# Early Childhood Development for the Poor:

# Impacting at Scale, Odisha, India

*Project PIs*:
Costas Meghir[⋆,ϒ], Orazio Attanasio[⋆,†], Britta Augsburg[⋆], Jere Behrman[⋆,Γ], Bet Caeyers[⋆], Monimalika Day[Ω], Sally Grantham-McGreggor[⋆], Smriti Pahwa[§], and Marta Rubio-Codina[⋆]

*We thank for important input*:
Akanksha Adya[Ω], Alison Andrew[⋆], Deepanjali Behera[ϱ], Jena Hamadani[Γ], Safija Jusupovic[⋆], Reema Kochar[Ω], Preeti Mahalwal[Ω], Fardina Mehrin[Γ], Maria Grazia Pastorello[ϱ], Edward Penington[⋆], Soonrita Sahasrabuddhe[ω], Karishma Vats[§]

*The stimulation and nutrition curricula were developed by*:
Akanksha Adya[Ω], Monimalika Day[Ω], Sally Grantham-McGreggor[⋆], Reema Kochar[Ω], Preeti Mahalwal[Ω], Smriti Pahwa[§], Marta Rubio-Codina[⋆], Karishma Vats[§] with input from Jena Hamadani[Γ], Fardina Mehrin[Γ] and the PI team

*Special thanks to Rukmini Banerji[§] and Venita Kaul[Ω] for their continuing support and input into the project*

# June 2016

[ϒ]Yale University, USA
[⋆]The Institute for Fiscal Studies, London, UK
[†]University College London, UK
[Γ]University of Pennsylvania, USA
[Ω]Centre for Early Childhood Education and Development (CECED), Ambedkar University Delhi, India
[§]Pratham Education Foundation- ASER Centre, India
[ϱ]JPAL-SA
[Γ]International Centre for Diarrhoeal Disease Research, Bangladesh
[ω]Independent Consultant

This baseline research was made possible through the generous support of the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (NIH); We also thank the Economic and Social Research Council (ESRC), The Population Studies Center at the University of Pennsylvania and Yale University for complementary funding. Responsibility for any errors is ours.

# Contents


**1 Executive Summary** **5**

**2 Overview** **8**

2.1 Introduction . . . . . 8

2.2 Project Background . . . . . 11

2.3 The Intervention . . . . . 12

2.3.1 Health and nutritional services link (*HNSL*): . . . . . 13

2.3.2 Individual Stimulation (*IS*) Curriculum . . . . . 13

2.3.3 Group Stimulation (*GS*) Curriculum . . . . . 14

2.3.4 Nutritional Education (NE) Curriculum . . . . . 15

2.4 Intervention implementation team . . . . . 15

2.4.1 Home visitors and group facilitators . . . . . 16

2.4.2 Training, mentoring and monitoring . . . . . 17

2.5 Our Evaluation Study . . . . . 18

**3 Methodology** **20**

3.1 Terminology . . . . . 20

3.2 Evaluation design . . . . . 21

3.2.1 Randomization . . . . . 21

3.2.2 Evaluation specification . . . . . 23

3.3 Sampling selection strategy . . . . . 24

3.3.1 Selecting villages . . . . . 24

3.3.2 Village mapping and census . . . . . 26

3.3.3 Sample of target children . . . . . 27

3.3.4 Sample of spillover children . . . . . 28

3.4 Outcome indicators . . . . . 29

3.4.1 Indicators of child development (primary outcomes) . . . . . 29

3.4.2 Indicators of child development inputs . . . . . 30

3.5 Instruments for data collection . . . . . 32

3.6 Data collection, sample size and data structure . . . . . 33

3.6.1 Time line . . . . . 33

3.6.2 Survey team structure . . . . . 34

3.6.3 Survey staff training . . . . . 35

3.6.4 Sample size . . . . . 35

3.6.5 Data structure . . . . . 36

3.7 Purpose of this baseline report . . . . . 38

**4 Baseline data – Community characteristics** **40**
4.1 Most important industries . . . 40
4.2 Sanitation and electricity . . . 41
4.3 Institutions . . . 42
4.3.1 Commercial institutions . . . 42
4.3.2 Social programs, clubs and other public institutions . . . 43
4.4 Education and health facilities . . . 44
4.4.1 Health facilities . . . 44
4.4.2 Educational facilities . . . 44

**5 Baseline data – household characteristics** **46**
5.1 Religion and Caste . . . 46
5.2 Household head and composition . . . 47
5.2.1 Household head . . . 47
5.2.2 Household composition . . . 49
5.3 Wealth indicators . . . 51
5.4 Economic status - asset, financial wealth and income . . . 53
5.4.1 Assets . . . 53
5.4.2 Income . . . 53
5.4.3 Income of main breadwinner, biological mother and father . . . 56
5.4.4 Savings and debt . . . 60
5.5 Consumption expenditures . . . 61

**6 Baseline data – Child characteristics** **63**
6.1 Age and sex . . . 63
6.2 Child development - ASQ . . . 65
6.2.1 Target children . . . 66
6.2.2 Spillover children . . . 69
6.3 Child development - anthropometrics . . . 70
6.3.1 Target children . . . 74
6.3.2 Spillover children . . . 76
6.4 Child development - morbidity . . . 79
6.4.1 Target children . . . 79
6.4.2 Spillover children . . . 79
6.5 Child Nutrition, Healthcare and Birth . . . 82
6.5.1 Birth and breastfeeding . . . 82
6.5.2 Nutrition . . . 82
6.6 Maternal Health and Education . . . 84

6.6.1 Maternal health . . . 85
6.6.2 Maternal education . . . 88
6.6.3 Maternal knowledge . . . 90
6.7 Quantity and Quality of Maternal Time . . . 91
6.8 Quality of home environment for stimulation . . . 92

**7 Conclusion** **99**

**A Training schedule** **103**

**B Age registration protocol** **104**
B.1 Age determination during and after census . . . 104
B.2 Age verification during baseline household roster administration . . . 104
B.3 Age verification prior to ASQ survey . . . 106
B.4 Age verification after survey . . . 106
B.5 Final age confirmation and updates during data cleaning . . . 106

# 1 Executive Summary

This report presents a detailed description of baseline data collection and intervention preparation activities conducted as part of the project 'Early Childhood Development for the Poor: Impacting at Scale', a project primarily funded by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (NIH) and the Strategic Impact Evaluation Fund (SIEF).

The project is a collaboration between Yale University (USA), The Institute for Fiscal Studies (UK), University College London (UK), University of Pennsylvania (USA), the Centre for Early Childhood Education and Development (India), Pratham Education Foundation- ASER Centre (India) and the Abdul Latif Jameel Poverty Action Lab SA (India).

Our research program has two main overarching objectives:

1. To improve the health and development of very young Indian children living in poor rural environments, hence contributing to breaking the intergenerational cycle of poverty.

2. To formulate a blueprint for early childhood development (ECD) strategies that will be applicable in poor communities, based on the use and empowerment of local resources, and capable of changing the developmental trajectories of children in a substantive way.

The project aims to establish scientifically the impact of Early Childhood Development policies that can form the basis for the design of low cost and sustainable ECD strategies to supplement current policy such as the ICDS.

To this end, the proposed research has two specific aims:

**Specific Aim 1**: To implement alternative service provision modes of an ECD intervention that combines a stimulation package and nutritional education, targeted at children younger than three years (i.e. before they attend the Anganwadi Centers) in communities of rural Odisha and that is delivered by local women. Alternative service provisions of the stimulation package include the delivery of the curriculum either in weekly individual home visits or during weekly group sessions.

**Specific Aim 2**: To evaluate by randomized control trial (RCT) the impacts of these interventions on child development and health, to investigate their scalability and the relative effectiveness of each mode of delivery, in comparison to an intervention solely based on nutritional education, and to identify the mechanisms whereby the interventions affect ECD outcomes.

The implementation of the intervention (Aim 1) and the collected data (Aim 2) establish the basis for the investigation of the research questions and research strategy.

This report discusses the steps understaken towards the implementation of the intervention, which started in the field in December of 2015, and discusses the baseline data collection, which took place between 31 August 2015 and 19 December 2015. A large part of the report is dedicated to describing the data

collected, providing a picture of the study communities and households, while at the same time discussing the balancedness of these characteristics between the different study arms.

Overall, we find that data collection was successfully implemented and that data are of the quality standards required. 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 randomization was successful at creating study groups (treatments and control) that appear similar on all dimensions. The only diffference will then be the interventions implemented.

**Snapshot of the study sample households and children**

The average household in our sample group is hindu and comprises five household members, with typically a male household head, in his mid-forties, who is able to read and write (~75%). The annual household income is INR 91,260 (US$ 1,372). With an average household size of 5.38, this implies that members of our study households live on average on US$ 0.70 per day (not applying any equivalent scale for children). This puts the average household member signicantly below the internationally acceptable poverty line of US$ 1.25 per person per day. In line, 65% of households own a ration card, making them eligible for government support.

The interventions target children are between the ages of 7 to 16 months at the time of the intervention start. We additionally collect data on children just below and above this age range to understand spillover effects. These children are very disadvantaged. Growing up in poverty establishes itself in stunted growth (having a height-for-age more than 2 SD below the median of the NCHS/WHO international reference): 13% of all our sample children (aged between 1 and 20 completed months at the time of measurement) are stunted (low height for age), a deficit is generally assumed to indicate exposure to an unhealthy environment, such as poor nutrition, lack of hygiene or disease. The probability of stunting significantly increases as the child grows older, especially after its first 6 months of life. One in five of the children 19-20 months old in our sample are stunted. In line, 76% of target children experienced at least one of the nine symptoms of bad health within the last two weeks: About 46% experienced fever and 55% coughing. One third of children experienced both of these symptoms, which is indicative of malaria, a predominant illness in the study areas. Also relatively common was diarrhoea, at 14%, vomiting (27%), and tiredness (17%) and paleness (19%).

Given this context, it comes at no surprise that the quality of the home environment in terms of the amount of stimulation these children receive is low. While the large majority of households have some toys for the children (~83%), the number and variety is low. Only about 10 percent of children are read to or shown picture books (~56% of households do not own any books), and 15% are told stories. The most common activity performed with children is singing or going to the market or park or other places (~71%)

**Comparison between study arms**

In terms of the balancedness of these characteristics between our study arms, we find no important differences in baseline indicators of child development and none in indicators of health and morbidity. This is important since it implies our treatment and control children are not fundametnally 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 treatments and control and most are small and do not provide evidence of systematic differences between the treatment and control group. An exception to this is the percentage of stunted children falling in the older spillover category. We find here that these 14-20 months old children are significantly shorter and significantly more likely to be stunted in treatment communities. We also find that the play activities household members perform with the child are not balanced between treatments and control throughout. We for example find that household members in communities allocated to the NE treatment arm are on average more likely to read with the child or look at picture books and this imbalance remains when considering the F-stat. We also find a small imbalance in the percentage of biological mothers having secondary schooling. However, on the vast majority of dimensions considered, the study groups are well balanced, this includes household composition, characteristics of households members, particularly the household head and the child's parents, income and wealth information, labour supply and dwelling characteristics.

We believe that the data collection process was successfully implemented and we find data quality to be at the standards required, implying that we are in a very promising position for the impact evaluation study.

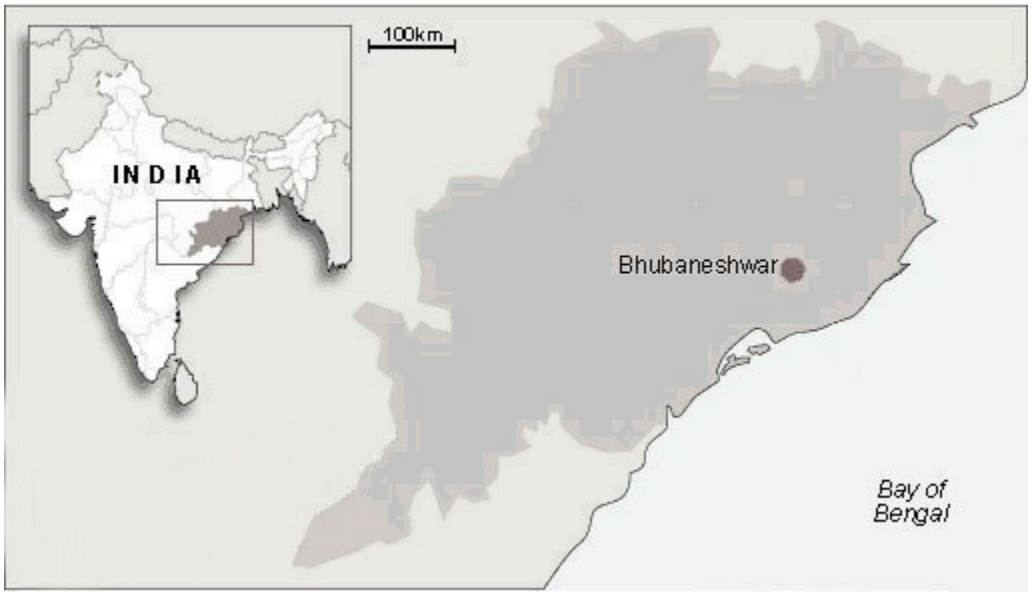
# 2 Overview

## 2.1 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 the first years as they develop skills and capabilities that affect lifetime outcomes as diverse as lifetime earnings, wellbeing and criminality. During this stage, gaps that open up between children, along familiar lines of wealth and income, 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, infections and unstimulating environments all contribute to children falling behind developmentally what they would have been capable of in a more supportive environments. Excitingly, however, a vibrant research agenda demonstrates that a child's development is highly malleable: it is markedly affected by their environment which can be altered by programs and policy. This creates a clear rationale for intervening early in life, especially for the most disadvantaged children.

It is in this context that researchers at the IFS, UCL, University of Pennsylvania and Yale University, in collaboration with Pratham, Centre for Early Childhood Education and Development (CECED) and JPal-SA, are running a trial of the relative effectiveness of three variants of an intensive two-year long Early Childhood Development (ECD) intervention in rural communities in three districts in Odisha, India. See Figure 1 for the location of Odisha in India. Each variant is based on a different combination and/or mode of delivery (home visits versus group sessions) of a childhood stimulation program and an interactive nutritional education intervention.

Figure 1: Map of Odisha



Small-scale studies have demonstrated that offering psychosocial and cognitive stimulation and nutrition to young children can have positive effects on short and long-term outcomes, such as cognition, self-esteem, depression and other important measures [3, 9, 13, 14, 19, 23, 24, 30]. In Jamaica, a home visiting program centred on increasing levels of psychosocial stimulation of children and strengthening the mother -child relationship (we use mothers as shorthand for principle caregiver)- had significant positive impacts on children's levels of cognitive functioning, mental health, social behaviour, educational and labour market outcomes both in the short, medium and longer term [10, 11, 12, 31]. Little evidence is available however on whether such interventions can be scaled-up to form the basis of sustainable and effective policy towards child development, particularly in poor contexts.

A larger scale program 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 [4]. This finding is important since it showed that an ambitious aim to improve child development through altering complex caregiver-child relationships and interactions, could be achieved in a scalable manner using local resources and through pre-existing institutions.

Another recent impact evaluation study of a home visit stimulation intervention in Cuttack, Odisha (India) built on that research agenda by investigating 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. Having local women deliver the intervention is key to designing low-cost and sustainable policies. Like the Colombian study, the program in Cuttack also made use of existing institutional infrastructure although this time leveraging the infrastructure of one of the country's largest non-governmental organisations, Pratham. This pilot study finds evidence of substantial and highly significant benefits of the home visiting programme at the end of the study.

In addition to the involvement of local women for the delivery of the child stimulation service, the formation of groups could further reduce costs and enhance the dissemination of good child rearing practices. A priori, both home visits and group meetings have advantages and disadvantages. Individual visits can be tailored to the developmental level of the child and adjusted to the child's progress. It may be easier for the visitor to develop good relationships with the mother and child. She will also be in a better position to understand the home environment and identify supporting and hindering aspects to the child's development, and work with or around them. In contrast, group meetings are less costly, easier to scale-up and are culturally appropriate in India. Mothers might additionally benefit from the social support provided by the group. On the other hand, mothers may not attend regularly, especially the highest risk mothers, and it probably requires more skill to engage and motivate groups rather than individuals and to identify the actual developmental level of each child.

The limited evidence on using groups to improve ECD outcomes, in developing countries [1, 7, 26] are mostly small, short term, efficacy trials and we are unaware of any long term follow-up of groups. However, group delivery has been shown to be effective for other programs. Tripathy et al. [29], for example, present

supporting evidence from a group-based intervention in Odisha relating to birth outcomes that emphasizes the importance of groups for community mobilization. Whether group delivery of ECD programs works if implemented based on community resources remains an under-researched but critically important question. The current study aims to shed light on this by comparing the effectiveness of adaptations of the Jamaican curriculum, delivered either through individual home visits or through group-based stimulation sessions, coupled to an interactive nutritional education intervention.

Each intervention will consist of 24 months of weekly home visits/group sessions organized by trained local women following a structured weekly curriculum based on the Jamaican curriculum. The aim of the stimulation intervention 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 nutritional education curriculum, attached to both home visits and group-based stimulation sessions, is designed to produce positive changes in food choice, preparation and storage and child health care practice, through interactive dialogue and discussion, and ultimately to lead to improved child nutrition and growth.

The impact evaluation of these interventions, which is based on a randomized control trial design with three different treatment arms and a control arm, will rigorously study the effectiveness of the group-based stimulation intervention relative to that of the home visiting stimulation intervention over a range of child development indicators. In addition, its design will allow us to compare the effectiveness of such stimulation interventions combined with nutritional education to the value of a package that is solely based on nutritional education. Further work will attempt to analyse the mechanisms through which the programs impacted (if at all) child development measures. Data will be collected on child developmental outcomes and detailed maternal- and household-level data before the start, halfway through and at the end of the two-year intervention.

This reports focuses on the baseline survey and offers a detailed description of sampling methodology, the practicalities of baseline data collection and descriptives of the data itself. Baseline data of beneficiary and non-beneficiary children and their caregivers were collected between 31 August 2015 and 19 December 2015, just before the (phased-in) intervention began in each district (Cuttack: 2 December 2015; Bolangir: 16 December 2015; Balasore: 11 January 2016). We provide descriptive statistics of our sample over many 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 households with young children living in rural villages in three districts (Balasore, Cuttack and Bolangir) of Odisha, India[1]. In terms of checking the validity of our prospective evaluation we test for any systematic differences between treatment and control groups which could undermine the argument that our randomization led to four groups (three treatment and one control group) balanced on observable and unobservable characteristics.

[1]Within each of these three districts we focus on one administrative block: block Salepur in Cuttack, block Soro in Balasore and block Bolangir in Bolangir.

## 2.2 Project Background

In 2011, India was home to more disadvantaged children under the age of five than any other country: More than 63 million under five met the criteria of disadvantage, defined as either being stunted, living in poverty or both [20]. These children face huge barriers, from poor nutrition 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 subsequently in the labour market. Partly driven by their own poverty parents are often unable to provide high quality care for their own children who will subsequently also have poor life chances. Not only do these 63 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 was one of the first countrues to create formal institutions responsible for early childhood development. In 1975 the Integrated Child Development Services (ICDS), now the world's largest integrated early childhood program, was created with the following stated objectives [22]:

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 [22]:

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.

Several states have taken initiatives to improve their preschool curriculum. In order to improve preschool outcomes, it is important to provide appropriate learning opportunities during the birth to three phase. This present study evaluates interventions that particularly focus, in a rigorous way, on a closer engagement with mothers and following a very structured curriculum. This model - which is designed with the potential of going to scale in mind - could 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 prove a valuable model in attempting to extend further the role of ICDS in this area.

## 2.3 The Intervention

In each one of 192 study villages in 3 districts in Odisha (Bolangir, Cuttack and Balasore) we implement either one of the following four intervention packages:

1. Health and nutritional services link (*HNSL*): A very basic one-off intervention that strengthens links with the existing community services will be available to all study arms, including the control group. The aim of offering the basic service to all is to create a baseline where the current policy framework is well understood. We will then measure the impact of our other interventions over and above a status quo, which encourages take-up of policy as is now.

2. *HNSL* + provision of nutritional education (*NE*): In addition to *HNSL,* local women hired for the project and trained in a specifically designed *NE* curriculum (henceforth referred to as *home visitors*) will conduct weekly visits to the home to deliver an interactive *NE* curriculum, which was designed to produce positive changes in food choice, preparation and storage and child health care practice;

3. *HNSL* + *NE* + Individual stimulation via home visits (*IS*): In addition to *HNSL,* local women hired and trained for the project (henceforth referred to as *home visitors*) will conduct weekly visits to the home to deliver both the *NE* curriculum *and* a specifically designed individual stimulation curriculum, that involves both mother and child in play and learning activities.

4. *HNSL* + *NE* + Group stimulation (*GS*). In this variant, in addition to *HNSL,* local women hired and trained for the project (henceforth referred to as *group facilitators*) will conduct weekly visits to the village to deliver a specially designed version of the *NE* curriculum *and* a specially designed version of

the stimulation curriculum to a group of mothers and children. The groups will meet weekly in their village and have a maximum of 8 mothers and children.

Below we provide more details on the activities, the curricula, the home visitors/group facilitators and their training and mentoring.

### 2.3.1 Health and nutritional services link (*HNSL*):

A very basic Health and Nutritional Services Link service will be provided by Pratham district coordinators in each of the 192 study villages, including the control group. The service will consist of a one-off one day visit (over the course of the two years of the project) of one district coordinator to each village, where she will mobilize child caregivers and village officials and discuss the availability and importance of public services (*other* than our intervention activities) available in the community, such as growth monitoring and food supplementation provided by Anganwadi Workers. The purpose of the *HSNL* service is to create an environment where the baseline policy framework is well understood.

### 2.3.2 Individual Stimulation (*IS*) Curriculum

The home visiting program evaluated in this study is based on the model and curriculum designed by Sally Grantham-McGregor for use in Jamaica, which was further developed and adapted by CECED. As described in Section 2.1 the Jamaican home visiting program 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 [10, 11, 12, 31]. Positive impacts have also been found when the program was adapted for other countries and contexts [4].

The core of the program is supporting the mother to promote her child's development using a 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 stacking blocks, doing puzzles and looking at books. The home visitor demonstrates the activities to the mother then helps her to do them giving feedback. The visitor then leaves the materials used in the visit with the mother and encourages her to continue the activities during the following week, before the next visit. At each visit the materials are exchanged for different play materials. In addition the mothers are encouraged to improve the quality of interactions with their child and use every day routine activities to chat and play with the child and teach them new words and concepts. 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 or other low-cost materials, often discarded objects such as empty plastic bottles (where available). Clearly, the original curriculum developed for Jamaica was not directly suitable for use in Odisha. In addition to translation it needed significant adaptations to make it as relevant as possible for poor rural households in Odisha. All the books were redrawn to reflect the local environment and local games and songs were included.

Some activities were included to increase the socio-emotional and gross motor content of the curriculum and to improve the mapping of the activities against current early learning standards for India.

### 2.3.3 Group Stimulation ($GS$) Curriculum

GS and IS differ in the mode of delivery. The Group Stimulation (GS) curriculum developed by CECED takes advantage of the group dynamics, interactions with other caregivers, and shared knowledge and experiences of child rearing. In contrast, the IS intervention delivery relies on the interaction between the home visitor, the caregiver and the child. The sessions start with free play to allow children to explore play materials and follow their interests and develop creativity. The group situation also allows certain socio-emotional activities such as sharing and taking turns for the older to be included. The group curriculum includes most of the activities in the individual curriculum and the key features and aims of the group session curriculum remain the same as those of the IS curriculum (e.g. structured curriculum, play and other developmental activities based on locally available or low-cost materials, focus on interaction between the mother and the child, and supporting mothers to promote their child's development).

The use of group sessions as a mode of administration, however, required substantial changes to the curriculum and implementation guidelines. For starters, the curriculum has to accommodate material suitable for children of different ages. Unlike the one-on-one individual home visit, each group session consists of 5 to 8 children, making it no longer feasible to tailor the activities to the exact age of each child. Instead, the curriculum describes activities for two groups: a 'younger group' (children aged 6-12 months at the start of intervention) and 'older group' (children aged 13-18 months at the start of the intervention)[2]. In case in a particular village a group consists of 'younger' children and 'older' children, the group is split into two and the group facilitator manages both groups simultaneously. It is clear that not only the curriculum but also the role of the group facilitator changes compared to that of the individual home visitor. In many ways the facilitation of a group session becomes more challenging. The group facilitator needs to find a venue that accommodates up to 8 caregivers and their children (which can be particularly difficult during the monsoon and hot season, when shelter is crucially important for smooth running of the session). She needs to manage group dynamics, such as encouraging all mothers to participate in group activities while preventing anyone from dominating the situation. Importantly, time management and making sure that all women and children remain actively involved, motivated and engaged is crucial for the session to be useful. The sessions start with free play to allow children to explore play materials and follow their interests and develop creativity. The group facilitator has to be flexible and innovative in her approach towards women and children. A lot of emphasis was put on these soft skills during training and will be closely monitored during implementation, especially during the early months of the intervention.

[2]Note that the curriculum allows for children slightly younger or slightly older than children in our target age range. This is to allow for material for children who are unusually more or less advanced than other children in their age category.

### 2.3.4 Nutritional Education (NE) Curriculum

The NE curriculum used in this study is the K.E.Y. (Knowing and Engaging for Young child food and feeding) curriculum, developed by Smriti Pahwa. It aims to create an interface for assessment as well as action around child nutrition at the household level. The key idea is to frequently involve the caregivers in looking at their young child's food and feeding along with other important domains such as immunization, growth monitoring and diarrhea-water-hygiene, using simple measurable and actionable indicators. This is achieved through (i) weekly 10 minute follow up sessions organised by home visitors or group facilitators at the home of the child and (ii) bi-monthly Focus Nutrition Sessions.

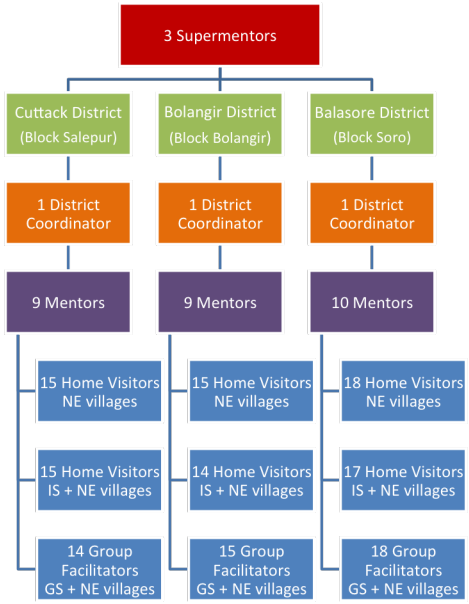
The weekly follow-up session is intended to sensitize child caregivers so that they understand the situation and take corrective action as needed at household level as well as in reaching out to the government provisions. The intention is to transform the caregivers from being beneficiaries of government provisions to becoming active stakeholders thereby creating bottom up traction and demand for services. The curriculum is designed as non-text heavy simple information booklets, which are used to orient the mentoring teams. In addition, the HVs/GFs use visual engagement formats to make the sessions as interactive as possible (e.g. story cards, recipe assessment cards, games and activity cards).

In addition to the weekly follow-up visits, bi-monthly intermittent Focus Nutrition Sessions (organized as individual home visits for the IS+NE and NE only treatment arms and organized in group for the GS+NE treatment arms) dedicate exclusive time for demonstration of regionally adapted low cost nutritious recipes for children, their quality assessment and other relevant discussions around important young child nutrition and health concepts.

## 2.4 Intervention implementation team

The intervention activities are delivered by a team of 175 locally recruited women from Odisha. Figure 2 gives an overview of the team structure: three supermentors, three district coordinators, 28 mentors and 141 home visitors and group facilitators, who each have a specific and important role to play over the course of two years of intervention. Pratham, the implementation partner, was responsible for recruiting the field staff. Many of them were recruited through Pratham's extensive network in the communities - many had previously been involved in other Pratham programs. This Section briefly describes the respective roles of each of these players and provides more details on their recruitment and selection procedure.

Figure 2: Intervention implementation team structure




### 2.4.1 Home visitors and group facilitators

The core of the work, the actual delivery of the curricula, is delivered by a team of 141 *home visitors* and *group facilitators*, local women who had no specific experience in child development, who were trained specifically to deliver the intervention (see Section 2.4.2 for more information about the trainings)[3]. Using paraprofessionals is an important design element of the program 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 Odisha who speak the local language 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, which can be particularly important in the context of rural India. This would suggest paraprofessionals might be better at encouraging families to take up the intervention and could be better at relating to the target mothers and making them more at ease.

Because the curricula 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 and group facilitator 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 and group facilitators having the background knowledge and experience needed for creating a successful home visit or group session without guidance.

[3]In general there is one home visitor or group facilitator per village, except for 4 villages in the IS treatment arm and two villages in the GS treatment arm where one HV/GF takes care of two adjacent villages.

The criteria for being eligible for selection as a home visitor or group facilitator were the same, as follows:

1. **Gender**: female
2. **Age**: Minimum 18 years
3. **Language proficiency**: Odiya
4. **Education**: 10th class completion preferable
5. **Residence location:** any one of the 144 treatment villages (max one per village)
6. **Experience:** No experience required

Table 1 summarizes some basic characteristics of the home visitors and group facilitators that were selected for the project.

Table 1: Characteristics home visitors and group facilitators

| | **NE** | **IS + NE** | **GS + NE** |
|---|---|---|---|
| Average age | 24 | 25 | 27 |
| Completed level of education (%): | | | |
| Completed 7-9th grade | 4 | 2 | 6 |
| Completed 10-12th grade | 73 | 85 | 70 |
| Completed Bachelor | 23 | 13 | 23 |
| Experience working with children (%) | 17 | 11 | 11 |
| Total number | 48 | 46 | 47 |

#### 2.4.2 Training, mentoring and monitoring

Home visitors and group facilitators were recruited, trained and (will be) supervised by 28 *mentors*. The selection criteria for the recruitment of the 28 mentors were as follows:

1. **Gender**: female
2. **Age**: Minimum 18 years
3. **Language proficiency**: Odiya
4. **Education**: 12th class completion preferable, 10th grade a must
5. **Residence location:** Commutable distance from the treatment villages but not from the control village
6. **Experience:** Prior experience of working with children preferable

Table 2: Characteristics mentors

| | |
|---|---|
| Average age | 25 |
| Completed level of education (%): | |
| Completed 10-12th grade | 21 |
| Completed Bachelor | 68 |
| Completed Master | 11 |
| Experience workig with children (%) | 43 |
| Total number | 28 |

Table 2 provides a summary of the characteristics of the mentors that were recruited.

The mentors, in turn, were trained and will be supervized themselves by three *supermentors* and by Pratham and senior staff members of the Centre for Early Childhood Education and Development at the Ambedkar University Delhi (CECED), all of whom have extensive work experience in the curricula of the intervention[4]. The three district coordinators, one in each of three district offices (Cuttack, Bolangir and Balasore), will be responsible for the implementation of the program in their own district. They will provide support and mentoring to the mentors in their district and will also take care of all administration and logistics involved with the project activities in their district (e.g. processing of monitoring data, accounting, etc.).

Supermentors, district coordinators, mentors, home visitors and group facilitators all received and will continue to receive intensive training in the curricula. Table 67 in Appendix A gives an overview of the time schedule of the completed and scheduled trainings and refresher trainings. By the end of the intervention in two years from now, each supermentor, mentor district-coordinator, HV and GF will have received around 30 (supermentors), 69 (mentors) and 40 (HVs/GFs/DCs) days of training in total, respectively. The home visitor and group facilitator trainings include toy making workshops. Home visitors that joined the program in case of replacements will be trained by the Pratham mentors at their convenience.

Given the novelty of various aspects of the project design, the project will rely on relatively intensive monitoring and supervision. District coordinators will meet with the mentors in their district at least once every month. Each mentor in turn will observe one home visit/group session per home visitor/group facilitator per week and provide feedback. In addition, each mentor will meet with each of the home visitors/group facilitators in her team on a bi-weekly basis to discuss the progress of each child, again providing an opportunity for continuous learning.

## 2.5 Our Evaluation Study

Our evaluation study will test the relative effectiveness of the following three variants of an early child development program delivered over 24 months, each of which were discussed in more detail in Section 2.3:

1. Nutritional Education intervention (NE)

[4] All of the trainers had themselves played a crucial active role in the Cuttack pilot project.

2. Nutritional Education + Individual Stimulation (NE + IS)

3. Nutritional Education + Group Stimulation (NE + GS)

In addition to assessing impacts on child development, we aim to study *how* these interventions affect child development by looking how these affect investments made by families into young children. The full list of outcomes we intend to study are detailed in Section 3.4. 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.

# 3 Methodology

## 3.1 Terminology

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

1. **Target children (TC)** are the key subjects of interest in the study. To have been selected into the study sample target children must meet full eligibility criteria: (1) being **between 7 and 16 months** at the start of the intervention, (2) residing in one of the 192 study villages (if temporarily absent, planning to return within six months), (3) not being a twin and (4) not having a physical or mental disability. Due to capacity constraints, not all eligible children in our study villages could participate in the study. Section 3.3.3 discusses how target children were selected amongst all eligible children in each village (at least 5 and maximum 8 children per village). The evaluation design will directly compare the outcomes of the target children in the intervention villages with those in the control after the end of the intervention to measure direct program impacts.

2. **Spillover children (SC)** are children that just fell short of meeting the target child eligibility criteria and are therefore not eligible for participation in the intervention. In particular, they meet the following criteria: (1) being **between 2 and 6 months or between 17-20 months** at the start of the intervention, (2) residing in one of the 192 study villages (if temporarily absent, planning to return within 6 months), (3) not being a twin and (4) not having a physical or mental disability. Our interest in spillover children stems from the hypothesised likelihood of observing *indirect* impacts of the intervention, 'spilled over' from target children on children who did not directly participate in the intervention but who are otherwise fairly similar to the target children. Section 3.3.4 explains how the sample of spillover children was selected for the study.

3. **Biological mothers** of the target and spillover 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.

4. **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 (184 out of 2170, or 8%).

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

6. **Villages** are the units at which randomization of the intervention treatment arms took place. We consider 192 study villages in three districts in Odisha: Bolangir, Cuttack and Balasore.

7. **Home visitors** are the 94 women identified and employed by Pratham, the implementing partner, to deliver the interventions that are based on home visits as the delivery mode. As described in Section 2.3, there is one home visitor for each one of the 48 *NE* villages and there is one home visitor for each one of the 48 *NE + IS* villages, except for two pairs of adjacent villages which share one home visitor between two villages.

8. **Group facilitators** are the 47 women identified and employed by Pratham, the implementing partner, to deliver the interventions that are based on group sessions as the delivery mode. As described in Section 2.3, there will be one group facilitator for each one of the 48 *NE + GS* villages, except for two adjacent villages which share one group facilitator.

## 3.2 Evaluation design

We are using a cluster randomized controlled trial (RCT) to evaluate the relative effectiveness of the three interventions outlined in Section 2.3 - *NE, NE + IS, NE + GS* - with the unit of randomization being the village.

### 3.2.1 Randomization

The 192 villages chosen to be part of the study (see Section 3.3 for details of sample selection) were randomly allocated to one of four possible treatment[5] states:

1. *NE* treatment: villages receiving the NE intervention (Section 2.3.4)

2. *NE + IS* treatment: villages receiving the ECD home-visiting stimulation intervention (Section 2.3.2) in combination with the nutritional education intervention (Section 2.3.4)

3. NE + GS treatment: villages receiving the ECD group-based stimulation intervention (2.3.3) in combination with the nutritional education intervention (2.3.3)

4. Control: villages receiving none of these interventions

Note that all villages, including the control villages, also receive a very basic one-off health and nutritional services link *(HNSL)* - see Section 2.3.1 - to ensure that existing public services were well understood.

Randomizing is vital for our evaluation approach since it ensures that the children and households in the treatment and control groups are, 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 four groups after the intervention can be attributed to the effects of the intervention. If we were to select which villages were to receive a particular intervention by some other means, for example selecting the poorest villages, then the four groups would look different in terms of

[5]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 do not receive any active treatment.

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 four 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 randomization.

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

1. over the course of the intervention efforts will be made to strengthen engagement with village officials, who play a key role in implementing government and state policies related to early childhood development;

2. randomizing 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 program through information or resources diffusing through social networks;

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

We used two levels of stratification. For one, we stratified our randomization of treatment status by district and second, on the number of children in the target age range in the village. For this latter level of stratification, we chose two strata: (i) villages that had more than eight children in the target age range (at that time 10-20 months) and (ii) villages that had fewer children in the target age range. Amongst the villages with more than eight children in the target age range we were only offering the intervention to a random eight children (details of this selection are given in Section 3.3.3), whereas in the smaller villages we were offering the intervention to all children. The reason for this approach is that our budget allowed us to employ only one home visitor per village and it was considered that nine to ten children is the maximum number of children a home visitor could work with without being stretched. Thus, anticipating that the work load of the home visitor might affect the effectiveness of the intervention, stratifying the randomization on this dimension will increase balance.

The experimental design is summarized in Table 3.

It is important to note that while the randomization happened simultaneous to the sample selection (see Section 3.3) all field staff working on the census and baseline survey did not know the result of the randomization. Therefore no selection bias could have been introduced through this.

Table 3: Random allocation of study villages to evaluation arms, by district

| | **Control** | **NE** | **NE + IS** | **NE + GS** | **Total** |
|---|---|---|---|---|---|
| Cuttack | 15 | 15 | 15 | 15 | 60 |
| Bolangir | 15 | 15 | 15 | 15 | 60 |
| Balasore | 18 | 18 | 18 | 18 | 72 |
| **Total** | **48** | **48** | **48** | **48** | **192** |

### 3.2.2 Evaluation specification

At midline (after 12 months) and at endline (after 24 months) 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 3.4), as well as supporting data on the household characteristics. For each outcome of interest we will estimate the relative impact of eligibility for the three different programs by running a regression of the following form:

$$y_{ij} = \alpha + \beta NE_j + \gamma IS_j + \delta GS_j + \eta X_{ij} + \nu_j + \epsilon_{ij} \tag{1}$$

where $y_{ij}$ is the outcome of interest for household (or child) $i$, in village $j$, $NE$ is a dummy variable equal to one if village $j$ received the nutritional education intervention (which is the case for each of the treatment arms) and equal to zero otherwise (control), $IS$ is a dummy variable equal to one if village $j$ was allocated to the treatment group receiving the home visiting stimulation intervention and equal to zero otherwise, $GS$ is a dummy variable equal to one if village $j$ was allocated to the treatment group receiving the group-based stimulation intervention and equal to zero otherwise, $X_{ij}$ is a vector of observed household, and village level, characteristics measured at baseline (including the baseline measure of the outcome of interest), $\upsilon_j$ is a cluster-specific unobserved effect and $\epsilon_{ij}$ is a random error term. Note that the error term, $\epsilon_{ij}$, cannot be assumed to be independent between households (or children) since households living in the same villages 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 village, allowing for arbitrary correlation between error terms of households in the same village.

In this regression framework parameters of interest are $\beta$, $\gamma$ and $\delta$. It is the size and significance of these parameters that will tell us the relative impacts of the different interventions on the outcome of interest and the degree of uncertainty associated with that estimate. Parameter $\beta$ measures the impact of participation in the nutritional education ($NE$) program. Parameters $\gamma$ and $\delta$ measure the additional impacts achieved by adding, respectively, the individual home stimulation ($IS$) program and the group-based stimulation program to the $NE$ program.[6]

[6]The alternative to using post-treatment data only approach is to 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.

It is important to note that this set-up is an *Intention to Treat* analysis - we are estimating the impact of being in a village that was allocated to a particular treatment group, and thus that the child was eligible for the respective programs. This may be different from the impact of actually receiving the intervention if some households decide not to participate in the program even though they were eligible, for example if they perceived the program 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 randomization. Secondly, the *unconditional* effect of a household being offered a program is arguably more useful from a policy perspective than the effect *conditional* on choosing to participate. The unconditional effect of being offered a particular program 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 effects $\beta$, $\gamma$ and $\delta$ 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 randomization was successful. In this report we check that the four 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 across the different treatment groups.

## 3.3 Sampling selection strategy

### 3.3.1 Selecting villages

Initially we had planned to work in 225 villages in 3 districts in Odisha, India: Bolangir, Cuttack and Boudh. These districts were selected in close collaboration with Pratham, taking their operations, as well as the desire to cover distinct areas in Odisha, into account.

When Pratham started their operations in these districts, they selected one block and within this block 100 villages to work in. Exact selection criteria have not been recorded at that time, but it is still known that issues around operationalization of the programs formed the basis for the choice of the localities.

We took these lists of 100 villages per block and matched them with official sources to get an understanding of where and how close they are located as well as how many pregnant women, babies and children we could

expect. We aimed to have 5-9 target children in each village and wanted to also ensure that villages would not be too close to minimize risk of spillover. Where secondary information on their GPS coordinates and expected number of pregnant women was not available[7], JPAL-SA staff went to the field to collect this information first-hand.

This exercise led to a number of observations. For one, we learned that Pratham is covering about 4-5 villages in ~70 gram panchayat's (GP) within the three selected districts. This implies that they are working in, on average, about one third of the villages within a GP. Importantly, we found that the average number of pregnant women and infants in the villages was lower than expected, particularly in Boudh. Therefore, the following decisions were taken around May 2015:

- To replace Boudh with a different district, which was decided to be Balasore;
- To extent the targeted age range at the start of the intervention 7-16 months (originally this was set to be 9-15 months);
- To cut the size of the control group; in Cuttack and Bolangir this had to be cut to 15 villages, that is, the same size as the treatment arms, and in Soro we were able to obtain 18 villages in the control group;
- Not to increase the average number of interviewed children per village in order to keep the same sample size in terms of children as planned in the proposal. The power analysis indicated that the gain would not be sufficient to make it worth it.

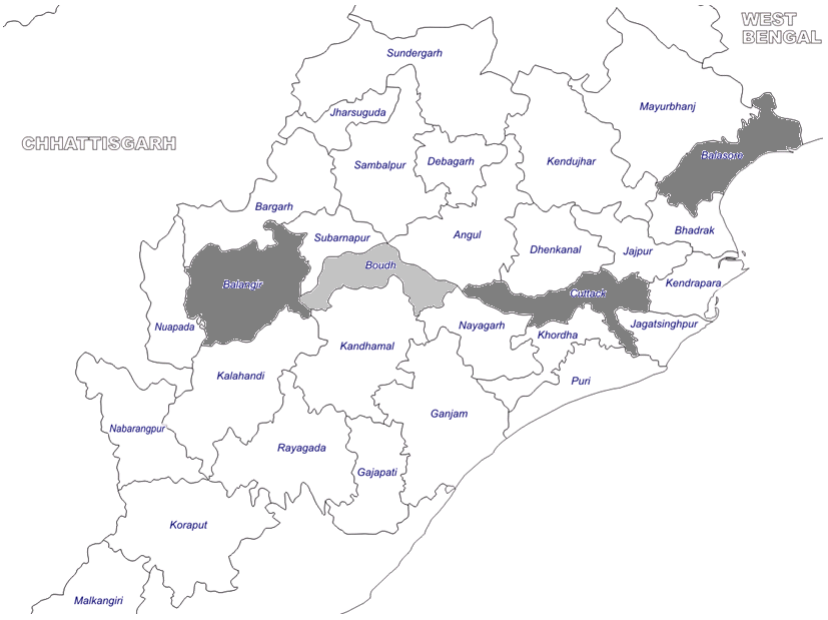
We note that, we had conducted the village randomization before some of these decisions had been made. The reason for such an early randomization was that Pratham needed the information for planning purposes. The process we had followed was to randomize more than the required sample size to the treatment arms. With more information on number of target children in these villages, we would then have dropped the smallest villages within each treatment allocation and district to get our required 15 villages per treatment arm, per district.

We decided to conduct a re-randomization for the following reasons:

- We had replaced Boudh with Balasore and still needed to conduct the randomization for Balasore;
- We had to drop 10 villages in Balangir as the residents were planned to be relocated due to a dam being built (by chance 6 of the 10 villages were in our control and we would therefore not have had enough villages with sufficient target children in this stratum);
- And, Pratham had not explicitly used the list at that point.

[7] We consulted different websites, including www.orissa.gov.in/panchayat/villagelist.xls and data on Birth to Three in Odisha villages, http://www.censusindia.gov.in/vital_statistics/AHSBulletins/AHS_Factsheets_2011_12/OdishaFactsheet_2011-12.pdf

Figure 3: Study area




*Source: http://www.d-maps.com*

### 3.3.2 Village mapping and census

An important component of the study was the implementation of a census, the purpose of which was twofold:

- To serve as a sampling frame for the selection of respondents for the baseline survey (see Section 3.3.3 and Section 3.3.4) - since the required information was otherwise not available. It was important to have a comprehensive and up-to-date sampling frame to ensure we would have a random sample of target and spillover children - including households that might have moved in recently, and covering all demographic groups;

- To collect GPS data that allow us, among other: (i) to check that all households interviewed are indeed situated within the selected villages, (ii) to make it easier to identify panel households during the follow-up survey, (iii) to calculate distances between neighbours which is required for the selection of respondents (see below), (iv) to construct social network measures to assess the degree of spillover effects across treatment arms.

First, each village 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.

Next, census teams carried out a digital door-to-door survey with every household in the village. 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, caste information

and the household's intention to remain in the village over the duration of the study. In addition, GPS coordinates were collected.

Appendix B provides details on the method we used to determine the age and TC/SC eligibility of the children in our study. In total in the 192 study villages we identified 2181 eligible target children (aged 7-16 months at start of intervention, living in the 192 study villages - if temporarily absent, planning to return within six months, excluding twins and children with physical or mental disabilities) and 1751 eligible spillover children (aged 2-6 or 17-20 months and otherwise similar inclusion criteria as target children).

### 3.3.3 Sample of target children

From the pool of all 2181 eligible target children, we obtained a baseline sample of 1427 target children - on average 7.5 per village - through the following procedure:

- If there were eight or fewer children in the target age range in a village, identified in the census above (and meeting the full inclusion criteria listed above), then we aimed to include all of them in our study (this was the case in 64 out of 192 villages).
- If in a village there were more than eight children in the target age range identified in the census above (and meeting the full inclusion criteria listed above), as was the case in 128 villages, we did the following:
    1. We (temporarily) dropped observations with obviously erroneous GPS coordinates (accuracy $>$ 1000 or recording $>$ 5 km from other villages) assuming that GPS errors happen at random;
    2. We randomly selected one *central child*, that is, a child in the village for which the 7th nearest eligible child lives less than 700 meters away;
    3. We allocated this randomly selected central child and her 7 nearest neighbours to the sample of target children in our study;
    4. By repeating this procedure we could select target children in such a way that the 7th nearest eligible child lives more than 700 meters away in only three villages, making the intervention logistically easier to implement;
    5. We allocated all eligible children that had not been selected as a target child to a reserve pool (capped at eight) in order of distance from the central child, which could be use in case one of the originally selected target children was not available.
    6. Finally, we added children with erroneous GPS coordinates (who had been dropped in the first step) to the bottom of the reserve list in case the cap of eight reserve children had not yet been obtained.
- This process resulted in an initial list of 1449 target children who we aimed to include in the study. This is 237 fewer than we had originally planned for (225*7.5). As discussed above, when the decision

to decrease the number of villages was made, it was also decided not to increase the number of children per village, given that this would again have affected the number of villages we could include in our sample. Power analysis indicated that the reduced number did not affect the sample to a worrying degree. The target child reserve list we obtained consisted of 732 children.

- We attempted to collect baseline data from all 1449 selected target children. However, in 206 cases (14%) no baseline data could be collected due to reasons that will be discussed in Section 3.6. These children had to be dropped from the sample. To replace them, we turned to the list of replacement children. In case there was not a sufficient number of replacement children available we approached children within the target age range that had migrated into the village after census (and therefore were not present in the original sample drawn from census).
- This gave us a total final baseline sample of 1427 target children (7.5 children per village on average).

#### 3.3.4 Sample of spillover children

In addition, from the pool of all 1751 eligible spillover children in our study area we obtained a baseline sample of 743 spillover children, using the following procedure:

- For each village we ordered the list of all eligible children as follows, from top to bottom (within each category giving priority to children living closest in distance to the central target child - see Section 3.3.3 for definition of a central child):

  1. A maximum of three 5-6 months old spillover children (depending on availability);
  2. A maximum of two 17-18 months old children (depending on availability);
  3. All other 5-6 months old, if any;
  4. All other 17-18 months old, if any;
  5. All 4 months old, if any;
  6. All 19 months old, if any;
  7. All 20 months old, if any;
  8. All 2 months old, if any;

  Note that priority in this order was given to children closest in age and geographic space to target children.

- For each village, we then sampled the first four children on this list for baseline survey and all others as reserve spillover children (capped at eight reserves), keeping the order defined above by which to approach the reserve children when necessary. Since we were 32 spillover children short of our target (from villages where there were less than four eligible spillover children) we chose nearest five spillover children in 33 villages (all villages where the total number of eligible spillover children was greater than 20).

- This resulted in an initial list of 767 spillover children who we aimed to include in the study. This is 133 fewer than we had originally planned for (225*4). We also obtained a list of 984 children that were included in the reserve list of target children to be used in case the originally sampled children were not available.
- We attempted to collect baseline data from all 767 selected spillover children. However, in 153 cases (20%) no baseline data could be collected due to reasons that will be discussed in Section 3.6. These children had to be dropped. To replace them, we turned to the spillover child reserve list. In villages where there was not a sufficient number of replacement spillover children we also considered children that had recently moved into the village (and therefore had not been considered during census).
- This gave us a total of 743 spillover children - on average 4 per village - for which we have baseline data available.

## 3.4 Outcome indicators

We are interested in how the psychosocial home visiting and group-based stimulation programs and the nutritional education program affect outcomes that fall into two categories: (1) indicators of "indicators of child development and child 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 between these two categories. The first category of outcome indicators evaluate *whether* the interventions achieve their intended aim of improving levels of child development. The second category of indicators allow us to analyse *how* (if at all) the programs affected child development. We call them indicators of child development inputs. They allow us to look at the mechanisms through which the programs work. For example, they might suggest that the home visiting stimulation program 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 program being effective. Further work, following on the impact evaluation, will involve using these 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 programs in particular populations to other populations and differently designed programs.

In the list below we also highlight outcome indicators that we label as primary and secondary (in accordance with the trial registry, registered on ISRCTN, study ID ISRCTN18111205).

### 3.4.1 Indicators of child development (primary outcomes)

To assessthe overall impact of the program on cognitive, communication, fine motor, gross motor and personal-social and language, we use two main measures: (i) the Bayley Scales of Infant and Toddler Development, Third Edition [5] - which are widely taken to be the gold standard of cognitive development for children under 42 months - and (ii) an adapted (see Section 6.2 for details of adaptation) version of the Ages and Stages Questionnaires, Third Edition [27], hereafter referred to as the ASQ-3.

At baseline, however, given budgetary and time constraints we have only used the ASQ-3. The ASQ-3 are a series of questionnaires to be completed by parents of young children about their child. Compared to the Bayley it has less items, a higher proportion of items are reported by the caregiver and it can be administered in the home. 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 [6] had a concurrent validity with a range of standardised measures was 85%[28]. It also displayed high test-retest reliability, inter-observer reliability, and internal consistency[28]. The test has subsequently been used in many contexts and languages across the world.

Personal social skills, which 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), were assessed at baseline using the Personal Social scale on the ASQ-3 [27] and. . motor development through two subscales of the ASQ-3[27] - the Fine Motor subscale and the Gross Motor subscale.

We further assess child health 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.

### 3.4.2 Indicators of child development inputs

**Quality of the home environment (Secondary outcome)** Poor stimulation within the home environment impacts negatively on many developmental domains [32]. 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 [15].

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.

**Child nutrient intake, healthcare and birth (Secondary outcome)** Good quality nutrition and access to healthcare services, both preventative and when sick, is crucial for children to grow up healthy. To

measure nutrition we asked 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 village level data we also know the type of healthcare services available in the village 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.

**Knowledge of child development (Secondary outcome)** 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 [21]. 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 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.

**Quality and quantity of maternal time (Secondary outcome)** 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 3.4.2). 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 program can increase stimulation within these constraints.

**Maternal and main caregiver wellbeing and education** 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 [25], 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.

**Economic resources** 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.

## 3.5 Instruments for data collection

The baseline data described in the previous section were collected through a series of questionnaires, divided into nine modules which addressed different members of the household on different topics. Table 4 provides an overview and brief description of these modules, along with the designated respondent and the average duration of each module[8].

[8]Non-sensical values for survey duration (e.g. negative and extreme outliers) - between 5-10 per module - were set to missing and excluded in the average survey duration calculations.

Table 4: Modules administered during baseline survey

| Module | Description | Respondent module | Averageduration |
|---|---|---|---|
| Household roster | Age verification, respondent identification and information on each household member, covering gender, age, biological mother and father, ability to read and write, school attendance, highest grade completed | Household head or most knowledgeable person | 22 |
| Household module | Topics covered: Workforce (labour supply of main breadwinner, biological mother and father), health of target child and its biological parents, information on the dwelling, caste, food preparation and fuel used, water and sanitation, prices, property and durable assets owned, savings, loans/debts, expenses, income and transfers, shocks experienced | Household head or most knowledgeable person | 72 |
| Mother and child diet tool | Captures child's dietary variety including caregiver's breastfeeding practice | Mother/primary caregiver of TC/SC | 20 |
| Target child | Interaction with parents, birth, lactation and feeding, handling of child feaces, immunisation, growth monitoring, morbidity, care-taking of the target child, expectation on its schooling, household environmental quality scale | Mother/primary caregiver of TC/SC | 46 |
| Biological mother | ASER literacy tool, maternal beliefs on involvement/stimulation/interaction, use of time, beliefs about feeding practices, knowledge of child development, wellbeing, social networks, empowerment, reproductive and contraceptive history, depression | Biological mother of TC/SC | 61 |
| Primary caregiver | family details, education and workforce, ASER literacy tool, beliefs on involvement/stimulation/interaction, use of time, beliefs about feeding practices and handling of child faeces, knowledge of child development | Main caregiver of TC/SC who is not the biological mother | 41 |
| Anthropometrics | Weight and height of both child and mother was collected following a strict protocol. The scales used were Seca 874. | Biological mother of TC/SC and TC/SC | 20 |
| ASQ | The ASQ-3 consists of five subscales, each measuring a different child developmental domain - problem solving, communication, personal-social, fine motor and gross motor. | Mother/primary caregiver TC/SC | 31 |
| Community | Asked about the location, population and infrastructure of the village; public services; transport and distances; health institutions located in the village; childcare institutions located in the village; educational institutions located in the village; commercial activity in the village and; social programs in the village. | Anganwadi worker (The most experienced anganwadi worker in case there are more than one anganwadi workers in the village) | 58 |

## 3.6 Data collection, sample size and data structure

### 3.6.1 Time line

Data collection in Cuttack, Bolangir and Balasore took place between 31 August 2015 and 19 December 2015. Given the relative remoteness and different profile of Bolangir compared to Cuttack and Balasore, it was decided to hire and train two separate survey teams: one team to conduct survey work first in Cuttack and then in Balasore, and another team to focus on Bolangir. Table 5 provides information on the timelime of the baseline data collection process, as per the planned and actual schedule[9].

[9]The survey teams experienced slight delays caused by lower than expected productivity in Cuttack in some of the survey modules which was resolved by training additional surveyors in Bolangir (where training started later) to join the Cuttack team.

Table 5: Timeline

| District | Start date data collection | | End date data collection | | Start date intervention | |
|---|---|---|---|---|---|---|
| | Planned | Actual | Planned | Actual | Planned | Actual |
| Cuttack | 31.8.2015 | 31.8.2015 | 18.10.2015 | 14.11.2015* | 23.11.2015 | 2.12.2015 |
| Bolangir | 5.10.2015 | 6.10.2015 | 5.12.2015 | 4.12.2015 | 7.12.2015 | 16.12.2015 |
| Balasore | 19.10.2015 | 28.10.2015 | 20.12.2015 | 19.12.2015 | 21.12.2015 | 11.01.2016 |

* Most of the data collection in Cuttack finished on the 19th of October. However, as few households had not been found during the main survey, some surveyors returned to the respondents' households on the November 13th and November 14 to complete the surveys

### 3.6.2 Survey team structure

The survey team in Cuttack/Balasore consisting of 25 surveyors and 3 supervisors whereas the team in Bolangir consisting of 17 surveyors and 3 supervisors. Since the modules administered in the survey were quite extensive surveyors were split into four survey teams, each separately trained and responsible for the administration of a particular set of module(s). The splitting of the modules was based on the time required to complete each module, the respondent, logistical constraints and the nature of the questions in each module. Table 6 shows how the modules were assigned to the four different survey teams.

Table 6: Modules by survey team

| | **Team 1** | **Team 2 (6-8 surveyors)** | **Team 3 (2 surveyors)** | **Team 4 (6-8 surveyors)** |
|---|---|---|---|---|
| Modules | Household roster<br>ASQ-3 | Household module<br>Diet tool<br>Community | Anthropometrics | Target child<br>Biological mother<br>Primary caregiver |
| No. of surveyors Balasore - Cuttack | 6 | 8 | 2 | 8 |
| No. surveyors Bolangir | 5 | 6 | 2 | 6 |

This survey structure meant that each household received four visits in total - one for each team - usually within a one week interval (except when the household was not available). Generally, the visits were organized as follows: Team 1 visited the household on day 1, Team 2 and Team 3 followed suit on day 2 whereas Team 4 visited the household on day 5 of the field work in a particular village[10]. The spacing was to allow the respondents rest time between surveys and to minimise refusals in between visits. Whereas Team 1 and Team 4 consisted of female surveyors only, team 2 was mixed. The anthropometrics module was completed by a team (Team 3) of two surveyors, 1 male and 1 female surveyor [11]. However, the team received support from members of Team 2 and Team 4 (who were also trained in anthropometrics measurement) whenever they were lacking behind in terms of productivity.

[10] At the start of the data collection, in Cuttack, the order of the visits was different but on 15th of September 2015 the order was changed to accommodate different productivity levels of the different teams.

[11] Anthropometrics was put into its own group because the size of the equipment made it cumbersome for field staff to walk with in the field. By having a pair on a bike it was easier to move the equipment between houses in the villages. Having one specialised anthropometrics team also increased the efficiency and accuracy of the administration of the module because the team gained a lot of practice.

### 3.6.3 Survey staff training

Training of the Cuttack/Balasore team took place in Cuttack and training of the Bolangir team took place in Bolangir. The training of Teams 2-4 was led by JPAL-SA and took 17 working days to complete in Cuttack (10 August - 30 August 2015) and 11 days in Bolangir (10 September - 4 October 2015). Team 1, i.e. the ASQ testers, received special training by CECED experts in the ASQ-3 questionnaire. This training took 11 days to complete in Cuttack (17-28 August 2015) and 11 days in Bolangir (21 September 2015 - 2 October 2015).

### 3.6.4 Sample size

As discussed in Section 3.3, after census a random sample of 2216 children was selected for baseline survey and 1716 children were selected as backup replacement children. When a targeted child could not be surveyed (see below), the surveyors were instructed to replace the child (following approval from the supervisor and the data collection project manager) with a child from the same category (TC/SC) from the reserve list of children. In particular, surveyors would look for replacement children when:

- After verification the sampled children did not (or no longer) meet the eligibility criteria (e.g. twins, child with permanent disability, unborn child);
- After age verification, sampled children turned out to have a different age then the one collected during census.
- Children were dead, unavailable or shifted permanently to some other place and not likely to return;
- Household members refused to participate in the survey (at the beginning or in between different modules);

Replacement households approached were sometimes not available for interview either. Table 7 reports on the reasons for why 436 households of all children approached for survey (originally sampled and replacement children) could not be surveyed . The most common reason for some households' non-participation was related to temporary or permanent migration of households since census (e.g. visit to maternal home, seasonal labour, migrant labour). Another important reason was re-categorization of children from Target Child category to Spillover Child category and vice versa after age verification (see below).

A total of 46 children of the original list of 2216 sampled children could not be surveyed nor could be replaced due to lack of sufficient available replacement children in some villages. Table 8 gives an overview (by district) of the number of TC and SC children originally targeted for survey (see Sections 3.3.3 and Section 3.3.4), the number of these children that was unavailable or no longer eligible for survey, the number of replacement children surveyed as well as the total number of children who were eventually surveyed.

Table 7: Reasons for non-participation of households to baseline survey (sampled + replacement)

| **Reason** | **Bolangir (%)** | **Cuttack (%)** | **Balasore (%)** | **Total (%)** |
|---|---|---|---|---|
| Change SC/TC category | 30 | 14 | 28 | 23 |
| Change eligibility criteria (twins, outside age range, disabled) | 1 | 3 | 2 | 2 |
| Temporarily migrated | 32 | 37 | 34 | 35 |
| Permanently migrated | 17 | 20 | 18 | 19 |
| Refusal (from beginning) | 0 | 11 | 7 | 6 |
| Refusal (between some modules) | 1 | 9 | 5 | 5 |
| Mortality | 11 | 3 | 3 | 5 |
| House not found | 7 | 3 | 1 | 3 |
| House locked | 0 | 1 | 1 | 1 |
| Total number of households approached for survey | 112 | 161 | 163 | 436 |

Table 8: Baseline sample size

| **District** | **Category** | **No. children originally sampled for interview** | **No. of unavailable sampled children** | **No. of surveyed replacement children** | **Total number of surveyed children** |
|---|---|---|---|---|---|
| Cuttack | Target children | 468 | 89 | 90 | 469 |
| | Spillover children | 248 | 48 | 35 | 235 |
| Bolangir | Target children | 434 | 51 | 41 | 424 |
| | Spillover children | 230 | 45 | 40 | 225 |
| Balasore | Target children | 547 | 66 | 53 | 534 |
| | Spillover children | 289 | 60 | 54 | 283 |
| Total | Target children | 1449 | 206 | 184 | 1427 |
| | Spillover children | 767 | 153 | 129 | 743 |

Note that in some cases there was more than one sampled child in a household that was surveyed. These children were either siblings or cousins but not twins. This means that the total number of households surveyed is smaller than the total number of children surveyed. Table 9 summarises this sample information.

Table 9: Baseline sample structure

| | **Bolangir** | **Cuttack** | **Balasore** | **Total** |
|---|---|---|---|---|
| No. of households with 1 sampled TC/SC child | 645 | 687 | 789 | 2021 |
| No. of households with 2 sampled TC/SC children | 2 | 7 | 14 | 23 |
| No. of households with 3 sampled TC/SC children | 0 | 1 | 0 | 1 |
| Total number of TC/SC children | 649 | 704 | 817 | 2170 |
| Number of households | 647 | 695 | 803 | 2145 |
| Number of villages | 60 | 60 | 72 | 192 |

Finally, Table 10 shows the distribution of the sample across the different treatment arms and Table 11 shows the distribution per age category.

### 3.6.5 Data structure

Whereas the total sample consists of 2170 children, for some children some modules could not be entirely completed due to temporarily unavailability of the household. As the survey had to be wrapped up and there

Table 10: Sample size by treatment arm

| | Control | NE | NE + IS | NE + GS | Total |
|---|---|---|---|---|---|
| Target Children | 353 | 353 | 370 | 351 | 1427 |
| Spillover Children | 179 | 184 | 190 | 190 | 743 |
| **Total** | **532** | **537** | **560** | **541** | **2170** |

Table 11: Sample size by age child (months)

| | 2-4 | 5-6 | 7-8 | 9-10 | 11-12 | 13-14 | 15-16 | 17-18 | 19-20 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Control | 24 | 87 | 55 | 62 | 94 | 76 | 64 | 56 | 14 | 532 |
| NE | 23 | 73 | 60 | 81 | 74 | 74 | 64 | 71 | 16 | 540 |
| IS + NE | 24 | 79 | 67 | 62 | 96 | 90 | 56 | 69 | 16 | 559 |
| GS + NE | 25 | 73 | 58 | 68 | 94 | 68 | 65 | 70 | 19 | 536 |
| **Total** | **96** | **312** | **240** | **273** | **358** | **308** | **249** | **266** | **65** | **2167** |

was no time to replace that child and collect data, those few households were still considered part of the final sample.

Table 12 summarises the structure of missing data for the 2,170 target children.

Table 12: Baseline data structure

| HH Roster | HH | TC | BM | PCG | ASQ | Anthro | DT | Comm | Number of cases | Notes |
|---|---|---|---|---|---|---|---|---|---|---|
| ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | 1,979 | No missing data (PCG not required) |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 170 | No missing data (PCG required) |
| ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | 9 | Missing anthropometrics module |
| ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | 4 | Missing diet tool module |
| ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | 2 | Missing anthropometrics and primary caregiver module (PCG required) |
| ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | 2 | No missing data (no biological mother module required) |
| ✓ | ✓ | ✓ | ✓ | | ✓ | | | ✓ | 1 | Missing anthropometrics and diet tool module |
| ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | 1 | Missing household module |
| ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | 1 | Missing household and diet tool modules |
| ✓ | | ✓ | ✓ | | ✓ | | | ✓ | 1 | Missing household, anthropometrics and diet tool modules |

Note: HH = Household; TC = Target Child; BM = Biological Mother; PCG = Primary Caregiver; DT = Diet Tool; Comm = Community

In addition to the structure of missing data detailed above, some observation counts will vary due to missing observations from field errors, specific refusals or special circumstances. For example, an interviewer

may have missed a section, or been interrupted at some point during the interview process. These variations and omissions are small and insignificant to the results portrayed.

### 3.7 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 of child development and its determinants in a rural 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 groups prior to the intervention starting. As discussed in Section 3.2 our evaluation methodology is based on the comparison of outcomes for children and households in villages allocated to receive a particular treatment 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 groups are similar in all respects, other than treatment. Since we randomized which villages were allocated to the intervention groups and which to the control group we expect that this will be the case - we know that it will be *in expectation.* The randomization removes sample selection bias so that, in theory, the only difference between the intervention and the control groups is eligibility for the stimulation and/or nutritional education program thus meaning any differences we observe between the two groups can be attributed to the program. This means we can estimate an unbiased effect of the programs on all outcomes of interest.

Balance tests rely on statistical probabilities of observing any differences between groups if in reality there is zero difference. If this probability is lower than 10%, then we conclude that there is a statistically significant difference between groups. However, as the number of comparisons increases (for instance, if you make comparisons over many variables, which is what we do in this report), it becomes more likely that the groups being compared will appear to differ in terms of at least one attribute. For example, if one test is performed at the 10% level, there is only a 10% chance of incorrectly concluding that there is a difference even if there is no true difference between two groups. However, for 100 comparisons, the expected number of incorrect rejections is 10. We therefore expect (and therefore accept) there to be around 1 statistically significant difference for every 10 comparisons we make.

At this stage in the evaluation, we check that the randomization did, indeed, give us a balanced allocation of treatment and control - i.e. four treatment/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 randomization 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 different treatment groups. This is what we formally test in this report. For all variables we report in this report we compare the mean values for each of the three intervention groups to those of the control unit. We conduct two tailed hypothesis

tests to see if any differences in mean values we observe are statistically significant at conventional levels[12]. We also report results of joint F-tests which test whether there any significant differences across any of the treatment/control groups. In all our analysis we allow for arbitrary correlation in unobservables for all units within the village (cluster) by using cluster robust standard errors.

The rest of the report proceeds as follows. We start in Section 4 by describing the communities in which the households and children in our study sample reside to better understand their living conditions (i.e. principal industries, sanitation, electricity and access to commercial, social, health and educational services). In Section 5 we characterize the households within those villages, that the children form part of, in terms of their religion, caste, household head's demographics and educational background, household composition and so forth. We also study these households' economic profile, proxied for instance by their ownership of ration cards, amenities, earnings (overall and separately for different household members) ownership of assets, consumption expenditures and debt and savings patterns. Section 6 then focuses on the profile of the children in our study sample. We first present the baseline results on our key child development outcomes of interest in this study (cognitive, non-cognitive, health and morbidity). This section also discusses various factors that we believe feed into child development outcomes such as feeding practice, maternal health, education and knowledge about child caring practices and other key aspects of child's home environment (e.g. quality and quantity of care).

[12]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 in fact 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).

# 4 Baseline data – Community characteristics

Before describing the households and children in our study sample, we briefly describe some key characteristics of the communities these individuals are living in. Note that the descriptive statistics summarized in this section are derived from information reported by the Anganwadi Worker in the community questionnaire. They are not based on information provided by the villagers themselves and may therefore not be entirely representative.

Throughout the remainder of the text, when describing the summary statistics provided in the tables that follow, we will focus on the statistic that applies to the control group, whilst highlighting differences with other groups in case these differences are statistically significant.

## 4.1 Most important industries

The large majority of households living in our study communities are either engaged in agriculture (as a farmer or agricultural labourer) or in non-agricultural labour. Table 13 shows the responses to the question of which are the main three activities that households are engaged in within the respective community.

Table 13: Percentage of villages with a given industry as one of their top three most important

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Farming | 97.92 | -8.33* | -6.25 | -0.00 | 192 | 1.59 |
| | (14.43) | (4.92) | (4.54) | (2.95) | | |
| Agricultural labour | 58.33 | -0.00 | 6.25 | 2.08 | 192 | 0.17 |
| | (49.82) | (10.17) | (10.02) | (10.13) | | |
| Non-agricultural labour | 95.83 | -4.17 | -6.25 | -4.17 | 192 | 0.58 |
| | (20.19) | (4.97) | (5.32) | (4.97) | | |
| Poultry | 6.25 | -0.00 | 4.17 | 4.17 | 192 | 0.36 |
| | (24.46) | (4.99) | (5.69) | (5.69) | | |
| Goats/sheep/dairy farming | 0.00 | 8.33** | 6.25* | 6.25* | 192 | 3.51** |
| | (0.00) | (4.03) | (3.53) | (3.53) | | |
| Owning shops | 12.50 | -6.25 | -6.25 | 2.08 | 192 | 0.96 |
| | (33.42) | (5.98) | (5.98) | (7.06) | | |
| Other | 29.17 | 8.33 | 2.08 | -12.50 | 192 | 2.11 |
| | (45.93) | (9.69) | (9.47) | (8.57) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

About 98% state farming, 58% agricultural labour, and 96% non-agricultural labour activities. Other activities, such as owning a shop and livestock rearing were much less frequently to rarely mentioned. We see some small imbalances in this table: we have somewhat less people engaged in farming in the NE study group

(significant at the 10% level) and we see that the control group has on average less communities in which livestock rearing (other than poultry) was mentioned as one of the main three activities. The low percentage of communities mentioning this category is likely to be driving this significance difference.

## 4.2 Sanitation and electricity

We next discuss the availability of water, sanitation and electricity in our study communities. In terms of water, we can see from Table 14 that in the large majority of communities (92%) households most commonly source their water from boreholes. Only 6% report piped water to be the most common source. In fact, 79% of communities report that none of their households has access to piped water.

Table 14: Sanitation and electricity availability (%)

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Main water source is borehole | 91.49 | -6.07 | -6.07 | -1.91 | 191 | 0.43 |
| | (28.21) | (6.59) | (6.59) | (6.06) | | |
| Main water source is piped | 6.38 | 8.20 | 1.95 | 4.03 | 191 | 0.61 |
| | (24.71) | (6.28) | (5.41) | (5.73) | | |
| No households have piped water | 78.72 | 0.44 | -9.97 | -8.51 | 190 | 0.74 |
| | (41.37) | (8.46) | (9.06) | (9.05) | | |
| Open defecation is common | 72.34 | -14.01 | -7.76 | -1.51 | 191 | 0.85 |
| | (45.22) | (9.76) | (9.60) | (9.35) | | |
| Whole village has access to electricity | 65.96 | 2.79 | -11.79 | -20.12** | 191 | 2.26* |
| | (47.90) | (9.72) | (10.08) | (10.08) | | |
| Electricity connection is very common | 36.17 | 3.41 | -7.00 | -4.92 | 191 | 0.47 |
| | (48.57) | (10.05) | (9.70) | (9.79) | | |
| Any access to sewerage system | 8.51 | 6.07 | 3.99 | 1.91 | 191 | 0.32 |
| | (28.21) | (6.59) | (6.34) | (6.06) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

We further learn that in two-thirds of our study communities (66%) open defecation is a common behaviour, indicating that many of the households do not have access to a toilet, as we will see also below. Relatedly, only 11% of communities have a sewage system.

58% of villages are reported to be served with electricity, whereas in only 34% of villages almost everybody has access to it. We observe only one imbalance here, with communities allocated to the group stimulation treatment arm being less likely than the control group to report that their whole village has access to electricity. This difference is significant at 5%.

## 4.3 Institutions

### 4.3.1 Commercial institutions

Except for general market shops and shops for the public distribution system, commercial institutions are very uncommon in our study area. As shown in Table 15, only 9% of communities have a bank, 4% a microfinance institution, 6% a credit cooperative and 6% have an agricultural cooperative. Shops are somewhat more common, with 19% of communities reporting to have clothes shop, 17% a shop where toys can be bought, 11% a bookshop, 15% a shop for alcohol ('wine shop'). Most common are general market shops (83%) and public distribution system (PDS) shops (62%). We find that GS communities are significantly more likely to have such PDS shops in their community, a difference that does not feed through to the general F-stat. We do see an overall difference in the availability of wine shops at the 5% significance level. All other variables are nicely balanced.

Table 15: Commercial Institutions

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Bank | 8.51 | -0.18 | 2.13 | 3.99 | 190 | 0.19 |
| | (28.21) | (5.76) | (6.13) | (6.34) | | |
| Microfinance institution | 4.35 | -0.18 | -2.22 | -2.26 | 189 | 0.23 |
| | (20.62) | (4.21) | (3.71) | (3.68) | | |
| Credit co-op | 6.38 | 1.95 | -4.26 | 4.03 | 190 | 1.37 |
| | (24.71) | (5.41) | (4.18) | (5.73) | | |
| Agricultural co-op | 6.38 | 6.12 | -2.13 | 4.03 | 190 | 0.92 |
| | (24.71) | (6.02) | (4.67) | (5.73) | | |
| Bookshop | 10.64 | -2.30 | 4.26 | -4.39 | 190 | 0.67 |
| | (31.17) | (6.08) | (6.94) | (5.76) | | |
| Toy shop | 17.02 | 10.06 | -4.26 | 1.73 | 190 | 1.05 |
| | (37.99) | (8.53) | (7.41) | (7.94) | | |
| Children's clothes shop | 19.15 | 1.68 | -4.26 | -0.40 | 190 | 0.21 |
| | (39.77) | (8.29) | (7.82) | (8.13) | | |
| Weekly market | 12.77 | -0.27 | 4.26 | 1.82 | 190 | 0.16 |
| | (33.73) | (6.89) | (7.41) | (7.12) | | |
| General market shop | 82.98 | 2.44 | 6.38 | -3.81 | 190 | 0.68 |
| | (37.99) | (7.56) | (7.17) | (8.11) | | |
| Wine shop | 14.89 | -10.73* | -0.00 | 3.86 | 190 | 2.70** |
| | (35.99) | (6.00) | (7.42) | (7.74) | | |
| Public distribution system | 61.70 | -3.37 | -6.38 | -22.12** | 190 | 1.88 |
| | (49.14) | (10.15) | (10.25) | (10.11) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 4.3.2 Social programs, clubs and other public institutions

Table 16 shows that social programs are relatively common in our study communities: Basically every single village has a self-help group, that is women joining forces to save and lend together, and 62% have a Mahila mandal, which are informal social service clubs formed to help rural women with problems of motherhood, as well as young girls from economically disadvantaged families to find education. 77% of communities have a youth club or centre, 74% a playground for children and basically all (96%) have a temple. About one third of our villages (32%) host the gram panchayat office. We see some small imbalances, all significant only at the 10% level, and none reflected in the overall F-stats (note that the F-stat cannot be calculated when the outcome considered displays no or hardly any variation, as is the case for complementary food for school children).

Table 16: Infrastructure and Institutions

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Mahila mandal | 61.70 | -5.45 | 4.26 | -1.29 | 190 | 0.32 |
| | (49.14) | (10.18) | (10.01) | (10.11) | | |
| Self-help groups | 100.00 | -2.08 | -2.13 | -2.08 | 190 | 1.00 |
| | (0.00) | (2.08) | (2.13) | (2.08) | | |
| Youth centre/club | 76.60 | 0.49 | 8.51 | 12.77* | 189 | 1.36 |
| | (42.80) | (8.75) | (8.16) | (7.72) | | |
| Panchayat office | 31.91 | -2.75 | -2.13 | -3.65 | 188 | 0.05 |
| | (47.12) | (9.55) | (9.63) | (9.61) | | |
| Temple | 95.74 | -10.33* | -4.26 | -10.96* | 188 | 1.67 |
| | (20.40) | (5.95) | (5.08) | (6.12) | | |
| Children's playground | 74.47 | 2.62 | 14.66* | 4.70 | 189 | 1.53 |
| | (44.08) | (8.88) | (7.93) | (8.74) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 17 shows statistics for a number of further social programs, including complementary food for school children, which none of our communities report to receive. However, 64% do report to receive advice on nutrition, often through public meetings, which are organized in all our communities.

Table 17: Social Programmes

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Complementary food for school children | 0.00 | 0.00 | 0.00 | 4.17 | 190 | . |
| | (0.00) | (0.00) | (0.00) | (2.92) | | |
| Advice on nutrition | 63.83 | -9.66 | -6.38 | -12.77 | 189 | 0.58 |
| | (48.57) | (10.15) | (10.16) | (10.22) | | |
| Organise public meetings | 100.00 | -4.17 | -8.51** | -2.08 | 190 | 2.44* |
| | (0.00) | (2.92) | (4.11) | (2.08) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

## 4.4 Education and health facilities

### 4.4.1 Health facilities

The Indian health service works with Primary Health Centres, each of which typically has around six sub-centres to reach out to more remote areas. 34% of our study communities have such a sub-centre, and/or an auxiliary nurse midwife (ANM), a village-level female health worker, typically seen as the first contact person between the community and the health services. This is shown in Table 18. ANMs are supported by up to five accredited social health activists (ASHAs), community health workers instituted by the government of India's Ministry of Health and Family Welfare. In line, we see that almost all (91%) of our study communities report to have such an ASHA worker.

Private health providers are less common. We can see that only 13% of communities have direct access to a private doctor or clinic, 15% have a quack doctor, and 26% a traditional birth attendance. Pharmacies are available in 26% of our study communities.

We see one imbalance in terms of the availability of these quack doctors, which is reflected in the F-stat, significant at 5%.

### 4.4.2 Educational facilities

The final community characteristics we discuss relate to educational institutions available within our study communities. These are displayed in Table 19.

All of the villages have an Anganwadi center. These centres (translated these mean "courtyard shelter"), are part of the Integrated Child Development Services program. While one of their main objectives is to combat child hunger and nutrition, their roles also include the organization of pre-school activities, providing health and nutrition education to families (especially pregnant women), as well as educating parents about child growth and development - and, importantly, they serve as a sort of kindergarten for children aged 3-5 years (and often younger and older children as well).

Table 18: Health Institutions

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Sub centre or ANM | 34.04 | -12.77 | 2.13 | -6.96 | 189 | 1.09 |
| | (47.90) | (9.23) | (9.95) | (9.53) | | |
| Private doctor/clinic | 12.77 | 3.90 | 12.77 | 1.82 | 190 | 0.90 |
| | (33.73) | (7.33) | (8.09) | (7.12) | | |
| Uncertified doctor | 14.89 | -12.77** | -2.13 | -0.31 | 189 | 3.61** |
| | (35.99) | (5.66) | (7.19) | (7.35) | | |
| ASHA | 91.49 | -10.64 | 2.13 | -1.91 | 189 | 1.20 |
| | (28.21) | (7.11) | (5.47) | (6.07) | | |
| Traditional birth attendant | 25.53 | -12.77 | -2.13 | -8.87 | 189 | 1.10 |
| | (44.08) | (8.10) | (8.96) | (8.42) | | |
| Chemist/pharmacy | 25.53 | -4.70 | 0.00 | -0.53 | 190 | 0.14 |
| | (44.08) | (8.74) | (9.09) | (9.01) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

Equally common in our communities are primary schools - every village that takes part in our study has a government primary school, some also a private or NGO-based one (30%) - other primary school, such as English medium ones, are rarely available 6%.

Ongoing education is less frequently available: whereas every village offers a primary school, only 45% of villages have a government secondary school, and only 6% offer higher secondary education.

As previously, we find only small imbalances, none of which are reflected in the F-stat.

Table 19: Education

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Anganwadi | 100.00 | -2.08 | 0.00 | -2.08 | 191 | . |
| | (0.00) | (2.08) | (.) | (2.08) | | |
| Private, NGO or donation-based primary | 29.79 | -2.70 | -6.87 | -8.95 | 191 | 0.40 |
| | (46.23) | (9.35) | (9.11) | (8.97) | | |
| Government primary | 100.00 | -8.33** | 0.00 | -4.17 | 191 | . |
| | (0.00) | (4.03) | (.) | (2.91) | | |
| English medium primary | 6.38 | -0.13 | 2.13 | 1.95 | 190 | 0.10 |
| | (24.71) | (5.05) | (5.47) | (5.41) | | |
| Non-English medium primary | 9.09 | 6.46 | 9.51 | -4.65 | 177 | 2.05 |
| | (29.08) | (7.01) | (7.43) | (5.37) | | |
| Government secondary | 44.68 | -6.38 | 6.38 | -5.10 | 189 | 0.63 |
| | (50.25) | (10.25) | (10.39) | (10.23) | | |
| Non-government secondary | 2.13 | 4.12 | 8.51* | 4.12 | 190 | 1.18 |
| | (14.59) | (4.12) | (5.02) | (4.12) | | |
| Government/private higher secondary | 6.38 | -0.13 | 4.26 | -2.22 | 190 | 0.48 |
| | (24.71) | (5.05) | (5.80) | (4.64) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 – household characteristics

Having a better overview of our study communities, we now zoom into our study households. We recall that these are not a random sub-set of our communities, but of households with children in our target age range (2-20 months), as discussed above.

## 5.1 Religion and Caste

Table 20 shows the breakdown in terms of religion as reported by the biological mother. The overwhelming majority of households is Hindu (94%), remaining households are predominantly Muslim, less than one percent of the sample is Christian. Of the households in our sample, 26% are members of scheduled castes and 6% are members of scheduled tribes. Together, scheduled castes and scheduled tribes make up roughly 32% percent of our sample, which is lower than the 40% recorded in the 2011 census. This is mainly due to the comparatively low level of scheduled tribes among the respondents we surveyed. Apart from scheduled castes and tribes, the most common single caste is Khandayata, which accounts for 17% of households, and is the state branch of the national Kshatriya caste group. In addition, a further 30% are listed as Other Backwards Castes.

Table 20: Religion and Caste

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Christian | 0.19 | 0.18 | 0.17 | 0.55 | 2166 | 0.47 |
| | (4.34) | (0.41) | (0.40) | (0.48) | | |
| Muslim | 6.21 | 0.69 | 2.57 | 3.40 | 2166 | 0.25 |
| | (24.16) | (4.16) | (4.70) | (4.49) | | |
| Hindu | 93.60 | -0.87 | -2.74 | -3.95 | 2166 | 0.31 |
| | (24.50) | (4.16) | (4.71) | (4.48) | | |
| Brahmin | 5.48 | -1.38 | -1.91 | -1.38 | 2161 | 0.34 |
| | (22.78) | (1.80) | (1.94) | (1.92) | | |
| Khandayata | 14.74 | 2.42 | 4.18 | 1.49 | 2161 | 0.32 |
| | (35.49) | (4.40) | (4.34) | (4.32) | | |
| Scheduled tribe | 3.59 | 4.62** | 2.84 | 2.38 | 2161 | 1.93 |
| | (18.63) | (2.19) | (1.95) | (2.04) | | |
| Scheduled caste | 27.03 | -1.66 | -3.28 | -0.17 | 2161 | 0.29 |
| | (44.45) | (4.61) | (4.48) | (4.63) | | |
| Other backwards caste | 32.70 | -3.04 | -5.74 | -2.48 | 2161 | 0.47 |
| | (46.96) | (5.35) | (4.96) | (5.14) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 Household head and composition

### 5.2.1 Household head

As shown in Table 21, the households are typically headed by a male, 44 year old who is able to read and write. While this is the average characteristic, it is worth pointing out that still 14% of households are headed by a female, often indicative of lower economic and social status in the context of this study; as well as that about a quarter of household heads is neither able to read or to write. In line, only about 84% of household heads attended school for any amount of time.

Table 22 gives information on the relationship between the household head and our target children. Most commonly (in 47% of cases), the household head is also the biological father of the child, or alternatively the head is the child's paternal grandfather. These statistics line up with the fact that most couples move in with the husband's family. We also saw above that 14% of households are headed by a female. Table 22 reveals that this female is in 7% of cases the target child's mother, in 4% the paternal grandmother and in less than one percent of households the maternal grandmother.

Table 21: Household Head Characteristics

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Female (%) | 14.42 | 2.68 | -0.63 | 2.21 | 2145 | 0.60 |
| | (35.16) | (3.06) | (3.16) | (3.36) | | |
| Age (years) | 43.94 | -0.81 | -1.69 | -0.23 | 2143 | 0.76 |
| | (15.53) | (1.21) | (1.25) | (1.23) | | |
| Can read (%) | 75.71 | 1.36 | 2.51 | -1.88 | 2145 | 0.76 |
| | (42.92) | (2.86) | (3.12) | (3.04) | | |
| Can write (%) | 75.33 | 0.75 | 1.98 | -2.25 | 2144 | 0.65 |
| | (43.15) | (2.79) | (3.09) | (3.03) | | |
| Attended school (%) | 83.91 | -1.41 | 0.60 | -1.51 | 2095 | 0.29 |
| | (36.78) | (2.59) | (2.85) | (2.73) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 22: Relation of Household Head to target child (%)

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Biological father | 46.99 | 2.73 | 6.40 | 1.25 | 2170 | 0.79 |
| | (49.96) | (4.61) | (4.51) | (4.57) | | |
| Biological mother | 7.33 | 0.30 | -1.80 | 0.25 | 2170 | 0.46 |
| | (26.09) | (1.95) | (2.02) | (2.24) | | |
| Maternal grandfather | 1.69 | -0.95 | -0.44 | -0.21 | 2170 | 0.82 |
| | (12.91) | (0.73) | (0.77) | (0.83) | | |
| Paternal grandfather | 24.25 | -3.21 | -2.64 | -0.77 | 2170 | 0.59 |
| | (42.90) | (2.92) | (3.11) | (2.95) | | |
| Maternal grandmother | 0.38 | 0.37 | 0.16 | 0.18 | 2170 | 0.17 |
| | (6.13) | (0.51) | (0.48) | (0.48) | | |
| Paternal grandmother | 4.14 | 0.33 | 1.58 | 1.23 | 2170 | 0.57 |
| | (19.93) | (1.27) | (1.40) | (1.47) | | |
| Other | 15.23 | 0.42 | -3.26 | -1.92 | 2170 | 0.64 |
| | (35.96) | (3.57) | (3.30) | (3.38) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 5.2.2 Household composition

Including the household head, our sampled households are typically made up of 5 or 6 people. Almost two of these household members are children under the age of 16, of which again the majority are in fact under the age of 6. This is to an extent a construct of our sample selection, given that we target households with young children. We know that by sample construction the lower limit of the number of children under the age of 6 years is one for our sample. Related is the fact that our study households have on average more adult women (1.85) than adult men (1.56); reflecting that, given the young age of the target children, the mother is likely to be present within the household. This is confirmed in Table 24, where we present information on the relationship between our target children and household members. It can be seen that in 99.8% of cases, the biological mother of our target children is present[13], whereas the the biological father is present only in 80% of households. This is often driven by out-migration of men in our study households. We know (not shown), that the father migrates to work in 21% of cases. Section 5.4.2, which discusses income sources of the study households further supports this notion, showing that 27% of households receive income from transfers and remittances.

Table 23: Household Size

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Number of people | 5.38 | -0.03 | 0.13 | 0.02 | 2145 | 0.32 |
| | (2.06) | (0.15) | (0.16) | (0.14) | | |
| Number of adult women (17 or older) | 1.87 | 0.02 | -0.01 | 0.02 | 2145 | 0.11 |
| | (0.91) | (0.07) | (0.07) | (0.07) | | |
| Number of adult men (17 or older) | 1.56 | -0.05 | 0.07 | 0.03 | 2145 | 0.76 |
| | (1.00) | (0.09) | (0.09) | (0.09) | | |
| Number of children (16 or younger) | 1.94 | 0.00 | 0.07 | -0.03 | 2145 | 0.68 |
| | (1.02) | (0.07) | (0.07) | (0.07) | | |
| Number of young children (6 or younger) | 1.60 | -0.02 | -0.00 | -0.01 | 2145 | 0.06 |
| | (0.73) | (0.05) | (0.05) | (0.05) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 also presents information on the presence of paternal and maternal grandparents of the target child. As mentioned before, it is cultural practice for brides to move into the household of their newly-wed husband. In line, we find that the presence of paternal grandparents is much more common than that of maternal ones. 32% of the children live with their paternal grandfather, 40% with their paternal grandmother, and in contrast only 2% and 3% with their maternal grandfather and grandmother respectively.

As is the case for households size and composition, also the presence of relatives in the household are

[13] In fact, for all but one child do we have that the biological mother is alive and lives in the household (not shown).

Table 24: Household Structure: relations of target child who live in household (%)

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Biological mother | 99.81 | 0.00 | -0.17 | 0.00 | 2170 | 0.13 |
| | (4.34) | (0.26) | (0.31) | (0.26) | | |
| Biological father | 79.89 | -0.37 | 6.36 | 1.26 | 2170 | 1.18 |
| | (40.12) | (4.57) | (4.41) | (4.50) | | |
| Paternal grandfather | 32.89 | -3.66 | -0.75 | -0.73 | 2170 | 0.66 |
| | (47.03) | (3.10) | (3.38) | (3.31) | | |
| Paternal grandmother | 40.23 | -1.68 | 1.02 | 1.92 | 2170 | 0.35 |
| | (49.08) | (4.04) | (3.96) | (4.19) | | |
| Maternal grandfather | 2.44 | -0.40 | -0.84 | -0.41 | 2170 | 0.35 |
| | (15.45) | (0.87) | (0.85) | (0.95) | | |
| Maternal grandmother | 3.01 | 0.16 | -0.51 | -0.23 | 2170 | 0.17 |
| | (17.10) | (1.11) | (1.08) | (1.12) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

nicely balanced between our different study arms.

## 5.3 Wealth indicators

Whilst in the next section we will discuss the (economic) wealth status of our study households in more detail below - primarily by looking at their assets, income and financial wealth - we want to start by providing a broad picture of their economic status. We do so based on typical indicators used in India and more widely, namely the households' ownership of a ration card and characteristics of the dwelling they live in.

Table 25: Household's ration card status

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Antyodaya/Annapurna card | 4.54 | 1.62 | 0.82 | 0.29 | 2164 | 0.46 |
| | (20.83) | (1.46) | (1.27) | (1.20) | | |
| Below Poverty Line card | 26.09 | 0.41 | 0.34 | 0.26 | 2164 | 0.01 |
| | (43.95) | (3.45) | (3.58) | (3.33) | | |
| Above Poverty Line card | 20.60 | -4.19 | -6.32** | -2.42 | 2164 | 1.55 |
| | (40.48) | (3.26) | (3.12) | (3.38) | | |
| Other ration card | 22.31 | 5.68 | 3.59 | 0.51 | 2164 | 1.17 |
| | (41.67) | (3.70) | (4.13) | (3.58) | | |
| Any ration cards | 65.03 | 1.20 | -1.81 | -0.46 | 2164 | 0.30 |
| | (47.73) | (3.25) | (3.53) | (3.50) | | |
| Multiple ration cards | 8.65 | 2.15 | 0.28 | -1.07 | 2170 | 0.71 |
| | (28.13) | (2.48) | (2.40) | (2.31) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 25 provides information on the ownership of different ration cards. The Indian ration card has three categories – extreme poverty level (Antyodaya), below poverty line (BPL) and above poverty line (APL). Each household can only receive one of these cards, providing it access to subsidized food and fuel. In addition it serves as an important subsistence tool for the poor, providing proof of identity and a connection with government databases.

Overall, 65% of our study households report to have a ration card (less than one percent say they have more than one). Five percent of our study households own an Antyodaya or Annapurna card, classifying them as extremely poor in the Government of India's system. Further 26% of households own a below poverty card, 17% an above poverty line card and 25% report to own some other ration card. We saw above in Section 4.3.1 that 62% of the villages in our sample have a PDS shop, which is where subsidized products can be obtained with these ration cards. Households living in the remaining 38% of villages would have to travel to other villages or gram panchayats to purchase key staple at reduced price. The IS group is significantly less likely to own an APL card, but this difference is not reflected in the joint F-stat.

Table 26 provides a picture on the immediate environment our study households and target children live

in. About half of the sample (51%) lives in houses of strong structure (pucca), and a further 30% live in semi-strong buildings. The number of rooms (occupied by 5 to 6 household members on average) varies, 27% of households live in only one room, 35% in two and 38% in three rooms. Less than half (47%) of the households do not have access to a toilet in their dwelling and in line, 61% of households report that it is common for household members to defecate in the open, which in fact implies that even when households own a toilet, open defecation is common (which is consistent with our earlier finding reported by the Anganwadi Worker in the community questionnaire, see Section 4.2.

Table 26: Home environment

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Dwelling has own toilet | 47.17 | -1.36 | 0.87 | -1.43 | 2167 | 0.07 |
| | (49.97) | (5.76) | (5.75) | (5.54) | | |
| Household members typically open defecate | 61.32 | 2.37 | -1.14 | 2.38 | 2167 | 0.21 |
| | (48.75) | (5.02) | (5.28) | (4.93) | | |
| Food prep. done in room with windows | 85.66 | 0.93 | -1.91 | 0.45 | 2167 | 0.44 |
| | (35.08) | (2.80) | (2.51) | (2.64) | | |
| Food prep. done in room with chimney/opening | 92.26 | 0.10 | -1.55 | -0.04 | 2167 | 0.30 |
| | (26.74) | (1.96) | (1.95) | (2.03) | | |
| Food cooked mainly over wood or charcoal | 72.83 | 2.31 | -1.40 | 1.98 | 2165 | 0.38 |
| | (44.53) | (3.62) | (4.15) | (3.67) | | |
| Pucca house (strong structure) | 50.94 | -0.11 | -2.19 | -3.54 | 2167 | 0.29 |
| | (50.04) | (4.36) | (4.36) | (4.49) | | |
| Semi-pucca house | 29.81 | 1.47 | 4.12 | 5.37 | 2167 | 0.88 |
| | (45.79) | (3.55) | (3.68) | (3.70) | | |
| Kutcha house (weak structure) | 19.25 | -1.37 | -1.92 | -1.84 | 2167 | 0.21 |
| | (39.46) | (2.75) | (2.92) | (2.66) | | |
| House has one room | 27.17 | -2.22 | 1.35 | -4.54 | 2156 | 1.56 |
| | (44.53) | (3.24) | (2.79) | (2.92) | | |
| House has two rooms | 35.09 | -0.95 | -3.33 | -3.00 | 2156 | 0.56 |
| | (47.77) | (3.10) | (3.15) | (2.92) | | |
| House has three or more rooms | 37.74 | 2.98 | 1.98 | 7.53* | 2156 | 1.32 |
| | (48.52) | (3.95) | (4.13) | (3.96) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 table also shows information on cooking facilities. About 86% of households cook their food in room with windows. If it does not have a window, it is reported to have a chimney or other type of opening that allows for smoke to escape. This is is known to be important for to reduce or avoid negative health impacts from the smoke, especially when the main fuel for cooking used is wool or charcoal, as is done by 73% of our

study households. None of these variables show any imbalances across treatment arms.

## 5.4 Economic status - asset, financial wealth and income

We already saw in Section 5.3 that our households are generally classified to be poor to extremely poor. We define this further in this section, by looking at their asset and financial wealth, as well as the income they earn.

### 5.4.1 Assets

Table 27 gives a breakdown of the types of durable assets our study households own. The average households owns 7.9 different types of these assets (out of 21). Furniture is relatively sparse in our household dwellings. Most households own two cots, often made of woven string, which they sit and sleep on and drag outside on warm days. Some households own mattresses, but these are less common (0.88 on average per household). Since it can get very hot in the selected districts - easily above 40 degrees celsius in the summer - it is not surprising that our study households also own two electric fans on average, which is made possible by the high electricity coverage we discussed in Section 4.2.

The average household has two chairs, but only about every second household owns a table (40% of households). Most households own either a pressure or gas cooker, only 19% own a fridge and 9% a sewing machine.

In terms of mobility we can see in the table that the average households owns a bike. Every second to third household owns a motorbike and cars are extremely uncommon, owned only by 4% of study households.

In line with the generally widespread ownership of mobile phones in India, we see that also our study households own 1-2 devices per household on average. Not every household has a TV (0.67 TVs are owned on average per household) and radio ownership is low at only 0.08 devices per household.

The final set of assets presented in Table 27 relates to agricultural equipment. We saw in our description of the study communities in Section 4.1 above that farming is one of the key activities. In line we find that 14% of households own some irrigation equipment, most likely those households that have their own farm. Further 42% of households own dairy assets. While dairy was not indicated as one of the top three industries, it is quite common for households to own a cow or buffalo for own milk consumption or to engage in milk-selling as an extra income generating activity. We will go into more detail of income sources of the households in the next section.

### 5.4.2 Income

We show in Table 28 the types of sources from which our study households received income over the past twelve months. It is extremely common for households to receive income from several sources, 3.2 on average (with a standard deviation of 1.16), not shown in the table. 35% of households received income from agricultural labour, 64% from non-agricultural labour, and 72% from their farm or other family business. It is very

Table 27: Ownership of Durable Assets

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| No. of Bicycles | 0.98 | 0.04 | 0.08 | 0.05 | 2166 | 0.52 |
| | (0.67) | (0.05) | (0.07) | (0.06) | | |
| No. of Motorbikes | 0.44 | 0.00 | 0.02 | -0.03 | 2165 | 0.35 |
| | (0.70) | (0.06) | (0.06) | (0.06) | | |
| Own Car (%) | 4.15 | -0.99 | -2.00 | -0.08 | 2166 | 1.38 |
| | (19.97) | (1.45) | (1.31) | (1.56) | | |
| Own Fridge (%) | 19.25 | -1.74 | 0.07 | -0.73 | 2166 | 0.14 |
| | (39.46) | (3.18) | (3.16) | (3.22) | | |
| No. of Electric Fans | 1.99 | 0.02 | -0.05 | -0.09 | 2167 | 0.15 |
| | (2.13) | (0.18) | (0.19) | (0.18) | | |
| No. of gas/pressure cookers | 1.05 | 0.10 | 0.07 | 0.07 | 2170 | 0.27 |
| | (1.27) | (0.12) | (0.12) | (0.12) | | |
| Own Sewing Machine (%) | 9.43 | -0.09 | 0.74 | 2.60 | 2165 | 0.75 |
| | (29.26) | (2.00) | (2.15) | (2.08) | | |
| No. of Tables | 0.44 | 0.08* | 0.05 | 0.02 | 2165 | 1.05 |
| | (0.67) | (0.05) | (0.05) | (0.05) | | |
| No. of Chairs | 2.08 | 0.03 | 0.04 | -0.03 | 2165 | 0.07 |
| | (2.47) | (0.16) | (0.18) | (0.16) | | |
| No. of Cots | 1.93 | 0.16 | 0.15 | 0.11 | 2167 | 0.26 |
| | (1.66) | (0.20) | (0.21) | (0.20) | | |
| No. of Mattresses | 0.88 | 0.10 | 0.10 | -0.01 | 2167 | 0.96 |
| | (1.00) | (0.09) | (0.09) | (0.07) | | |
| No. of TVs or Computers | 0.70 | -0.01 | 0.02 | -0.00 | 2170 | 0.11 |
| | (0.67) | (0.05) | (0.06) | (0.05) | | |
| No. of Mobile Phones | 1.44 | 0.08 | 0.08 | 0.09 | 2166 | 0.32 |
| | (1.21) | (0.11) | (0.12) | (0.10) | | |
| No. of Radios | 0.08 | 0.00 | -0.01 | -0.01 | 2167 | 0.21 |
| | (0.45) | (0.02) | (0.02) | (0.02) | | |
| Own Irrigation Equip. (%) | 13.58 | -1.06 | 1.08 | -2.66 | 2164 | 0.83 |
| | (34.30) | (2.63) | (2.68) | (2.37) | | |
| Own Dairy Assets (%) | 41.70 | -4.43 | -1.34 | -1.51 | 2164 | 0.33 |
| | (49.35) | (4.58) | (4.58) | (4.80) | | |
| Other Equipment (%) | 6.23 | 2.17 | 0.38 | 3.59* | 2166 | 1.29 |
| | (24.19) | (2.01) | (1.71) | (2.04) | | |
| No. of Furniture Assets | 5.29 | 0.39 | 0.37 | 0.12 | 2170 | 0.63 |
| | (4.54) | (0.33) | (0.38) | (0.32) | | |
| No. of Electrical Assets | 4.39 | 0.09 | 0.06 | -0.01 | 2169 | 0.04 |
| | (3.95) | (0.35) | (0.38) | (0.33) | | |
| No. of Asset Types (/21) | 7.90 | 0.20 | 0.26 | 0.15 | 2142 | 0.18 |
| | (3.77) | (0.36) | (0.38) | (0.35) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

common to also receive income from some sort of support or benefit program: 85% of study households received government benefits, 4% did public relief work and 30% received pension or retirement income. 27%

of households further received income from transfers or remittances. These income sources are balanced across treatment arms.

Table 28: Income Sources in the Past 12 Months

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Agricultural labour | 35.09 | -2.57 | -2.84 | -1.33 | 2162 | 0.22 |
| | (47.77) | (3.79) | (4.07) | (3.66) | | |
| Non-agricultural labour | 64.46 | 5.71 | 3.22 | 0.10 | 2161 | 0.97 |
| | (47.91) | (4.12) | (4.57) | (4.36) | | |
| Public relief work | 3.78 | 0.91 | 1.58 | 2.34 | 2161 | 0.84 |
| | (19.09) | (1.23) | (1.31) | (1.68) | | |
| Family business/farm | 71.68 | -4.63 | -2.39 | -3.32 | 2128 | 0.64 |
| | (45.10) | (3.40) | (3.78) | (4.00) | | |
| Traditional occupation | 3.77 | 0.15 | 0.53 | 0.69 | 2161 | 0.12 |
| | (19.07) | (1.33) | (1.32) | (1.31) | | |
| Pension/Retirement income | 30.00 | -0.71 | -2.27 | -2.17 | 2164 | 0.27 |
| | (45.87) | (3.11) | (3.08) | (3.03) | | |
| Government benefits | 85.01 | -3.24 | -1.68 | -3.89 | 2152 | 0.68 |
| | (35.73) | (3.01) | (2.67) | (3.01) | | |
| Dowry | 0.76 | 0.74 | -0.04 | 0.54 | 2165 | 0.72 |
| | (8.67) | (0.66) | (0.50) | (0.64) | | |
| Transfers/Remittances | 27.55 | 0.12 | -6.30 | -2.55 | 2165 | 0.79 |
| | (44.72) | (5.41) | (4.94) | (5.25) | | |
| Income from rent | 0.57 | 1.86* | 0.68 | 0.73 | 2166 | 1.49 |
| | (7.51) | (1.02) | (0.65) | (0.61) | | |
| Other income sources | 4.37 | 0.32 | -0.43 | 1.92 | 2157 | 0.39 |
| | (20.47) | (1.60) | (1.73) | (2.21) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 the amount of income earned, we see from Table 29 that the average annual income of our study households is INR 91,260. Taking the December 2015 average exchange rate of INR 66.52 to 1 US$, this average income translates into US$ 1,372 per year. With an average household size of 5.38, we find that our study households earn on average US$ 0.70 per person per day (not applying any equivalent scale for children). This puts the average household member significantly below the internationally acceptable poverty line of US$ 1.25 per person per day.

Table 29: Income (INR) from Sources in the Past 12 Months

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Agricultural labour | 2,177.38 | -12.06 | -286.13 | -188.09 | 2116 | 0.26 |
| | (5,164.18) | (444.53) | (380.07) | (359.38) | | |
| Non-agricultural labour | 40,242.87 | 4,550.91 | 4,729.98 | -5,363.95 | 2097 | 1.33 |
| | (126,561.12) | (8,321.99) | (8,332.65) | (6,984.22) | | |
| Public relief work | 88.40 | 140.29* | 147.73** | 100.65 | 2144 | 2.34* |
| | (693.70) | (73.17) | (69.80) | (71.78) | | |
| Family business/farm | 24,526.94 | -3,884.90 | 1,512.04 | 3,919.50 | 2019 | 1.07 |
| | (63,891.98) | (4,751.64) | (5,267.81) | (6,207.87) | | |
| Traditional occupation | 1,087.85 | -392.85 | 343.95 | 483.89 | 2156 | 0.78 |
| | (9,717.80) | (517.26) | (890.59) | (857.76) | | |
| Pension/Retirement income | 8,430.82 | -1,027.88 | -1,384.32 | -3,587.91* | 2159 | 1.60 |
| | (32,879.45) | (2,220.70) | (2,288.91) | (2,017.95) | | |
| Government benefits | 4,489.09 | -593.15 | 98.16 | -245.24 | 2124 | 0.79 |
| | (7,747.63) | (543.24) | (591.75) | (656.19) | | |
| Dowry | 680.53 | -190.81 | -170.69 | -188.88 | 2162 | 0.06 |
| | (9,957.77) | (479.41) | (535.18) | (485.97) | | |
| Transfers/Remittances | 10,721.07 | 761.85 | -3,816.59 | -879.96 | 2150 | 1.57 |
| | (32,316.99) | (2,711.66) | (2,471.64) | (2,663.16) | | |
| Income from rent | 7.36 | 396.78 | 93.23 | 416.72* | 2161 | 2.07 |
| | (104.54) | (257.34) | (89.59) | (251.36) | | |
| Other income sources | 2,649.03 | 1,060.76 | -355.76 | 3,840.14 | 2151 | 0.58 |
| | (20,976.38) | (1,451.35) | (1,676.46) | (3,830.22) | | |
| Total Income | 91,259.81 | 1,783.81 | 1,200.79 | -751.85 | 2167 | 0.03 |
| | (144,715.80) | (10,297.22) | (10,615.71) | (9,596.27) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 5.4.3 Income of main breadwinner, biological mother and father

We collected more detailed information on earnings of the main breadwinner in the household, as well as (if different from the main breadwinner), the biological father and mother of the target child.

Table 30 provides information on the main breadwinner. Most often (in 64% of households) the household head is the main breadwinner and in 17% of households the main breadwinner, who is almost exclusively male, lives outside the household.

He works on average 66 hours in a typical week, earning about INR 70,189 (US$ 1,055) over the year, which implies an hourly wage rate of INR 35 (US$ 0.53). When excluding those above the top 98th and those below the bottom 2nd percentiles (including some outliers), these amounts drop to an annual income

of INR 60,980 (US$ 917) and an hourly rate of INR 29 (US$ 0.44).

Table 30: Work and Earnings of Main Breadwinner

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Breadwinner is Household Head (%) | 64.27 | -0.59 | 6.62 | 1.96 | 2165 | 1.33 |
| | (47.97) | (4.50) | (4.06) | (4.08) | | |
| Breadwinner lives outside household (%) | 17.39 | 0.67 | -6.14 | -3.85 | 2165 | 1.45 |
| | (37.94) | (4.17) | (3.89) | (3.99) | | |
| Breadwinner is female (%) | 0.02 | 0.00 | -0.00 | 0.01 | 1839 | 0.53 |
| | (0.14) | (0.01) | (0.01) | (0.01) | | |
| Averages Hours per Week | 66.10 | -0.04 | -0.04 | -1.34 | 1842 | 0.29 |
| | (23.69) | (1.79) | (1.93) | (1.77) | | |
| Earnings Last Year (Rs.) | 70,189.12 | -1,960.69 | -5,003.97 | -1,392.16 | 1761 | 0.27 |
| | (80,388.50) | (6,176.63) | (6,264.13) | (6,612.68) | | |
| Earnings Last Year - trimmed | 60,979.58 | 2,561.03 | 1,710.50 | 798.29 | 1700 | 0.15 |
| | (43,087.27) | (4,084.06) | (3,946.61) | (3,940.41) | | |
| Hourly Wage (Rs.) | 35.24 | -1.80 | -0.78 | -1.99 | 1753 | 0.15 |
| | (52.33) | (3.18) | (4.08) | (3.49) | | |
| Hourly Wage - trimmed | 28.98 | 2.05 | 0.61 | -0.16 | 1685 | 0.60 |
| | (20.91) | (1.86) | (1.89) | (1.74) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 shows the type of occupation the main breadwinner is typically involved in. Most (43%) work as daily labourer, about 18% are self-employed and own a business or farm, 15% are private employees on a fixed income, 11% are doing work that falls under the category of petty trader/vendor. We then have a few households where the main breadwinner is a government employee (3%), auto or taxi driver (3%) or doing other types of jobs (7%).

Table 31: Occupation of the Main Breadwinner

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Private employee (fixed income) | 0.15 | -0.01 | -0.04 | -0.01 | 2115 | 0.96 |
| | (0.35) | (0.02) | (0.03) | (0.03) | | |
| Government job | 0.03 | 0.01 | 0.00 | -0.00 | 2115 | 0.32 |
| | (0.18) | (0.01) | (0.01) | (0.01) | | |
| Daily labourer | 0.43 | 0.08** | 0.04 | 0.03 | 2115 | 1.80 |
| | (0.49) | (0.04) | (0.04) | (0.03) | | |
| Auto/taxi driver | 0.03 | 0.00 | -0.01 | 0.00 | 2115 | 0.67 |
| | (0.17) | (0.01) | (0.01) | (0.01) | | |
| Petty trader vendor | 0.11 | -0.04* | -0.01 | -0.01 | 2115 | 1.36 |
| | (0.32) | (0.02) | (0.02) | (0.02) | | |
| Boss/owner of a farm or business | 0.18 | -0.03 | 0.01 | -0.00 | 2115 | 1.66 |
| | (0.38) | (0.02) | (0.02) | (0.02) | | |
| Other occupation | 0.07 | -0.01 | -0.01 | -0.00 | 2115 | 0.14 |
| | (0.26) | (0.02) | (0.02) | (0.02) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

We repeat the same information on hours worked and earnings for the biological father in Tables 32 and 33 see that fathers, who are in 75% of the cases also the main breadwinner, work on average 75 hours per week, hence more than the main breadwinner, but they earn on average a bit less at INR 68,238 (US$ 1,026). The biological mother works on average 41 hours per week, earning INR 18,692 (US$ 281) over the year, implying an hourly wage of INR 17 (US$ 0.25).

Overall, we see only few imbalances in variables related to earnings, none of which are reflected in the F-statistic.

Table 32: Work and Earnings of Biological Father

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Father is main breadwinner (%) | 74.53 | 5.36* | 1.54 | 1.21 | 2167 | 1.54 |
| | (43.61) | (2.86) | (3.21) | (3.01) | | |
| Averages Hours per Week | 64.44 | 1.78 | 3.16* | 0.20 | 1790 | 1.32 |
| | (23.51) | (1.81) | (1.80) | (1.82) | | |
| Earnings Last Year (Rs.) | 68,237.68 | -2,325.85 | -5,371.64 | -2,090.40 | 1713 | 0.39 |
| | (73,838.12) | (5,767.23) | (5,658.17) | (5,941.48) | | |
| Earnings Last Year - trimmed | 60,615.26 | 688.77 | 835.28 | -260.80 | 1648 | 0.05 |
| | (39,005.83) | (3,412.01) | (3,309.00) | (3,165.27) | | |
| Hourly Wage (Rs.) | 34.89 | -2.76 | -3.37 | -3.63 | 1705 | 0.47 |
| | (45.66) | (3.09) | (3.56) | (3.20) | | |
| Hourly Wage - trimmed | 29.67 | -1.59 | -0.86 | -0.96 | 1637 | 0.32 |
| | (19.17) | (1.64) | (1.64) | (1.57) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 33: Work and Earnings of Biological Mother

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Averages Hours per Week | 41.95 | 0.58 | -0.17 | 0.24 | 2127 | 0.22 |
| | (14.23) | (0.93) | (1.10) | (0.98) | | |
| Earnings Last Year (Rs.) | 18,692.08 | -326.54 | -3,303.64 | 1,718.62 | 92 | 0.20 |
| | (24,004.18) | (8,510.91) | (7,630.65) | (7,607.24) | | |
| Earnings Last Year - trimmed | 18,692.08 | -5,850.94 | -2,636.02 | 1,718.62 | 90 | 0.88 |
| | (24,004.18) | (6,474.55) | (7,808.59) | (7,610.91) | | |
| Hourly Wage (Rs.) | 16.77 | 21.11 | 11.74 | 11.63 | 93 | 1.31 |
| | (19.71) | (15.47) | (7.40) | (9.11) | | |
| Hourly Wage - trimmed | 16.77 | 8.61 | 11.74 | 11.63 | 92 | 1.02 |
| | (19.71) | (8.78) | (7.40) | (9.11) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 5.4.4 Savings and debt

We now consider the financial assets of the households. Table 34 shows that 27% of households report to have no savings, 37% report to have up to INR 5000 (US$ 75), 10% between INR 5,000 and INR 10,000 (US$ 75-150), and 26% of households report to have more than INR 10,000 in savings.

In terms of debt we can see in the same table that just about one quarter of households (24%) have no outstanding debt, 18% owe up to INR 5,000 (US$ 75), 11% of households owe between INR 5,000 and 10,000 and 38% have more than INR 10,000 in debt outstanding.

Table 34: Savings and debt

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Household has no savings | 0.27 | -0.00 | 0.03 | -0.01 | 2019 | 0.50 |
| | (0.44) | (0.04) | (0.04) | (0.03) | | |
| More than zero, less than Rs.5000 in savings | 0.37 | -0.01 | -0.03 | -0.01 | 2019 | 0.36 |
| | (0.48) | (0.03) | (0.03) | (0.03) | | |
| Between Rs.5000 and Rs.10000 in savings | 0.10 | 0.02 | 0.00 | 0.01 | 2019 | 0.44 |
| | (0.31) | (0.02) | (0.02) | (0.02) | | |
| More than Rs.10000 in savings | 0.26 | -0.01 | 0.00 | 0.01 | 2019 | 0.12 |
| | (0.44) | (0.03) | (0.04) | (0.03) | | |
| Household has no debt | 0.24 | 0.06** | -0.00 | 0.03 | 2157 | 2.73** |
| | (0.43) | (0.03) | (0.03) | (0.03) | | |
| More than zero, less than Rs.5000 in debt | 0.18 | 0.01 | 0.02 | -0.01 | 2157 | 0.39 |
| | (0.38) | (0.02) | (0.02) | (0.03) | | |
| Between Rs.5000 and Rs.10000 in debt | 0.11 | -0.02 | -0.03 | -0.02 | 2157 | 0.67 |
| | (0.32) | (0.02) | (0.02) | (0.02) | | |
| More than Rs.10000 in debt | 0.38 | -0.04 | 0.03 | 0.04 | 2157 | 2.70** |
| | (0.48) | (0.03) | (0.04) | (0.04) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 Consumption expenditures

We discuss here the expenses of the household over different time periods[14], which often is a more reliable measure of household's income than household's self-reported income levels discussed in the previous sections. We can see from Table 35 that the total expenses of the average study household amount to about INR 123,000 (US$ 1849) in a year. If we exclude the value of home grown food, we get to on average about INR 102,000 (US$ 1,533), which closely aligns with the average annual income of about 91,000 (US$ 1,368). This is one possible explanation why we saw above in Section 5.4.4 that a lot of households have no savings, and only few have no debt.

Since every household has non-zero expenditures on food items (bought as well as home grown), we can calculate from this information the share that households spend on food. Doing so we learn that for the average study households 50% of expenditures go towards food.

[14]The abbreviations 7d, 30d, 6m, and 12m in Table 35 stand for last seven days, last 30 days, last 6 months, and last 12 months respectively.

Table 35: Recorded expenses (INR)

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Bought food (7d) | 849.40 | -23.52 | 16.34 | -28.96 | 2104 | 0.30 |
| | (653.29) | (54.79) | (58.67) | (49.91) | | |
| Home-grown food consumed (7d) | 336.67 | -35.66 | -26.04 | 354.57 | 380 | 0.54 |
| | (601.81) | (89.47) | (76.70) | (318.44) | | |
| Tobacco and alcohol (7d) | 85.02 | -22.81 | -26.91 | -12.31 | 2027 | 1.35 |
| | (351.41) | (18.45) | (17.62) | (18.00) | | |
| Transport (30d) | 696.44 | -78.59 | -35.87 | -10.75 | 2099 | 0.28 |
| | (1,251.60) | (93.55) | (88.27) | (94.12) | | |
| Water/electricity/etc. (30d) | 288.97 | 5.08 | 32.60 | 14.27 | 2047 | 0.26 |
| | (369.65) | (34.28) | (39.04) | (29.64) | | |
| Fuel (30d) | 265.07 | -25.95 | 32.96 | -28.25 | 2138 | 1.09 |
| | (408.81) | (31.05) | (41.02) | (28.03) | | |
| Salaries for workers/aid (30d) | 988.34 | -175.58 | -346.50* | -98.03 | 2128 | 1.27 |
| | (3,053.49) | (223.87) | (205.78) | (215.00) | | |
| Education (30d) | 257.96 | 32.03 | 19.39 | 50.82 | 2144 | 0.24 |
| | (1,018.60) | (58.67) | (56.84) | (63.42) | | |
| Health expenses (30d) | 1,064.27 | 391.76 | 395.80* | -69.00 | 2149 | 1.76 |
| | (2,853.36) | (404.09) | (236.01) | (149.93) | | |
| Services (30d) | 72.95 | -2.39 | 9.13 | -0.16 | 2011 | 0.53 |
| | (94.10) | (7.61) | (9.16) | (7.07) | | |
| Hygiene products (30d) | 302.79 | 11.73 | 24.70 | -16.56 | 2114 | 0.87 |
| | (365.79) | (27.03) | (32.29) | (24.35) | | |
| Children's hygiene (30d) | 148.60 | 1.73 | 17.72 | 2.60 | 2119 | 0.32 |
| | (193.25) | (15.48) | (19.58) | (15.36) | | |
| Toys for children under 6 (6m) | 183.82 | 11.74 | 24.38 | 11.45 | 2131 | 0.24 |
| | (313.51) | (22.47) | (32.63) | (20.43) | | |
| Books for children under 6 (6m) | 35.83 | 6.72 | -8.86 | 0.95 | 2146 | 0.70 |
| | (199.94) | (12.86) | (11.51) | (14.68) | | |
| Clothes for children under 6 (6m) | 657.67 | -25.08 | 17.86 | 45.35 | 2128 | 0.57 |
| | (719.23) | (45.10) | (61.85) | (61.96) | | |
| Clothes for other HH members (6m) | 1,369.61 | 176.14 | 124.50 | 243.36 | 2112 | 0.90 |
| | (1,856.56) | (168.56) | (212.76) | (154.22) | | |
| Household repairs (12m) | 4,022.28 | 3,218.60 | 986.92 | 441.64 | 2043 | 0.52 |
| | (17,565.23) | (2,995.33) | (1,239.30) | (1,129.89) | | |
| Weddings (12m) | 11,264.60 | -1,035.03 | -3,482.95 | 1,558.93 | 2110 | 1.10 |
| | (57,095.61) | (2,843.33) | (2,748.22) | (3,595.98) | | |
| Festivals (12m) | 3,606.96 | 1,495.66 | -5.85 | 1,169.84 | 2119 | 0.80 |
| | (11,718.66) | (1,478.12) | (695.52) | (1,107.45) | | |
| Funerals (12m) | 1,800.78 | 172.02 | -104.23 | 949.77 | 2131 | 0.66 |
| | (7,448.12) | (590.13) | (451.13) | (773.62) | | |
| Total expenses over 12 months | 122,786.74 | 3,509.13 | -788.95 | 5,327.05 | 2170 | 0.20 |
| | (120,900.72) | (11,344.08) | (9,147.79) | (9,072.29) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 – Child characteristics

This section focuses on the profile of the children participating in this study. We will for the most part report combined statistics for target and spillover children, as we have done when describing the households they live in. We make an exception when reporting our main measures of their development, namely scores on the Ages and Stages Questionnaire (ASQ), anthropometrics measures, and proxies of morbidity.

## 6.1 Age and sex

We start by giving an overview of the age and gender distribution of the children we have data on. These are displayed in Table 36. As expected, two thirds of the sample are those children that are eligible to become part of the intervention, i.e. those falling in the age range 7-16 months at the start of the intervention. The remaining - those between 2-6 and 17-20 months at the start of intervention) are what we refer to as 'spillover children', which are again split into children just below and just above the intervention start target age range. The split between younger and older is 56% and 44% respectively. This is consistent with our sample strategy giving slight priority to younger children (see Section 3.3). Importantly, we do not see any important difference between the study groups, except that target children in the IS group are somewhat more likely to be female, which is however only significant at the 10% level.

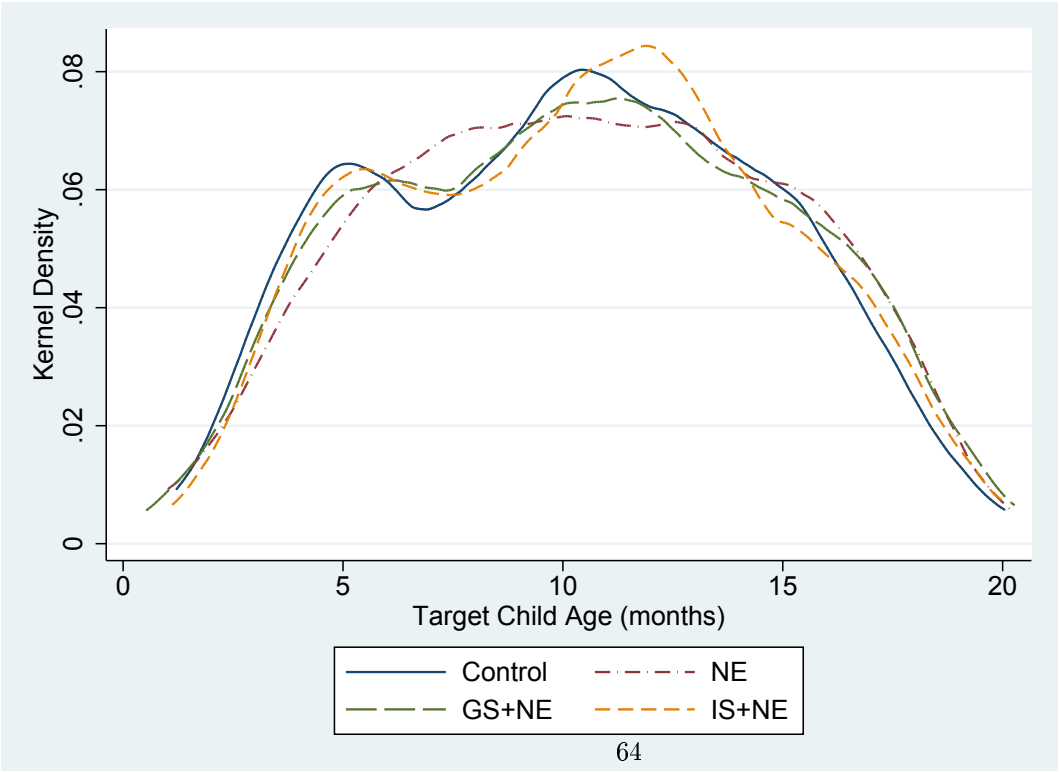
Note that the baseline survey started 2 months prior to the start of the intervention and finished right before the intervention started, therefore the age range of the children in the baseline dataset is 0-20 months. Figure 4 shows the age distribution of all children by treatment group.

Table 36: Age and sex of target children

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Target child (%) | 66.35 | -0.62 | -0.28 | -1.47 | 2170 | 0.49 |
| | (47.29) | (1.30) | (1.23) | (1.40) | | |
| Spillover child (%) | 33.65 | 0.62 | 0.28 | 1.47 | 2170 | 0.49 |
| | (47.29) | (1.30) | (1.23) | (1.40) | | |
| Younger Spillover - Female (%) | 55.36 | 5.47 | -8.19 | -3.87 | 416 | 1.07 |
| | (49.94) | (7.75) | (7.39) | (7.72) | | |
| Younger Spillover - Age in months | 4.21 | -0.02 | 0.13 | -0.06 | 416 | 0.39 |
| | (1.14) | (0.21) | (0.18) | (0.20) | | |
| Target Children - Female (%) | 47.31 | -0.85 | 7.02* | 0.84 | 1427 | 1.73 |
| | (50.00) | (4.15) | (3.78) | (3.63) | | |
| Target Children - Age in months | 10.80 | -0.15 | -0.13 | -0.20 | 1427 | 0.30 |
| | (2.73) | (0.23) | (0.21) | (0.24) | | |
| Older Spillover - Female (%) | 49.25 | 0.17 | -5.21 | -3.19 | 327 | 0.22 |
| | (50.37) | (8.06) | (8.27) | (8.19) | | |
| Older Spillover - Age in months | 16.95 | 0.02 | 0.07 | 0.03 | 327 | 0.03 |
| | (1.19) | (0.23) | (0.25) | (0.25) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 age by treatment group



## 6.2 Child development - ASQ

Our main baseline measure of levels of child development was the Ages and Stages Questionnaires (third edition). 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[15].

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.

We first look at the frequency of response options chosen. Respondents were given three options for responding to questions - "yes" (scoring 10 ), "not yet" (scoring 0) and "sometimes" (scoring 5). We can see from Table 37 shows that "yes" was throughout the answer most frequently chosen by the mothers, combined with "sometimes" adding to almost eighty percent in each of the domains measured.

Table 37: ASQ - frequency of answers

| | 0. No | 5. Sometimes | 10. Yes |
|---|---|---|---|
| Communication | 22.4 | 11.3 | 66.3 |
| Gross motor | 25.7 | 4.8 | 69.5 |
| Fine motor | 19.2 | 7.7 | 73.1 |
| Problem solving | 18.7 | 7.2 | 74.1 |
| Personal social | 26.0 | 11.5 | 62.5 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

Before discussing the scores achieved by target and spillover children, we present a measure of internal

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

reliability. In Table 38 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 [16]. Looking at Table 38 we see that about forty percent of our alphas (for each subscale-age-specific test combination) fall short of that criterion (about 16% fall below 0.6). 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 we will mainly rely on the Bayley-III test as our main tool to measure progress on child development outcomes at endline.

Table 38: ASQ - Cronbach's Alpha by 12 Question Subscales

| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Communication | 0.8 | 0.7 | 0.6 | 0.5 | 0.6 | 0.7 | 0.7 | 0.8 | 0.6 | 0.8 |
| Gross motor | 0.9 | 0.9 | 0.7 | 0.8 | 0.8 | 0.8 | 0.9 | 0.9 | 0.7 | 0.7 |
| Fine motor | 0.7 | 0.8 | 0.8 | 0.7 | 0.6 | 0.6 | 0.7 | 0.6 | 0.5 | 0.8 |
| Problem solving | 0.8 | 0.8 | 0.8 | 0.7 | 0.7 | 0.8 | 0.7 | 0.7 | 0.5 | 0.8 |
| Personal social | 0.8 | 0.7 | 0.5 | 0.5 | 0.6 | 0.7 | 0.7 | 0.6 | 0.5 | 0.6 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 6.2.1 Target children

We start with a discussion of the baseline performance of the target children on the adapted ASQ-3 test. Table 39 presents these children's raw scores, 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 fairly balanced over treatment status across four domains (gross motor, fine motor, problem solving and personal social). We observe a slight imbalance in terms of communication, however, with children in the NE and IS+NE treatment arms scoring 4 to 5 points higher relative to control. We will account for this imbalance in our impact assessment analysis at endline.

Table 42 shows the mean and standard deviation of each subscale, broken down by each age-specific test (in months).[16] 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

[16] Note that for questionnaire 2 (first column), fewer questions/items are considered since this is the first questionnaire. This explains the difference in means to the other columns.

Table 39: ASQ - raw scores from 12 question subscales, target child

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Communication | 82.63 | 3.78* | 4.82** | 1.87 | 1427 | 2.33* |
| | (20.88) | (2.00) | (1.98) | (2.07) | | |
| Gross motor | 85.85 | 1.63 | 1.64 | 2.08 | 1427 | 0.27 |
| | (27.13) | (2.31) | (2.45) | (2.42) | | |
| Fine motor | 93.34 | 0.99 | 1.93 | 1.44 | 1427 | 0.48 |
| | (18.69) | (1.67) | (1.68) | (1.59) | | |
| Problem solving | 92.82 | 2.79 | 1.60 | 2.92 | 1426 | 1.04 |
| | (21.41) | (1.92) | (1.90) | (1.83) | | |
| Personal social | 77.49 | 3.30* | 2.90 | 2.42 | 1427 | 1.10 |
| | (21.88) | (1.94) | (1.83) | (2.00) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

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.

Table 40: ASQ - mean and sd of 12 questions subscales by age category

| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Communication | 58.7 | 95.4 | 85.1 | 85.2 | 81.4 | 90.5 | 83.6 | 83.1 | 90.0 | 88.3 |
| | (20.2) | (19.4) | (15.7) | (15.9) | (20.1) | (20.2) | (22.5) | (23.4) | (19.1) | (24.2) |
| Gross motor | 37.7 | 75.5 | 69.5 | 82.5 | 84.2 | 87.2 | 98.4 | 105.7 | 101.9 | 100.4 |
| | (19.9) | (32.1) | (20.4) | (22.9) | (25.2) | (26.5) | (27.2) | (22.0) | (15.2) | (16.3) |
| Fine motor | 50.4 | 77.2 | 92.3 | 100.6 | 99.5 | 90.9 | 87.4 | 96.9 | 98.3 | 101.7 |
| | (17.2) | (25.8) | (22.1) | (16.3) | (15.5) | (16.3) | (19.0) | (16.6) | (14.7) | (17.0) |
| Problem solving | 42.7 | 77.0 | 83.7 | 98.4 | 89.8 | 94.3 | 98.5 | 105.3 | 103.9 | 105.0 |
| | (22.2) | (23.2) | (23.0) | (17.6) | (20.3) | (23.1) | (20.1) | (16.5) | (13.2) | (18.0) |
| Personal social | 53.3 | 81.7 | 81.0 | 80.3 | 75.4 | 75.5 | 83.9 | 92.3 | 99.7 | 98.3 |
| | (21.1) | (21.5) | (18.0) | (16.0) | (18.2) | (22.9) | (23.9) | (18.9) | (15.8) | (16.5) |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 these scores are not standardized 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 these data will be to compare children within the sample which will become very useful when we analyze 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 report on the means and standard deviations for the middle six questions (i.e. the original questions) on each subscale, first for all ages combined (see Table 41 ) and then by age group (see Table 42).

Table 41: ASQ - raw scores from 6 question subscales, target child

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Communication | 43.84 | 2.27** | 2.53** | 1.26 | 1427 | 2.18* |
| | (12.57) | (1.10) | (1.09) | (1.08) | | |
| Gross motor | 44.48 | 1.90 | 1.27 | 2.18 | 1427 | 0.93 |
| | (16.39) | (1.34) | (1.47) | (1.41) | | |
| Fine motor | 49.42 | 1.03 | 1.55* | 0.97 | 1427 | 1.00 |
| | (10.55) | (0.91) | (0.90) | (0.92) | | |
| Problem solving | 48.56 | 0.76 | 0.76 | 0.95 | 1427 | 0.32 |
| | (12.15) | (1.07) | (1.10) | (1.03) | | |
| Personal social | 38.61 | 3.27*** | 2.58** | 1.92 | 1427 | 2.91** |
| | (14.72) | (1.14) | (1.14) | (1.20) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 42: ASQ - mean and sd of 6 questions subscales by age category, target child

| | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Communication | 34.9 | 47.9 | 47.1 | 47.6 | 44.1 | 45.2 | 44.8 | 41.8 | 41.0 | 42.5 |
| | (16.2) | (10.3) | (8.6) | (9.1) | (12.5) | (12.7) | (13.7) | (12.8) | (12.2) | (15.9) |
| Gross motor | 32.7 | 38.1 | 39.5 | 47.3 | 43.5 | 44.6 | 49.0 | 52.9 | 55.2 | 53.3 |
| | (11.2) | (16.7) | (13.8) | (15.1) | (16.5) | (16.2) | (16.0) | (13.7) | (7.5) | (10.4) |
| Fine motor | 44.3 | 39.6 | 52.1 | 53.7 | 49.8 | 49.0 | 47.8 | 52.5 | 52.9 | 50.8 |
| | (10.9) | (14.3) | (12.6) | (7.7) | (9.3) | (9.4) | (11.1) | (10.4) | (8.4) | (10.1) |
| Problem solving | 36.4 | 46.3 | 44.1 | 50.7 | 47.9 | 48.8 | 50.1 | 53.3 | 51.4 | 51.7 |
| | (16.0) | (13.9) | (13.5) | (10.3) | (12.0) | (13.3) | (13.1) | (10.5) | (8.6) | (9.5) |
| Personal social | 42.8 | 39.7 | 41.8 | 49.5 | 34.3 | 36.9 | 41.2 | 45.6 | 52.0 | 49.4 |
| | (13.0) | (13.6) | (12.1) | (10.2) | (11.0) | (14.9) | (13.9) | (12.3) | (10.9) | (9.8) |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 6.2.2 Spillover children

Similar results apply when we focus the discussion of ASQ-3 test performance on spillover children (aged 2-6 or 17-20 months). Table 43 and Table 44 present findings pooling all age categories together and show no statistically significant differences across treatment groups.

Table 43: ASQ - raw scores from 12 question subscales, spillover child

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Communication | 87.04 | -1.85 | 0.83 | -1.72 | 743 | 0.44 |
| | (21.31) | (2.90) | (2.69) | (2.88) | | |
| Gross motor | 80.98 | 1.55 | 1.97 | 2.71 | 743 | 0.22 |
| | (31.13) | (3.50) | (3.29) | (3.60) | | |
| Fine motor | 84.53 | 3.30 | 1.16 | 0.08 | 743 | 0.51 |
| | (24.87) | (2.90) | (2.65) | (3.11) | | |
| Problem solving | 84.13 | 3.15 | 1.76 | 2.37 | 743 | 0.35 |
| | (28.38) | (3.19) | (2.86) | (3.25) | | |
| Personal social | 83.07 | 2.06 | 1.61 | 1.24 | 742 | 0.22 |
| | (23.11) | (2.65) | (2.54) | (2.78) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 44: ASQ - raw scores from 6 question subscales, spillover child

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Communication | 44.33 | -1.39 | 0.17 | -0.83 | 743 | 0.45 |
| | (12.11) | (1.62) | (1.45) | (1.53) | | |
| Gross motor | 42.82 | 1.12 | 1.26 | 2.44 | 743 | 0.71 |
| | (15.43) | (1.65) | (1.63) | (1.70) | | |
| Fine motor | 46.62 | 2.43 | 0.49 | 0.22 | 743 | 1.00 |
| | (13.78) | (1.58) | (1.45) | (1.54) | | |
| Problem solving | 46.65 | 0.99 | 0.40 | 0.48 | 743 | 0.15 |
| | (14.34) | (1.53) | (1.46) | (1.58) | | |
| Personal social | 42.96 | 1.39 | 0.64 | 0.46 | 743 | 0.35 |
| | (14.06) | (1.45) | (1.47) | (1.44) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 Child development - anthropometrics

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 this section, we present the raw height and weight of the children in our sample, separately for target children and spillover children. We also report the following:

1. **Weight for age (Underweight)**. This particularly measures shorter-run nutritional 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, and is especially useful 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 and is the most common assessment of longer-run or chronic nutritional status. Height is a product of the full history of the child's health and nutritional inputs. As such, a deficit in height for age is generally assumed to indicate exposure to an unhealthy environment, such as poor nutrition, lack of hygiene or disease.

More specifically, the above variables are reported in terms of z-scores: a value denoting a child's placement in the distribution of some reference population. In this case, we are talking about how weight-for-age, height-for-age and weight-for-height are distributed in a reference population provided by the WHO, and how the children in our study compare. The children in this reference population are 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 [33]. The z-score tells us how many standard deviations (measured in standard deviations of the reference population) the child in question is away from the median of the reference population (and in what direction). For example, a height-for-age z-score of -1 would tell us that the child is one standard deviation smaller than the mean child in the reference population of healthy children.

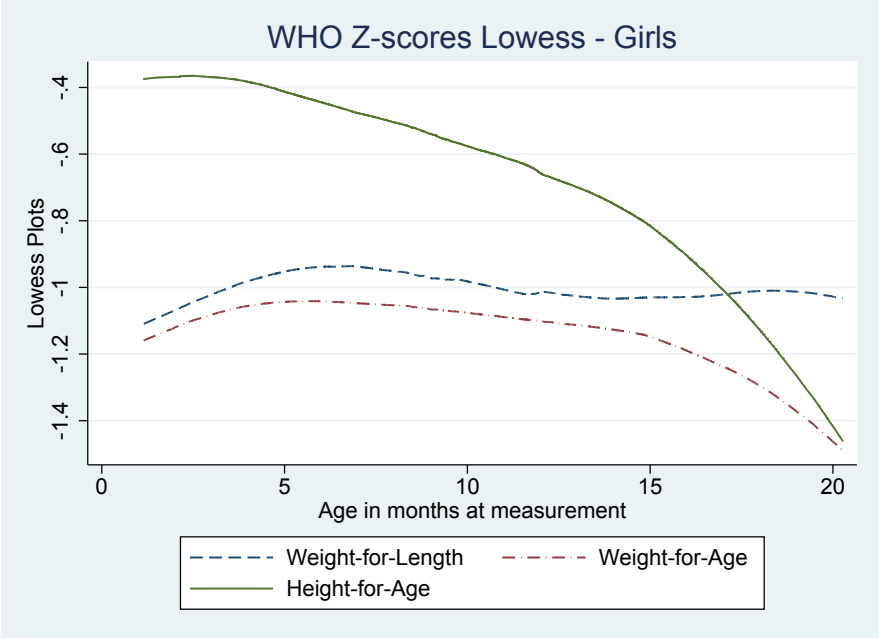
Using the WHO distributions, and the z-score which they provide, we are further able to classify children as stunted, wasted or underweight if they have a low height-for-age, weight-for-height or weight-for-age respectively, i.e. a z-score of less than -2. These are standard markers that show severe levels of disease, malnutrition or restrictions on growth. In the reference population, fewer than 3% of children would be expected to be marked as stunted, wasted or underweight, so higher levels of stunting, wasting or underweight are indicative of an overall less healthy environment for child growth.

Before discussing balance across treatment arms separately for treatment and spillover children, we present a few figures that summarize the anthropometric profile of the children in our sample, by age group. Figure 5 and Figure 6 plot lowess curves for weight-for-length, weight-for-age and height-for-age z-scores for boys and girls, respectively. Interestingly, we observe slight improvements in z-scores at the beginning of the child's life

(especially for boys) up to 6 months old, which is the age until which 45% of the children in our sample are exclusively breastfed. From 6 months onwards, we find deteriorating z-scores for height-for-age and weight-for-age as the child grows older. It is possible that the increase in malnutrition is, at least partially, related to the switch to nutrition other than breastfeeding and the related change in behaviour and practices.

Figure 5: Anthropometrics boys


Figure 6: Anthropometrics girls



Figures 7 and Figure 8 in turn show equivalent patterns for boys and girls' stunting and underweight rates, which in line with the previous figures, increase after the age of six months..

Figure 7: Stunting

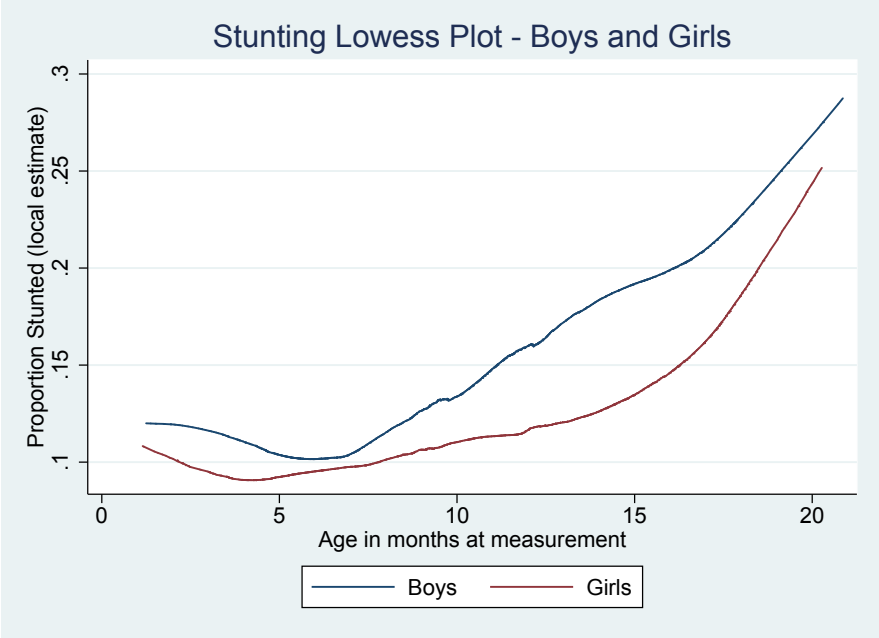
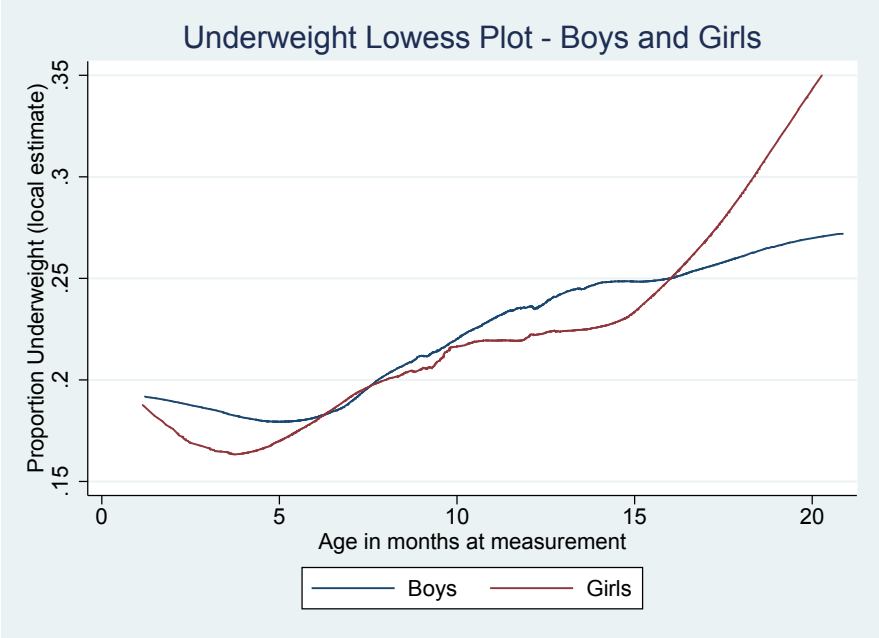



Figure 8: Underweight




Finally, Figures 9 and Figure 10 show how up until the age of 6 months the pattern of height-for-age

and weight-for-age (for boys and girls combined) indeed depends on whether the child is being exclusively breastfed or not.

Figure 9: Height-for-Age and exclusive breastfeeding

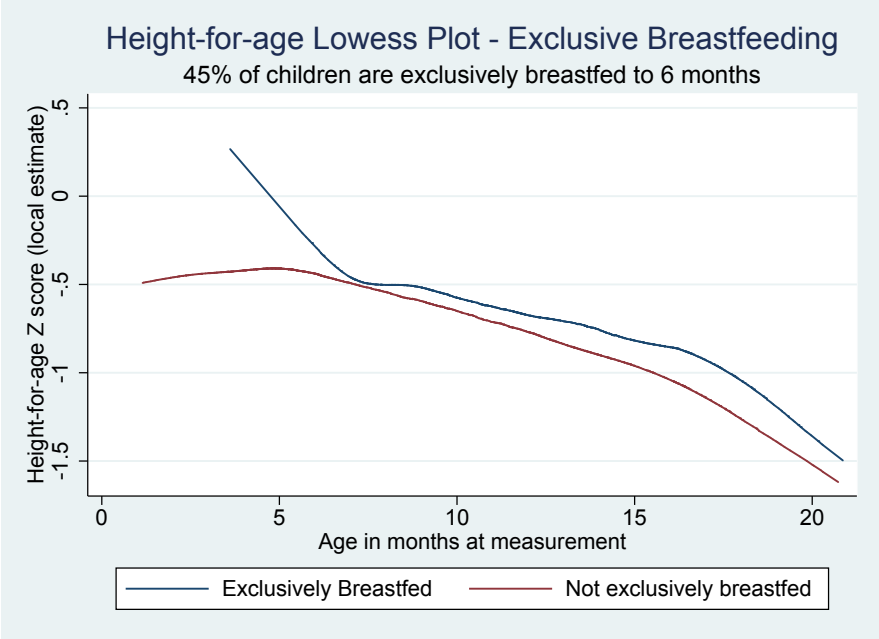



Figure 10: Weight-for-Age and exclusive breastfeeding

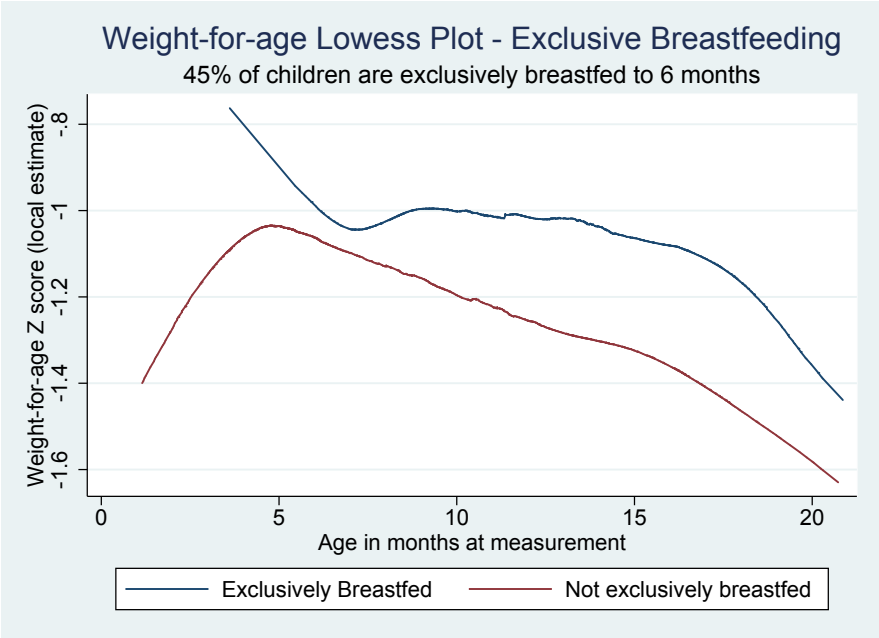


Table 45: Proportion of children who are classified as underweight, stunted or wasted, target child

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Weight-for-age | -1.11 | -0.04 | -0.02 | 0.02 | 1416 | 0.12 |
| | (1.18) | (0.11) | (0.11) | (0.12) | | |
| Height-for-age | -0.69 | 0.03 | -0.02 | 0.10 | 1408 | 0.40 |
| | (1.27) | (0.11) | (0.12) | (0.12) | | |
| Weight-for-height | -0.99 | -0.09 | -0.02 | -0.04 | 1406 | 0.27 |
| | (1.14) | (0.11) | (0.10) | (0.10) | | |
| Underweight (%) | 21.78 | 0.95 | 1.93 | -1.09 | 1416 | 0.23 |
| | (41.33) | (3.55) | (3.72) | (3.79) | | |
| Stunted (%) | 12.00 | -0.83 | 3.62 | 1.37 | 1408 | 0.99 |
| | (32.54) | (2.51) | (2.92) | (2.88) | | |
| Wasted (%) | 18.68 | -0.91 | -3.61 | -1.24 | 1406 | 0.49 |
| | (39.03) | (3.43) | (3.18) | (3.39) | | |
| Raw height (cm) | 72.03 | -0.14 | -0.34 | -0.04 | 1409 | 0.33 |
| | (4.40) | (0.37) | (0.37) | (0.41) | | |
| Raw weight (kg) | 8.06 | -0.08 | -0.10 | -0.05 | 1416 | 0.27 |
| | (1.29) | (0.12) | (0.12) | (0.12) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 6.3.1 Target children

In Table 45 we find that our target children (aged 5-16 months at the time of baseline survey) are significantly shorter and lighter than the WHO's reference population of healthy children. On average, our target children our roughly one standard deviation (of the reference population) lighter and 0.7 standard deviation shorter than children of their age from the WHO reference population. They are roughly, on average, 1 standard deviation lighter than children of their height in the WHO reference population. Moreover, 22% of our target child sample are classified as underweight, 12% are classified as stunted and 19% are classified as wasted.

Figures 11, 12 and 13 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. We also see the relative sizes of the area in each distribution which is classified as underweight, stunted or wasted. This is indicative of poor nutritional status amongst our study population.

Figure 11: Distribution of weight-for-age Z-Scores of target children compared with WHO reference population

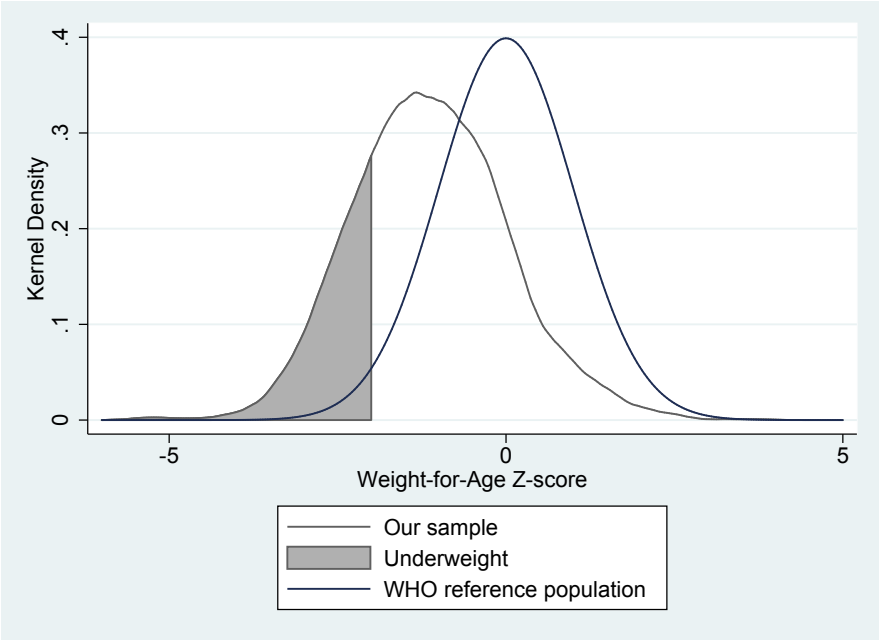



Figure 12: Distribution of height-for-age Z-Scores of target children compared with WHO reference population

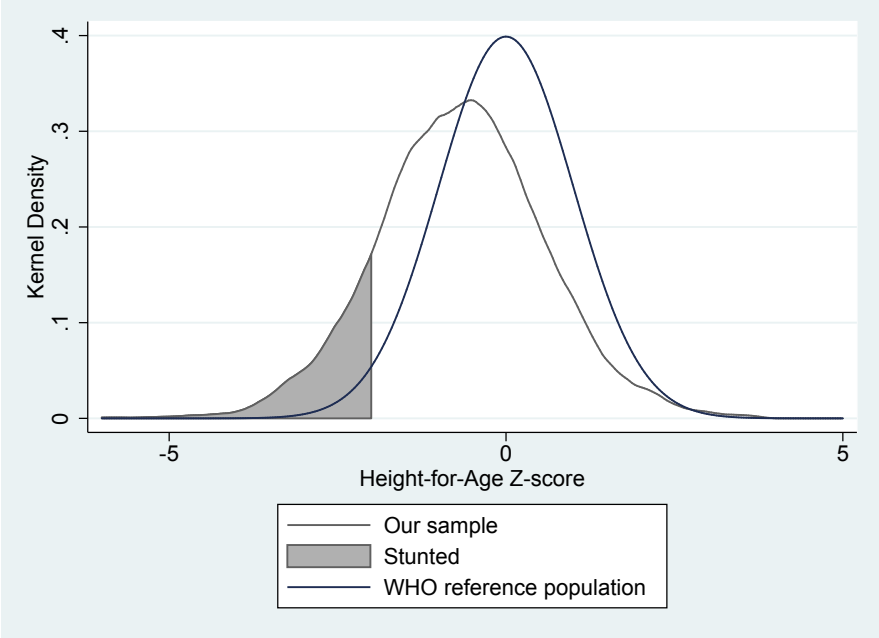
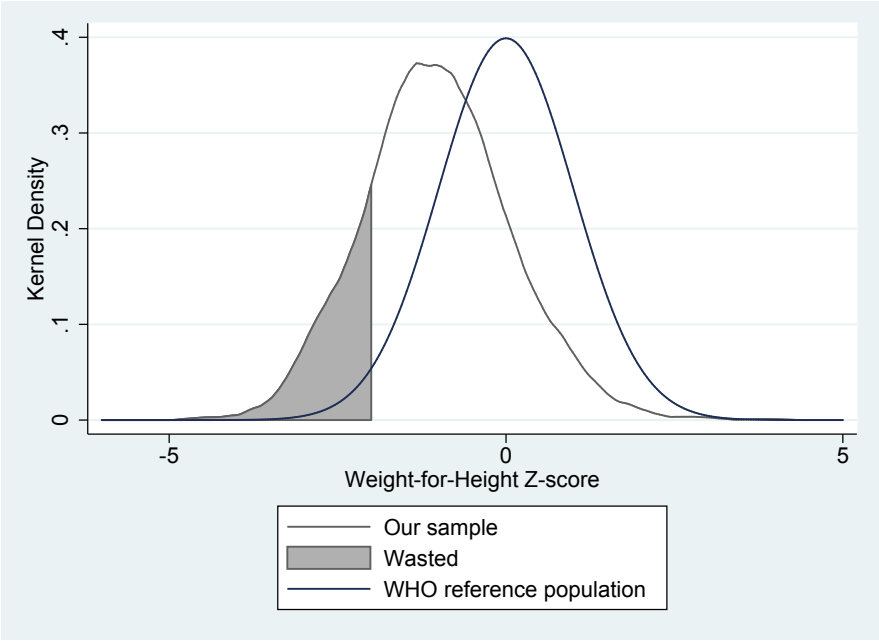


Figure 13: Distribution of weight-for-height Z-Scores of target children compared with WHO reference population




### 6.3.2 Spillover children

The anthropometric patterns observed for spillover children (aged 0-6 and 14-20 at the time of survey) in Table 46 are similar to that of target children except that here the children in the control group seem to be significantly taller and less stunted than the children in the treatment arms, the older group in particular.

Figures 14, 15 and 16 show visually the distribution of height-for-age, weight-for-age and weight-for-height z-scores relative to that of the WHO reference population. The curious shape of the sample distribution in Figure 15 is indicative of the fact that spillover children represent two separate age groups with a large gap in the middle (i.e. the target children). Since height-for-age is representative of cumulative nutritional status, we see a different lower height-for-age for those children who have lived for longer in a nutritional environment with restrictive qualities.

Table 46: Proportion of children who are classified as underweight, stunted or wasted, spillover child

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Young - Weight-for-age | -1.05 | 0.10 | 0.02 | -0.08 | 414 | 0.38 |
| | (1.12) | (0.17) | (0.16) | (0.17) | | |
| Young - Height-for-age | -0.39 | -0.05 | 0.05 | -0.19 | 410 | 0.62 |
| | (1.18) | (0.17) | (0.18) | (0.18) | | |
| Young - Weight-for-height | -0.96 | 0.21 | -0.05 | 0.05 | 409 | 0.84 |
| | (1.22) | (0.17) | (0.18) | (0.16) | | |
| Young - Underweight (%) | 14.29 | -0.88 | 4.95 | 6.51 | 414 | 0.80 |
| | (35.15) | (5.19) | (5.51) | (5.72) | | |
| Young - Stunted (%) | 7.27 | 2.10 | 0.42 | 8.73* | 410 | 1.38 |
| | (26.09) | (3.77) | (3.67) | (4.62) | | |
| Young - Wasted (%) | 18.18 | -5.68 | -0.71 | -0.18 | 409 | 0.63 |
| | (38.75) | (5.24) | (5.74) | (5.23) | | |
| Young - Raw height (cm) | 62.93 | -0.26 | 0.54 | -0.34 | 410 | 0.92 |
| | (3.69) | (0.63) | (0.56) | (0.62) | | |
| Young - Raw weight (kg) | 6.14 | 0.01 | 0.09 | -0.04 | 415 | 0.21 |
| | (1.09) | (0.19) | (0.16) | (0.18) | | |
| Old - Weight-for-age | -1.02 | -0.09 | -0.25 | -0.32 | 324 | 1.06 |
| | (1.02) | (0.20) | (0.20) | (0.20) | | |
| Old - Height-for-age | -0.71 | -0.17 | -0.48** | -0.45** | 318 | 2.62* |
| | (1.09) | (0.20) | (0.20) | (0.21) | | |
| Old - Weight-for-height | -0.93 | -0.01 | -0.03 | -0.16 | 319 | 0.29 |
| | (1.13) | (0.20) | (0.19) | (0.20) | | |
| Old - Underweight (%) | 19.40 | 7.03 | 9.87 | 9.01 | 324 | 0.85 |
| | (39.84) | (6.95) | (6.91) | (7.02) | | |
| Old - Stunted (%) | 5.97 | 11.89** | 17.78*** | 18.17*** | 318 | 5.39*** |
| | (23.87) | (5.31) | (5.61) | (6.16) | | |
| Old - Wasted (%) | 14.93 | 5.31 | 2.36 | 4.61 | 319 | 0.24 |
| | (35.90) | (7.16) | (6.33) | (6.63) | | |
| Old - Raw height (cm) | 78.78 | -0.40 | -1.39** | -1.13* | 319 | 2.50* |
| | (3.02) | (0.53) | (0.58) | (0.58) | | |
| Old - Raw weight (kg) | 9.35 | -0.04 | -0.19 | -0.27 | 324 | 0.65 |
| | (1.12) | (0.21) | (0.22) | (0.22) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 14: Distribution of weight-for-age Z-Scores of spillover children compared with WHO reference population

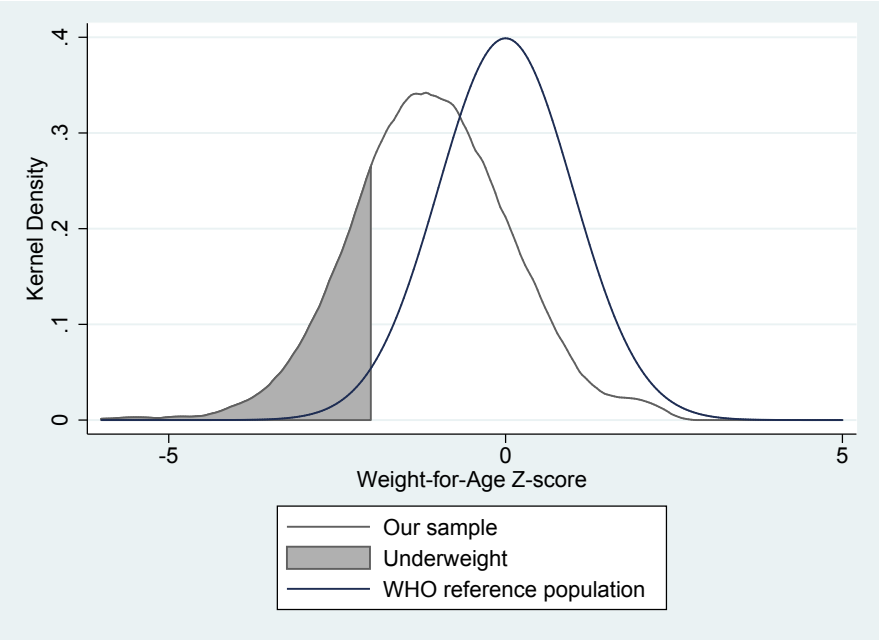



Figure 15: Distribution of height-for-age Z-Scores of spillover children compared with WHO reference population

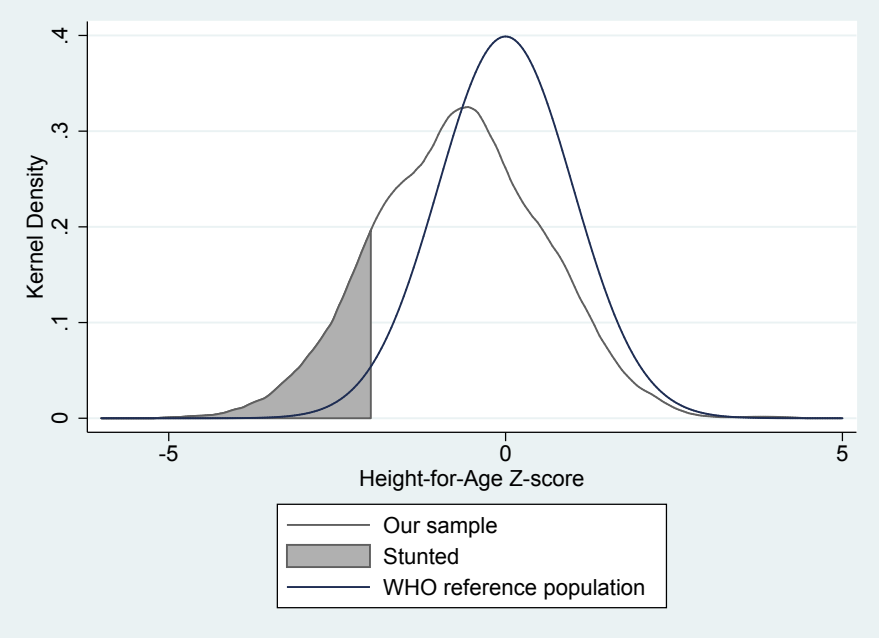
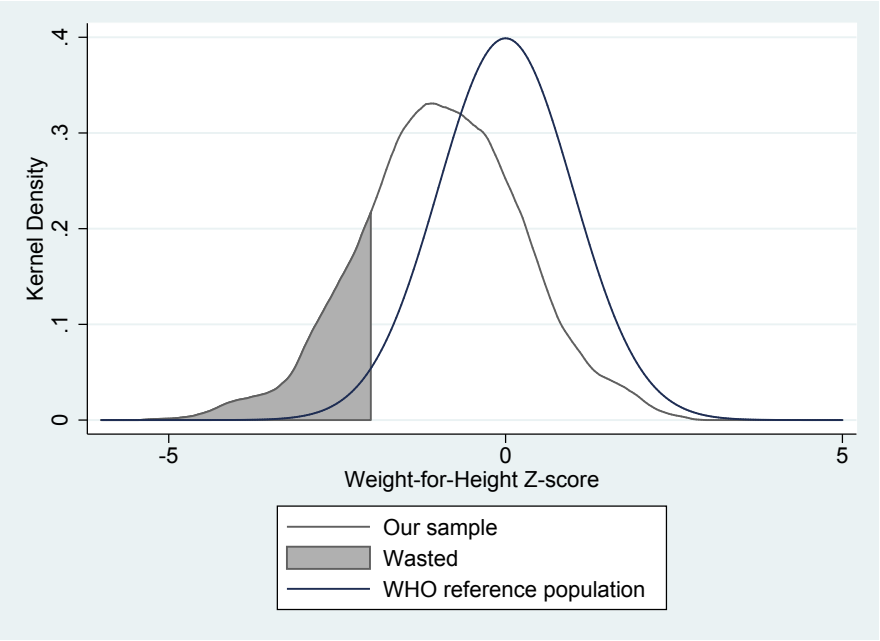


Figure 16: Distribution of weight-for-height Z-Scores of spillover children compared with WHO reference population




## 6.4 Child development - morbidity

We ask whether the target and spillover children in our sample experienced any of nine symptoms in the past two weeks, including fever, diarrhoea and coughing, among other.

### 6.4.1 Target children

Table 47 shows that 76% of target children experiences at least one of the nine symptoms asked about within the last two weeks. About 46% experienced fever and 55% coughing. One third of children experienced both of these symptoms, which is indicative of malaria, a predominant illness in the study areas. Also relatively common was diarrhoea, at 14%, vomiting (27%), and tiredness (17%) and paleness (19%). Except for one minor difference in the variable 'vomitting' (significant only at 10%), these symptoms are balanced across the study groups.

### 6.4.2 Spillover children

We show the same information for spillover children in Table 48. While we find that the same percentage of children experienced any of the symptoms, it appears that spillover children where somewhat less likely to have experienced several symptoms over the last two weeks. Still, occurrences remain high, with 39% having experienced fever, 51% coughing, 32% vomited and 11% had diarrhoea. All variables are balanced.

Table 47: Frequency (%) of Symptoms in the Last Two Weeks - target children

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Fever | 45.61 | -3.28 | 2.50 | -0.03 | 1426 | 0.81 |
| | (49.88) | (3.71) | (4.00) | (3.80) | | |
| Diarrhea | 13.96 | -1.71 | -3.88 | -2.79 | 1418 | 0.77 |
| | (34.71) | (2.77) | (2.68) | (2.50) | | |
| Cough | 54.96 | -5.10 | 2.61 | -1.11 | 1427 | 1.21 |
| | (49.82) | (4.30) | (4.32) | (4.28) | | |
| Vomit | 26.63 | -6.80* | -7.17* | -1.56 | 1427 | 1.60 |
| | (44.26) | (4.05) | (4.05) | (4.18) | | |
| Skin Rash | 15.86 | -0.28 | -1.54 | 3.22 | 1427 | 1.05 |
| | (36.59) | (3.32) | (2.98) | (2.97) | | |
| Itching | 3.69 | -1.43 | -1.25 | 1.15 | 1425 | 1.25 |
| | (18.89) | (1.39) | (1.33) | (1.61) | | |
| Stomach Pain | 5.38 | 0.03 | 0.34 | 1.82 | 1418 | 0.38 |
| | (22.60) | (1.81) | (1.91) | (1.91) | | |
| Tiredness | 16.71 | -1.98 | -2.31 | 2.94 | 1425 | 0.94 |
| | (37.36) | (3.54) | (3.50) | (3.76) | | |
| Paleness | 28.61 | -3.40 | -4.29 | 1.67 | 1426 | 1.21 |
| | (45.26) | (3.48) | (3.67) | (3.80) | | |
| At least one of the above symptoms | 75.64 | -0.57 | 1.93 | 2.14 | 1427 | 0.33 |
| | (42.99) | (3.62) | (3.37) | (3.48) | | |
| Cough combined with fever | 33.99 | -4.73 | 4.65 | -0.38 | 1426 | 1.95 |
| | (47.44) | (3.76) | (4.15) | (3.79) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 48: Frequency (%) of Symptoms in the Last Two Weeks - spillover children

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Fever | 38.55 | 6.02 | 2.51 | -1.71 | 743 | 0.90 |
| | (48.81) | (4.69) | (5.36) | (5.29) | | |
| Diarrhea | 10.67 | 3.69 | 3.08 | 3.16 | 736 | 0.61 |
| | (30.97) | (3.12) | (3.43) | (3.23) | | |
| Cough | 50.84 | -1.38 | 0.21 | -1.36 | 743 | 0.06 |
| | (50.13) | (5.23) | (5.29) | (5.17) | | |
| Vomit | 32.40 | -7.40 | -8.72 | -7.14 | 743 | 0.97 |
| | (46.93) | (5.21) | (5.53) | (5.18) | | |
| Skin Rash | 13.97 | 0.16 | 2.88 | 2.88 | 743 | 0.36 |
| | (34.76) | (3.28) | (3.88) | (3.87) | | |
| Itching | 5.59 | -1.78 | -1.90 | 0.73 | 743 | 0.69 |
| | (23.03) | (2.20) | (2.18) | (2.50) | | |
| Stomach Pain | 6.78 | 0.32 | -0.40 | -0.99 | 738 | 0.10 |
| | (25.21) | (2.49) | (2.40) | (2.42) | | |
| Tiredness | 14.53 | 1.78 | -2.29 | -0.84 | 741 | 0.36 |
| | (35.33) | (4.09) | (3.77) | (3.76) | | |
| Paleness | 24.58 | -5.02 | -0.37 | -4.05 | 743 | 0.66 |
| | (43.18) | (4.48) | (4.43) | (4.18) | | |
| At least one of the above symptoms | 75.98 | -3.15 | -3.87 | -6.50 | 743 | 0.56 |
| | (42.84) | (4.61) | (4.78) | (5.08) | | |
| Cough combined with fever | 28.49 | 0.86 | 2.03 | 0.46 | 743 | 0.06 |
| | (45.26) | (5.04) | (4.86) | (4.65) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

## 6.5 Child Nutrition, Healthcare and Birth

### 6.5.1 Birth and breastfeeding

The World Health Organization recommends that all children (except for those with certain medical conditions) be breastfed exclusively for the first six months of life. This means that they do not consume any substance other than breastmilk (including water) until they are 6 months old. In Table 49, we see that only 42% of children in our study sample satisfy this criteria. We also see, however, that the vast majority of children (99%) are breastfed at least at some point, even though many are introduced to water, milk and infant formula before the first 6 months of life are over.

Table 49: Breastfeeding and Birth

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Born in Hospital | 95.11 | -2.19 | -3.51* | -1.77 | 2170 | 1.36 |
| | (21.58) | (1.52) | (1.99) | (1.54) | | |
| Birth Weight (grams) | 2,867.66 | -30.04 | -0.44 | -22.55 | 2145 | 0.30 |
| | (486.35) | (43.92) | (38.07) | (35.17) | | |
| Low weight ($<$ 2500 grams) | 29.41 | 1.73 | -0.89 | -2.83 | 2145 | 0.75 |
| | (45.61) | (3.55) | (3.56) | (3.33) | | |
| Child was breastfed | 99.44 | -0.18 | -0.15 | 0.38 | 2170 | 1.14 |
| | (7.50) | (0.48) | (0.46) | (0.36) | | |
| Exclusively breastfed to 6 months | 41.54 | 4.64 | 2.03 | 5.78 | 2170 | 0.72 |
| | (49.33) | (4.29) | (4.23) | (4.38) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 weight at time of birth is recorded for almost all children, and provides insight into their genetic endowment, their development in the womb and whether or not they were born preterm. The data do not allow for detailed descriptions of prevalence of pre-term births, but Table 49 does show that 30% of children are born with a weight below 2,500 grams.

We can attribute our good knowledge of birth weight in part to the fact that over 90% of target children were born in a hospital.

### 6.5.2 Nutrition

Good nutrition is vital for the physical, cognitive and non-cognitive development of a child. Whilst anthropometric measurements are a good indicator of long term nutritional status, it is useful to also directly capture information on children's diets. To track nutritional inputs, we asked the respondent to recall all the food and liquid given to the target child in the past 24 hours, and use this as representative of the child's diet

Table 50: Percentage of children consuming foods in past 24 hours

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Rice/Bread/Grains | 67.99 | 2.03 | -0.67 | 2.64 | 2163 | 0.63 |
| | (46.69) | (2.73) | (2.66) | (2.42) | | |
| Roots/Potatos/Tubers | 35.23 | 3.13 | 0.13 | 3.25 | 2163 | 0.65 |
| | (47.81) | (3.16) | (3.29) | (3.52) | | |
| Visible fat/Ghee/Butter | 27.46 | 0.84 | 2.72 | 2.65 | 2163 | 0.16 |
| | (44.67) | (4.83) | (4.95) | (4.72) | | |
| Nuts | 12.88 | 1.09 | 1.94 | 1.81 | 2163 | 0.16 |
| | (33.53) | (3.34) | (3.16) | (3.30) | | |
| Sugars | 42.42 | 3.76 | 2.22 | 2.74 | 2163 | 0.45 |
| | (49.47) | (3.30) | (3.49) | (3.68) | | |
| Pulses/Beans | 45.45 | 5.20 | 1.69 | 0.46 | 2163 | 0.87 |
| | (49.84) | (3.52) | (3.39) | (3.44) | | |
| Milk and milk products | 21.02 | -0.17 | 0.58 | 1.28 | 2163 | 0.06 |
| | (40.79) | (4.03) | (3.96) | (4.00) | | |
| Mother's milk | 77.84 | 3.72 | 4.30* | 2.08 | 2163 | 1.25 |
| | (41.57) | (2.58) | (2.38) | (2.61) | | |
| Egg | 1.14 | 0.17 | 2.44*** | 1.47* | 2163 | 3.48** |
| | (10.61) | (0.73) | (0.85) | (0.78) | | |
| Meat and meat products | 3.22 | -0.99 | -0.54 | -0.06 | 2163 | 0.35 |
| | (17.67) | (1.19) | (1.14) | (1.22) | | |
| Green leafy vegetables | 7.20 | 1.56 | 2.45 | 3.21 | 2163 | 0.53 |
| | (25.87) | (2.70) | (2.60) | (2.83) | | |
| Dark coloured vegetables | 32.95 | 5.59 | 4.37 | 1.06 | 2163 | 1.13 |
| | (47.05) | (3.44) | (3.67) | (3.50) | | |
| Other vegetables | 23.48 | 5.38 | 3.12 | 3.65 | 2163 | 0.74 |
| | (42.43) | (3.84) | (3.78) | (3.65) | | |
| Dark coloured fruits | 20.45 | 3.38 | 0.44 | 1.29 | 2163 | 0.34 |
| | (40.38) | (3.68) | (3.44) | (3.68) | | |
| Other fruits | 11.74 | 4.46* | 1.65 | -0.03 | 2163 | 1.26 |
| | (32.22) | (2.61) | (2.83) | (2.63) | | |
| Commercial formula feeds | 24.05 | 2.95 | 1.66 | 1.78 | 2163 | 0.20 |
| | (42.78) | (3.80) | (3.77) | (3.78) | | |
| Formula milk | 7.77 | -1.43 | -0.44 | -2.93 | 2163 | 0.90 |
| | (26.79) | (1.81) | (2.10) | (1.91) | | |
| Powder milk | 6.44 | -1.41 | -0.19 | -2.35 | 2163 | 1.02 |
| | (24.57) | (1.68) | (1.88) | (1.61) | | |
| Sweet & salted snacks | 47.35 | 6.10* | 9.97** | 7.30** | 2163 | 2.70** |
| | (49.98) | (3.45) | (3.85) | (3.34) | | |
| Unpackaged foods (e.g. samosa) | 2.08 | 0.15 | 0.95 | 2.38** | 2163 | 1.53 |
| | (14.30) | (0.96) | (1.20) | (1.20) | | |
| Water | 75.57 | 3.20 | 3.72 | 1.76 | 2163 | 0.57 |
| | (43.01) | (3.85) | (2.99) | (3.13) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

more broadly. Table 50 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, which include rice (consumed by 68% of children in the past 24 hours), and pulses (consumed by 46%), yet relatively low in foods high in proteins and heme iron. Sugars as well as sweet and salted snacks were also commonly consumed (by 42 and 47% of children respectively).

We use this detailed information to construct an indicator of dietary diversity. Dietary diversity, defined as the number of food groups 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 [2] which was shown to correlate well with broad measures of nutritional status. We place these above food groups into the seven larger food groups[17], scoring each child as a 1 if they consumed some food in this food group in the past 24 hours and as a 0 if they did not.

We then construct dietary diversity scores by simply summing the total number of these seven food groups consumed by the child in the past 24 hours. Arimond and Ruel*[2]* 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 food groups, those who consumed 3 to 4 groups and those who consumed 5 to 7 groups.

Table 51 presents our results. 71% of target children had consumed 'starchy staples' in the past 24 hours, 45% legumes yet only 20% had consumed dairy (excluding breast milk), and even less (4%) had consumed meat, fish or egg. 42% on the other hand consumed Vitamin A rich fruit and vegetables and a further 32% other fruits and vegetables.

In terms of the aggregate scores of diet diversity, 51% - half of our of target children - had consumed only two or less distinct food groups in the past 24 hours, which Arimond and Ruel term as low diet diversity. 29% of target children had a middle diet diversity score of 3 to 4 whilst 20% had a high diversity score of 5 or more. These measures were balanced across treatment and control (we only observe two small imbalances in the consumption of fruits and vegetables).

## 6.6 Maternal Health and Education

This section concentrates on the mothers or the study target children. We will discuss aspects such as maternal health, education, empowerment, and knowledge, all of which have been shown important in determining child outcomes, such as health and development.

[17] *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.[2]*

Table 51: Dietary Diversity Measures

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Starchy Staples | 70.64 | 0.86 | -1.36 | 1.85 | 2163 | 0.44 |
| | (45.58) | (2.62) | (2.54) | (2.45) | | |
| Legumes/Pulses | 45.45 | 5.20 | 1.69 | 0.46 | 2163 | 0.87 |
| | (49.84) | (3.52) | (3.39) | (3.44) | | |
| Dairy (excluding breast milk) | 20.45 | -0.34 | 0.44 | 0.55 | 2163 | 0.02 |
| | (40.38) | (4.04) | (3.98) | (4.06) | | |
| Meat, fish, egg | 4.36 | -0.82 | 1.54 | 0.48 | 2163 | 0.98 |
| | (20.43) | (1.43) | (1.45) | (1.46) | | |
| Vitamin A rich fruit and vegetables | 42.05 | 6.00* | 2.42 | 1.82 | 2163 | 1.09 |
| | (49.41) | (3.39) | (3.52) | (3.60) | | |
| Other fruit and vegetables | 32.20 | 7.28** | 4.05 | 2.93 | 2163 | 1.76 |
| | (46.77) | (3.22) | (3.37) | (3.24) | | |
| Foods made with oil, fate or butter | 27.46 | 0.84 | 2.72 | 2.65 | 2163 | 0.16 |
| | (44.67) | (4.83) | (4.95) | (4.72) | | |
| Dietary diversity score of 0 to 2 | 51.14 | -3.84 | -2.03 | -2.25 | 2163 | 0.44 |
| | (50.03) | (3.41) | (3.31) | (3.51) | | |
| Dietary diversity score of 3 to 4 | 28.60 | 2.50 | -1.63 | 0.58 | 2163 | 0.75 |
| | (45.23) | (2.82) | (2.56) | (2.82) | | |
| Dietary diversity score of 5 to 7 | 20.27 | 1.34 | 3.66 | 1.67 | 2163 | 0.43 |
| | (40.24) | (3.34) | (3.23) | (3.06) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

### 6.6.1 Maternal health

Maternal physical and mental health and wellbeing are a crucial determinant of a mother's relationship and interaction with her child. We start by discussing some physical attributes of the mother, before turning in more detail to her mental health status.

The typical study target child has a mother that is 1.5m tall and weighs 47kg, as shown in Table 52, implying a healthy body mass index (BMI) of 20.6. This contrasts somewhat the results on stunting and wasting we saw above.

Despite a healthy BMI, a relatively large percentage of mothers has difficulties with some basic tasks, as is shown in Table 53. 23% report that they have difficulty walking for one kilometer, 20% have difficulties bowing, squatting and/or kneeling, and 14% find it hard to carry a heavy load for 20m. We find that 4% of mothers have difficulties to bathe without assistance. These are balanced across treatment arms.

There is substantial evidence that maternal depression affects parenting behaviours and child outcomes

Table 52: Mother's anthropometry

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Mother's height (cm) | 150.87 | 0.22 | -0.34 | 0.18 | 2151 | 0.95 |
| | (5.46) | (0.32) | (0.37) | (0.38) | | |
| Mother's weight (kg) | 47.03 | -0.28 | -0.58 | -0.31 | 2150 | 0.27 |
| | (8.55) | (0.72) | (0.65) | (0.67) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 53: Any difficulty performing the following tasks (%)

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Walk 1km | 22.68 | 0.19 | 0.79 | -0.44 | 2151 | 0.06 |
| | (41.92) | (3.05) | (3.09) | (3.36) | | |
| Bathe without assistance | 4.33 | -1.53 | 0.32 | -0.45 | 2167 | 1.03 |
| | (20.38) | (1.13) | (1.47) | (1.26) | | |
| Bow, squat and kneel | 20.23 | -0.12 | 0.74 | 2.14 | 2165 | 0.22 |
| | (40.21) | (3.06) | (3.13) | (3.11) | | |
| Carry a heavy load for 20m | 14.31 | 3.04 | 3.40 | 3.80 | 2167 | 0.84 |
| | (35.05) | (2.88) | (2.82) | (2.82) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

[18][17]. Evidence suggests that the most significant negative impacts of depression are for mothers of infants, the age of our target children[18].

We measured depressive symptoms in our sample of biological mothers using a shortened version of the Center for Epidemiologic Studies Depression Scale [25], a short self-report scale that is useful in study settings when full clinical assessments would be infeasible. Mothers were asked ten questions on whether they experienced different symptoms of depression over the last seven days. The original questionnaire offers the mothers one of four options: (1) 'almost never or never (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)'.

Piloting of these questions revealed that the instrument is a very difficult one to implement in our study setting. 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. To facilitate the administration, we therefore decided to reduce the response options, asking the respondent only to state that the symptom was experienced or not. While this approach made the administration easier, concerns about the quality of the translation remains and we caution to interpret this information as indicative at best. In general we 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.

Table 54 displays the ten questions mothers were asked, and indicates how many mothers responded in the positive to them. We can see that typically around 30-40% of mothers responded in the affirmative to questions, such as whether they are usually felt bothered by things within the past week, whether they had trouble keeping their minds off of things, or felt depressed. At the same time, we also see that almost half (47%) of mothers report to feel hopeful about the future and 81% felt happy in the past week.

Table 54: Maternal depression questions: felt the following in the past week (%)

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Unusually bothered by things? | 26.37 | 1.75 | 5.48* | 0.44 | 2168 | 1.61 |
| | (44.10) | (3.00) | (2.95) | (2.88) | | |
| Trouble keeping your mind on things? | 34.72 | 7.37* | 7.50** | 0.10 | 2166 | 2.68** |
| | (47.65) | (3.78) | (3.75) | (3.67) | | |
| Feel depressed? | 42.37 | 0.54 | 2.17 | -2.82 | 2167 | 0.66 |
| | (49.46) | (4.06) | (3.55) | (3.81) | | |
| Feel pain in everyday work? | 47.65 | 2.45 | 3.87 | 3.65 | 2167 | 0.40 |
| | (49.99) | (4.21) | (3.88) | (3.97) | | |
| Feel hopeful about the future? | 57.25 | -0.64 | -0.62 | -0.87 | 2167 | 0.02 |
| | (49.52) | (3.41) | (3.27) | (3.55) | | |
| Feel fearful? | 28.63 | 0.05 | -0.49 | -1.45 | 2167 | 0.11 |
| | (45.24) | (3.09) | (3.19) | (3.13) | | |
| Sleep with worry? | 35.78 | 1.72 | -1.08 | 0.26 | 2167 | 0.26 |
| | (47.98) | (3.33) | (3.05) | (2.88) | | |
| Feel happy? | 81.13 | 3.23 | 3.10 | 2.60 | 2166 | 0.75 |
| | (39.16) | (2.32) | (2.40) | (2.52) | | |
| Feel lonely? | 36.98 | -3.03 | -2.39 | -1.68 | 2165 | 0.33 |
| | (48.32) | (3.15) | (3.24) | (3.31) | | |
| Felt not active? | 38.98 | -2.98 | 2.34 | 0.09 | 2166 | 0.74 |
| | (48.82) | (3.66) | (3.36) | (3.60) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

We sum over all the responses to create a total score which takes ten as the maximum value, reverse-scoring questions about experiencing positives states of mind. The higher the score, the higher measured depression. Table 55 shows that the depression score is balanced across treatment arms.

Table 55: Maternal Depression Scores (%)

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Depression Score (/10) | 3.53 | 0.05 | 0.15 | -0.03 | 2168 | 0.37 |
| | (2.41) | (0.20) | (0.19) | (0.20) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 addition to specifically measuring the symptoms of clinical depression, we have also measured the emotional and physical wellbeing of mothers more generally. In Table 56, we list the questions asked to gauge a mother's welfare, self-worth and positive outlook on life. Respondents are asked whether they have felt the given emotion or thought in the past 15 days, with the option of responding either "Yes" or "No". In Table 56, those who have responded "Don't Know", or refused a given question, are excluded. We see from the significant variation in "Yes" and "No" answers that the questions are effective at addressing varied aspects of emotional wellbeing. Also of interest is the fact that for one question in particular, "Did you feel like people loved you?", the number of "Don't Know" answers and refusals jumped sharply, suggesting that either the question is an inherently difficult one or that some respondents did not feel comfortable with their interviewers. If we use Cronbach's alpha to calculate the level of internal validity, we get a value of 0.6839, which is close to the benchmark of 0.7 we apply to the ASQ tests.

Using the answers to the questions in Table 56, we calculate a maternal welfare score out of 14 by adding up the number of "Yes" answers. Those who answered "Don't Know" or refused are considered to have answered "No" for the purposes of this scoring. This measure of maternal wellbeing is arbitrary, but gives insight into the distribution of the emotional capacity for motherhood beyond the explicitly clinical. The distribution of these scores is presented in Table 57, which shows that the mean score for the control group is 9.77, and that this is distributed fairly evenly over treatment groups, with only NE group differing at the 10% level. This means that, of the 14 questions asked to measure maternal wellbeing, mothers on average answered all but 4 in a positive way. The subsequent breakdowns in 57 show that for all but 12% of mothers, more than half the questions were answered in a positive way. It is important to reiterate that this measure does not tell us anything specific or important about the mothers in our sample, but its even distribution does suggest that the basic emotional capacity necessary for raising children does not differ significantly by treatment group.

### 6.6.2 Maternal education

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 [8].

In this section we provide more detailed information on education of the main caretaker of the target

Table 56: Answers to Maternal Welfare Questions (%)

| | No | Yes |
|---|---|---|
| Did you feel your life would be happy in the future? | 30.4 | 69.6 |
| Did you feel your existence helped others in some way? | 45.8 | 54.2 |
| Have you been feeling without-worry? | 56.9 | 43.1 |
| Did you wish to mix with people? | 10.9 | 89.1 |
| Did you feel you had energy for anything else apart from work? | 34.0 | 66.0 |
| Think about problems had, did you face them well? | 24.5 | 75.5 |
| Were you able to think clearly about things? | 15.0 | 85.0 |
| Did you feel good about yourself? | 19.6 | 80.4 |
| Did you feel you could talk about your feelings with someone? | 44.8 | 55.2 |
| Do you feel that you have the ability to do your current and future duties? | 17.9 | 82.1 |
| Did you take decisions on your own, for anything? | 55.1 | 44.9 |
| Did you feel like people loved you? | 14.9 | 85.1 |
| Were you interested in doing new things? | 33.5 | 66.5 |
| Did you ever feel cheerful? | 9.3 | 90.7 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 57: Maternal Welfare Scores (%)

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Maternal welfare score | 9.77 | 0.30* | -0.00 | -0.05 | 2168 | 2.01 |
| | (2.69) | (0.16) | (0.17) | (0.18) | | |
| Between 0 and 6 (%) | 12.05 | -1.62 | -0.96 | 2.37 | 2168 | 1.26 |
| | (32.59) | (1.90) | (2.04) | (2.17) | | |
| Between 7 and 11 (%) | 60.64 | -4.03 | 2.69 | -4.63 | 2168 | 2.42* |
| | (48.90) | (3.00) | (2.92) | (3.11) | | |
| Between 12 and 14 (%) | 27.31 | 5.65** | -1.73 | 2.27 | 2168 | 2.37* |
| | (44.60) | (2.66) | (2.62) | (2.97) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

child, which is in 92% of the cases the biological mother (not shown).

We see from Table 58 that she is on average 25 years of age. About 10% have basically no education (5% none, 4% basic literacy/numeracy only, and 1% did not complete class 1). Almost one third (32%) completed

Table 58: Highest Level of Education Completed - child's main caregiver (%)

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Age (years) | 25.15 | 0.19 | 0.34 | 0.25 | 2163 | 0.53 |
| | (4.11) | (0.28) | (0.27) | (0.29) | | |
| None | 5.08 | -0.61 | 1.17 | 0.84 | 2170 | 0.49 |
| | (21.97) | (1.41) | (1.80) | (1.59) | | |
| Basic literacy/numeracy only | 4.14 | -0.04 | 1.76 | 0.49 | 2170 | 0.47 |
| | (19.93) | (1.35) | (1.66) | (1.42) | | |
| Less than Class 1 | 1.32 | -0.38 | -0.24 | -0.02 | 2170 | 0.14 |
| | (11.41) | (0.72) | (0.73) | (0.70) | | |
| Class 1-5 | 16.54 | -3.69 | -4.58* | -4.71** | 2170 | 1.64 |
| | (37.19) | (2.46) | (2.54) | (2.22) | | |
| Class 6-8 | 16.35 | 2.27 | 4.54* | 4.35* | 2170 | 1.53 |
| | (37.02) | (2.35) | (2.43) | (2.54) | | |
| Class 9 | 30.45 | -0.84 | 0.08 | -1.80 | 2170 | 0.17 |
| | (46.06) | (3.13) | (3.23) | (2.90) | | |
| Matriculation | 11.84 | 5.10** | 1.73 | 2.02 | 2170 | 1.76 |
| | (32.34) | (2.27) | (2.27) | (2.02) | | |
| Higher Secondary | 7.71 | -1.00 | -3.60** | -0.87 | 2170 | 2.89** |
| | (26.69) | (1.61) | (1.40) | (1.62) | | |
| Graduate and higher | 6.20 | -0.99 | -1.20 | -0.47 | 2170 | 0.25 |
| | (24.14) | (1.58) | (1.53) | (1.49) | | |
| Vocational | 0.38 | 0.18 | 0.34 | -0.01 | 2170 | 0.29 |
| | (6.13) | (0.41) | (0.43) | (0.36) | | |
| Don't Know | 0.00 | 0.00 | 0.00 | 0.18 | 2170 | . |
| | (0.00) | (0.00) | (0.00) | (0.19) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

not more than primary education (class 1 to 8), further 30% completed class 9, the first year of secondary education. Almost 12% completed matriculation and 7% higher secondary school. Less than 7% of mothers completed any higher degree of schooling. These statistics are generally balanced across treatment arms, particularly considering the F-stats.

### 6.6.3 Maternal knowledge

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 hypothesize the home-visiting intervention may work. Therefore, it is important to understand the level of mothers' knowledge of this area and to assess how it will be affected by the program.

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 [21]. 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". The questions and responses given are listed in Table 59.[18] Some interesting observations from this table include that almost half (46%) of mothers believe it to be true that too much love and attention spoil a child, and 62% believe it to be true that carrying a baby when it cries results in the baby wanting to be carried all the time. Mothers further believe child-to-child interaction to be very important, 96% say it is true that it is important for a busy mother to spend time with a young child and yet 56% believe that it is better for children to play on their own than with parents.

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.

Table 60 presents the percentage scores of the mothers on this instrument for each domain and the the instrument as a whole.[19] 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".

## 6.7 Quantity and Quality of Maternal Time

In section 3.4 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 61 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 (about twenty minutes). Instead, a large proportion of their time was spend in housework activities. Four to five hours were spent on things like cooking, cleaning and washing

[18] One question we added is the last one, whether "It is important for children to play with their father as well". The overwhelming majority of mothers (93%) report this statement to be true.

[19] We did not include here the question we added on the importance of fathers interaction with the child.

Table 59: Answers to Development Knowledge Questions (%)

| | Not True | Partly True | True |
|---|---|---|---|
| Too much love and attention spoil a child | 39.8 | 14.1 | 46.0 |
| Babies carried when crying want to be carried all the time | 31.2 | 7.2 | 61.6 |
| Child intelligence is difficult to change after birth | 63.0 | 4.7 | 32.3 |
| It is better for children to play on their own than with parents | 28.6 | 13.5 | 57.9 |
| Young children only understand words they can say | 48.6 | 7.9 | 43.4 |
| Children who know lots of words learn to read earlier | 5.2 | 3.8 | 90.9 |
| Praising a child too much will make her a show-off | 22.9 | 6.3 | 70.7 |
| It is important for a busy mother to spend time with a young child | 2.4 | 2.0 | 95.6 |
| Mother's interaction with a young child affects their future intelligence | 3.3 | 1.6 | 95.1 |
| Children who know more words do better in school | 3.9 | 4.9 | 91.3 |
| Hitting your child might be a good way of teaching them | 78.6 | 5.3 | 16.1 |
| Children who play more do better in school (?) | 20.8 | 7.7 | 71.5 |
| When your child hits another child, you should hit him back | 58.0 | 6.1 | 36.0 |
| A one-year-old knows right from wrong | 89.6 | 3.1 | 7.2 |
| It is important to teach one-year-olds the alphabet | 47.2 | 3.7 | 49.1 |
| Parents do not need to teach talking, babies learn themselves | 81.2 | 6.0 | 12.8 |
| Children that perform better in school earn more as adults | 8.3 | 13.3 | 78.4 |
| It is important for children to play with their father as well | 3.7 | 2.8 | 93.4 |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

clothes. Mothers spent, on average, three hours looking after small children, 72 minutes of the day playing with them and 14 minutes reading or telling stories to them. This is generally balanced between treatment and control communities. One imbalance observed is in the reported time spent watching TV, which is on average one hour in the control group, and between 5-13 minutes more in treatment groups.

We collected more detailed information on what play activities household members performed with the target children. This was done as part of the Family Care Indicators (FCI) questionnaire developed by UNICEF, which we will discuss in the next section.

## 6.8 Quality of home environment for stimulation

We will discuss these, as well as the quality of the home environment for stimulation in more detail in this section.

The quality of the home environment in terms of the amount of stimulation it provides is crucial for

Table 60: Knowledge of Infant Development Inventory Score

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Attention/Praise (/9) | 5.08 | 0.01 | 0.11 | 0.12 | 2168 | 0.51 |
| | (1.84) | (0.13) | (0.13) | (0.13) | | |
| Hitting (/6) | 4.84 | -0.09 | 0.04 | 0.05 | 2168 | 0.48 |
| | (1.38) | (0.12) | (0.12) | (0.14) | | |
| Schooling (/12) | 9.41 | 0.05 | 0.16* | 0.15 | 2168 | 1.45 |
| | (1.30) | (0.09) | (0.09) | (0.09) | | |
| Maternal interaction (/9) | 7.47 | 0.11 | 0.14* | 0.06 | 2168 | 1.33 |
| | (1.05) | (0.07) | (0.08) | (0.07) | | |
| Age expectations (/18) | 13.40 | 0.23 | 0.25 | 0.02 | 2168 | 0.35 |
| | (3.01) | (0.30) | (0.32) | (0.33) | | |
| Total (/54) | 38.57 | 0.22 | 0.59 | 0.30 | 2168 | 0.42 |
| | (4.85) | (0.51) | (0.53) | (0.55) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

understanding the driving forces behind child development. 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.*[15]:

1. **Sources of play materials** (maximum score of 2)**:** 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.

2. **Variety of play materials** (maximum score of 7)**:** the number of different types of play materials (types listed in Table 63) the child has played with in the past 30 days.

3. **Play activities** (maximum score of 7)**:** the number of different play activities (listed in table 64) the child has done with a household member over the age of 15 in the past 3 days.

4. **Household books** (maximum score of 6)**:** number of books for adults in household (not including school books).

5. **Household newspapers and magazines** (maximum score of 6)**:** number of newspapers and magazines in household.

We will briefly discuss each subscale Tables 62 to 65 show the distribution of its constituent parts.

Table 62 provides information on the sources of play materials for our target children over the past 30 days. Bought toys are the most common toy, used by 83% of target children. Almost half of the sample (46%) has four or more such bought toys. Homemade toys are very uncommon - only about 4% of households

Table 61: Mother's time use in minutes during the last working day

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Working for a wage | 20.22 | -3.21 | -1.57 | 0.38 | 2166 | 0.38 |
| | (57.45) | (3.81) | (4.38) | (3.98) | | |
| Travelling | 12.85 | -0.18 | 0.01 | 1.31 | 2168 | 0.10 |
| | (49.38) | (3.08) | (3.18) | (3.03) | | |
| Cooking | 205.91 | 0.80 | -0.84 | 5.63 | 2167 | 0.41 |
| | (93.98) | (6.76) | (6.55) | (5.97) | | |
| Washing/cleaning clothes | 49.10 | -0.98 | -0.33 | 0.12 | 2168 | 0.03 |
| | (41.25) | (3.97) | (2.54) | (2.60) | | |
| Cleaning house | 55.90 | 1.13 | 2.17 | -0.46 | 2168 | 0.30 |
| | (37.36) | (4.82) | (2.99) | (3.08) | | |
| Collecting/carrying water | 15.30 | 1.75 | -0.88 | 0.99 | 2167 | 0.26 |
| | (23.51) | (3.27) | (2.46) | (2.37) | | |
| Other household activities | 13.73 | 2.08 | -1.05 | 3.08 | 2168 | 3.00** |
| | (21.67) | (1.50) | (1.37) | (1.94) | | |
| Taking children to school | 2.19 | 0.69 | -1.13 | 0.25 | 2168 | 1.99 |
| | (17.59) | (1.17) | (0.80) | (1.05) | | |
| Looking after small children | 180.79 | -5.40 | 2.12 | -9.25 | 2167 | 0.59 |
| | (114.61) | (9.08) | (9.66) | (9.63) | | |
| Playing with small children | 71.97 | -2.89 | -2.61 | 3.07 | 2168 | 0.56 |
| | (65.25) | (4.66) | (4.06) | (5.16) | | |
| Reading/telling stories to children | 14.00 | 3.27 | -1.68 | 1.92 | 2167 | 1.81 |
| | (30.71) | (2.73) | (2.04) | (2.19) | | |
| Looking after ill household member | 2.97 | -0.02 | 2.57* | 1.19 | 2168 | 1.25 |
| | (15.99) | (1.14) | (1.52) | (1.28) | | |
| Watching TV | 60.75 | -10.54** | -13.34*** | -5.76 | 2168 | 3.14** |
| | (70.85) | (4.66) | (4.72) | (4.42) | | |
| Religious activities | 22.51 | 1.85 | 0.25 | 2.27 | 2168 | 0.36 |
| | (28.56) | (2.61) | (2.67) | (2.71) | | |
| Socialising | 47.72 | 2.13 | -0.35 | 0.09 | 2166 | 0.23 |
| | (48.38) | (3.65) | (3.11) | (3.81) | | |
| Other activities | 1.39 | 0.66 | 1.01 | 1.18 | 2168 | 1.15 |
| | (7.27) | (0.60) | (0.77) | (0.82) | | |
| Sleeping | 445.20 | -1.03 | -0.36 | -5.70 | 2166 | 0.24 |
| | (85.55) | (7.25) | (7.57) | (7.45) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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.

report that their children play with such. Household objects are on the other hand relatively frequently used for playing in our setting (58%). We only see a minor imbalance (at 10%) in the variable homemade toys, possibly driven by the low use.

Table 62: Sources of play materials in past 30 days

| | | **Difference from Control** | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Any homemade toys (%) | 3.95 | 2.76* | 0.70 | 2.71* | 2170 | 1.72 |
| | (19.49) | (1.48) | (1.35) | (1.58) | | |
| Any bought toys (%) | 82.67 | 2.24 | 0.18 | 3.09 | 2169 | 0.47 |
| | (37.88) | (3.39) | (3.41) | (3.09) | | |
| Four or more bought toys (%) | 46.33 | 2.28 | 0.28 | -0.49 | 2169 | 0.13 |
| | (49.91) | (4.99) | (4.94) | (4.70) | | |
| Any household objects (%) | 57.89 | -1.28 | -0.39 | 1.25 | 2170 | 0.05 |
| | (49.42) | (5.97) | (5.43) | (6.22) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 second subscale refers to the variety of play materials. Table 64 lists what types are owned. Almost half of the children have access to toys that induce physical movement, about 17% have toys that play music and 14% have things for drawing and/or writing. Picture books for children are extremely rare. They are only owned by less than one percent of study households.

Table 64 indicates what types of play activities household members performed with the target children over the last three days, relating to the third subscale. It can be seen that most commonly (in 71% of cases), the children were taken out to the market, park or other place. 64% of our sample children were also sang to within the last three days and in 45% of cases a household member played together with the child and its toys. Other activities asked about are quite uncommon: Only ten percent of children were read to by any household member, 15% were told stories, and naming objects or colours was played in 22% of cases. We observe some imbalances in these reported activities, that also feed through to the F-stat. These are in terms of reading with the child and playing together with its toys. We find here that the NE group is more likely to do so in both cases, and the GS groups is similarly more likely to play with its toys. These differences are significant at the 5 and 10% level.

Constituent parts for the 4th and 5th subscale are shown in Table 65. A staggering 58% of households do not own any books, further 15% have one or two, 12% between three and five and only 15% of households own more than six books. Similarly, 88% of households do not have a newspaper or magazine in their house, 8% have one or two and the remaining 3-4% have more than two. These statistics are balanced across treatment arms.

Finally, Table 66 shows the scores for each subscale, broken down by treatment and control. We see that

Table 63: Variety of play materials in past 30 days

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Toys to play music (%) | 17.86 | 3.19 | 1.07 | 2.29 | 2170 | 0.35 |
| | (38.34) | (3.46) | (3.10) | (2.99) | | |
| Toys to assemble or build (%) | 5.64 | 1.25 | -1.89 | -1.94 | 2170 | 2.75** |
| | (23.09) | (1.56) | (1.32) | (1.34) | | |
| Things for drawing and/or writing (%) | 14.10 | 2.48 | -2.31 | 1.43 | 2170 | 1.19 |
| | (34.83) | (3.14) | (2.76) | (3.40) | | |
| Toys that induce physical movement (%) | 49.06 | 2.15 | -3.35 | 3.07 | 2170 | 0.97 |
| | (50.04) | (3.81) | (3.77) | (4.06) | | |
| Dolls and other objects that aid fantasy games (%) | 10.15 | 4.37* | 0.74 | 1.49 | 2170 | 1.06 |
| | (30.23) | (2.55) | (2.24) | (2.19) | | |
| Picture books for children (%) | 0.75 | -0.01 | -0.22 | -0.20 | 2170 | 0.13 |
| | (8.65) | (0.50) | (0.47) | (0.47) | | |
| Toys to learn shapes and/or colours (%) | 1.13 | -0.01 | 0.12 | 0.54 | 2170 | 0.18 |
| | (10.57) | (0.76) | (0.67) | (0.80) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 sample appears balanced over treatment and control on measures of sources of play materials, variety of play materials, household books and newspapers and magazines. We do see a slight imbalance on the play activities - the NE and GS treatment groups appear to do slightly more of these, as we could already see in the break-down, shown in Table 64. A rough comparison of these results with summary statistics presented for an urban Oriya population (from the slums of Cuttack) shows that our population appears to have access to significantly fewer sources and variety of play material but similar scores for the number of play activities and newspapers and magazines in the household. Interestingly, we find households in rural areas to own more books than those in urban areas.

Table 64: Play activities performed with household members in past 3 days

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Read or looked at picture books | 10.34 | 5.12** | -1.05 | 2.25 | 2169 | 3.66** |
| | (30.47) | (2.00) | (1.92) | (2.06) | | |
| Told stories | 14.69 | -0.35 | -1.81 | -2.65 | 2167 | 0.73 |
| | (35.43) | (2.13) | (2.37) | (2.25) | | |
| Sang | 64.29 | 3.68 | 0.36 | -0.40 | 2169 | 0.73 |
| | (47.96) | (3.05) | (3.46) | (3.13) | | |
| Went out to market, park or other place | 71.05 | 3.62 | -0.75 | 1.22 | 2169 | 0.71 |
| | (45.39) | (3.21) | (3.30) | (3.65) | | |
| Played together with child's toys | 45.68 | 8.70** | 5.22 | 9.78** | 2170 | 2.58* |
| | (49.86) | (3.86) | (3.85) | (3.92) | | |
| Made drawings, paintings or writing | 11.09 | 1.57 | -0.91 | 2.61 | 2169 | 0.99 |
| | (31.43) | (1.92) | (1.97) | (2.47) | | |
| Played naming objects or colours, or counting | 22.56 | 3.42 | -0.02 | 5.54 | 2167 | 0.89 |
| | (41.83) | (4.23) | (3.76) | (4.10) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 65: Books, newspapers and magazines in household

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| No books | 57.69 | -1.71 | -2.23 | 1.46 | 2163 | 0.15 |
| | (49.45) | (6.45) | (6.50) | (6.39) | | |
| Between 1 and 2 books | 15.37 | -0.44 | 0.73 | 1.45 | 2163 | 0.14 |
| | (36.10) | (2.82) | (2.71) | (3.02) | | |
| Between 3 and 5 books | 12.14 | 1.66 | 0.91 | -0.31 | 2163 | 0.17 |
| | (32.70) | (3.28) | (2.54) | (2.68) | | |
| 6 or more books | 14.80 | 0.50 | 0.58 | -2.60 | 2163 | 0.61 |
| | (35.54) | (3.46) | (3.48) | (3.04) | | |
| No newspapers or magazines | 88.35 | 1.23 | 1.48 | -1.47 | 2170 | 0.50 |
| | (32.12) | (2.88) | (2.64) | (2.90) | | |
| Between 1 and 2 newspapers or magazines | 8.08 | -0.45 | -0.23 | 0.24 | 2170 | 0.03 |
| | (27.28) | (2.57) | (2.50) | (2.59) | | |
| Between 3 and 5 newspapers or magazines | 0.94 | 0.18 | -0.05 | 0.54 | 2170 | 0.19 |
| | (9.66) | (0.64) | (0.66) | (0.85) | | |
| 6 or more newspapers or magazines | 2.63 | -0.96 | -1.20 | 0.70 | 2170 | 1.41 |
| | (16.02) | (1.07) | (1.02) | (1.30) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 66: Family Care Indicator subscales

| | | Difference from Control | | | | |
|---|---|---|---|---|---|---|
| | Control Mean | NE | IS | GS | N | F-stat |
| Sources of play materials (/2) | 0.50 | 0.05 | 0.01 | 0.02 | 2169 | 0.40 |
| | (0.53) | (0.05) | (0.05) | (0.05) | | |
| Variety of play materials (/7) | 0.99 | 0.13 | -0.06 | 0.07 | 2170 | 1.31 |
| | (1.13) | (0.11) | (0.09) | (0.10) | | |
| Play Activities (/7) | 2.40 | 0.26*** | 0.01 | 0.18* | 2170 | 3.25** |
| | (1.46) | (0.09) | (0.10) | (0.10) | | |
| Household books | 1.58 | 0.10 | 0.09 | -0.11 | 2163 | 0.29 |
| | (2.26) | (0.30) | (0.29) | (0.26) | | |
| Newspapers and Magazines | 0.28 | -0.05 | -0.07 | 0.07 | 2170 | 1.37 |
| | (1.05) | (0.08) | (0.07) | (0.09) | | |

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$: 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 Conclusion

This baseline report has provided an overview of the key findings from the first round of data collection conducted under the project Early Childhood Development fpr the Poor: Impacting at Scale, a project primarily funded by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (NIH) and the Strategic Impact Evaluation Fund (SIEF). In addition to providing a picture of the study communities and households and checking for balancedness in these observable characteristics, this report serves to document key project activities with respect to the intervention and evaluation.

Overall, we find that data collection was successfully implemented and that data are of the quality standards required. 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 study groups (treatments and control) that appear similar on all dimensions. The only diffference will then be the interventions implemented.

We find no important differences in baseline indicators of child development and none in indicators of health and morbidity. This is important since it implies our treatment and control children are not fundametnally 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 treatments and control and most are small and do not provide evidence of systematic differences between the treatment and control group. An exception to this is the percentage of stunted children falling in the older spillover category. We find here that these 14-20 months old children are significantly shorter and significantly more likely to be stunted in treatment communities. We also find that the play activities household members perform with the child are not balanced between treatments and control throughout. We for example find that household members in communities allocated to the NE treatment arm are on average more likely to read with the child or look at picture books and this imbalance remains when considering the F-stat. We also find a small imbalance in the percentage of biological mothers having secondary schooling. However, on the vast majority of dimensions considered, the study groups are well balanced, this includes household composition, characteristics of households members, particularly the household head and the child's parents, income and wealth information, labour supply and dwelling characteristics.

This provides us a good foundation for analysing the impacts of the home visiting programme.

# References


[1] FE Aboud and S. Akhter. A cluster-randomized evaluation of a responsive stimulation and feeding intervention in bangladesh. *Pediatrics*, 127:e1191–97, 2011.

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

[3] O. Attanasio, V. Oppendisano, and M. Vera-Hernandez. Should cash transfers be conditional? conditionality, preventive care, and health outcomes. *American Economic Journal: Applied Economics*, 7(2): 1–19, 2015.

[4] Orazio P Attanasio, Camila Fernández, Emla O a Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, and Marta 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):g5785, January 2014. ISSN 1756-1833. doi: 10.1136/bmj.g5785.

[5] Nancy Bayley. *Bayley scales of infact and toddler development, third edition*. Pearson Education , Inc., 2006.

[6] Diane Bricker, Jane Squires, and Linda Mounts. *Ages and stages questionnaires, first edition*. Paul H. Brookes Publishing Co., Inc., 1995.

[7] S. Chang, S. Grantham-McGregor, C.A. Powell, M. Vera-Hernandez, F. Lopez Boo, H. Baker-Henningham, and S. Walker. Integrating a parenting intervention with routine primary health care: A cluster randomized trial. *Pediatrics*, 136(2), 2015.

[8] S Desai and S Alva. Maternal education and child health: is there a strong causal relationship? *Demography*, 35:71–81, 1998. ISSN 0070-3370. doi: 10.2307/3004028.

[9] S. Eickmann, A. Lima, M. Guerra, M. Lima, P. Lira, S. Huttly, and A. Ashworth. Improved cognitive and motor development in a community-based intervention of psychosocial stimulation in northeast brazil. *Developmental Medicine and Child Neurology*, 45:536–541, 2003.

[10] Paul Gertler, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M Chang, and Sally Grantham-McGregor. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science (New York, N.Y.)*, 344(6187):998–1001, May 2014. ISSN 1095-9203. doi: 10.1126/science.1251178.

[11] Sally M Grantham-McGregor, Susan P Walker, Susan M Chang, and Christine 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):247–253, August 1997. ISSN 00029165.

[12] S.M Grantham-McGregor, C.A Powell, S.P Walker, and J.H Himes. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *The Lancet*, 338(8758):1–5, July 1991. ISSN 01406736. doi: 10.1016/0140-6736(91)90001-6.

[13] S.M. Grantham-McGregor, C.A. Powell, S.P. Walker, and J.H. Himes. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the jamaican study. *The Lancet*, 338:1–5, 1991.

[14] J.D. Hamadani, S.N. Huda, F. Khatun, and S.M. Grantham-McGregor. Psychosocial stimulation improves the development of undernourished children in rural bangladesh. *Journal of Nutrition*, 136: 2645–52, 2006.

[15] Jena D Hamadani, Fahmida Tofail, Afroza Hilaly, Syed N Huda, Patrice Engle, and Sally M Grantham-McGregor. Use of family care indicators and their relationship with child development in Bangladesh. *Journal of health, population, and nutrition*, 28(1):23–33, February 2010. ISSN 1606-0997.

[16] Paul Kline. *Handbook of psychological testing*. Routledge, second edition, 2013.

[17] S Kurstjens and 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):623–636, 2001. ISSN 0021-9630.

[18] M. Christine Lovejoy, Patricia A. Graczyk, Elizabeth O'Hare, and George Neuman. Maternal depression and parenting behavior: A meta-analytic review. *Clinical Psychology Review*, 20(5):561–592, 2000. ISSN 02727358. doi: 10.1016/S0272-7358(98)00100-7.

[19] B. Lozoff, JB Smith, KM Clark, CG Perales, F Rivera, and M. Castillo. Home intervention improves cognitive and social-emotional scores in iron-deficient anemic infants. *Pediatrics*, 126(4):e884–94, 2010.

[20] C. Lu, M. Black, and LM Richter. Young children exposed to risks for poor development in low- and middle-income countries: Global, regional and country prevalence. *The Lancet (under review)*, Under Review.

[21] 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.

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

[23] B. Nahar, M.I. Hossain, J.D. Hamadani, T. Ahmed, S.N. Huda, and S.M. et al. Grantham-McGregor. Effects of a community-based approach of food and psychosocial stimulation on growth and development of severely malnourished children in bangladesh: a randomised trial. *European Journal of Clinical Nutrition*, 66:701–9, 2012.

[24] C. Powell and S. Grantham-McGregor. Home visiting of varying frequency and child development. *Pediatrics*, 84:157–64, 1989.

[25] L. S. Radloff. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population, 1977. ISSN 0146-6216.

[26] Aboud FE Singla DR1, Kumbakumba E2. Effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural uganda: a community-based, cluster randomised trial. effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural uganda: a community-based, cluster randomised trial. *The Lancet*, 3 (8):458–69, 2015.

[27] Jane Squires and Diane Bricker. *Ages and stages questionnaires, third edition (ASQ-3)*. Paul H. Brookes Publishing Co., Inc., 2009.

[28] Jane Squires, Diane Bricker, and LaWanda Potter. Revision of a parent-completed developmental screening tool: Ages and stages questionnaires. *Journal of Pediatric Psychology*, 22(3):313–328, 1997. ISSN 01468693. doi: 10.1093/jpepsy/22.3.313.

[29] P. Tripathy, N. Nair, S. Barnett, R. Mahapatra, J. Borghi, S. Rath, R. Gope, D. Mahto, R. Sinha, R. Lakshminarayana, V. Patel, C. Pagel, A. Prost, and A. Costello. Effect of a participatory intervention with women s groups on birth outcomes and maternal depression in jharkhand and orissa, india: a cluster-randomised controlled trial. *Lancet*, 375(9721):1182–92, 2010.

[30] S. Walker, S. Chang, C. Powell, and S.M. Grantham-McGregor. Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-retarded jamaican children: a prospective cohort study. *The Lancet*, 366(1804-7), 2005.

[31] Susan P Walker, Susan M Chang, Christine A Powell, Emily Simonoff, and Sally 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):472, September 2006. ISSN 1756-1833. doi: 10.1136/bmj.38897.555208.2F.

[32] Susan P Walker, Theodore D Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A Wasserman, Ernesto Pollitt, and Julie A Carter. Child development in developing countries 2 Child development : risk factors for adverse outcomes in developing countries. pages 145–157, 2007.

[33] WHO Multicentre Growth Reference Study Group. WHO Child Growth Standards based on length/height, weight and age. *Acta paediatrica. Supplementum*, 450:76–85, 2006. ISSN 0803-5326. doi: 10.1080/08035320500495548.

# A Training schedule

Table 67: Training schedule

| Trainees | Training material covered | No. days | Dates |
|---|---|---|---|
| Supermentors | IS curriculum (all intervention months) | 7 | 15 Sept – 21 Sept 2014 |
| | NE curriculum (all intervention months) | 5 | 23 Sept – 27 Sept 2014 |
| | GS curriculum (intervention months 1 – 10) | 5 | 1 Dec 2014 – 5 Dec 2014 |
| | GS curriculum (intervention months 11 – 18) | 5 | 9 Feb 2015 – 13 Feb 2015 |
| | Refresher training IS + GS + NE curriculum (intervention months 8 – 18) | 3 | 18 April – 20 April 2016 |
| | IS + GS + NE curriculum refresher training + GS curriculum training intervention months 19 - 25 | 5 | December 2016 (to be decided) |
| Mentors | IS + GS + NE Intervention months 1- 8 (ages 6 – 20 months) | 18 | 4 Aug – 21 Aug 2015 (all districts) |
| | Practice period | 11 | 1 Sept – 13 Sept 2015 (all districts) |
| | IS + GS + NE intervention months 9 – 15 (ages 21 – 32 months) | 18 | 18 Sept – 9 Oct 2015 (Cuttack)<br>7 Oct – 19 Oct 2015 (Bolangir)<br>29 Oct – 3 Nov 2015 (Bolangir – continued)<br>29 Oct – 17 Nov 2015 (Balasore) |
| | IS + GS curriculum: Refresher training intervention months 8 – 15 + training intervention months 16 - 18 | 9 | 2 May – 10 May 2016 (all districts) |
| | NE curriculum: Refresher training intervention months 8 – 15 + training intervention months 16 - 18 | 3 | 11 May – 13 May 2016 (all districts) |
| | IS + GS + NE Refresher training months 16-18 + GS training intervention months 19 - 25 (including practice) | 10 | Jan 2017 (to be decided) (all districts) |
| HVs/GFs | IS + GS + NE: Intervention months 1 – 7 (ages 6 – 24 months) (including 5 days of toy making) | 18 | 29 Oct – 23 Nov 2015 (Cuttack)<br>13 Nov – 9 Dec 2015 (Bolangir)<br>26 Nov – 20 Dec 2015 (Balasore) |
| | IS + GS: Refresher training intervention month 7 + training intervention months 8 – 18 (ages 25 – 35 months) (including 2 days of toy making) | 9 | 16 May – 24 May 2016 (Cuttack)<br>30 May – 7 June 2016 (Bolangir)<br>20 June – 28 June 2016 (Balasore) |
| | NE: Refresher training intervention month 7 + training intervention months 8 – 18 (ages 25 – 35 months) | 3 | 25 May – 27 May 2016 (Cuttack)<br>9 June – 11 June 2016 (Bolangir)<br>29 June – 1 July 2016 (Balasore) |
| | IS + GS + NE Refresher training months 8 - 18 + GS training intervention months 19 – 25 (ages 36 – 42 months) (including toy making) | 10 | Feb – March 2017 |

# B Age registration protocol

Given the importance of the age of the child for the design and implementation of this study, special care and effort was taken to collect accurate ages of the children in our sample. This Appendix describes in detail the protocol that was followed for this purpose.

## B.1 Age determination during and after census

The census survey recorded two sources of information on the child's age: (i) Date of birth (verified using official documents where available) and (ii) age of the child (in completed months) at the time of the census survey, as reported by the respondent. On the basis of these census data three types of children were identified:

1. Children not yet born at the time of census, i.e. for which the mother was still pregnant (30% across sample);
2. Children that were born at the time of census but for whom there was a significant difference between reported age and age derived from the date of birth recorded during census survey (2% across sample). When defining eligibility for TC/SC (which is based on the age of the child at the expected start of the intervention), the loosest of the two age variables was used so that if the child fell in the eligible age range for any one of these two age variables then the child was categorised as eligible (priority given to TC in case the choice was between TC and SC);
3. All remaining children;

Children in categories 1 and 2 were flagged for careful age verification during baseline survey. Prior to the baseline survey visit, for children in category 3 the cover sheet of the baseline household roster module (the first module to be administered of each household) was pre-populated by the supervisor with the child's date of birth, the expected age at the start of intervention and the SC/TC categorisation as per the census. For children in categories 1 and 2 these fields were left blank and the surveyor was asked to provide this information. This process is explained in the next section.

## B.2 Age verification during baseline household roster administration

Upon administering the household roster, survey Team 1 was instructed to do the following:

- Note whether the date of birth as per census had been pre-filled or not;
- In the event that the child had passed away, was no longer available for the survey or the mother was still pregnant then this was noted in the replacement tracking sheet and a replacement child was selected from the replacement list for the category that the child was listed in;
- In the event that the age field was not pre-filled but the child was present the respondent was asked to do the following:

- Provide the date of birth of the child and to provide an official document for verification (method of verification was also recorded by the surveyor);
- After establishing the date of birth of the child the surveyor was asked to calculate the age of the child at the expected time of the start of the intervention in their respondents' respective districts (see Table 5 ). The answer was noted in completed months with no rounding since ASQ modules are based on strict age categories; The following instructions were given to the enumerator to calculate the age at the time of intervention:
  - Count the number of complete months from the date of birth to the date of the start of intervention or the day of the interview
  - Treat every month as if it had 30 days. For example: Feb 15 to March 15 is 1 month.
  - Incomplete days are included which means that the day of the interview is counted as 1 day. For example: 5 months and 1 day means that we are interviewing on the first day of the 5th month.
  - Complete month means that the current month is between 1 and 30 days running. For example: 5 months and 1 day is 5 complete months, 5 months and 30 days is 5 complete months but 5 months and 31 days is 6 months and 1 day.
- The calculated age at the time of intervention was subsequently recorded in the household roster;
- Based on the calculated age of the child at the expected start date of the intervention the child was grouped into one of three categories:
  - Younger Spillover 2-6 months at the time of intervention start
  - Target Child 7-16 months at the time of intervention start
  - Older Spillover $17-20$ months at the time of intervention start
- In case there was a mismatch between the categorisation calculated here and the category that the child had been assigned to after census then this was logged accordingly in the interviewer tracking sheet and the field supervisor was consulted;
- The field supervisor then verified the age calculation and confirmed the appropriate age category for the child. If the child was not in the correct age category then the supervisor provided the surveyor with details of a suitable replacement child (of the appropriate category). There were two situations when a sampled child was replaced:
  - In Sample but wrong category: The category (target, spillover etc. as described above) of the child had to be updated and the child was put in the reserve list for that category. In case there were no other reserve children and there were less than the required number for that category in the village ($<$ 8 target children, $<$ 4 spillover children) the replacement child was added directly to the list of those to be surveyed and the interview continued as planned;

* Not in Sample: The child was found not to belong to any of the TC/SC age categories eligible for survey. In this case the child was removed from the sample. A replacement child was found from the reserve list for the appropriate age group from which the child was removed and reason for replacement was recorded in the survey tracking sheet;

## B.3 Age verification prior to ASQ survey

For all children (including those in group 3 in Section B.1), the ASQ survey team was asked to calculate the expected age of the child at the time of the interview based on the date of birth provided (pre-filled) on the ASQ questionnaire cover sheet, in order to administer the age appropriate ASQ module.

## B.4 Age verification after survey

After completion of the survey, the survey scrutinizer (responsible for quality assurance) verified the information provided in the household roster and ASQ module and took special care to make sure that the calculations and selection of each child's category was correct. Whatever action was taken was duly noted in the survey master tracking sheet as well as the effect this had on the sample size of each age category. In addition, if there had been a replacement all the identification details of the replacement and the replaced households were noted in the master tracking sheet along with the reason for the replacement.

## B.5 Final age confirmation and updates during data cleaning

During data cleaning, data on ages were analyzed and checked for consistency across different modules. For the final categorization of children to target child and spillover child categories (used in this baseline report), we used the ages recorded on the cover sheet of the household roster as a base for the calculation of the child's age at the expected start of the intervention in his or her district[20]. In case of nonsensical entries of this variable (e.g. negative, missing or outlier values), the hard copies of the questionnaires were consulted for confirmation. If the issue could not be resolved, other information in the questionnaires (e.g. ages recorded in other modules, interview dates, etc.) were used to update the final age variable.

[20] Note that there was a slight delay in the start of the intervention (see Table **??**) so the actual age of the children at the actual start of the intervention ended up being slightly older. For the purpose of consistency with the sampling frame, however, we stuck to the originally scheduled timing of the intervention for the categorisation of children to TC/SC categories.