56.2	418	53.1	59.4	0.440
Roof material - sheet/tin (%)	23.9	418	26.5	21.3	0.406
Roof material - thatch (%)	16.5	418	18.5	14.5	0.512
Dwelling connected to electricity (%)	98.6	418	98.6	98.6	0.980
Dwelling has piped water (%)	58.4	418	57.3	59.4	0.810
Dwelling connected to sewage system (%)	0.0	416	0.0	0.0	.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

6 Baseline data – Sahis

A final data collection instrument we used was a sahi survey. This contained general information on the sahi and was answered by person who was knowledgeable about the sahi as a whole - often the Anganwadi worker.

6.1 Population

In terms of the population of the study sahis table 48 shows that the average estimated number of households in the sahi was 323.6. Of these respondents estimated that roughly one third belonged to a Scheduled Caste and another third belonged to an Other Backward Class. We see no imbalance between treatment and control over these variables.

Table 48: Population of sahi

	Mean (SD)	N	Control Mean (SD)	Treatment Mean (SD)	P-value
Number of households in the sahi	323.6	54	346.1	305.6	0.612
	(290.5)		(361.3)	(208.1)	
Number of Scheduled Caste families	107.8	54	130.4	89.26	0.256
	(132.9)		(137.1)	(128.6)	
Number of Scheduled Tribe families	7.909	54	7.741	8.370	0.920
	(22.70)		(18.21)	(27.13)	
Number of Other Backward Class families	101.2	54	101	105.2	0.914
	(143.1)		(131.5)	(157.5)	
Number of general families	101.8	54	102.9	97.11	0.898
	(165.3)		(198.0)	(131.0)	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

6.2 Facilities within the sahi services

Moving on to the type of facilities that are available within the sahi we see, from table 49, that all the sahis have at least one private health facility (defined as a a private clinic, private hospital, private doctor/quack) whereas only 12.7% contained a government health facility (defined as a sub-centre, a primary health centre, a community health centre, a government dispensary or a government hospital). Significantly more control sahis contained government health facilities than treatment sahis. In terms of institutions and facilities aimed at early childhood education 85.5% of sahis contained a government Anganwadi whereas only 16.4% contained a private nursery or crèche. For older children, 47.3% of sahis contained a primary school whereas just 7.3% contained a secondary school.

In terms of facilities at which to purchase food, just 16.4% of sahis had a market within the sahi yet 45.5% had a public distribution system shop. For buying goods that enrich the quality of the home environment, in terms of early child development, 14.5% of sahis contained a bookshop and 10.9% a toyshop. These are balanced across treatment arms.

Table 49: Facilities within the sahi

	Whole Sample	N	Control	Treatment	P-value
Government health facility (%)	12.7	54	22.2	3.7	0.042*
Private health facility (%)	100.0	54	100.0	100.0	.
Private nursary/creche (%)	16.4	54	18.5	14.8	0.719
Government anganwadi (%)	85.5	54	85.2	85.2	1.000
Government primary school (%)	47.3	54	40.7	51.9	0.418
Government secondary school (%)	7.3	54	7.4	7.4	1.000
Chemist (%)	40.0	54	44.4	37.0	0.584
Bank (%)	5.5	54	0.0	11.1	0.074
Bookshop (%)	14.5	54	11.1	18.5	0.449
Micro-finance institution (%)	21.8	54	22.2	18.5	0.739
Toy shop (%)	10.9	54	11.1	11.1	1.000
Co-operative society (%)	3.6	54	3.7	3.7	1.000
Childrens' clothes shop (%)	14.5	54	7.4	22.2	0.127
Playground (%)	14.5	54	11.1	18.5	0.449
Market (%)	16.4	54	11.1	18.5	0.449
Wine shop (%)	21.8	54	29.6	14.8	0.193
Public distribution system shop (%)	45.5	54	40.7	48.1	0.589

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

6.3 Social programmes within the sahi

Table 50 shows that 90.9% of sahis had social programme to provide free, micro-nutrient enriched food to school and/or Anganwadi children. 90.9% also provided advice on nutrition to women whilst 72.7% organised public meetings/activities. The sample was balanced across treatment and control for each of these social programmes.

Table 50: Social programmes (government or other) within the sahi

	Whole Sample	N	Control	Treatment	P-value
Provide complementary food for school/ Anganwadi children (%)	90.9	54	88.9	92.6	0.643
Provide micronutrient enriched food for school/ Anganwadi children (%)	90.9	54	88.9	92.6	0.643
Provide advice on nutrition to women (%)	90.9	54	88.9	92.6	0.643
Provide advice on nutrition to pregnant women (%)	89.1	54	85.2	92.6	0.391
Organise public meetings/activities (%)	72.7	54	77.8	66.7	0.367

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

7 Baseline data – home visitors

In this section we move onto information collected on the home visitors - the 27 women selected and trained by Pratham to carry out the home visiting programme. Naturally, we only have home visitors in the 27 treatment sahis so there is no comparison to be made between treatment and control. We simply provide descriptives on the characteristics of these women. In the event that home visitors drop out of the programme, Pratham will recruit replacements who we will record equivalent information for. It should be noted that all information on home visitors was collected prior to training beginning but after selection so answers should not have been influenced by changes in knowledge or expectations due to training. Likewise, home visitors should have had no incentive, at this stage, to overstate their qualification for the role.

7.1 Background and status

From table 51 we see that the majority, 63.0%, of home visitors were under the age of 25, just over half, 55.6%, were married and just under half, 48.1%, had their own children. 22.2% of home visitors had a Below Poverty Line (BPL) card, which is a higher proportion than for the sample households (see table 46). However, we do not see this comparison as particularly meaningful since the possession of cards is often not strongly related with higher levels of poverty (Ram et al., 2009). In terms of education levels, to be eligible to become a home visitor women should have completed 10th standard and preferably 12th. However, when we look at the self-reported education data from home visitors, in table 51, we find that only 74.1% had completed 10th standard. We will look into whether this discrepancy arose because of applicants misreporting their educational level or because of the criteria being relaxed in certain cases. However, we do see that almost all, 92.6%, of recruited home visitors had completed ninth standard whilst a third had completed higher secondary school. This is far higher than the educational levels of the sample mothers. Correspondingly, all home visitors could read a newspaper and write a letter. Two thirds could send a text message. In the week before taking the job as a home visitor, 55.6% had mostly done household work whilst 37.0% had mostly worked.

7.2 Knowledge of child development

We measured home visitors' knowledge of child development, prior to them engaging in any training for the home visiting programme, using the same tool as we used for biological mothers. However, due to an error in compiling different versions of the questionnaires the final version of this instrument in the home visitors' questionnaires had a four-point scale rather than the three-point one used in the biological mothers' questionnaires. Unfortunately, this means the scores of the two groups are not comparable.

Interestingly, however, whilst the mothers scored highest on the 'school readiness and expectations' section the home visitors scored the lowest on this section (see table 52). This might be because only around half the home visitors had children of their own. At the other end of the scale, the home visitors, on average, scored best on the 'punishing child' domain which included statements like "Hitting your child might be a good way of teaching her/him what she/he can and cannot do."

Table 51: Background characteristics of home visitors

	Mean	
	%	N
25 years old or older (%)	63.0	27
35 years old or older (%)	33.3	27
45 years old or older (%)	7.4	27
Married (%)	55.6	27
Has own child(ren) (%)	48.1	27
Has Below Poverty Line ration card (%)	22.2	27
Can read a newspaper (%)	100.0	27
Can write a letter (%)	100.0	27
Can send a SMS (%)	66.7	27
Completed ninth standard schooling (%)	92.6	27
Completed tenth standard schooling (%)	74.1	27
Completed higher secondary schooling (%)	33.3	27
Completed graduate level schooling (%)	18.5	27
Mostly household work in week before becoming HV (%)	55.6	27
Mostly worked in week before becoming HV (%)	37.0	27

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

Table 52: Home visitors' knowledge of child development

	Mean	
	%	N
Praising/paying attention to child (% score)	68.5	23
Punishing child (% score)	88.0	23
School readiness and expectations (% score)	60.6	23
Importance of maternal interactions and play (% score)	72.5	23
Age appropriate expectations (% score)	80.3	23
Total child development knowledge score (% score)	69.7	23

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: significance level at which we can reject the hypothesis that the treatment and control mean are equal. All p-values adjusted for clustering at the slum level.

8 Conclusions

This report has provided an in-depth look at the baseline data collected for the randomised evaluation of a home visiting intervention in Cuttack, India, aimed to improve early childhood development. We have shown formal tests comparing all important characteristics collected at baseline, across treatment and control. This is an important exercise since it allows us to see whether, indeed, the randomisation was successful at creating two groups (treatment and control) that appear similar on all dimensions other than one (treatment) will be eligible to receive the intervention.

The results from the exercise are generally encouraging. We find few significant differences in key variables across treatment and control units. We find no significant differences in baseline indicators of child cognitive, non-cognitive and motor development and none in indicators of health and morbidity. This is important since it implies our treatment and control children are not significantly different in terms of their starting level of development, prior to the intervention starting. In terms of inputs into child development we only occasionally find imbalances across treatment and control and most are small and do not provide evidence of systematic differences between the treatment and control group. An exception to this might be the educational and literacy of the biological mother where mothers in the treatment group appear slightly more educated and more literate than those in the control although the magnitude is small. The other possible exception is on some dimensions of wealth where treatment sahis appear to be slightly better off than their control counterparts. On four categories of asset ownership the treatment group appears to have a higher ownership rate than the controls. In addition, the treatment group has higher estimated earnings of the father as well as more households with high levels of debt. However, over other dimensions of wealth there appears to be no significant difference between the groups - for example household expenditures, dwelling characteristics and ration card status.

Overall, though, our sample units do appear balanced across treatment and control which lays down a good foundation for analysing the impacts of the home visiting programme.

9 Future directions

At the time of this report's publication the home visiting intervention is ongoing in Cuttack. Due to unforeseen delays in follow-up data collection the length of the home visiting intervention has been extended so, in total, the intervention will run for around 18 months in all treatment sahis, instead of the original 12 months. Preparations are currently being finalised for follow-up data collection which will begin in May 2015. If the anticipated timeline is kept, a full impact evaluation report will be published by mid-2016.

10 Bibliography

References

[1] S. Grantham-McGregor, Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp, Developmental potential in the first 5 years for children in developing countries., Lancet 369 (9555) (2007) 60–70. doi:10.1016/S0140-6736(07)60032-4.

[2] Ministry of Women & Child Development, Integrated Child Development Services (ICDS) Scheme.

[3] P. Gertler, J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, S. Grantham-McGregor, Labor market returns to an early childhood stimulation intervention in Jamaica., Science (New York, N.Y.) 344 (6187) (2014) 998–1001. doi:10.1126/science.1251178.

[4] S. Grantham-McGregor, C. Powell, S. Walker, J. Himes, Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study, The Lancet 338 (8758) (1991) 1–5. doi:10.1016/0140-6736(91)90001-6.

[5] S. M. Grantham-McGregor, S. P. Walker, S. M. Chang, C. A. Powell, Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children, American Journal of Clinical Nutrition 66 (2) (1997) 247–253.

[6] S. P. Walker, S. M. Chang, C. A. Powell, E. Simonoff, S. M. Grantham-McGregor, Effects of psychosocial stimulation and dietary supplementation in early childhood on psychosocial functioning in late adolescence: follow-up of randomised controlled trial., BMJ (Clinical research ed.) 333 (7566) (2006) 472. doi:10.1136/bmj.38897.555208.2F.

[7] O. P. Attanasio, C. Fernández, E. O. a. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, 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 (Clinical research ed.) 349 (September) (2014) g5785. doi:10.1136/bmj.g5785.

[8] N. Bayley, Manual for the Bayley scales of infant and toddler development, third edition, Pearson Education , Inc., 2006.

[9] N. Bayley, Bayley scales of infact and toddler development, third edition, Pearson Education , Inc., 2006.

[10] J. Squires, D. Bricker, Ages and stages questionnaires, third edition (ASQ-3), Paul H. Brookes Publishing Co., Inc., 2009.

[11] D. Bricker, J. Squires, L. Mounts, Ages and stages questionnaires, first edition, Paul H. Brookes Publishing Co., Inc., 1995.

[12] J. Squires, D. Bricker, L. Potter, Revision of a parent-completed developmental screening tool: Ages and stages questionnaires, Journal of Pediatric Psychology 22 (3) (1997) 313–328. doi:10.1093/jpepsy/22.3.313.

[13] S. Bowes, H. Gintis, Schooling in Captalist America Revisited, Sociology of Education 75 (1) (2002) 1–18. doi:10.2307/3090251.

[14] G. Farkas, Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes, Annual Review of Sociology 29 (2003) 541–562. doi:10.1146/annurev.soc.29.010202.100023.

[15] J. J. Heckman, J. Stixrud, S. Urzua, The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior (2006). doi:10.1086/504455.

[16] J. J. Heckman, Y. Rubinstein, The benefits of skill: The importance of noncognitive skills: Lessons from the GED testing program, American Economic Review 91 (2) (2001) 145–154. doi:10.1257/aer.91.2.145.

[17] C. Lleras, Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings, Social Science Research 37 (2008) 888–902. doi:10.1016/j.ssresearch.2008.03.004.

[18] L. M. Gutman, I. Schoon, The impact of non-cognitive skills on outcomes for young people: Literature review (2013).

[19] M. Haga, The relationship between physical fitness and motor competence in children., Child: care, health and development 34 (2008) 329–334. doi:10.1111/j.1365-2214.2008.00814.x.

[20] K. Adolph, Motor development, 7th Edition, Vol. 2, Wiley, New York, 2015.

[21] S. P. Walker, T. D. Wachs, J. M. Gardner, B. Lozoff, G. A. Wasserman, E. Pollitt, J. A. Carter, Child development in developing countries 2 Child development : risk factors for adverse outcomes in developing countries (2007) 145–157.

[22] J. D. Hamadani, F. Tofail, A. Hilaly, S. N. Huda, P. Engle, S. M. Grantham-McGregor, Use of family care indicators and their relationship with child development in Bangladesh., Journal of health, population, and nutrition 28 (1) (2010) 23–33.

[23] D. MacPhee, The Nature of Parents' Experiences with Knowledge About Infant Development, in: Biennial Meeting of the Society of Research Development: The Study and Implications of Parental Knowledge, Perceptions, Reasoning and Beliefs., 1983.

[24] L. S. Radloff, The CES-D Scale: A Self-Report Depression Scale for Research in the General Population (1977). doi:10.1177/014662167700100306.

[25] P. Kline, Handbook of psychological testing, 2nd Edition, Routledge, 2013.

[26] WHO Multicentre Growth Reference Study Group., WHO Child Growth Standards based on length/height, weight and age., Acta paediatrica. Supplementum 450 (2006) 76–85. doi:10.1080/08035320500495548.

[27] World Health Organisation, Infant and young child feeding (2014).

[28] M. Arimond, M. T. Ruel, Dietary diversity is associated with child nutritional status: evidence from 11 demographic and health surveys., The Journal of nutrition 134 (August) (2004) 2579–2585. doi:0022-3166/04.

[29] M. C. Lovejoy, P. A. Graczyk, E. O'Hare, G. Neuman, Maternal depression and parenting behavior: A meta-analytic review, Clinical Psychology Review 20 (5) (2000) 561–592. doi:10.1016/S0272-7358(98)00100-7.

[30] S. Kurstjens, D. Wolke, Effects of maternal depression on cognitive development of children over the first 7 years of life., Journal of child psychology and psychiatry, and allied disciplines 42 (5) (2001) 623–636.

[31] S. Desai, S. Alva, Maternal education and child health: is there a strong causal relationship?, Demography 35 (1998) 71–81. doi:10.2307/3004028.

[32] K. Cunningham, M. Ruel, E. Ferguson, R. Uauy, Women's empowerment and child nutritional status in South Asia: a synthesis of the literature., Maternal & child nutrition 11 (2015) 1–19. doi:10.1111/mcn.12125.

[33] F. Ram, S. K. Mohanty, U. Ram, Understanding the Distribution of BPL Cards: All-India and Selected States, Economic and Political Weekly xliv (7) (2009) 6671.