Title

Child development in a global context: Innovations, interventions, and implementation at scale

Author

Pitchik, Helen Osborne

Publication Date

2021

Peer reviewed|Thesis/dissertation

Child development in a global context: Innovations, interventions, and implementation at scale

by

Helen Osborne Pitchik

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Lia C.H. Fernald, Chair
Professor Alan E. Hubbard
Professor Stephen P. Luby

Spring 2021

Child development in a global context: Innovations, interventions, and implementation at scale

Helen Osborne Pitchik

Abstract

Child development in a global context: Innovations, interventions, and implementation at scale

by

Helen Osborne Pitchik

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Lia C.H. Fernald, Chair

Globally over 249 million children in low- and middle-income countries (LMICs) are at risk of not meeting developmental milestones on the motor, cognitive, language or socio-emotional domains. This compromised development may result in less educational attainment and lower economic earnings later in life. The nurturing care framework, developed by the World Health Organization, summarizes the inputs required to promote healthy child development. These inputs include good health, nutrition, responsive caregiving, access to early learning, and safety and security. In pursuit of improved child development, caregiver support interventions have been developed and implemented in numerous LMICs, and have succeeded at improving child language and cognitive development. Most of these interventions, however, were designed to be implemented on a small scale and focus primarily on access to early learning and responsive caregiving.

This dissertation aims to inform the design and evaluation of large-scale interventions to improve child development over the life-course in low-resource settings. In Chapter 1, I evaluate the effects of a multicomponent intervention, that covers early learning, responsive caregiving nutrition, water, sanitation, hygiene, caregiver mental health and lead exposure prevention on child development and risk factors for poor child development. I use data from a cluster randomized controlled trial of an intervention that was delivered primarily in groups in Kishoreganj, Bangladesh. This research contributes to existing literature on the effects on caregiver support interventions on early child development and expands the scope of these interventions to include additional components that may result in larger and more sustained impacts on child outcomes. In Chapter 2, I assess the concurrent correlation between two measures of child development: (i) the Bangladeshi-adapted Ages and Stages Questionnaire: Inventory and (ii) the Bayley Scales of Infant Development-III. The Ages and Stages Questionnaire: Inventory is primarily a caregiver report measure that is feasible to implement to evaluate intervention effects when financial, human, and time resources

are constrained, as may be the case with large-scale evaluations. The Bayley Scales of Infant Development-III is a direct assessment measure for child development that has been adapted for use in Bangladesh, and is considered to be less biased than caregiver report in evaluating the effects of early child development interventions. I use endline data from the same cluster randomized trial used in Chapter 1 to examine correlations overall, by age, and by intervention status. This work illustrates the potential of the Bangladeshi-adapted Ages and Stages Questionnaire: Inventory for use in the evaluation of large-scale child development interventions. Finally, in Chapter 3, I describe how the distribution of risk factors for poor child development shifted in response to experiences during the COVID-19 pandemic in Chatmohar, Bangladesh. I analyze data from the evaluation of a large-scale multicomponent child development intervention in Chatmohar, Bangladesh. This analysis highlights the shift of intervention targets for improving population-level early child development in the context of the COVID-19 pandemic. Taken together, this dissertation contributes to the evidence for the design and measurement of effective and scalable interventions to improve child development in LMICs.

For my parents and my sister.

Contents

List of Figures

List of Tables

Acknowledgments

I am immensely thankful for the community of people who have supported me throughout this work. Firstly, I am extremely grateful for the mentorship of my advisor and dissertation chair, Lia Fernald. Her encouragement of new ideas, care for my well-being, and constant support has been instrumental to my experience at Berkeley. Lia's mentorship has allowed for a rewarding academic experience. She is a remarkable researcher, role model, and person.

I would like to thank Steve Luby and Peter Winch both for their integral role in the research presented in this dissertation and for their thoughtful guidance through the implementation, analysis, and interpretation of this work. They have each provided consistent guidance throughout my doctoral degree and have enabled me to be a better researcher. I would also like to thank Alan Hubbard for his excellent statistical guidance, and for his kind and enthusiastic support for my dissertation work.

I have been privileged to work with many collaborators at icddr,b in Bangladesh who have been invaluable to my growth as a researcher. This dissertation would not have been possible without such a dedicated, smart, and thoughtful team. In addition to their excellent contributions to research and implementation of the projects discussed in this dissertation, I would also like to thank them for making my time in Bangladesh such a joy, and for introducing me to Bangladeshi food and culture.

My friends and family have been central to my doctoral experience. At UC Berkeley I have had the good fortune to be surrounded by a wonderful community of students and faculty and an enriching research group. My brilliant fellow doctoral students and friends Catherine Duarte, Carrie Fahey, Yoshika Crider, Lina Montoya, Liz McQuade, and Anoop Jain have been a consistent source of methodological support and laughter. My parents Carolyn Pitchik and Martin Osborne, and my sister Vivian Pitchik have provided unwavering support and encouragement, and I am so lucky to have them by my side. I am also incredibly thankful to Roger LeMesurier for being such an immense source of joy over the last couple of years.

I am also incredibly thankful to the field staff and participants in Kishoreganj and Chatmohar, Bangladesh who made this work possible. I am so grateful for their time, energy, and willingness to participate in our research. I very much appreciate they funding provided by the Subir and Malini Chowdhury Center for Bangladesh Studies and the Center for Global Public Health for summer travel and a student fellowship.

Finally, I would like to thank my co-authors on the manuscripts that have come out of this dissertation: Mahbubur Rahman, Fahmida Tofail, Stephen Luby, Peter Winch, Lia Fernald, Jenna Forsyth, Fahmida Akter, Md. Reza Hasan, Kendra Byrd, Farzana Yeasmin, Jesmin Sultana, Md. Khobair Hossain, Laura Kwong, Rizwana Khan, Tarique Md. Nurul Huda, Natasha Kashaul, Tania Jahir, Abul Kasham Shoab, Jyoti Bhushan Das, Esther Chung, Malay Mrindha, and Md. Ruhul Amin.

Introduction

Importance of child development

Early motor, cognitive, and socioemotional development are associated with later life outcomes, including educational attainment and economic earnings [1–3]. Several mechanisms have been posited to explain how early life experiences shape developmental outcomes. Early in life, a young child's brain is in a sensitive period for development, making risk factors during this early period especially salient for later development outcomes and well-being [4]. During this sensitive period, the brain is especially responsive to events or experiences, with positive experiences are more likely to contribute to the development of synaptic connections important for optimal developmental trajectories, while negative experiences can shift a child off the optimal developmental trajectory [4, 5]. Thus, interventions to reduce risk factors for impaired development in early life are critical to promoting positive developmental trajectories and later life outcomes.

The Nurturing Care framework developed by the World Health Organization summarizes the components needed to promote development during early life [6, 7]. These components include adequate nutrition, good health, security and safety, responsive caregiving, and exposure to opportunities for early learning. By contrast, experiences of poverty may directly – or indirectly via poor caregiver mental health – compromise caregivers' ability to provide nurturing care [6]. In low- and middle-income countries (LMICs), children experience a disproportionately high burden of risk factors for delayed development, including limited early learning and stimulation, as well as nutritional deficiencies [8]. An estimated 80.8 million (or 33% of) three and four-year old children in LMICs fail to meet key milestones in cognitive or socioemotional development, and 249 million children under 5 in LMICs are at risk for poor development [9, 10].

Child development has been widely acknowledged as a fundamental right. In 1989, the United Nations Convention for Rights of the Child – a human rights treaty, since ratified by a majority of the world's countries, that protects and promotes the rights of children – declared "Parties shall ensure to the maximum extent possible the survival and development of the child." [11]. In 2015, child development was explicitly included in the Sustainable Development Goals, with Goal 4.2 declaring that, "By 2030,. . . all girls and boys have access

to quality early childhood development, care and pre-primary education so that they are ready for primary education" [12]. This particular goal has brought an increased focus on the measurement of child development outcomes and the development of interventions to improve child development at scale.

Interventions to improve child development

As children under the age of three consistently rely on their parents and caregivers for care, parent or caregiver support programs can provide a useful platform for the promotion of early child development. Most parent support programs focus on responsive caregiving and enhanced play and learning opportunities within the home. Some programs also include a component on the promotion of maternal and child nutrition or the provision of micro or macro-nutrients. Parent support programs that promote children's play and learning opportunities, also referred to as responsive stimulation interventions, have been implemented and evaluated in many countries [8, 13]. These responsive stimulation interventions aim not only to teach caregivers about the developmental importance of caregiver-child interactions, but also to coach caregivers to practice responsive stimulation with their young children. Responsive stimulation interventions have shown effects on child development outcomes in many LMIC settings worldwide, including Bangladesh [14–18]. A meta-analysis of psychosocial stimulation interventions in LMICs found that these interventions have a notable effect on short-term child cognitive and language development (0.42 and 0.47 standard deviations respectively) [13]. A majority of these interventions were delivered in home visits where heath workers had individualized contacts with caregivers, or in a combination of home visits and group sessions [13]. The established effectiveness of small-scale responsive stimulation interventions has prompted a transition to a focus on the sustainability and scalability of these interventions. In recent years there has been a growing interest in the delivery of parent support programs exclusively in group sessions in order to reduce the program cost and personnel time needed and therefore increase scalability [19–21]. In addition to reducing costs, group-based intervention delivery has the potential to improve maternal mental health by formalizing social support networks in the community [19]. Despite consistent effects on short-term child development outcomes following intervention completion, few studies have examined the sustainability of effects in middle and late childhood and beyond. Those that have followed-up children as they age have not consistently found effects [22].

Multiple known risk factors for poor child development are often excluded from parent support interventions, such as access to clean water, sanitation and hygiene (WASH), nutrition, exposure to environmental toxins, and caregiver mental health. Including these components may contribute to more holistic changes in children's early environments and promote sustained intervention effects as children grow. However, adding these myriad components may also dilute the effects of the intervention contents focused on stimulating caregiving [23]. To study these issues, Chapter 1 of this dissertation evaluates the impact of a group-based multicomponent intervention in rural Bangladesh on stimulating caregiving, child development,

maternal mental health, and indicators of WASH and nutrition status. At the time of implementation, group-only intervention delivery of a responsive stimulation intervention had not been directly compared to delivery that also included home visits. To minimize confounding by the use of group-only sessions, we included two intervention arms, group-only delivery and delivery through a combination of group sessions and individual home visits.

Measurement of interventions to improve child development

As the research on early child development interventions transitions to focus on interventions to improve population-level child development at scale, parallel research has sought to ascertain how to appropriately measure population-level intervention impacts from large-scale interventions. Two primary methods to measure child development include (1) direct assessment, where an enumerator administers test items directly to the child, and (2) caregiver report, where the child's caregiver answers a questionnaire about the child's developmental status [24]. Direct assessments of child development are thought to be less prone to bias when used for the evaluation of interventions, and more precise than caregiver report. However, direct assessment tests are typically longer and require enumerators that have more experience and training compared to caregiver report. Direct assessment is therefore challenging to implement in low-resource settings, where cost, availability of enumerator training, and time burdens to administrators, enumerators, and respondents pose substantial barriers. One caregiver report tool that has been used in at least 20 LMIC settings is the Ages and Stages Questionnaire-3 (ASQ-3). When evaluated against the Bayley Scales of Infant Development-III, a direct assessment test, in different upper-middle-income country settings (two in rural China, and one in urban Colombia), low to moderate concurrent correlations were found [25–27](25–27). Adaptations of the ASQ-3 to improve its performance were made, and the Ages and Stages Questionnaire: Inventory (ASQ:I) was created [28]. The ASQ:I has been used in the evaluation of an intervention in Madagascar and to measure child development in rural Kenya [29, 30]. In Bangladesh, a subset of the items of the ASQ:I as adapted to be administered through direct assessment with the use of low-cost and simple materials. If found to represent the same constructs captured by tools that use direct assessment, the primarily caregiver report ASQ:I tool could be a feasible and reliable alternative to direct assessment when financial, human, or time resources do not allow for direct assessment. In Chapter 2, data from the endline evaluation of the intervention evaluated in Chapter 1 are used to estimate concurrent correlations between domains of child development assessed with the Ages and Stages Questionnaire Inventory and the Bayley Scales of Infant Development-III.

Child development during COVID-19

In early 2020, countries around the globe began to respond to the Coronavirus Disease 2019 (COVID-19) pandemic by initiating lockdowns that restricted indoor and outdoor gatherings, business operations, government offices, and in-person schooling. While these public health strategies were a necessary, urgent respond to the immediate demands of a rapidly

evolving and catastrophic global pandemic, they were not without adverse consequences. In many LMICs the pandemic and subsequent mitigation strategies resulted in not only large economic losses but also reductions in access to health care and child vaccination rates [31]. The economic hardship and difficulty accessing food and routine healthcare incurred by families could create or exacerbate existing barriers for promoting child development. Families experiencing disruption in their daily lives, including financial loss and poor mental health, may experience a disruption in their ability to provide nurturing care for young children [32]. Chapter 3 analyzes data from the evaluation of a large-scale child development intervention in rural Chatmohar, Bangladesh to describe how the burden of risk factors for poor child development, including poor caregiver mental health and the quantity of stimulation in the home, shifted in response to experiences during the COVID-19 pandemic.

Chapter 1

A holistic approach to promoting early child development: a cluster randomised trial of a group-based, multicomponent intervention in rural Bangladesh

1.1 Abstract

Introduction In low- and middle-income countries, children experience multiple risks for delayed development. We evaluated a multicomponent, group-based early child development intervention including behavioral recommendations on responsive stimulation, nutrition, water, sanitation, hygiene, mental health and lead exposure prevention.

Methods We conducted a 9-month, parallel, multiarm, cluster-randomized controlled trial in 31 rural villages in Kishoreganj District, Bangladesh. Villages were randomly allocated to: group sessions ('group'); alternating groups and home visits ('combined'); or a passive control arm. Sessions were delivered fortnightly by trained community members. The primary outcome was child stimulation (Family Care Indicators); the secondary outcome was child development (Ages and Stages Questionnaire Inventory, ASQi). Other outcomes included dietary diversity, latrine status, use of a child potty, handwashing infrastructure, caregiver mental health and knowledge of lead. Analyses were intention to treat. Data collectors were independent from implementers.

Results In July—August 2017, 621 pregnant women and primary caregivers of children < 15 months were enrolled (group $n = 160$, combined $n = 160$, control $n = 301$). At

endline, immediately following intervention completion (July–August 2018), 574 participants were assessed (group $n = 144$, combined $n = 149$, control $n = 281$). Primary caregivers in both intervention arms participated in more play activities than control caregivers (age-adjusted means: group 4.22, 95% CI 3.97 to 4.47; combined 4.77, 4.60 to 4.96; control 3.24, 3.05 to 3.39), and provided a larger variety of play materials (age-adjusted means: group 3.63, 3.31 to 3.96; combined 3.81, 3.62 to 3.99; control 2.48, 2.34 to 2.59). Compared with the control arm, children in the group arm had higher total ASQi scores (adjusted mean difference in standardized scores: 0.39, 0.15 to 0.64), while in the combined arm scores were not significantly different from the control (0.25, −0.07 to 0.54).

Conclusion Our findings suggest that group-based, multicomponent interventions can be effective at improving child development outcomes in rural Bangladesh, and that they have the potential to be delivered at scale.

1.2 Introduction

Early motor, cognitive and socioemotional development affect later life outcomes, including educational attainment and economic earnings [3]. In low- and middle-income countries (LMICs), children experience a disproportionately high burden of risk factors for delayed development when compared with children in high- income countries. An estimated one-third of 3-year-old and 4-year-old children in LMICs—80.8 million children in total—did not meet basic developmental milestones in 2010 [10]. Factors that promote development during early life include responsive caregiving, maternal and child nutrition, caregiver's mental health, exposure to opportunities for early learning and avoidance of infection [6].

Across many different countries, cultures and contexts, caregiver-support programmes have improved short-term early child development (ECD) outcomes by encouraging responsive caregiving and stimulation through the promotion of age-appropriate caregiver-child interactions [13]. Interventions addressing other risk factors for poor child development including maternal mental health, nutrition and water, sanitation and hygiene (WASH) have also been shown to improve parental investments for children [33], or ECD outcomes [34, 35], though effects are smaller than for interventions that include responsive stimulation. Additionally, lead exposure has been associated with impaired cognitive development and can occur through exposure to contaminated turmeric and lead- soldered food storage cans [36–38]. Globally up to 800 million children, mostly in LMICs, have elevated lead exposure, but interventions have not assessed the impact of lead exposure reduction on ECD outcomes [39].

Integrated interventions targeting multiple risk factors have been recommended in the WHO guideline for improving ECD outcomes [40]. Integration of contents on multiple components may allow for efficiencies in intervention delivery through economies of scope, and may re-

sult in more holistic changes in the early environment, resulting in improved outcomes and cost savings. For example, an integrated stimulation, nutrition and health intervention in rural China showed positive effects on multiple outcomes beyond child development, including caregiver-reported child health, nutrition and diarrhea prevalence [41]. However, the evaluation of an intervention in rural India demonstrated that it is possible that integrating many intervention components may take caregivers' focus away from stimulating caregiving practices and disperse behavior changes across multiple domains [23].

Sustainability and scalability of ECD interventions are critical to their ultimate success, and this has driven the push to explore group-based delivery mechanisms. Many responsive stimulation interventions were originally developed to be delivered in home visits, which allow for more personalized coaching and problem-solving when compared with group sessions; however, delivery at-scale may be easier to attain with groups [42]. Group-based intervention delivery may also improve maternal mental health by facilitating the development of formalized social support networks in the community, and may contribute to sustained changes in community norms [43]. Thus, groups may promote longer term intervention effects, an important consideration given the recent demonstrations of a fade-out of initial promising effects of scaled-up home-visiting programs [44].

In this study, we aimed to evaluate the effects of a multicomponent group-based responsive stimulation, nutrition, WASH, maternal mental health and lead exposure prevention intervention in rural Bangladesh on stimulating caregiving practices, child development and multiple other risk factors for poor child development. We tested two delivery mechanisms, one that consisted of only group sessions and one that combined group sessions and home visits. We hypothesized that the integrated multicomponent intervention would improve caregiving practices, child development and caregiver mental health through both delivery mechanisms.

1.3 Methods

Study design and participants

The Research on Integrated Nutrition, ECD and WASH (RINEW) intervention was a three-arm cluster-randomized controlled trial conducted in the Katiadi and Kuliarchar subdistricts of Kishoreganj District, Bangladesh. Trial arms were (1) community group sessions (group arm), (2) alternating community group sessions and home visits (combined arm) and (3) passive control. As group sessions were community-based, villages were used as the unit of randomization to avoid spillover of intervention contents across arms.

All villages in the Katiadi and Kuliarchar subdistricts with populations between 200-800 households were considered for inclusion except for those in Masua union, where formative work was conducted. Villages were excluded if their basic demographic factors (ie, literacy,

electricity status) were more than 1.5 standard deviations (SDs) higher or lower than district averages. This was done to decrease the probability of chance imbalances in the intervention arms at baseline, which would decrease precision of effect estimates. Remaining villages were included if they were at least 2 km apart. Exceptions to these criteria are described in the appendix (table A.1).

Eligible participants were women living in the selected villages who were in their second or third trimester of pregnancy or primary caregivers of a child under 15 months of age. All participants were eligible for all 18 intervention sessions. The pregnant woman's in-utero child, or the youngest child of the primary caregiver (for participants who were not pregnant) was considered to be the child enrolled in the RINEW study. All participants gave verbal and written informed consent before being enrolled in the study.

Randomisation and masking

Each village was a cluster, and randomization was stratified by subdistrict. Clusters within each subdistrict were randomly allocated into one of two active intervention arms, or an oversized control arm, by an investigator at the University of California, Berkeley (HOP). The allocation ratio was 5:5:8 in Katiadi, and 3:3:7 in Kuliarchar for the group, combined and control arms, respectively. We used an oversized control arm to improve precision when comparing each intervention arm to the control arm. Participants were randomly selected from eligible participants in each cluster. Participants were informed of their intervention assignment following the baseline survey. Study participants and community health workers (CHWs) were not masked to intervention arm as the control arm participants were not invited to sessions with CHWs and only the combined arm included household visits. To mask data collectors to group status they were independent from CHWs and were not made aware of the study design or intervention components. Though items from the intervention sessions (toys, books) may have been visible to data collectors, they were not made aware that these contents were part of the intervention.

Intervention

The RINEW intervention took place between September 2017 and May 2018. All participants in villages randomized to either the group or combined intervention arms were invited to attend 18 intervention sessions delivered by CHWs every 2 weeks for 9 months. The integrated multicomponent intervention curriculum was developed through a year-long piloting process. Each of the individual intervention components was developed and refined by adapting existing curricula based on field testing and feedback from community members and CHWs [45]. Group sessions took 45–60 min and home visits took 20–25 min. Those in the group arm received 18 group sessions delivered every 2 weeks in a location close to their homestead with 3–6 pregnant women and caregiver-child dyads. Those in the combined arm received nine group sessions alternating with nine individual home visit sessions, with an

intervention session every 2 weeks. Groups were assembled based on geographic proximity. The material covered was equivalent across the delivery mechanisms. In home visit sessions, facilitators discussed the age-specific recommendations presented in the group sessions that were applicable to the household. CHWs did not visit the control communities.

Each intervention session included age-specific material on responsive stimulation. For caregivers with children this portion included a brief interactive discussion about the importance of play, review of activities from previous sessions, the introduction of new developmentally appropriate games with low-cost toys made from recycled materials, a local song and activities with a simple picture book. The main aim of the simulation component in each session was to encourage caregivers to participate in responsive caregiving and create learning opportunities through positive interaction, and to teach pregnant women how to engage in responsive stimulation with their newborn children. This intervention component was adapted from the Jamaican Reach Up Programme [46], with materials added for pregnant women and caregivers with children under 6 months of age. Each session also included material on one or more of the integrated components which included nutrition, WASH, lead exposure prevention, and caregiver mental health (table 1.1, table A.2). A tablet application was used to guide CHWs through the age-relevant curriculum depending on who was recorded present in each session, using the CommCare software platform. Pregnant participants were encouraged to watch and learn from the activities conducted with caregiver-child dyads.

For sessions 9 and onwards, other caregivers were invited to attend sessions, with a focus on assisting with childcare during the parts of the session not focused on caregiver-child interaction. In addition, concurrently with the 15th and 16th intervention sessions fathers were invited to attend two separate group sessions, with 10–12 peers. These sessions primarily focused on components that required support from household decision makers, including upgrading WASH infrastructure, purchasing lead-free food storage containers and unpolished turmeric and improving the diversity of food purchased for the household.

CHWs were 18–38-year-old women (mean 28 years) from the selected villages who had completed secondary school education. Many of the CHWs (75%) had previously worked in education or health. CHWs received 8 days of basic training, 4 days of refresher and tablet training immediately prior to the start of the intervention, and 9 additional 2–3 day trainings during the 9-month intervention. Trainings included didactic sessions, in-class practice, and field practice where CHWs were given feedback and practiced observing and giving feedback to others. At least one group session (or three individual home visits) per CHW was supervised during each 2-week period. Supervisors filled out session monitoring sheets and provided feedback to CHWs.

Assessments

After enrollment, baseline data were collected on demographic information for all participants and child-related measures for children over 6 months of age. A team of university-educated enumerators who were not involved in intervention delivery conducted endline data collection during two home visits immediately following intervention completion. The first visit included assessments of the home environment, child development and maternal mental health, and enumerators received 12 days of training; the second visit included assessments of WASH, nutrition, and lead, and enumerators received 6 days of training. Training for both modules included interactive discussion, role play and field testing in non-intervention sites followed by interobserver reliability testing, feedback, and refresher trainings. Interviews were conducted in Bengali, and data was collected using a tablet computer with CommCare software.

Outcomes

The pre-specified primary quantitative outcome of this trial was the Family Care Indicators (FCI), a caregiver report questionnaire with an observation component used to assess stimulation in the home [47]. This outcome contains two primary subscales, stimulating caregiving practices, and the variety of play materials available in the home. The stimulating caregiving practices subscale has questions about the variety of stimulating caregiving activities that any adult has engaged with the child in the previous 3 days (six items). We analyzed data on stimulation provided by the primary caregiver who was invited to attend the intervention sessions. The variety of play materials subscale includes observations of the variety of play materials in the home that the caregiver reported the child played with in the previous 30 days (six items). During the FCI interview caregiver responsiveness and the child's environment were observed and recorded. This observation scale includes items from the Infant Toddler Home Observation for Measurement of the Environment about caregiver responsiveness and interactions with the child and two items on the safety of the home environment (table A.3) [48]. The prespecified secondary outcome was child development as assessed by the Ages and Stages Questionnaire Inventory (ASQi). The ASQi is primarily a caregiver report measure used to assess attainment of milestones in the communication, gross motor, fine motor, problem-solving and personal social domains of development for children between 1 and 54 months. The ASQi includes direct assessment items for a subset (50) of the questions across five domains (table A.4). The ASQi was piloted by our study team on 60 children not included in this study sample, to ensure appropriate ranking of questions. In addition, an inventory developed following the principles of the Macarthur-Bates Communicative Development Inventories (CDI) was used to capture language development in both the expressive and receptive domains. Raw ASQi and CDI scores for each domain were internally age-standardized to the control arm using age-conditional means and SDs. Children with standardized scores over 4 were excluded. Total ASQi scores were created by summing raw scores across the five domains before standardizing.

Other outcomes included maternal dietary diversity assessed using the Minimum Dietary Diversity for Women score, an indicator of adequate dietary diversity when at least 5 of 10 food groups are consumed in the previous 24 hours [49]. Dietary diversity in young children is a similar indicator, and the cut-off for achieving the minimum is the consumption of at least 5 of 8 food groups, including breastmilk [50]. Maternal depressive symptoms were measured with the 20-question Center for Epidemiologic Studies Depression scale (CES-D). Maternal depressive symptoms scores were analyzed with the continuous 60-point CESD-D score. Maternal knowledge about lead was assessed by asking if respondents had ever heard of lead, and household WASH status was assessed through the observed presence of a handwashing station with water and soap or a soapy water bottle and of a clean, functional, hygienic latrine in the household.

Ongoing inter-rater reliability was conducted during the endline assessment for 4.7% of the sample ($n = 27$). Inter-rater reliabilities for the ASQi domains and the home observation subscale were high (the intraclass correlation for ASQi domains was ≥ 0.99 for all domains except for personal social, where it was 0.93; for the home observation subscale the ICC was 0.92)

As a supplementary analysis, we collected data on a direct-assessment measure of child development, the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III), for a stratified random subset of 16 villages from those that had children of both sexes in each age group (8 control, 4 group, 4 combined); 254 children ($n = 134$ control; $n = 120$ intervention) were randomly selected after stratifying by age group (6-12, 13-18, and 19-24 months) and sex.

Statistical analysis

Sample size calculations were conducted for the total FCI score (range 0–13), based on a difference of 2.0 in mean total score between each intervention arm and the control arm, and an SD of 3.3 [16]. The calculations assumed an intracluster correlation of 0.20, power of 0.80, and type 1 error of 0.05. With 20 participants per cluster, the sample size calculations indicated that 15 control arm villages and 8 villages in each of the intervention groups were required. The study was not powered to detect differences between the two intervention delivery methods.

All analyses were conducted according to the randomized intervention arm at enrollment (intention to treat), without considering session attendance. The primary analysis consisted of age-adjusted mean differences between the control arm and each of the group and combined intervention arms and at endline for the primary outcomes (FCI play activities and play materials subscales). Secondary analyses include mean differences (for continuous outcomes) or prevalence differences (for binary outcomes) for child development and other outcomes. Potential covariates for inclusion in adjusted models were selected based on the

child development literature, and included parental education, child age and sex, household income, household wall material, household assets and the outcome of interest measured at baseline. Interviewer was also included as a potential covariate in the adjustment set for child development and observed home environment outcomes. For each outcome, covariates were prescreened using a likelihood ratio test, and all covariates with $p< 0.20$ were included in adjusted analyses. Adjusted analyses were done with parametric g-computation using linear regression for continuous outcomes and logistic regression for binary outcomes to generate mean differences and prevalence differences for these outcomes, respectively [51]. CIs were generated with bootstrapped samples clustered by village (1000 samples). For each outcome except for the Bayley-III, two comparisons to the control arm were made, one for each intervention arm. For the Bayley-III assessment, only one comparison to the control arm was made, with children in any intervention arm combined due to small sample sizes. No adjustments were made for multiple comparisons [52]. Analyses were performed in Stata V.14 and R (V.4.0.1, Vienna, Austria), with the riskCommunicator package [53].

1.4 Results

Between July and August 2017, fieldworkers enrolled 621 pregnant women and primary caregivers of children under 15 months of age located in 31 villages in the RINEW trial. At intervention endline, 47 (7.6%) participants were lost to follow-up, and 6 participants had only 1 day of data collection complete, resulting in full data collection on 574 (91.4%) participants (figure 1.1); the majority of those lost had migrated. Loss to follow-up was not statistically significantly different across study arms (control 6.6%; group 10.0%; combined 6.9%), or demographic variables collected from participants at baseline. Intervention arms were similar when compared with the control arm across many baseline values (table 1.2). At intervention endline, the mean age of the children assessed was 16.5 months (range 3.9—26.4, SD 5.4), and the primary caregiver was the target child's mother for 570 (99%) participants (4 interviews were done with other female primary caregivers of the child).

The mean number of the 18 sessions attended was similar across arms, with 14.2 (SD 4.0) in the group arm, and 15.4 (SD 3.2) in the combined arm. Participants in both intervention arms had a higher prevalence of any children's picture books in the home at intervention endline (control 19%, group 85%, combined 93%), an indication that participants kept the books they received in intervention sessions.

Home stimulation and child development outcomes

Children in the group and combined intervention arms received significantly more stimulating activities in the past 3 days from their primary caregiver (age-adjusted means: group 4.22 (95% CI 3.97 to 4.47); combined 4.77 (4.60 to 4.96); control 3.24 (3.05 to 3.39)), had a larger variety of stimulating play materials in the home (age-adjusted means: group: 3.63

(3.31 to 3.96); combined: 3.81 (3.62 to 3.99); control 2.48 (2.34 to 2.59)), and had improved scores for the observation of caregiver responsiveness and the child's environment scale (age-adjusted means: group: 8.82 (8.59 to 9.10); combined: 8.93 (8.67 to 9.18); control 8.26 (8.05 to 8.45)) when compared with the control arm (figure 1.2, unadjusted means and adjusted mean differences in table 1.3). For comparison with other work, we calculated the unadjusted Cohen's d effect size for the stimulating caregiving activities outcome for the group (0.66 (0.45 to 0.87)) and combined (1.08 (0.87 to 1.29)) arms (results not shown). The stimulation activities 'played with', 'read books to', and 'sang songs to' were the stimulation items that had the highest prevalence differences when comparing the intervention groups to the control (figure A.1).

Children in the group and combined arms scored higher than the control arm on all domains of the ASQi and CDI assessments. Differences for the group arm were significant for all domains except for problem-solving and receptive language, and differences for the combined arm were not significant for any domains (adjusted mean differences for standardized total ASQi score; group vs. control 0.39 (0.15 to 0.64); combined vs. control 0.25 (−0.07 to 0.54)) (figure 1.3, table A.5). In the supplementary analysis comparing both intervention arms to the control arm for the five domains of the Bayley-III assessment, the expressive communication and total Bayley-III scores were significantly higher for children in any intervention group compared to control (adjusted mean differences: expressive communication 0.33 (0.02 to 0.64); total Bayley-III score 0.38 (0.06 to 0.74; figure A.2).

Maternal mental health, nutrition, WASH and lead outcomes

In both intervention arms, participants reported fewer depressive symptoms as compared with the control arm, indicated by lower CES-D scores, with a control arm mean of 15.01 (SD 8.96). The adjusted mean differences were significant for the group arm −2.05 (95% CI −3.23 to 0.66), but not the combined arm −1.34 (−3.12 to 0.41); table 1.3). Minimum dietary diversity for mothers and children was improved in the combined intervention arm with adjusted prevalence differences of 0.14 (0.04 to 0.22) for mothers and 0.14 (0.05 to 0.19) for children. There were no significant differences between the group and control arms for maternal or child dietary diversity. There was no difference in the presence of a functional, clean, and hygienic latrine for either intervention arm when compared with the control, and only the group arm had higher prevalence of a handwashing station with soap and water (adjusted prevalence difference: 0.12 (0.01 to 0.24)). Participants in both intervention arms had greater potty use when compared with the control, with differences significant for the combined arm (adjusted prevalence differences: combined 0.10 (0.00 to 0.21); group 0.13 (−0.02 to 0.28). Knowledge of lead was significantly higher in both intervention arms, with a control arm prevalence of 0.24 and adjusted prevalence differences of 0.51 (0.41 to 0.61) and 0.52 (0.39 to 0.63) for the group and combined arm, respectively.

1.5 Discussion

We found beneficial intervention impacts on our primary outcome of stimulation in the home, our secondary outcome of child development, as well as across a range of risk factors for child development addressed by the multicomponent intervention including caregiver depressive symptoms, caregiver and child dietary diversity, WASH, and knowledge of lead. We observed impacts on play activities, play materials, and observations of the home environment in both intervention arms. However, as the source of each play material was not asked, the results cannot be disaggregated by materials provided in the sessions and materials caregivers obtained on their own. Child development scores, as assessed by the ASQi and CDI were higher in both intervention arms when compared to the control, with differences for 6 of the 7 domains significant for the group arm, but not for the combined arm. Differences in standardized child development scores for the group intervention arm are between 0.18 to 0.39, similar to those from a group intervention in rural India which found significant improvements in cognition scores of 0.28 SDs [54], and an integrated home visiting program in rural China, with intervention effects of 0.24 SDs [41]. The results from a supplementary analysis on a subset of participants demonstrate improved receptive communication and total Bayley-III scores among those who received any intervention compared with the control. Though the current study was not powered to compare the group and combined arms directly, recent work from rural India finds similar effects on child development outcomes from group and individual home-based sessions [54]. Further, recent work from rural Kenya indicates that group sessions may outperform combined delivery in some settings [55]. Our work supports the delivery of multicomponent, mixed-age group sessions to improve risk factors for poor child development in rural Bangladesh.

Maternal depression is a risk factor for delayed child development, and has been associated with poor behavioral and developmental outcomes [56]. We found fewer reported depressive symptoms from caregivers in both intervention arms, with a significant difference for the group arm. We hypothesize that the focus on maternal depression and the social support facilitated during the group intervention sessions contributed to the reduced depressive symptoms in both intervention arms. The effect may have been stronger in the group arm because this delivery mode offered structured peer social support 18 times over the course of the intervention compared to nine times in the combined arm. A meta-analysis of the effect of child stimulation interventions on caregiver depressive symptoms found no significant effect (-0.20 (-0.23 to 0.03)) [57]. However, the only group-based intervention that included contents on mental health found an effect size of -0.54 (-0.76 to -0.32). This is higher in magnitude than the effect in the group arm of the current study (unadjusted Cohen's d: -0.22 (-0.42 to -0.02)) [19].

We found improvements in nutrition, WASH and knowledge of lead in the intervention group. In the combined arm, a higher proportion of children and caregivers had a more diverse diet, though no difference was found for the group arm. As improving dietary diversity require

changes in both purchasing and meal preparation, it may be that an approach where CHWs can respond to individual needs of families is required. There was no difference in presence of hygienic latrines in either intervention arm when compared to the control, more families in the group arm had a handwashing station with soap and water, and more caregivers in both arms reported that their child used a potty regularly, though the difference was only statistically significant for the combined arm. We do not know if the 17 families who were provided a potty would have purchased and used one in the absence of it being provided. We hypothesize that a subsidy may be required to improve hygienic latrine status in this low-income community given the investment required to upgrade WASH infrastructure. Caregiver's knowledge of lead was improved in both arms, with large effect sizes. Lead is an invisible toxin unknown to the majority of the population at baseline, thus knowledge of lead is the first step towards reducing exposure. Secondary analyses will further investigate the intervention effects on behaviors related to lead exposure.

The unadjusted Cohen's d effect sizes for stimulating caregiving practices in both intervention arms (group 0.66 (95% CI 0.45 to 0.87); combined 1.08 (0.87 to 1.29)) are slightly larger than the pooled effect sizes (0.57 (0.37 to 0.77)) from a recent meta-analysis of the effect of stimulation interventions on stimulation in the home, measured by the FCI and Infant Toddler Home Observation for Measurement of the Environment [57]. The interventions in the meta-analysis included at most two additional components in additional to child stimulation, whereas our study included four additional components. Further, none of the interventions in the meta-analysis were delivered to both pregnant women and caregivers of children under 24 months of age. Thus, our findings suggest that the effects on caregiver-related outcomes were not diminished with the inclusion of multiple integrated intervention components, nor intervention delivery across both pregnant women and caregivers of young children.

In the RINEW intervention, groups were based on geographical proximity to reduce barriers to attendance, and included pregnant women and mixed-age, caregiver-child dyads. Other group-based ECD interventions are delivered to groups with children of similar ages, to allow for the presentation of age-specific materials relevant to the whole group [41]. In addition to reducing barriers to attendance, grouping participants with others who they may interact with daily may increase the potential for continued social support for intervention activities outside sessions. However, delivering sessions to a mixed-age group may increase session duration, and include less engaging components for some participants (ie, pregnancy contents for non-pregnant participants) [58]. In future work, the trade-offs of mixed-age delivery and geographic proximity should take into account the geographic density of eligible participants and accessibility of session locations.

This study has several strengths, including the focus on a group-based intervention to address multiple risk factors for poor ECD, which makes this approach more scalable than one-on-one home visiting programs. In addition, the intervention was delivered simultaneously to

both pregnant and lactating women with children of mixed-ages, an approach that is easier to scale than more narrowly focused programs. We used a tablet application to facilitate session delivery, which enabled the inclusion of multiple age-specific intervention components. Another unique feature of this program was the integration of information on the reduction of heavy metal exposure, in addition to standard messages about nutrition, health, and hygiene. Finally, we included a set of outcomes that spans a broad range of influences in early life and development in order to gain a more comprehensive understanding of the impact of the intervention on a child's development and well-being.

The current study has important limitations. First, due to budget constraints, the sample was not powered to detect small differences on many of the secondary outcomes, or differences between the two active intervention arms. As such, we are only able to interpret the direction and magnitude of these effects. Second, the FCI, ASQi, CDI, dietary diversity, depressive symptoms, and knowledge of lead assessments are primarily based on caregiver-report, allowing for the possibility that caregiver responses about behaviors could be influenced by knowledge and social desirability, or caregivers' mental health status. The risk of respondent bias was minimized through extensive training of survey enumerators, the use of direct-observation items within the ASQi, an observation scale to complement the FCI, and follow-up questions to confirm reported lead knowledge. We found significant improvements in observed caregiver responsiveness and the caregiving environment, highlighting that changes were found for observed behavior in addition to caregiver report. Additionally, as this was the first time implementing such an intervention curriculum in Bangladesh, there were some adjustments to the strategies used to build group cohesiveness and encourage attendance, and intervention modules were refined as the sessions progressed. The results do not represent the impacts of the intervention that may be achieved with further refinements, and the current estimates may be a lower bound on the possible impact. Finally, we were not able to examine the cost effectiveness of group compared with individual or combined delivery mechanisms for integrated interventions.

This intervention illustrates the feasibility of locally recruited CHWs delivering a group-based, mixed-age, multicomponent ECD intervention in rural Bangladesh. The feasibility of scaling such a group-based intervention through a government health system, or the large-scale implementation through a regional or national non-government organization is not known, and warrants exploration. A promising, recent study found that child stimulation sessions delivered through Government of Bangladesh community clinics to pairs of mother-child dyads, had large impacts on child development [59]. Notably, these clinics serve as regular point of care, routinely providing maternal and neonatal healthcare as well as nutrition and health education. Differences in CHW workload, session attendance, and intervention impacts with each of the delivery mechanisms will inform the design of scalable and impactful child development interventions.

The long-term impact of the RINEW intervention, or similar integrated interventions tar-

geting multiple risk factors for ECD, will be critical to understand the scope of intervention impact. Although many stimulation interventions have shown impacts on child development outcomes at intervention endline, there is mixed evidence on the later impacts of these early interventions [44, 60]. It is possible that integrated interventions addressing multiple risk factors may contribute to sustained intervention impacts on child development as they more holistically improve children's early-life caregiving and health environments. Medium and long-term follow-up of children enrolled in multicomponent interventions is required to examine this hypothesis.

1.6 Conclusion

In conclusion, we found that a carefully designed group-based multicomponent intervention delivered by well-trained CHWs can address multiple additional risk factors for child development beyond stimulating caregiving, and demonstrate similar effects on stimulating caregiving as interventions with fewer integrated components. CHWs were able to deliver the complex multicomponent RINEW intervention for 9 months and community members regularly attended intervention sessions regardless of delivery platform. This multicomponent approach may be used as a template to design a scalable and impactful intervention to improve child well-being in low-income settings.

1.7 Figures and Tables

Figure 1.1: Trial profile

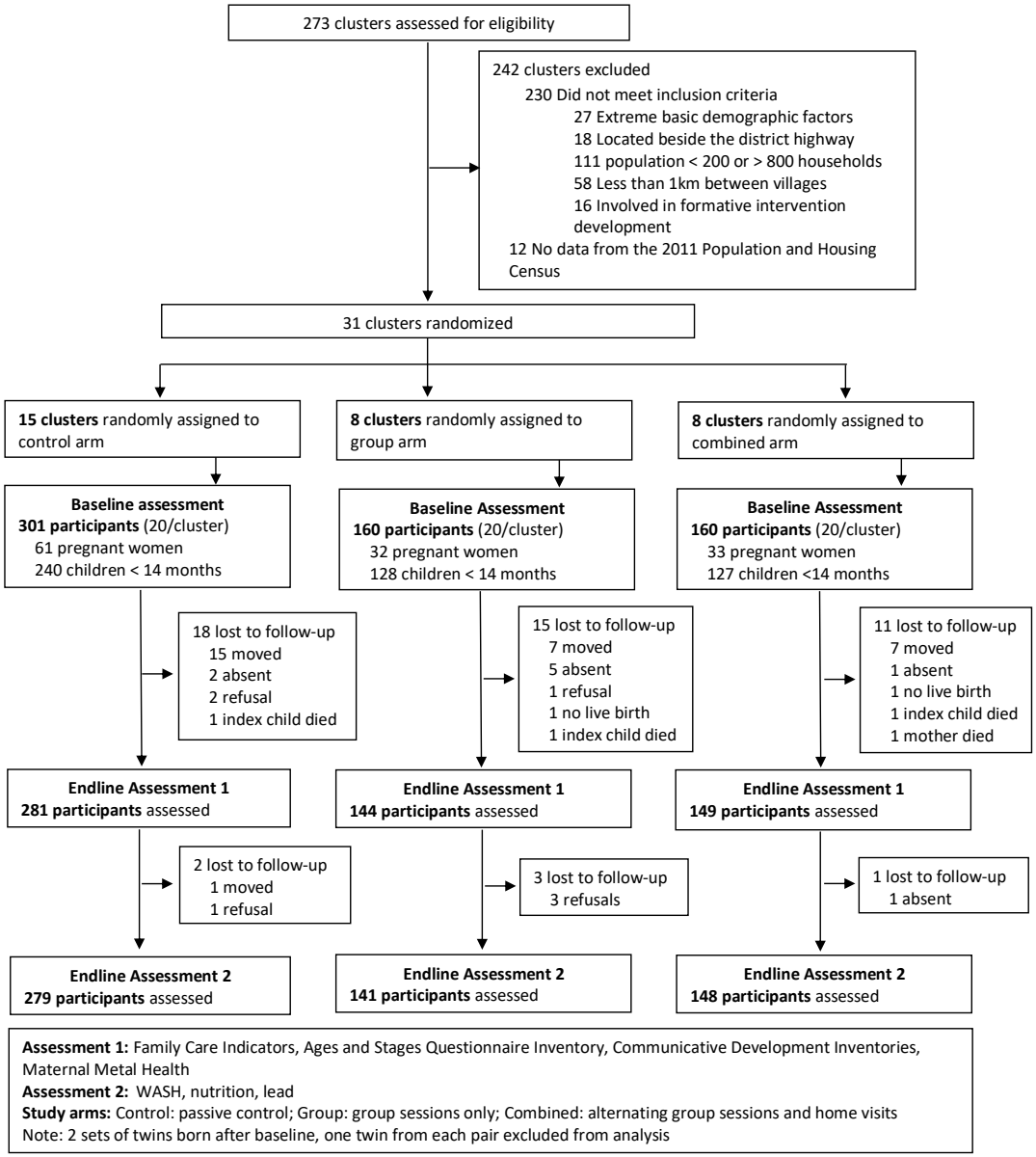
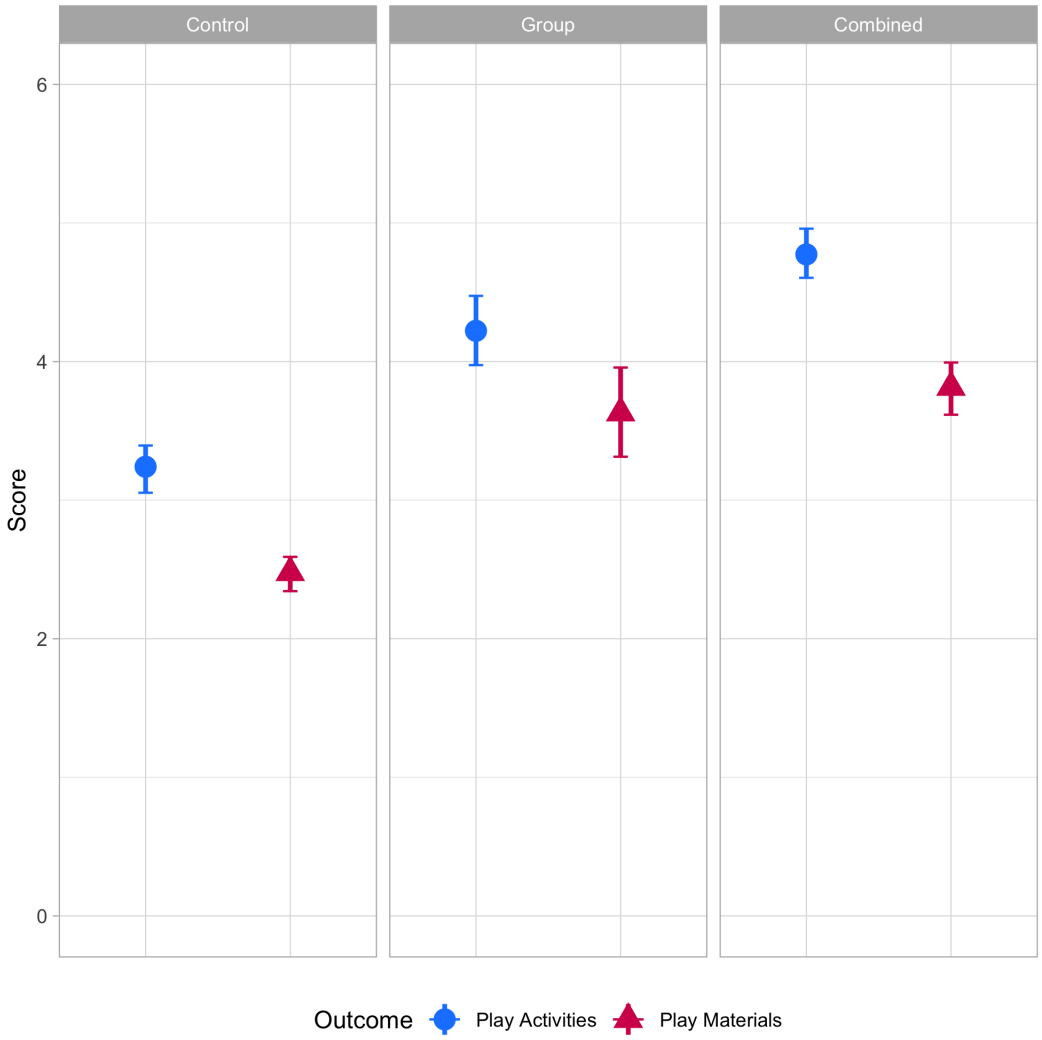
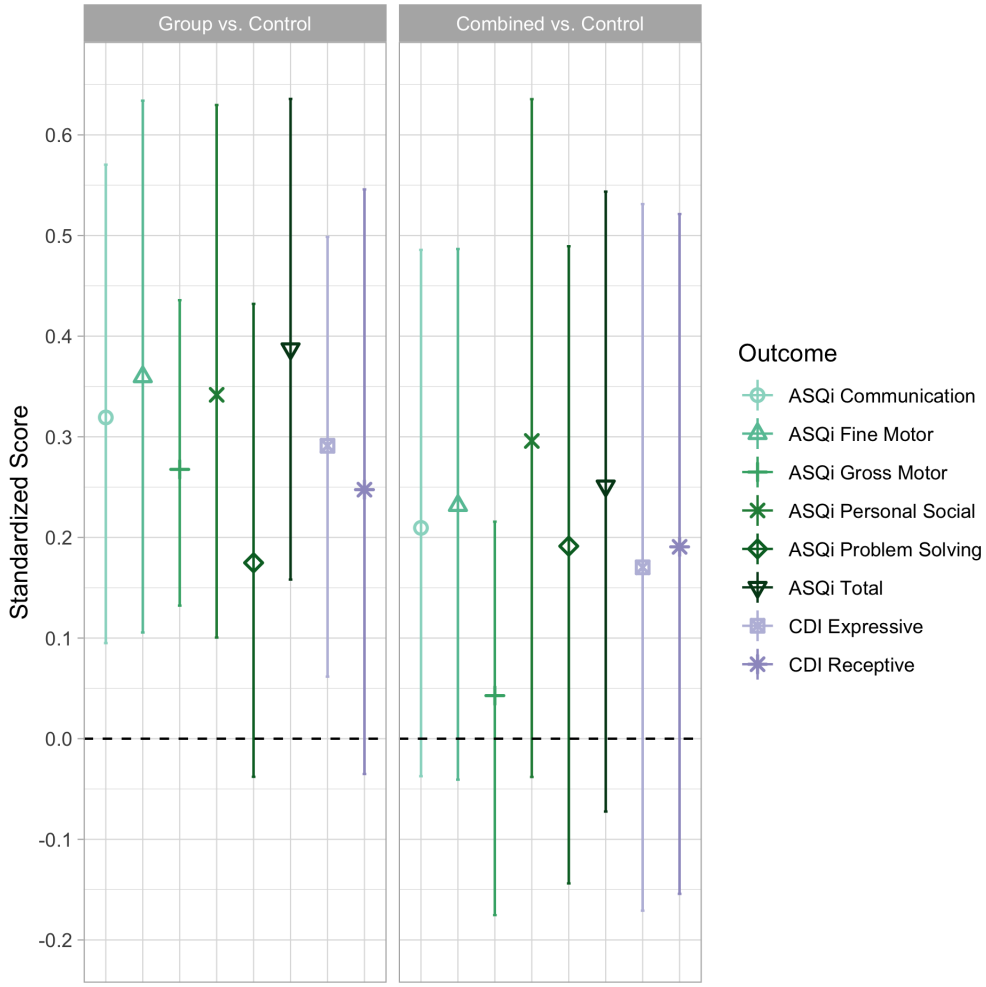


Figure 1.2: Mean stimulation in the home by study arm at endline



Points represent mean scores in each intervention arm, adjusted by child age at endline. Bars represent 95% CIs. Play activities (0-6): number of play activities that the primary caregiver engaged in with the child in the last three days. Individual items summed, and include: read books or looked at picture books; told stories; sang songs; took outside the home; played; named, counted or drew. Play Materials (0-6): number of varieties of play materials observed in the home, and reported that the child played with in the last 30 days. Individual items summed, and include things: that play music; for drawing or writing; for pretending; used when running or jumping; for teaching shapes; for stacking

Figure 1.3: Adjusted mean differences in age-standardized Ages and Stages Questionnaire Inventory (ASQi) and Communicative Development Inventory (CDI), by intervention arm



Results for all domains internally age-standardized to the control arm, points represent mean differences in standardized scores between each intervention arm and the control arm, lines represent 95% CIs. Sample size by domain, after removing outliers and missing data: ASQi communication, $n = 566$; ASQi fine motor, $n = 559$; ASQi gross motor, $n = 563$; ASQi problem solving, $n = 563$; ASQi personal social, $n = 550$; ASQi total, $n = 532$; CDI receptive $n = 573$; CDI expressive $n = 498$ (only including children over 9 months old)

Table 1.1: Intervention components

Component	**Description of the intervention components delivered in the group and combined intervention arms**
Nutrition	Contents: This component was adapted from the WASH-Benefits intervention [34]. The nutrition component included specific recommendations for each trimester of pregnancy, recommendations for lactation, and recommendations for the complementary feeding period. Supplements: Nutritional supplements were distributed to participants depending on age and child nutritional status as indicated by mid-upper arm circumference (MUAC). • Pregnant women and caregivers of children under 6 months of age were given multiple micronutrients • Children with a MUAC 11.5-12.49 cm were given small-quantity lipid-based supplements (SonaMoni) • Children over 6 months with a MUAC 12.5 cm and greater were given multiple micronutrient powder (Pushtikona). • Children over 6 months with a MUAC under 11.5 cm were referred to a health facility (control arm children were also referred)
WASH	This component was adapted from the WASH-Benefits intervention and included activities to coach participants to identify changes they could make in their own environments [34]. Soapy water bottles were provided to all households
Lead	This component included teaching participants about the harms of lead and encouraging changes in their household to reduce lead exposure from previously identified lead sources: turmeric and lead-soldered cans [38, 61].
Mental Health	This component was adapted from the thinking healthy program [62]. Through field piloting the strategies were simplified, integrated with other intervention material and incorporated behavioral activation [45].
Targeted households	• Participants with a MUAC under 12.5 were considered at-risk: 17 at-risk participants who did not have access to their own hygienic latrine at baseline received WASH infrastructure (a child potty, a handwashing station, and a dual pit latrine). • Pregnant women received foot measurement sticks: participants who gave birth to a child who had a foot length of < 7cm received a visit from the CHW who confirmed the foot length and provided (1) a session to teach the mother to provide Kangaroo Mother Care (KMC) to the baby; and (2) gave the mother a KMC kit that included three KMC pouches, one hat and one pair of socks.
Toys and books	All participants in sessions received low-cost picture books and toys made out of recycled materials for children over 6 months of age

WASH: water, sanitation and hygiene; MUAC: Mid-upper arm circumference

Table 1.2: Characteristics of the sample at baseline

	Study arm, n (%) or mean ± SD		
	Control (n=301)	**Group (n=160)**	**Combined (n=160)**
Caregiver Characteristics			
Age	25 ± 5.6	25 ± 6.4[1]	25 ± 6.2
Completed primary education	173 (57%)	86 (54%)	101 (63%)
Pregnant woman enrolled	61 (20%)	32 (20%)	33 (21%)
CES-D score (0-60)	12.4 ± 8.6	12.8 ± 8.7[1]	13.4 ± 9.7
Knowledge of lead	67 (22%)	47 (29%)	49 (31%)
Child Characteristics (n=496)			
Age (in months)[2]	7.0 ± 3.9	6.7 ± 3.9	7.5 ± 4.0
Female	134 (56%)	78 (61%)	66 (52%)
FCI Play activities subscale (0-6)[3]	3.3 ± 1.7	3.8 ± 1.4	3.4 ± 1.6
FCI Play materials subscale (0-6)[3]	2.1 ± 1.1[8]	2.2 ± 1.0	2.2 ± 1.2
Home observation subscale (0-11)[3]	7.9 ± 1.5	8.1 ± 1.4	8.2 ± 1.3
1+ children's book(s) present in home (n=288)[3,4]	16 (11%)	9 (14%)[9]	6 (8%)
MUAC <12.5 cm (n=296)[3]	11 (8%)[10]	6 (8%)[9]	7 (9%)
Household Characteristics			
Household size	5.2 ± 2.2	5.3 ± 2.6	5.2 ± 2.0
Number of children 2-15 years	1.3 (1.2)	1.2 (1.1)[1]	1.2 (1.2)
Has cement floor	65 (22%)	28 (18%)	26 (16%)
Has brick walls	74 (25%)	27 (17%)	26 (16%)
Has electricity	243 (81%)	150 (94%)	139 (87%)
WASH			
Access to a handwashing station with water and soap or soapy water	62 (21%)	36 (23%)	33 (21%)
Access to a hygienic latrine[4]	102 (34%)	59 (37%)	43 (27%)
Use of potty[3,5] (n=297)	35 (24%)[10]	16 (22%)	22 (28%)
Nutrition			
Maternal number of food groups	5.0 ± 1.3	4.9 ± 1.3	5.0 ± 1.4
Maternal minimum dietary diversity[6]	182 (60%)	88 (55%)	101 (63%)
Child number of food groups (n=272)[3]	3.8 ± 1.5	3.7 ± 1.4[8]	4.1 ± 1.4
Child minimum dietary diversity[3,7] (n=272)	45 (35%)	19 (29%)[8]	28 (37%)[11]

CES-D: Center for Epidemiologic Studies 20 Question Depression questionnaire, scores range from 0–60, with higher scores indicating more depressive symptoms experienced; FCI: Family Care Indicators, the play activities subscale is a sum score of the number of play activities that the caregiver participated in with the child in the previous three days (0-6), the play materials subscale is the number of varieties of play materials observed in the home, and reported that the child played with in the last 30 days (0-6); MUAC: Mid-upper arm circumference; WASH, water, sanitation and hygiene [1]$n = 159$, 1 participant did not respond [2]Including index children born as of the baseline assessment [3]Index children > 6 months of age at baseline included ($n = 296$, control = 144, group = 73, mixed = 79), unless otherwise indicated. (notes continued on next page)

Table 1.2: (continued notes from previous page) [4]Clean, functional, Hygienic latrine. Government of Bangladesh National Sanitation Strategy, 2005 definition of hygienic latrine: Flush or pour-flush toilet/latrine to (1) Piped sewer system, (2) Septic tank; Pit latrine with slab and water seal; Pit latrine with slab and lid, no water seal; Pit latrine with slab and flap, no water seal; Ventilated Improved Pit latrine; Composting latrine. [5]Use of potty for $> 50\%$ of defecation events in last week [6]Mother reported eating 5 or more food groups in the last 24 hours, out of the following 10 groups: grains, roots, and tubers, pulses, nuts and seeds, dairy products, animal flesh foods, eggs, dark green leafy vegetables, other vitamin A rich fruits and vegetables, other vegetables, other fruits [7]Children > 6 months reported eating 5 or more food groups in the last 24 hours, out of the following groups: breast milk, grains, roots, and tubers, legumes and nuts, dairy products, animal flesh foods, eggs, vitamin A rich fruits and vegetables, other fruits and vegetables [8]$n = 143$ [9]$n = 66$ [10]$n = 145$ [11]$n = 76$

Table 1.3: Child development, maternal mental health, nutrition, water, sanitation and hygiene and lead outcomes at endline

Indicators	Unadjusted mean ± SD or n (%) by arm			Adjusted mean difference or risk difference vs. control (95% CI)	
	Control	Group	Combined	Group	Combined
FCI					
Activities	3.2 ± 1.5	4.2 ± 1.5	4.8 ± 1.3	1.05 (0.72 to 1.34)	1.56 (1.33 to 1.78)
Materials	2.5 ± 1.4	3.6 ± 1.7	3.9 ± 1.6	1.18 (0.88 to 1.51)	1.36 (1.18 to 1.51)
Observation	8.3 ± 1.5	8.8 ± 1.2	8.9 ± 1.5	0.56 (0.26 to 0.89)	0.67 (0.35 to 0.99)
ASQi					
Communication	56.0 ± 16.4	58.4 ± 15.3	58.6 ± 16.3	0.32 (0.10 to 0.57)	0.21 (-0.04 to 0.49)
Fine Motor	50.2 ± 11.3	51.8 ± 10.1	51.6 ± 11.6	0.36 (0.11 to 0.63)	0.23 (-0.04 to 0.49)
Gross motor	58.6 ± 15.1	60.8 ± 14.1	59.3 ± 16.1	0.27 (0.13 to 0.44)	0.04 (-0.18 to 0.22)
Problem-solving	55.6 ± 15.8	56.9 ± 14.6	58.2 ± 15.6	0.18 (-0.04 to 0.43)	0.19 (-0.14 to 0.49)
Personal Social	55.3 ± 16.3	57.7 ± 15.5	58.0 ± 16.4	0.34 (0.10 to 0.63)	0.30 (-0.04 to 0.64)
Total	275.7 ± 71.2	284.6 ± 65.2	285.0 ± 72.6	0.39 (0.16 to 0.64)	0.25 (-0.07 to 0.54)
CDI					
Receptive	44.9 ± 23.7	49.0 ± 22.4	49.0 ± 23.3	0.25 (-0.04 to 0.55)	0.19 (-0.15 to 0.52)
Expressive	16.8 ± 17.2	18.9 ± 17.9	19.2 ± 18.3	0.29 (0.06 to 0.50)	0.17 (-0.17 to 0.53)
Depressive symptoms					
CES-D score	15.0 ± 9.0	13.2 ± 7.0	14.1 ± 9.1	-2.06 (-3.23 to -0.66)	-1.34 (-3.12 to 0.41)
Minimum dietary diversity					
Child[1]	146 (54%)	85 (61%)	98 (68%)	0.07 (-0.03 to 0.17)	0.14 (0.04 to 0.22)
Maternal[2]	177 (63%)	89 (63%)	113 (76%)	0.03 (-0.07 to 0.12)	0.14 (0.05 to 0.19)
WASH					
Hygienic latrine[3]	96 (34%)	49 (35%)	51 (35%)	-0.03 (-0.19 to 0.11)	0.02 (-0.12 to 0.15)
Handwashing station with soap and water	59 (21%)	44 (31%)	36 (24%)	0.12 (0.01 to 0.24)	0.04 (-0.08 to 0.19)
Use of potty[4]	55 (20%)	42 (30%)	44 (30%)	0.13 (-0.02 to 0.28)	0.10 (0.00 to 0.21)
Lead					
Knowledge of lead	68 (24%)	103 (73%)	110 (74%)	0.51 (0.40 to 0.60)	0.52 (0.39 to 0.63)

Activities: Number of play activities that the mother participated in with the child in the last three days (out of six); Materials: Number of varieties of play materials observed available in the home (out of six); Observation: 11 observation items about caregiver-child interactions during the interview, and observations of the home environment (A.3 for details) For ASQi and CDI results: unadjusted mean values are raw values before standardization, all adjusted mean differences use scores which are internally age-standardized to the control arm, and point estimates represent SDs away from the control arm mean. Adjusted analyses are adjusted the following potential covariates: interviewer (for FCI, ASQi and CDI outcomes), maternal and paternal education, child age, child sex, household income above the median, household wall material, presence of electricity in the home, the presence of household assets (wardrobe, table, chair, watch/clock, television, bicycle, sewing machine) and the measure assessed at baseline (if assessed in the whole population). Covariates with $p < 0.20$ from a likelihood ratio test for each outcome are included in adjusted analyses. (notes continued on next page)

Table 1.3: (continued notes from previous page) n for ASQi and CDI outcomes (excluding outliers ± 4 SD from the control arm mean): Communication $n = 566$; Fine Motor $n = 559$, Gross Motor $n = 563$; Problem-solving $n = 563$; Personal Social $n = 550$; Total $n = 532$; Receptive $n = 573$; Expressive (only children over 9 months of age included) $n = 498$ ASQi: Ages and Stages Questionnaire Inventory; CDI: Communicative Development Inventories CES-D: Center for Epidemiologic Studies 20 question Depression scale; WASH: Water, Sanitation and Hygiene [1]Children > 6 months reported eating 5 or more food groups in the last 24 hours, out of the following groups: breast milk, grains, legumes, dairy products, flesh foods, eggs, vitamin A rich fruits and vegetables, other fruits and vegetables ($n = 552$) [2]Mother reported eating 5 or more food groups in the last 24 hours, out of the following 10 groups: grains, legumes, nuts and seeds, dairy products, flesh foods, eggs, vitamin A rich fruits and vegetables, other vitamin A rich fruits and vegetables, other vegetables, other fruits [3]Hygienic latrine (according to Government of Bangladesh National Sanitation Strategy 2005): Flush or pour-flush toilet/latrine to (i) Piped sewer system, (ii) Septic tank; pit latrine with slab and water seal; pit latrine with slab and lid, no water seal; pit latrine with slab and flap, no water seal; ventilated improved pit latrine; composting latrine [4]Use of potty for $> 50\%$ of defecation events for the index child in the last 7 days

Chapter 2

Concurrent validity of the Ages and Stages Questionnaire Inventory and the Bayley Scales of Infant Development-III in rural Bangladesh

2.1 Abstract

Background Accurate measurements of early child development outcomes are necessary for the design of effective interventions, programs, and policies to improve early child outcomes. Direct assessment tools are considered more appropriate than tools that rely on caregiver report because they avoid the possibility of caregiver report bias. One widely used direct assessment measure for cognitive, motor, and language development in children under two in low- and middle-income countries (LMICs) is the Bayley Scales of Infant and Toddler Development III (Bayley-III). The Bayley-III, however, is an expensive tool that requires extensive training, a controlled environment for administration, and a lengthy administration time. An alternative assessment tool, the Ages and Stages Questionnaire Inventory (ASQ:I), is primarily a caregiver-report questionnaire which was adapted to the Bangladeshi context to include a subset of direct observation items that can be administered more quickly, inexpensively, and with less training than the Bayley-III. We aimed to assess the concurrent validity of the Bangladeshi-adapted ASQ:I with the Bayley-III in children 4-27 months of age in rural Bangladesh

Methods We collected data in July and August 2019. The analytic sample included 244 children aged 4-27 months in Kishoreganj, Bangladesh. The sample was comprised of a sub-sample of endline data from an evaluation of an integrated early child development intervention (52%, $n = 128$ received the intervention). The Bayley-III and ASQ:I testers

were separately trained and independently assessed children on separate days on average 13 days apart. All children received the ASQ:I assessment first. We internally age-standardized domain-specific and total scores for each measure, and assessed concurrent validity between domains using Pearson correlations for the full sample as well as stratified by age group and intervention status. Finally, we assessed correlations between test scores and variables theoretically related to child development including maternal education and stimulation in the home.

Results The overall correlation between total ASQ:I and Bayley-III total scores was moderate ($r = 0.42$ 95% CI: 0.30–0.53). There were no systematic differences in correlations between concordant domains of ASQ:I and Bayley-III assessments for the group that had received the early child development intervention compared to the control. Across the full age range, concurrent validity was highest for the gross motor domains of both assessments ($r = 0.51$, 0.40–0.60), and lowest for the fine motor domains of both assessments ($r = 0.20$, 0.04–0.33). Total ASQ:I and Bayley-III scores were both significantly positively correlated with stimulation in the home and maternal education (correlations ranged from 0.17 to 0.43 for total scores).

Conclusion The Bangladeshi-adapted ASQ:I is a low-cost tool that can be feasibly administered in rural Bangladesh, is moderately correlated with the Bayley-III, and can be used to evaluate intervention effects and monitor child development over time when human, time, or financial resources are constrained.

2.2 Introduction

Over 249 million children in low- and middle- income countries were at risk for poor development in 2010 [9]. Child development is a global priority, as demonstrated by the explicit inclusion of child development in the sustainable development goal 4.2, "By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education" [12]. The development and evaluation of early child development interventions is transitioning from small-scale efficacy trials to large-scale delivery through routine health systems. Valid and feasible measurement of child development is important to understanding which interventions work to improve child development outcomes at-scale and track child development at the population level [63]. The tools used to evaluate child development outcomes following small scale efficacy trials may no longer be feasible for the evaluation of large-scale interventions where resource, personnel, or time resources are limited.

Tools to evaluate the effects of child development interventions in low- and middle-income countries typically aggregate responses on individual questions to estimate children's underlying developmental status on different developmental domains including cognitive, motor,

and language development [24]. Direct child assessments include test items that are administered directly to the child by a trained enumerator, and are less biased and more precise compared to caregiver report, especially when the assessments are used for the evaluation of intervention impacts [24]. This is because assessments that are administered directly to the child avoid the potential bias from differential caregiver report depending on intervention status. Direct assessments may also be more accurate in determining children's developmental status in the case of milestones or abilities that caregivers may not yet notice have developed [24]. If enumerators are masked to intervention status, direct assessments can avoid differential scoring based on intervention status on the part of enumerators, however they may still represent biased effect estimates due to children's differential comfort with enumerators. Children who did not receive the intervention may be more reserved with strangers and therefore the assessment may underestimate their true ability as compared to children who received the intervention, and are more used to interacting with strangers. This bias is considered to be smaller than the potential bias due to caregiver report [24], and as such direct assessment measures are considered to have less bias in assessment of intervention effects.

One such direct assessment measure that has been used in over 20 countries globally is the Bayley Scales of Infant and Toddler Development III (Bayley-III), a tool that includes direct assessment for the evaluation of cognitive, motor, and language development between the ages of one and forty-two months [24, 64, 65]. The Bayley-III, however, is a copyrighted tool that comes with a large initial cost, as well as a high cost per assessment that requires skilled enumerators, more extensive training compared to caregiver report, a controlled environment for administration, and a lengthy administration time [24]. Thus, the Bayley-III is not only difficult to use in settings where financial, human, or time resources are constrained, but can also be prohibitively time and resource intensive for large-scale evaluations in low-resource settings.

A caregiver report assessment is a questionnaire that is either filled in directly by the child's primary caregiver, or administered as an interview with the child's primary caregiver. The Ages and Stages Questionnaire (ASQ) assessments is primarily caregiver a report and has been used in over 20 LMICs [66]. These caregiver report assessments are cheaper than the Bayley-III and can be administered more quickly with less training. The version of the ASQ most used to assess child development in low- and middle-income contexts to date is the ASQ-3 [66], which has been used primarily as a caregiver-report survey with inclusion of observation items when the parent is unable to answer a question. The ASQ-3 includes the administration of 6 questions for each domain that depend on the child's age group. It was designed as a short screening tool to detect developmental delay and is used in well-baby visits. This assessment may include instructions to caregivers about what milestones to watch for before the next visit.

In the last two decades the ASQ-3 has also been used to evaluate the impacts of early inter-

ventions on child development in low-income contexts. Two previous studies that examined the concurrent validity of the ASQ-3 against the Bayley-III in upper-middle-income countries, in rural China [26, 27] and urban Colombia [25], found low to medium correlations amongst children under 24 months of age. The ASQ-3 has been adapted for use in research studies to avoid ceiling effects that occur because of the small number of questions asked per domain, and to include direct observation of some items that caregivers might not observe. Adaptations to the ASQ-3 include the extended Ages and Stages Questionnaire (EASQ), which not only includes a subset of direct assessment items, but also extends the number of questions asked to children in each age range by adding three questions from both the previous and subsequent age groups. The EASQ was adapted from the ASQ-3 by researchers, and has been used to evaluate programs in multiple LMICs including Bangladesh and Kenya [34, 67, 68]. A further adaptation, the ASQ:Inventory (ASQ:I), expands on the EASQ and is administered as a continuous measure with starting rules that depend on the child's age, and stopping rules that depend on the child's performance [28]. The ASQ:I reduces the potential for the ceiling effects that are found in ASQ-3 and EASQ which limit the number of questions for each domain to 6 or 12, respectively. The ASQ:I has been used in the evaluation of an intervention in Madagascar and in a longitudinal cohort of children in Kenya [29, 30]. The ASQ:I was adapted for use in China as a measure for population monitoring of child development. Its performance was compared to the Beijing Gesell Development Schedule, and it was found to have adequate psychometric properties [69]. In Bangladesh the ASQ:I was adapted to include a subset of items that are administered through direct assessment with inexpensive and common materials. Due to the continuous nature of the Bangladeshi-adapted ASQ:I and its explicit subset of direct assessment items, we hypothesized that it would have stronger concurrent validity with the Bayley-III as compared to the ASQ-3 and EASQ. In this study we aim to assess the concurrent correlation between the Bangladeshi-adapted ASQ:I against the Bayley-III in rural Bangladesh.

2.3 Methods

Participants

Data were collected between July and August 2019. Participants are a subset of those from the endline assessment of a cluster randomized controlled trial of an early child development intervention in Kishoreganj District, Bangladesh [70]. In the main trial, children with visible physical or cognitive disabilities were excluded, as was one randomly selected child for each pair of twins. At intervention endline, all children were assessed on the ASQ:I ($n = 574$ from 31 villages, 15 control and 16 intervention). For the validation study, a stratified subset of 16 villages (8 control, 8 intervention) were selected for the Bayley-III assessment from those that had children of both sexes in each age group (6–12, 13–18, and 19–24 months).

Measures

The Bayley-III assessment consists of five subscales (cognitive, gross motor, fine motor, receptive language and expressive language) that are administered through direct assessment. During scoring, these subscales can be combined into three composite domains that are externally standardized to a US sample: cognitive, motor, and language. For this analysis we examine the raw scores on the more disaggregated subscales as opposed to the composite domain scores in order to ensure the scores were comparable to those on the ASQ:I. This analysis is in line with previous work [25, 27]. The Bayley-III also includes two domains assessed through caregiver report (adaptive behavior and socio-emotional), which were not administered as part of this study. The Bayley-III was translated to Bengali and culturally adapted to the Bangladeshi context through a replacement of culturally inappropriate pictures without changing the order of the items or their underlying concepts. This cultural adaptation was previously validated in Bangladesh [65]. The Bayley-III served as the criterion measure in this analysis.

The adapted ASQ:I assessment for in this study was first piloted by members of the study team with 50 children in the Hossainpur subdistrict of Kishoreganj, Bangladesh in 2010. During this pilot, some items were adapted to ensure they were culturally appropriate, and direct assessment using inexpensive and common materials was piloted for a subset of the questions. The 7-day test-retest reliability of the assessment during this pilot ($n = 28$) was between 0.97-0.99 for all domains. The direct assessment items were further piloted with 453 children in 21 villages in Keraniganj subdistrict of Dhaka district, Bangladesh [71]. The ASQ:I consists of five domains: problem solving, gross motor, fine motor, communication, and personal social. The majority of the adapted ASQ:I items were assessed through caregiver report, with 16% of items assessed through direct assessment of the child (between 8-32% depending on the domain), using low-cost stimuli (table B.1; figure B.1). To ensure appropriate ordering of questions in order of increasing difficulty to justify stop rules, the ASQ:I was also piloted on 60 children not included in this study sample just prior to the start of the current study. Both the Bayley-III and ASQ:I have an age-based start rule, and a stopping rule that depends on the child's performance or reported ability.

In previous work, child development has been correlated with maternal education and stimulation in the home [68, 72]. In this study maternal education was assessed by asking the mother the number of years of education she had received. Stimulation in the home was assessed by the play activities and play materials subscales of the Family Care Indicators (FCI) [47]. The play activities subscale consists of six questions about the variety of stimulating play activities the child has participated in with an adult over the past 3 days. The play materials subscale consists of an observation of the variety of play materials that a child has played with in the past 30 days.

Assessments

The enumerators who assessed children on the ASQ:I and Bayley-III were trained separately. ASQ:I enumerators had completed a minimum of a bachelor's degree, and received 7 days of training on the tool. Bayley-III enumerators had a minimum of a master's degree and received 15 days of training. Training for both groups included interactive discussion, role play, and field testing in non-intervention sites followed by inter-observer and reliability testing, feedback, and refresher trainings. Participants were assessed in their own homes, and assessments were conducted in Bengali. During the assessments, inter-rater reliability was conducted for 4.7% of the ASQ:I sample used in this analysis ($n = 12$). If needed, feedback or correction was given immediately following these assessments.

Statistical analysis

Scores for each domain on both assessments were internally standardized using local-mean standardization by age in days to the control group sample that was included in this analysis. Observations with scores greater than 4 standard deviations from the control group mean were excluded, and remaining observations were re-standardized. Total ASQ:I and Bayley-III scores were created by summing raw scores across all domains before standardizing. As a measure of internal consistency, Cronbach's alpha was calculated with raw item scores for each domain of the ASQ:I and Bayley-III assessments. Items with no variability in the sample were excluded prior to calculation of Cronbach's alpha.

We calculated Pearson correlations for the ASQ:I and Bayley-III assessments by domain, both across the full sample, and stratified by intervention arm (any intervention vs. control) and by child age group (4–12, 12–18 and 18–26). For all correlations we constructed quantile-based confidence intervals with 1000 bootstrap samples clustered at the village level. We classified correlations as low ($r = 0.20-0.39$), medium ($r = 0.40-0.59$) or high ($r \geq 0.60$) [73]. Throughout we focused the results and interpretation on correlations between concordant domains across assessments (table 2.1). Concordant domains were designed to cover similar or the same underlying constructs and so should theoretically be the most correlated across tests. We also present correlations between all non-concordant domains between tests in the results tables to be consistent with prior work [25, 27]. As we did not administer the caregiver report domains as part of the Bayley-III assessment, there is no concordant domain for the ASQ:I Social-emotional domain. We present the correlation between this domain and Bayley-III domains but do not interpreted or highlight this result. We then computed the concurrent correlations between scores on each domain for each assessment and maternal education and stimulation activities in the home, variables known to be related to child development in other work. Analyses were performed in Stata v14, and R (V.4.0.1, Vienna Austria). Ethical approval was obtained from the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), and the University of California, Davis.

2.4 Results

Of the total sample of children in the 16 villages selected for the Bayley-III assessment, 300 ($n = 151$ control; $n = 149$ intervention) received the ASQ:I assessment. A total of 244 (81%) of these children were assessed on the Bayley-III ($n = 128$ control; $n = 116$ intervention). Children in the sample were on average 16.2 (SD 5.4) months of age, with 30% ($n = 73$), 35% ($n = 86$), and 35% ($n = 85$) in the 4–12, 13–18, and 19–26 month age groups respectively (table 2.2). Just under half (45%) of the children in the sample were female. All primary caregivers except 3 were the child's biological mother. Mother's education was on average 6.4 (SD 3.2) years. Demographic characteristics did not differ across the participants sampled in the control and intervention arms. The scores for the FCI subscales and the scores on the ASQ:I and Bayley-III were higher amongst children in the intervention arm (table 2.2).

Assessments occurred on average 13 days apart. We had a total of 3 participants who had a score that was an outlier on either assessment, and 21 participants who had missing data on one or more domains, resulting in a total of between 220 and 243 participants observed on each of the ASQ:I and Bayley-III domains. The Cronbach's alpha was between 0.78-0.81 for the ASQ:I domains, and between 0.90-0.95 for the BSID-III domains (table 2.3).

Concurrent validity

The concurrent validity for concordant domains ranged from 0.24 (fine motor) to 0.54 (gross motor), with an average correlation between total scores of 0.42 (table 2.4). Concurrent validity of the total score did not systematically differ by age, however there were suggestions of trends to increased correlation between scores by age for the fine motor and communication domains, and decreased correlation by age for the gross motor and cognitive domains (figure 2.1; table B.2). Concurrent validity did not differ systematically when comparing the intervention group to the control group (table 2.5). The two concordant comparisons with correlations that differed by more than 0.05 across intervention and control arms were the correlation between receptive language (Bayley-III) and communication (ASQ:I) which was higher for the intervention arm (0.44 vs. 0.32), and the correlation between gross motor scores on both measures, which was higher for the control arm (0.63 vs. 0.50).

Correlations with other variables

All domains of the Bayley-III and ASQ:I were positively correlated with maternal education, and with FCI play activities and play materials subscales (table 2.6). Correlations were highest, on average, between domains on both measures and FCI play materials (correlations ranged from 0.17 to 0.43) and FCI play materials (range from 0.08 to 0.37) compared with maternal education (range from 0.03 to 0.20). Most correlations between individual domains and each of these measures were statistically significant at $p < 0.05$.

2.5 Discussion

The Bangladeshi-adapted ASQ:I is a low-cost tool that can be feasibly administered in rural Bangladesh. We found moderate correlations between the adapted ASQ:I and Bayley-III assessments for the gross motor domain and total score, and low, but significant correlations between the cognitive/problem solving, language, and fine motor domains in a sample of children aged 4-27 months in rural Bangladesh. The lower correlation between the Bayley-III cognitive domain and the ASQ:I problem solving domain was expected as the ASQ:I Problem Solving domain only covers a subset of the cognitive domain captured in the Bayley-III. We did not find systematic differences across domains between concurrent validity of the ASQ:I across intervention status, or by age. We observed significant correlations with variables that have been previously shown to correlate with better child development outcomes including the variety of play activities that an adult has participated in with the child in the last 3 days, and the variety of toys that the child has played with in the last week for most domains of both the ASQ:I and Bayley-III. We also found acceptable internal consistency (Cronbach's alpha > 0.75 for all domains) for the ASQ:I in our sample.

The concurrent validity of ASQ-3 has been assessed by domain in two upper-middle-income country settings, one in urban Colombia and two in rural China (one smaller and one larger study in the Qinba mountain region) [25–27]. Our concurrent correlations with the Bayley-III were higher than those found in the studies from Colombia and China for children under 30 months for the majority of the comparisons. The study in Colombia did not recommend use of the ASQ-3 for children under 31 months of age, as they found the majority of correlations between concordant domains to be below 0.25 [25]. The continuous nature of the ASQ:I, and the inclusion of direct assessment may contribute to the stronger correlation with the direct assessment measure, when compared to that of the ASQ-3 in other settings. Differences in the populations included in each study may also contribute to the differences in correlations with direct assessment measures. For example, the caregivers in our sample had less education with a mean number of years of 6.4 (SD 3.2) compared to 10.3 (SD 3.4) in Colombia and 9.2 (SD 2.7) in the larger study in China (the smaller study in China did not report years of education) [25–27]. Though previous work in India found that correlations between another caregiver report measure and the Bayley-III did not to differ by caregiver education, the large difference in caregiver education may be indicative of other differences between the communities [74]. The concurrent correlation between the Chinese adapted ASQ:I and the Beijing Gesell Development Schedule was assessed in 53 children 11 and 12 months of age in an urban setting in the city of Kunshan, China. Correlations were between 0.74 and 0.89 for fine and gross motor, personal-social and problem solving/adaptive domains, and 0.29 for the communication/language domain. These correlations are higher than found in our work, but we note that their assessments were done by pediatricians and that some pediatricians preferred to observe the child before interviewing the caregiver. As such this comparison was not strictly between a primarily caregiver report measure and a direct assessment and the way in which it was administered (by a pediatrician, sometimes through observation) is

not comparable to our work. The researchers present the ASQ:I as a promising tool for a secondary screening measure for developmental delay, but do not discuss its use to evaluate intervention effects.

We did not find systematic differences in correlations between the intervention and control study arms. One disadvantage presented for conducting assessments that employ parent report as part of the evaluation of child development interventions is that there may be recall bias induced by the intervention [24]. This is to say that caregivers in the intervention arms may differentially report on their children's developmental status [24]. Two reasons for this have been presented in the literature. The first is that caregivers in the control arm may underestimate children's development status compared to those in the intervention arm because those who did not receive the intervention may be less attentive to their child's development status, and so may not notice the achievement of milestones that are caught by caregivers who received the intervention. The second is that caregivers in the intervention arm are taught about the importance of play and child development they may be more prone to overstating their child's developmental status as part of social desirability bias [24]. In both cases, the intervention effects would be upwardly biased. If this were true in our sample, we would expect the correlation between the Bayley-III and the ASQ:I to be lower in the intervention group. In the current study we do not observe such a bias, as the concurrent validity of the ASQ:I does not systematically differ for children in the intervention groups compared to the control. This apparent lack of bias indicates that differential reporting of child development across intervention arms does not seem to be an issue in this work, and bolsters the overall validity of using caregiver report assessments in the evaluation of early child development interventions.

Both the study in urban Colombia, and the larger study in rural China found that the concurrent validity increased by age, which we were not powered to detect in our sample. Since we did not see consistent increases in the magnitude of the correlations by age, however, we have some indication that there was no systematic increase in correlation by age. We saw increased correlations between concordant measures over age for fine motor and expressive communication domains, decreased correlations for the gross motor domain, and no consistent change for cognitive, receptive communication or total scores. In Colombia, the more pronounced differences by age may be due to the fact that the age groups were larger and included older children (6-18, 19-30 and 31-42 month age groups) [25]. The larger study in China used similar age groups to the ones in the current study (5-12, 13-18 and 19-24), and found very low correlations between concordant domains for the 5-12 month age group, $r = 0.07$ to 0.34, compared to our 0.19 to 0.62. This lower starting point may have contributed to the patterns of increased correlation over age [27]. The smaller study in rural China, which looked at correlations in 5-11, 12-17, and 18-23 month age groups, also did not found a pattern of increasing correlation between the ASQ-3 and the Bayley-III domains as children got older [26]. There is additional evidence from a upper-middle-income country with high levels of education (17.7 (SD 2.6) years) in Chile that an age gradient maybe

present only when the age range is extended [75]. They found that the concurrent validity of the ASQ-3 with the Bayley-III total score was 0.55, 0.56 and 0.75 for 8, 18, and 30 month old children, respectively [75]. The lack of change in correlation between total scores on both measures at 8 and 18 months is consistent with what we see in our work.

This study contributes to the literature on the measurement of child development in LMIC contexts and has multiple strengths. It provides more information on tools that can be feasibly used in the evaluation of large-scale interventions in low-resource contexts. It also contributes to the ASQ-specific literature with information on the performance of the ASQ:I which can be compared to the concurrent validity of the ASQ-3 in previous work. Further this study allowed for comparisons of concurrent correlations between ASQ:I and Bayley-III across intervention arms, allowing us to address a common concern with caregiver report assessments in the context of an intervention.

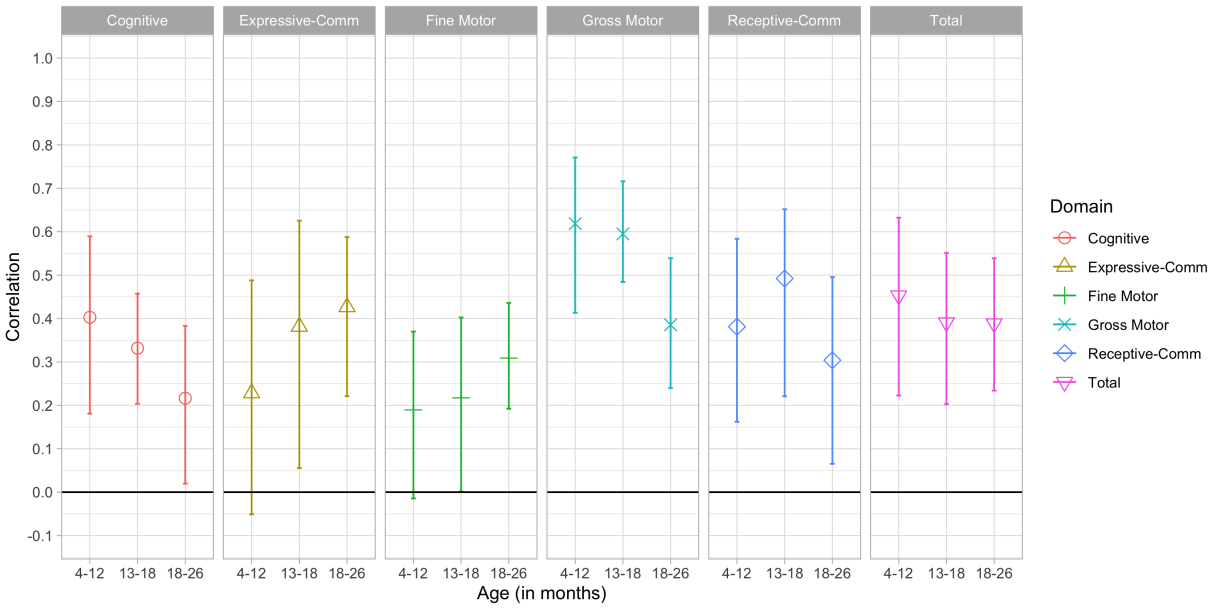
There are a few limitations of this work. The relatively small sample size in each age group means we have limited statistical power to detect differences in correlations by child age. Previous work has not used statistical inference when comparing concurrent validity in different groups [25–27], and, as such, we also interpret the magnitude of the differences as opposed to their statistical significance. We also did not assess children over 27 months, and so we are not able to determine the concurrent correlations between ASQ:I and Bayley-III for children across the full age range for which the tool was developed. Additionally, in this study we assume that direct assessment is more accurate in identifying underlying abilities in children. We acknowledge, however, that even with appropriate training of skilled enumerators, direct assessment may be affected by the child's current state (including hunger, shyness, and tiredness) and thus may underestimate the child's true ability, or be biased in assessment of intervention effects [24]. Thus, though direct assessment is considered more accurate and less biased in measurement of child development outcomes following interventions, it has limitations beyond the resources required. Finally, both assessment tools were originally developed in the United States and though both were culturally adapted and piloted in rural Bangladesh, they may be biased in identifying the underlying developmental status of children.

2.6 Conclusion

The ASQ:I has several benefits. It is low-cost, can be administered by enumerators who have completed a bachelor's degree and have received a 7-day training, and has the potential to capture intervention effects following a child development intervention. Research teams should apply the tool that best correlates with culturally appropriate direct assessment, given resource constraints. We recommend comparing culturally appropriate direct assessment tools to the ASQ:I and other options that are feasible in their setting. The validity of the ASQ:I in older age groups and across socio-economic gradients warrants future examination.

2.7 Figures and Tables

Figure 2.1: Correlation between Bayley-III and ASQ:I assessments by child age and domain



Pearson correlations presented only for concordant domains, with 95% confidence intervals calculated with cluster boostrap. Bayley-III: Bayley scales of infant development-III; ASQ:I: Ages and Stages Questionnaire: Inventory; Receptive-Comm is the correlation between Bayley-III receptive language and ASQ:I communication; Expressive-Comm is the correlation between Bayley-III expressive language and ASQ:I communication; maximum n for each age: $4-12 = 73$, $12-18 = 86$, $18-26 = 85$

Table 2.1: Characteristics of assessment tools

	Bayley Scales of Infant and toddler Development III (Bayley-III)	**Ages and Stages Questionnaire-Inventory (ASQ:I)**
Type of assessment	Direct assessment of ability	Parent report with 16% of items direct assessment
Cost	$1025 kit + $4.89 per child	Currently available through contacting publisher
Qualifications of trainer	Master's degree in psychology degree; A degree or license to practice in the healthcare or allied healthcare field; or formal training specific to assessing children. They can require 3 or more weeks of training	Easy to use by providers with varying levels of education and expertise, and they generally require 1-2 weeks of training
Duration	60-120 min	45-60 minutes
Domains assessed (concordant domains shown in the same row)	Cognitive	Problem solving
	Fine Motor	Fine Motor
	Gross Motor	Gross Motor
	Receptive Language	Communication
	Expressive Language	Communication
		Personal-Social

Table 2.2: Characteristics of the sample

	Analytic sample (n=244)	**Control (n=116)**	**Intervention (n=128)**
Caregiver Characteristics			
Age[1]	25 ± 5.8	25.0 (5.6)	25.7 (6.0)
Primary caregiver is biological mother[1]	99% (n=240)	98% (n=125)	99% (n=115)
Mother's education (years)	6.4 ± 3.2	6.1 (3.4)	6.7 (3.0)
Mother completed primary education	63% (n=154)	58% (n=80)	69% (n=74)
Father's education (years)[2]	4.6 ± 4.1	4.6 (3.9)	4.7 (4.2)
Father completed primary education[2]	34% (n=81)	34% (n=41)	35% (n=40)
Child Characteristics			
Age at Bayley-III assessment (months)	16.21 (5.4)	15.6 (5.2)	16.9 (5.4)
4-12 months	30% (n=73)	36% (n=42)	24% (n=31)
13-18 months	35% (n=86)	42% (n=49)	29% (n=37)
19-26 months	35% (n=85)	32% (n=37)	38% (n=48)
Female	45% (n=111)	44% (n=56)	47% (n=55)
FCI Play activities subscale (0-6)	3.7 ± 1.6	3.1 (1.5)	4.5 (1.4)
Household Characteristics			
Household size	5.5 ± 2.1	5.4 ± 1.9	5.7 ± 2.3
Has cement floor	18% (n=43)	18% (n=23)	17% (n=20)
Has brick walls	21% (n=50)	24% (n=31)	16% (n=19)
Has electricity	88% (n=215)	82% (n=23)	95% (n=110)
Bayley-III Raw Scores			
Cognitive[3]	46 (9.6)	44.8 (9.5)	47.8 (9.5)
Receptive Language	17.3 (5.4)	16.3 (5.2)	18.3 (5.4)
Expressive Language[1]	19.3 (6.3)	18.0 (6.0)	20.6 (6.4)
Fine Motor[3]	31.4 (6.0)	30.7 (6.2)	32.2 (5.7)
Gross Motor[4]	46.8 (9.7)	46.0 (9.7)	47.8 (9.6)
Total score[5]	161.4 (35.2)	156.1 (35.0)	167.0 (34.9)
Bayley-III Composite Scores			
Cognitive[3]	94.8 (11.7)	93.6 (13.3)	96.0 (9.7)
Language[1]	96.5 (14.4)	94.6 (16.6)	98.6 (11.3)
Motor[6]	100.0 (13.8)	99.3 (14.7)	100.7 (12.8)
ASQ:I Raw Scores			
Problem solving[7]	54.3 (15.9)	57.6 (15.4)	51.3 (15.8)
Communication[1]	55.7 (16.1)	52.9 (15.3)	58.8 (16.5)
Fine Motor[5]	49.8 (11.5)	48.3 (11.3)	51.5 (11.5)
Gross Motor[8]	58.1 (16.5)	55.9 (16.3)	60.5 (16.4)
Personal Social[9]	54.0 (16.6)	50.9 (15.9)	57.4 (16.8)
Total[10]	271.8 (72.7)	260.0 (70.3)	284.4 (73.4)

FCI: Family Care Indicators. Maximum scores for Bayley-III domains: cognitive=91; receptive language=49; expressive language=48; fine motor=66; gross motor=72; total=326. Maximum scores for ASQ:I domains: problem solving=68; communication=63; fine motor=63; gross motor=65; personal social=50; total=309. [1]$n = 243$; [2]$n = 236$; [3]$n = 241$; [4]$n = 242$ [5]$n = 237$; [6]$n = 240$; [7]$n = 239$; [8]$n = 238$; [9]$n = 234$; [10]$n = 225$.

Table 2.3: Internal Consistency by domain

Domain	**Cronbach's alpha**
Bayley-III	
Cognitive	0.93
Fine Motor	0.91
Gross Motor	0.95
Receptive Language	0.90
Expressive Language	0.91
ASQ:I	
Problem Solving	0.78
Fine Motor	0.79
Gross Motor	0.79
Communication	0.77
Personal Social	0.81

Table 2.4: Correlation between ASQ:I and Bayley-III domains in the full sample

	Bayley-III domain					
ASQ:I domain	**Cognitive (n=237)**	**Receptive Language (n=243)**	**Expressive Language (n=242)**	**Fine Motor (n=235)**	**Gross Motor (n=237)**	**Total (n=220)**
Problem Solving	**0.308 (0.23,0.38)**	0.287 (0.14,0.41)	0.265 (0.16,0.36)	0.286 (0.14,0.40)	0.247 (0.14,0.35)	0.346 (0.23,0.44)
Communication	0.220 (0.07,0.35)	**0.398 (0.26,0.51)**	**0.345 (0.20,0.47)**	0.219 (0.06,0.38)	0.242 (0.10,0.36)	0.327 (0.19,0.47)
Fine Motor	0.199 (0.09,0.28)	0.209 (0.09,0.31)	0.219 (0.12,0.30)	**0.241 (0.11,0.35)**	0.217 (0.03,0.35)	0.236 (0.08,0.37)
Gross Motor	0.186 (0.08,0.28)	0.206 (0.09,0.33)	0.146 (0.07,0.22)	0.165 (0.04,0.29)	**0.545 (0.45,0.63)**	0.388 (0.29,0.47)
Personal Social	0.172 (0.03,0.31)	0.260 (0.12,0.38)	0.188 (0.11,0.27)	0.233 (0.12,0.33)	0.134 (0.01,0.25)	0.245 (0.13,0.36)
Total	0.291 (0.17,0.39)	0.356 (0.24,0.45)	0.292 (0.17,0.41)	0.274 (0.11,0.42)	0.342 (0.19,0.48)	**0.423 (0.29,0.54)**

Estimates are Pearson correlations with 95% confidence intervals calculated with cluster bootstrap. Concordant domains across measures are represented in bold, n's represent the n used for each the concordant correlation bolded in each column

Table 2.5: Correlation between Bayley-III and ASQ:I assessments by intervention status and domain

ASQ:I domain	**Bayley-III domain**					
Control	**Cognitive**	**Receptive Language**	**Expressive Language**	**Fine Motor**	**Gross Motor**	**Total**
Problem Solving	**0.312 (0.19,0.41)**	0.108 (-0.08,0.27)	0.198 (0.00,0.33)	0.268 (0.07,0.41)	0.272 (0.12,0.40)	0.339 (0.19,0.48)
Communication	0.108 (-0.14,0.33)	**0.319 (0.07,0.51)**	**0.300 (0.08,0.46)**	0.232 (-0.01,0.43)	0.244 (0.14,0.35)	0.270 (0.06,0.44)
Fine Motor	0.190 (0.06,0.30)	0.070 (-0.07,0.20)	0.161 (0.02,0.30)	**0.209 (-0.01,0.38)**	0.246 (0.10,0.39)	0.211 (0.06,0.34)
Gross Motor	0.099 (-0.07,0.25)	0.161 (0.01,0.32)	0.115 (0.03,0.21)	0.171 (-0.04,0.35)	**0.630 (0.55,0.70)**	0.377 (0.29,0.47)
Personal Social	0.061 (-0.14,0.24)	0.161 (0.01,0.29)	0.137 (0.04,0.24)	0.134 (-0.01,0.26)	0.229 (0.09,0.41)	0.192 (0.06,0.37)
Total	0.212 (0.03,0.39)	0.230 (0.04,0.39)	0.196 (0.06,0.32)	0.241 (0.02,0.44)	0.461 (0.31,0.64)	**0.375 (0.24,0.56)**
Intervention						
Problem Solving	**0.281 (0.18,0.37)**	0.412 (0.29,0.53)	0.286 (0.14,0.44)	0.280 (0.10,0.46)	0.253 (0.07,0.41)	0.318 (0.13,0.47)
Communication	0.290 (0.18,0.41)	**0.444 (0.30,0.59)**	**0.352 (0.16,0.53)**	0.182 (0.01,0.39)	0.260 (0.02,0.46)	0.340 (0.12,0.55)
Fine Motor	0.197 (0.03,0.31)	0.301 (0.16,0.43)	0.248 (0.12,0.34)	**0.260 (0.12,0.39)**	0.207 (-0.09,0.39)	0.241 (-0.01,0.43)
Gross Motor	0.240 (0.11,0.38)	0.225 (0.03,0.41)	0.149 (0.02,0.26)	0.144 (-0.01,0.30)	**0.497 (0.33,0.62)**	0.384 (0.24,0.51)
Personal Social	0.218 (0.07,0.40)	0.291 (0.09,0.50)	0.171 (0.06,0.30)	0.281 (0.17,0.40)	0.098 (-0.09,0.28)	0.224 (0.06,0.38)
Total	0.332 (0.22,0.46)	0.427 (0.32,0.54)	0.328 (0.10,0.50)	0.277 (0.06,0.48)	0.278 (0.04,0.47)	**0.431 (0.22,0.60)**

Estimates are Pearson correlations with 95% confidence intervals calculated with cluster bootstrap. Concordant domains across measures are represented in bold, Maximum n per group: intervention=116; control=128. Bayley-III: Bayley scales of infant development-III; ASQ:I: Ages and Stages Questionnaire: Inventory

Table 2.6: Correlations between Bayley-III and ASQ:I and measures of the home environment and maternal education

	FCI: Play activities (0-6)	**FCI: Play materials (0-6)**	**Maternal education (years)**
Bayley-III domain			
Cognitive	0.195 (0.03,0.28)	0.227 (0.14,0.32)	0.084 (0.00,0.17)
Fine Motor	0.106 (-0.01,0.20)	0.170 (0.03,0.30)	0.029 (-0.11,0.15)
Gross Motor	0.076 (-0.02,0.20)	0.215 (0.10,0.34)	0.115 (-0.01,0.24)
Receptive Language	0.202 (0.06,0.32)	0.299 (0.20,0.40)	0.103 (0.01,0.21)
Expressive Language	0.175 (0.03,0.30)	0.265 (0.14,0.40)	0.148 (0.08,0.23)
Total	0.215 (0.07,0.32)	0.345 (0.23,0.17)	0.167 (0.07,0.25)
ASQ:I domain			
Problem Solving	0.305 (0.15,0.43)	0.304 (0.22,0.38)	0.111 (0.00,0.21)
Fine Motor	0.247 (0.12,0.36)	0.257 (0.17,0.33)	0.187 (0.05,0.29)
Gross Motor	0.175 (0.09,0.25)	0.282 (0.17,0.38)	0.097 (-0.03,0.21)
Communication	0.298 (0.19,0.40)	0.347 (0.25,0.44)	0.185 (0.07,0.28)
Social Emotional	0.340 (0.23,0.45)	0.315 (0.20,0.42)	0.172 (0.04,0.30)
Total	0.370 (0.24,0.48)	0.428 (0.34,0.50)	0.196 (0.07,0.32)

Estimates are Pearson correlations with 95% confidence intervals calculated with cluster bootstrap.
FCI: Family Care Indicators; Bayley-III: Bayley scales of infant development-III; ASQ:I: Ages and Stages Questionnaire: Inventory

Chapter 3

Effects of the COVID-19 pandemic on caregiver mental health and the child caregiving environment in a low-resource, rural context

3.1 Abstract

Young children are impacted by the COVID-19 pandemic through caregiving and home environments. This study estimates the effects of the COVID-19 pandemic and subsequent mitigation strategies on mental health, caregiving practices, and movement outside the home among female Bengali caregivers of children 6-27 months (50% female), in rural Bangladesh. The primary sample is a cohort (n = 517) assessed twice (May-June, 2019 and July-September, 2020). Caregivers who experienced more food insecurity and financial loss during the pandemic reported larger increases in depressive symptoms (0.26SD, 95%CI 0.08-0.44; 0.21SD, 0.04-0.40) compared to less affected caregivers. Stimulating caregiving and freedom of movement results were not consistently related to food insecurity and financial loss. These findings suggest that caregiver mental health was affected by the COVID-19 pandemic, and this may have consequences for child development.

3.2 Introduction

The COVID-19 pandemic threatens child development throughout the world as nutrition, routine healthcare, and the ability to provide nurturing care for young children are disrupted [32]. In many low- and middle-income countries (LMICs), COVID-19 lockdowns have resulted in difficulties accessing essential health care, reductions in child vaccination rates, and economic effects that are disproportionately felt by poor families [31].

Studies on the effects of the COVID-19 pandemic on young children in LMICs to date have been limited to primarily outcomes of mortality and morbidity. The disruptions in routine healthcare and access to food in LMICs induced by the COVID-19 pandemic have been estimated to result in at least 9.8% additional child deaths per month [76]. In addition to mortality, reduced income and food insecurity induced by the COVID-19 pandemic and resulting government-mandated mitigation measures including national and regional lockdowns, may lead to increases in nutrition-related morbidities including poor dietary intake, higher disease incidence with longer durations, lasting effects on child growth and development, and higher risk of compromised maternal health and intergenerational transfer of poor nutrition [77, 78].

Given that young children rely on their parents or other adults for care, they experience the pandemic and subsequent lockdowns through their parent's ability to provide nurturing and responsive care [32]. There are multiple risk factors for compromised child development that may be directly or indirectly affected by the COVID-19 pandemic and subsequent lockdowns. Exposure to nurturing care, including health, nutrition, security and safety, responsive caregiving, and opportunities for early learning, is especially important during the first few years of life [8]. Poor maternal mental health and lack of opportunities for child stimulation are both important risk factors for poor child development outcomes [6].

Maternal mental health, and the ability to provide stimulating and responsive care are adversely affected when families are facing economic hardship [6, 56, 79]. Parenting quality is a mediator between maternal mental health status and child outcomes [80, 81]. Poor maternal mental health can affect children through multiple pathways including altered maternal-child interactions, decreased early childhood attachment, and increased harsh punishment [56]. Mothers who are depressed are less responsive to their infant children, and less likely to be engaged in responsive stimulation [80, 82]. Further, the relationship between maternal mental health and stimulating parenting behaviors is cyclic, and responds dynamically to child temperament and behavior. Mothers who engage in responsive stimulation may be at a decreased risk of developing depressive symptoms, possibly due to improved mental health following rewarding experiences with the child [19].

Women's empowerment affects both maternal and child health and is comprised of multiple dimensions including resources, agency, and achievements [83, 84]. Freedom of movement is one indicator of women's empowerment that may have been impacted as a result of COVID-19 lockdowns and their sequalae [85, 86]. Freedom of movement means women are willing and able to travel to health centres, friends and relatives' houses, outside of the village, and to the market on their own. However, during the COVID pandemic more frequent trips outside of the home may also represent risky behaviors.

There have been notable gains in maternal and child health in Bangladesh in the 21st century, including large decreases in under five, infant, and maternal mortality as well as increases in

life expectancy at birth [87]. These gains have been partially attributed to the structure and reach of the country's health system [87]. However, Bangladesh still faces a high prevalence of poverty, income inequality, undernutrition, and stunting. Bangladesh is a country with a population of over 160 million, 23% of whom live in poverty [88, 89]. A substantial number of children in Bangladesh are stunted (31%), underweight (22%), or wasted (8%), and 40% of women between age 15-40 experience iron-deficiency anemia [90, 91]. Despite exposure to multiple risk factors for poor child development, 72% of 3- and 4- year old rural children are developmentally on track in the social-emotional domain as assessed by the early child development index [90]. In terms of literacy and numeracy, however, only 29% of 3- and 4-year old children are developmentally on track [90].

The United Nations Convention for Rights of the Child is a human rights treaty that protects and promotes the civil, political, economic, social, health, and cultural rights of children [11]. Bangladesh signed onto the convention in 1990, and is joined by the vast majority of the world's nations as signatories. The convention states "States Parties shall ensure to the maximum extent possible the survival and development of the child." [11]. Further, the Sustainable Development Goals explicitly include quality child development as a focus in goal 4.2, which is "By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education" [92]. The current work is guided by these principles, and we focus on identifying children who may be at increased risk for altered developmental trajectories in the context of the COVID-19 pandemic. Though not all children experiencing risk factors have compromised development, the duration, co-occurrence, and magnitude of these risks contributes to increased probability of delay [93].

In response to the pandemic, the Government of Bangladesh shut down schools, business, and other institutions between March and May 2020. Some businesses and other institutions re-opened, but schools and educational institutions continued to be closed until May 23rd, 2021 with a re-opening planned for thereafter. Recent work demonstrated that in May and June 2020 the median income of families in a rural area just outside the capital city of Dhaka fell to just over 25% of what it was one year previously, severe food insecurity increased from 6% to 36%, and maternal depressive symptoms increased [94]. Intimate partner violence is a risk factor for maternal depression, and in May and June over half of women who reported experiencing intimate partner violence reported an increase since the shut-down for COVID-19 began [94, 95].

The present study

The primary aim of this research was to assess the effects of the COVID-19 pandemic and related lockdowns on risk factors for poor early child development including maternal mental health and caregiver stimulation in the home in a large sample of families in rural Bangladesh. Understanding the effects of the COVID-19 pandemic and related lockdowns on risk factors

for caregivers will contribute to the understanding of the impact of COVID-19 on child development. A secondary aim was to explore the relationship between the pandemic and caregiver freedom of movement.

3.3 Methods

Sample recruitment and selection

All data come from assessments done as part of the Research on Integration of Nutrition Early Childhood Development and WASH intervention delivered through the Government health system in Bangladesh (RINEW-G) study, which was set in Chatmohar. Chatmohar sub-district is located in Pabna district of Central Bangladesh, consists of 11 rural unions and one urban municipality, and had a population of 291,121 recorded at the most recent population census in 2011 [96]. The primary occupation of employed residents is in agriculture, other less common occupations are small business owners, or salaried government, private business or non-governmental organization workers. In the most recent census less than half of the population over 7 years old in Chatmohar was literate (46%), and 83% of the population lived in houses with no permanent structure ('katcha') [96]. The majority of the population of Chatmohar is Muslim (96%), and 96% of the households own the dwelling they live in [97].

The study team collected data at two time points: a baseline ("pre-COVID"; May 18 to June 22, 2019), and a follow-up, four months after the first case of COVID was detected in Bangladesh ("mid-COVID"; July 11 to September 16, 2020). The in-person pre-COVID assessments consisted of a sample of 1635 primary caregivers of children 6-24 months old living in 109 rural villages of Chatmohar, selected through multi-stage sampling (figure 3.1 and Appendix C). Exclusion criteria were caregivers who were not planning on living in Chatmohar for the following year and caregivers or children with physical or cognitive disabilities. At the mid-COVID assessment, the study team contacted 754 caregivers from the original sample who had children under 24 months of age in July 2020 (figure 3.1). An additional cross-sectional sample was also recruited in order to gather information during the mid-COVID timepoint across the original age range of 6-24 months. The majority of this cross-sectional sample were caregivers of children between 6 to 18 months of age, and lived in the same villages sampled during the pre-COVID assessment (figures C.1 and C.2). This mid-COVID cross-sectional sample was used in the present study in a supplementary analysis examining correlations between exposures and outcomes across the larger sample and age range.

Before data collection, all participants provided written consent during the pre-COVID in-person visit, and verbal consent during the mid-COVID phone survey.

Measures

Pre-COVID assessments were conducted in person in participant's homes. Mid-COVID assessments were conducted over the phone. Measures collected at both time points are described below.

Training for data collection lasted 15 days for the pre-COVID assessment, and 8 days for the mid-COVID assessment. Training included instruction, role-play practice with peers, and practice with sample participants. During both assessment time points supervisors conducted ongoing quality control for 5% of the sample, with supervisors observing assessments (either watching in person or listening on the phone) to provide feedback following the assessment if needed.

COVID-19 experiences

The questionnaire administered during the mid-COVID assessment included modules to assess experiences during the COVID pandemic (Table C.1). Questions included changes in the number of household members, interactions with the target child, changes made in the household due to COVID-19, indicators of food insecurity, changes in household economic status and coping mechanisms, and sickness in the household. These questions were developed through an iterative process including qualitative interviews and pre-testing, and used some questions from a survey previously done in Bangladesh (survey development described in Appendix C) [98].

Using the survey responses, we characterized the effects of COVID-19 across three domains: food security, economic status, and health. Participants were split into categories of more or less affected in each of these three domains. For the food security domain participants were categorized as more affected if they answered that they were negatively affected in response to two or more of the following questions: 1) Was your household able to buy essential food items over the past 7 days; 2) in the last 7 days are you consuming less food when compared to the same time last year, and is that reduction due to COVID-19 pandemic; 3) in order to cover household's basic needs participants have had to reduce the number or size of meals for some household members, or 4) rely on less preferred and less expensive foods. For the economic impact domain, participants were characterized as more affected if both 1) since April 2020 (the start of the COVID-19 lockdown in Bangladesh), the main earning member of the household lost their job/income source, and 2) the status of the household's current income was reported as none or reduced compared to April 2020. For the health domain, households were categorized into ones where either the respondent or any household member had been sick with symptoms of fever, cough, cold, loss of taste/smell, shortness of breath or difficulty breathing, or COVID-19 between April 2020 and the time of the mid-COVID assessment.

Outcomes

Outcomes were measured at both the pre-COVID and mid-COVID time points. Maternal depressive symptoms were assessed using the Center for Epidemiologic Studies Depression Scale (CES-D), which consists of 20 questions asking about the number of days each participant experienced depressive symptoms in the past week. Each question was scored on a Likert scale, including 0 (0 days), 1 (1–2 days), 2 (3–4 days) and 3 (5–7 days), and individual question scores were summed to a total range of 0-60, with higher scores representing more depressive symptoms [99]. The scale reliability of the CES-D score in the sample was relatively high (alpha=0.83 at the pre-COVID assessment and 0.84 at the mid-COVID assessment for the cohort sample).

Child stimulation in the home was assessed with the Family Care Indicators (FCI), which has been used previously in Bangladesh [47], and consist of two subscales, stimulating caregiving practices and the variety of play materials available in the home. The stimulating caregiving practices subscale asks if the child's mother, father, or any adult (> 15 years old) has engaged with the child in six stimulating activities in the previous three days. We conducted the primary analysis of this outcome with maternal reported participation in stimulating caregiving practices (range 0–6) as the mother was the respondent and we hypothesized that this would be the most accurate report. The variety of play materials subscale consists of questions about the variety of play materials that the child has played with in the past 30 days (range 0–6). The correlation between the play activities and play materials subscales of the FCI was 0.30 during the baseline assessment and 0.32 at the mid-COVID assessment.

Freedom of movement for the primary caregiver was measured with questions about experiences going to the market, a medical facility, outside of the village, and to a paternal home or friends or relatives' home in the last six months. Each participant was given a freedom of movement score, which included one point if the respondent had attended each place in the last 6 months, and an additional point if the respondent had been to this place alone in the past 6 months, with a with a total range of 0–8. These questions were adapted from a scale used previously in Bangladesh [100]. The scale reliability of freedom of movement scores were 0.61 and 0.64 during the pre- and mid-COVID assessments respectively.

Covariates

Information on the family's socioeconomic status was collected at both time points with questions about monthly income (separated into tertiles), household assets, household size (categorized into 2-3, 4, 5, 6, 7+ household members), maternal education (categorized into none, less than primary school, completed school, and completed secondary school), and housing materials (indicators were made for if the household had a concrete floor or brick walls). Caregivers were asked if they get to spend their own money independently, and the responses were categorized into a binary response where responses of "no" and "no independent money" were combined. Health care access and health seeking behavior

was measured through the number of antenatal care visits that each participant reported attending prior to their most recent pregnancy, and was categorized into above or below 4, the recommended number of visits.

At the pre-COVID assessment we also measured some indicators of nutrition and WASH status which were used in the comparison of pre-COVID characteristics for the difference-in-differences samples. Maternal and child dietary diversity was assessed by asking about all foods consumed in the previous 24 hours. Caregivers who consumed at least 5 of 10 food groups were considered having met minimum dietary diversity. Children's dietary diversity was assessed with 8 food groups, including breastmilk, and were considered having met the minimum dietary diversity threshold if they had consumed 5 or more of the 8 groups [49, 50]. The status of WASH facilities in the household was assessed through observation, and the presence of a handwashing station with water and soap or a soapy water bottle within 20 meters of the cooking facility, and of a clean, functional, hygienic latrine in the household was assessed.

Statistical analysis

Descriptive statistics included summaries of the distribution and histograms for continuous variables, and tabulations of proportions for categorical variables. Our primary analysis to examine the effects of the COVID-19 pandemic and subsequent lockdowns on risk factors for poor child development was a difference-in-differences analysis with the longitudinal cohort assessed at both the pre-COVID and mid-COVID time points. Our supplementary analysis consisted of a cross-sectional comparison for all participants in the mid-COVID sample. For both analytic strategies we examined four outcomes of interest: our primary outcomes of maternal depressive symptoms, stimulating caregiving practices, and varieties of play materials, and a secondary outcome of caregiver freedom of movement.

For the difference-in-differences analysis we compared changes in outcomes over time for those the cohort sample who were more vs. less affected on the food security, economic, and health domains at the mid-COVID assessment. We use generalized estimating equations (GEEs) controlling for an indicator of village membership, and accounting for repeated observations within the cohort sample.

For the supplementary analysis on the full cross-sectional mid-COVID assessment we conducted an adjusted cross-sectional comparison of outcomes at the mid-COVID assessment, comparing families that reported being more vs. less affected on the food security, economic, and health domains at the mid-COVID assessment. We estimate differences in outcomes between groups using a generalized linear model with standard errors adjusted for clustering by village.

In both analyses, we controlled for theoretical confounders of child age, maternal education,

income, antenatal care, control over assets, household size, child sex, and housing materials. For the difference-in-differences analyses we controlled for baseline values of these variables, and for the cross-sectional analyses we controlled for concurrent values as the sample included some cross-sectional participants who do not have baseline measurements. In all analyses we additionally control for attendance at intervention sessions. Further, in the difference-in-difference analyses with maternal depressive symptoms or home stimulation outcomes we control for baseline caregiver freedom of movement.

For the maternal depressive symptoms outcome, we expected the distribution of scores to be non-normal, with a right skew. We conducted sensitivity analyses with median regression in addition to our primary analysis. As GEE models allowed us to account for repeated measures we prefer the GEE results and present these as our main findings.

In order to estimate the effects on child development, we stratified the population into more and less depressed using a cutoff value of 16 for the CES-D score. In the U.S. a CES-D score cutoff of 16 has been validated to indicated higher risk of clinical depression [101]. As this cutoff has not been validated in Bangladesh, we adopt the continuous CES-D score in our primary analyses. We took the cutoff value as a proxy for more depressive symptoms to compare our results to the literature looking at the relationship between caregiver depression and child development. We conducted a difference-in-differences analysis with a binary outcome indicating CES-D score 16 and above or 15 and under. We estimate this model with a GEE with an identity link to estimate risk differences, accounting for repeated observations by village. Using our results from this analyses and meta-analyzed estimates of the relationship between maternal depressive symptoms and child development outcomes, we estimated the burden of increased risk for poor child development in rural Bangladesh. Statistical analyses were performed in Stata 14 (College Station, TX) and R (Version 4.0.2, Vienna, Austria).

Given the longitudinal study design, and multiple analyses undertaken for robustness, we viewed the analyses on maternal depressive symptoms, stimulating activities with the child, and variety of stimulating play materials work as confirmatory. Due to the use of only a single measure to draw inference on a larger construct, we considered the analyses that include the outcome of caregiver freedom of movement to be exploratory.

3.4 Results

Participants

The study team collected data on 1635 caregivers of children 6–24 months of age during the pre-COVID assessment, and 754 of these children were under 24 months in July 2020, and followed-up (figure 3.1). A total of 523 households from the pre-COVID sample were reached for complete mid-COVID data collection (69% of those attempted). Six participants were

excluded from the analytic sample, three because their child was over 27 months of age at the time of assessment, and three for having a different caregiver assessed at the second time point. In all, 517 caregivers comprise the cohort sample (figure 3.1). The cohort sample that was followed up was not different from those who were attempted to be followed-up and were not reached across multiple demographic factors (table C.2). An additional cross-sectional sample of 1,176 participants was collected during the mid-COVID assessment, 54% of those contacted, and are used in a supplemental analysis (figure C.1).

Most caregivers in the cohort sample had completed primary education, the mean child age was 8.55 months (SD 1.8), and approximately one fifth of the sample had a cement floor, brick walls, and a refrigerator (table 3.1).

COVID-19-related responses during the mid-COVID assessment

When asked what changes they had made due to the COVID-19 pandemic (no prompt given), most respondents reported that they were washing their hands more (66%, $n = 334$), and cleaning the household and outdoor space more (61%, $n = 316$) (table 3.2). Some said they were interacting less with people outside the household (19%, $n = 98$), had restricted their movement (16%, $n = 83$), or were wearing a mask when going outside the house (15%, $n = 80$) (table 3.2). About a fifth (16%, $n = 81$) of participants said they had made no changes (table 3.2). Most participants reported that they interacted with their child (talking, playing, or spending time) the same amount compared to before the COVID-19 lockdown (53%, $n = 274$), while 41% ($n = 210$) said they interacted more, and 6% ($n = 33$) said less (table 3.2). Of those who said they interacted more, the most common reason (77%, $n = 162$) was spending more time in the household due to restricted movement outside the house (data not shown). Of those that said they interacted less, the majority (48%, $n = 16$) said this was because they had more household chores (data not shown). Just under a quarter of participants (22%, $n = 114$) had the number of household members change since April, with most of these participants having their household size increased with an adult or child moving in (87%, $n = 99$) (table 3.2). Most of these moves were temporary with 17% of the families who reported having additional members staying still staying at the time of the survey (data not shown). Patterns of COVID-related responses in the full mid-COVID sample were similar to those in the cohort (data not shown).

Ten percent of families ($n = 55$) reported that the family had earned no income since April, and an additional 71% ($n = 366$) reported a reduced income compared to prior to the start of the pandemic (table 3.2). Over a third (42% $n = 219$) of participants reported that since the COVID-19 shutdown there was a change of employment where the main earning member of the household lost their job or income source (table 3.2). A total of 40% ($n = 208$) of families experience both a loss of job for the main earning member in the household, and a reduction in income (table 3.2). In the seven days prior to the assessment just less than one third of families (29%, $n = 152$) reduced the number or size of their meals to cover basic household

needs, over a third (39%, $n = 202$) relied on less preferred and less expensive foods, and almost a quarter (23%, $n = 23$) were not able to buy essential food items (table 3.2). Almost half (44%, $n = 230$) of participants reported consuming less food when compared to the same time last year, and attributed the change to the COVID pandemic (table 3.2). Overall, 41% ($n = 214$) of participants reported two or more of these indicators of food insecurity, and 67% of those who experienced more food insecurity also experienced a loss of job and a reduced income (table 3.2). When participants were asked if they or household members had been sick with symptoms of fever, cough, cold, loss of taste or smell, shortness of breath or difficult breathing, or COVID-19, 15% ($n = 77$) or participants reported being sick, and 27% ($n = 138$) reported that their family members had been sick (table 3.2). Overall 35% ($n = 181$) of participants reported either they or their household members had been sick with these symptoms since April 2020 (table 3.2).

When comparing the pre-COVID and mid-COVID time points in the cohort sample, on average, after adjusting for covariates there was no difference in caregiver depressive symptoms between the two time points (-0.55, 95% CI: -1.30 to 0.21) (table 3.3). Play activities and play materials scores were both increased over time (mean differences for play activities: 1.25 (1.08 to 1.42); play materials 1.66 (1.52 to 1.79)), and the caregiver freedom of movement score decreased over time (-0.41, -0.58 to -0.25)) (table 3.3).

Difference-in-differences analysis

We compared changes over time in risk factors for poor child development between those in the cohort who reported being more and less affected across the domains of food security, economic status, and health during the mid-COVID assessment through difference-in-differences analyses. The risk factors we looked at were caregiver mental health, stimulating play activities and play materials, and caregiver freedom of movement. There were some differences in observed baseline characteristics between the more and less exposed groups on each of the three domains, with the most differences in pre-COVID characteristics for the comparison between the caregivers more and less food insecure at the mid-COVID assessment (table 3.4). We adjusted for differences at the caregiver, child, and household levels in the difference-in-differences analyses.

A kernel density plot of unadjusted CES-D scores at the pre-COVID and mid-COVID time points stratified by experiences of food security during the mid-COVID assessment displays differences in the distributions for each group over time, illustrating a shift in the distribution towards higher scores in the more food insecure group compared to the less food insecure group at follow-up (figure 3.2).

Depressive symptoms increased more between the pre- and mid-COVID timepoints for those experiencing more food insecurity or a loss of income and job for the primary earner of the household during the mid-COVID assessment, after adjustment for covariates at the child,

caregiver, and household level (average increases in CES-D scores of 2.29, 95% CI 0.72 to 3.86 and 1.93, 0.31 to 3.35 respectively) (table 3.5, figure 3.3).

Caregiver depressive symptoms also increased more between timepoints when a household member was sick with COVID-19 like symptoms between April and July–Sept 2010 (average increase 1.19, 95% CI −0.47 to 2.84), though this difference was not significant at the $p< 0.05$ significance threshold. The greater the number of domains in which a caregiver was affected related to larger differences in CES-D scores over time (table C.3).

For families who experienced a reduction of income and a loss of job for the primary earner in the household, there was an increase in reported play activities in the home between the pre- and mid-COVID timepoints compared to families who did not experience this financial effect (0.34, 95% CI 0.01 to 0.68) (table 3.5, figure 3.3). Having a household member sick was associated with a larger variety of play materials (0.44, 0.16 to 0.73) and more freedom of movement (0.41, 0.12 to 0.71) at follow-up (table 3.5, figure 3.3). There were no other differences in outcomes between groups.

We found that the prevalence of CES-D scores over 16 was higher among caregivers who experienced more impacts on food security, economic status, and health in the household at follow-up (prevalence differences from difference-in-difference estimates: 0.06, −0.05 to 0.17; 0.06, −0.05 to 0.16; and 0.08, −0.03 to 0.19, respectively) (results not shown). For the sensitivity analyses with caregiver depressive symptoms as an outcome re-run with median regression all estimates for were within 0.50 from the GEE estimates, and inference did not change for any analyses (table C.4). Inference also did not change for sensitivity analyses with caregiver freedom of movement as an outcome that were re-run with a freedom of movement score that did not include questions about visiting a health facility (0–6) (table C.5). A comparison of individuals in the full cross-sectional sample at the mid-COVID timepoint found similar results across most contrasts (table C.6)

3.5 Discussion

In this paper, we report the widespread experiences of food insecurity and loss of employment and income among low-income families in rural Bangladesh during the COVID-19 pandemic, and the associated increased depressive symptoms amongst caregivers of young children. Specifically, primary caregivers had more depressive symptoms if they lived in families that experienced more food insecurity, both job loss of the main earner of the household and a loss of household income, or had sickness in the family after the country-wide lockdown due to the COVID-19 pandemic, compared to those who did not have these experiences. Our findings relating to stimulating caregiving practices and freedom of movement were inconsistent, with more stimulating activities in the participants more affected by job loss and loss of income, more play materials and freedom of movement in the households that had at least one household member sick, and otherwise null results.

A major strength of this study is the longitudinal sample, with baseline data collected before the COVID-19 pandemic. Another strength is that the information about household-level experiences during the pandemic is available across multiple domains, and this information is used to stratify the sample and examine differences between subgroups with a difference-in-differences design. This approach is unbiased if the trends in outcomes over time due to unobserved factors are the same in each set of comparison groups.

Despite its strengths, the current study does have limitations. The greatest threat to validity is potential bias from loss-to-follow-up in the cohort sample. We are optimistic about the validity of the sample, however, because the retained and lost samples are similar across all observed characteristics at the pre-COVID assessment. In addition, we were not able to test the "parallel trends" assumptions for the differences-in-differences analysis as we do not have data from additional prior time points to examine trends in outcomes prior to the pre-COVID assessment. We address this concern by controlling for covariates that are theoretically related to the outcomes and differed during the pre-COVID time point across groups. Further, as our follow-up survey was done over the phone, we were not able to conduct any observations of the quality of caregiver-child interactions or other observations of the home environment.

Compared to other recent work, we found a relatively lower proportion of our sample experienced a loss of income for the family, with 81% of the current sample reporting any loss of household income and 10% reporting a complete loss of income. A previous study in Bangladesh reported that 96% of mothers had a reduction in paid work for the family and 39% experienced a complete loss of income [47]. This may have been because the other study was conducted immediately after the lockdown when all income sources (factories, industries, offices, and transportation of agricultural products from rural to urban areas) were suddenly stopped. Other possible explanations for these differences are that the population in the earlier sample was more universally affected by the pandemic due to their proximity to the capital city Dhaka, and the effects on employment and income diminished over time, with our results from experiences in July-September, compared to earlier work done in May-June. Our results on changes in food habits to cope with financial losses are consistent with recent work from Bangladesh which found that in a cross-sectional sample ($n = 106$) in Matlab, Bangladesh in April and May 2020 44% of the rural population sampled had consumed less food or changed their food habit to cope with financial losses due to the pandemic, compared to 48% in the current sample [102].

Between the pre-COVID and mid-COVID time points, depressive symptoms increased more in the groups more affected on the food security, economic, and health domains during the mid-COVID time point. The increases in depressive symptoms were largest when comparing caregivers who experienced more food insecurity to those who experienced less food insecurity. In a previous study set in Bangladesh, maternal mental health was assessed prior to the pandemic (between 2017 and February 2020) and in May-June 2020 and depressive

symptoms scores were worse overall during the latter time point [94]. Depressive symptoms were assessed with a modified 7-question CES-D scored using total number of days per week of symptoms reported as opposed to a Likert scale, and as such, their estimates cannot be compared directly to ours [94]. In our analyses, we do not find increases in depressive symptoms overall, only within subgroups of caregivers who experienced more food insecurity and financial loss during the pandemic. A meta-analysis including studies published before July 5, 2020 found 3 studies that compared maternal mental health before and during the pandemic for post-partum women, and found a pooled effect size of 0.40 (-0.05 to 0.96) for depressive symptom scores [103]. None of these studies were from LMICs, with populations from Italy, China, and Canada. When our estimate for the more food insecure group is converted to a measure of standard deviations of the population pre-COVID, we find that the difference-in-difference estimate is 0.26SD (0.08 to 0.44). Previous work has linked poverty and food insecurity to poorer mental health, this work adds to the literature on poverty and mental health as well as the literature on the magnitude and duration of the effects of the pandemic and subsequent lockdowns on caregiver mental health, and suggests that examining effects in subgroups may be critical to understanding the burden of poor mental health [104, 105]. Participants in the less affected groups, on average, experienced improvement in depressive symptom scores over time. These improvements may have been partially a result of children getting older over time, and risk of depression in the postpartum time period being higher than risk of depression when children are older [106].

We found that caregivers who had less than a primary school education, lived in a home without a permanent wall or concrete floor, were in the lowest income tertile, or had fewer assets had a higher probability of being more food insecure at the mid-COVID time point. It may be that families who were better off financially were better able to use their resources to adapt to the sudden economic shutdown. Additionally, ownership of agricultural land and education of household members have been found to be associated with reduced food insecurity in Bangladesh and may have played a role to mitigate experiences of food insecurity and economic status for families during the COVID pandemic [107]. Further, the availability and use of social and community support systems may have contributed to both financial and social support for caregivers and the heterogeneity in experiences of food security and financial impacts as well as mental health. Individual social capital has been linked to improved mental health outcomes, and in Indonesia, individual's trust in their community was found to be positively related to mental health independent of poverty [105, 108].

We found that caregivers who had the primary earning member of their family lose a job and had reduced income due to the pandemic participated in a larger variety of play activities with their child mid-COVID when comparing those to those who had not. We speculate that job loss may have resulted in having more family members at home to help with household chores and thus may allow the primary caregiver to spend more time interacting with their young child. This finding was contrary to expectations relating to effects of maternal depressive symptoms on decreased stimulation in the home [56]. We do not find any corresponding

negative associations between compromised food security, economic, or health status and the quantity of play activities that the caregiver participated in, or the variety of play materials in the home. The lack of correspondence between the results on maternal mental health and those on play activities may be both a result of both the increased number of family members at home during the pandemic and that our assessments of stimulation in the home measured the quantity, but not the quality of stimulation. Though we did not find decreases in the variety of stimulating play activities, these interactions may have been more difficult for the caregiver to engage with due to their increased depressive symptoms, and may have been reduced in quality.

In families where one or more household member was sick with COVID-like symptoms during the mid-COVID assessment, there was a larger variety of play materials at home, and caregivers in these households had increased freedom of movement compared to households where no members were sick during this period. The latter relationship held even when travel to health care facilities was excluded from the freedom of movement score. Thus, it may be that the caregiver was required to leave the house for reasons related to the sickness or for tasks that would routinely fall on the sick household members, or that the symptoms were related to the caregiver having to participate in activities outside of the home. In the context of the COVID pandemic the freedom of movement measure may not be a robust indicator of female autonomy.

The timing and duration of caregiver depressive symptoms matters for both caregiver and child well-being, and the increase in depressive symptoms may have implications both for the child who was part of this study, as well as future children of the caregiver [56, 109]. Further, simultaneous effects on both economic status and caregiver mental health as are seen in this sample may be especially detrimental as they perpetuate the cycle of poverty and poor mental health [79]. Financial assistance or other means of economic empowerment are needed to ensure families are able to meet their basic needs when their regular income has been compromised by the COVID-19 pandemic. Only 3% of the sample relied on government or NGO assistance to cover basic household needs, and this was 6% amongst those who reported no income (3% amongst those who reported some income). Taken together, these findings suggest that either there is very little assistance offered in Chatmohar, or it is difficult to obtain for rural families. The unavailability of financial support to families may results in long-term effects on caregivers, children, and future generations.

We estimated the magnitude of effect that the COVID-19 pandemic and subsequent lockdowns may have on population-level child development outcomes through increased depressive symptoms among caregivers experiencing more food insecurity during the COVID pandemic. A meta-analysis on the relationship between maternal depressive symptoms and child cognitive development found that children under 56 months whose mother's had postnatal depressive symptoms had -0.27 standard deviations (95% CI -0.43 to -0.11) lower cognitive scores [110]. There are approximately 8.8 million children under 5 living in rural areas

in Bangladesh [111, 112]. If the proportion of children affected by increased food insecurity in the rural population of children under 5 in Bangladesh is the same as in our cohort sample (41.4%), an increase of caregiver depression of 6 percentage points in this group would represent over 218,000 children who are at increased risk of poor development. The effect size of -0.27 standard deviations on these 218,000 children would come close to reversing the effects of stimulation interventions to improve child development outcomes ($d = 0.42$, 0.36 to 0.48) [13, 110]. This has the potential to set children back to a more disadvantaged starting point for future interventions, and have long-term implications for their well-being. As such, financial assistance or other means of economic empowerment could be pivotal to reducing the population impacts on child development and caregiver mental health.

Future work should address the critically important range of other outcomes in addition to depressive symptoms that comprise mental health. Anxiety was assessed cross-sectionally during May and June in Bangladesh, and it was found that 14% of participants reported moderate or severe anxiety, and when asked about if their symptoms had changed since the lockdown began, 99% stated that they experienced increased anxiety [94].In addition, the meta-analysis by Hessami et al. found three studies that looked at anxiety scores pre- vs. during the pandemic and found an increased in anxiety scores post with a mean difference of 0.82 (0.49 to 1.16) [103]. Therefore, there are some mental health effects that we did not capture, and the full impact on caregiver mental health is likely to be larger than what was captured in this study. Further, future work on the concordance between the quantity and quality of child stimulation for caregivers with increased depressive symptoms would be helpful to further investigate our results for reported variety of stimulating activities and materials. Finally, follow-up work to understand the duration of the adverse conditions experienced during the COVID pandemic will contribute to the understanding of the magnitude and duration of impact on children's development. Finally, in addition to financial assistance, interventions that focus on caregiver mental health and the quality of responsive and nurturing care during times of crisis may be helpful in supporting children and caregivers and mitigate the negative impacts of the COVID-19 pandemic in rural Bangladesh.

3.6 Conclusion

We add to the existing literature the effects of the COVID-19 pandemic and subsequent lockdowns on children 6-27 months through a more nuanced understanding of the effects of the pandemic on caregiver depressive symptoms, and estimate the potential impact this may have on child development outcomes. Future work on the concordance between the quantity and quality of child stimulation for caregivers with increased depressive symptoms would be helpful to further investigate our results for reported variety of stimulating activities and materials. Our research underscores the urgency of financial and mental health interventions in rural Bangladesh to mitigate the long-term effects of the COVID-19 pandemic on caregivers and children, promote positive developmental trajectories, and improve later life outcomes.

3.7 Figures and Tables

Figure 3.1: Study sample

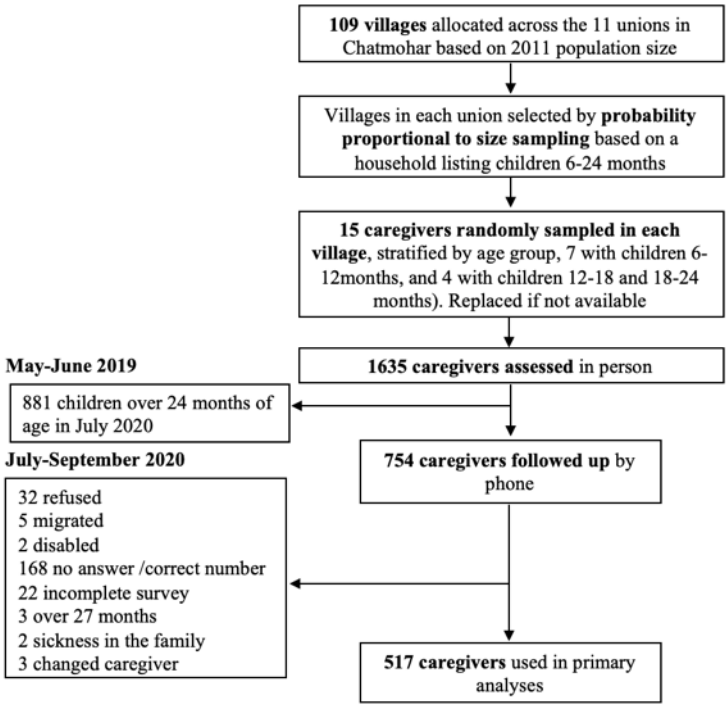
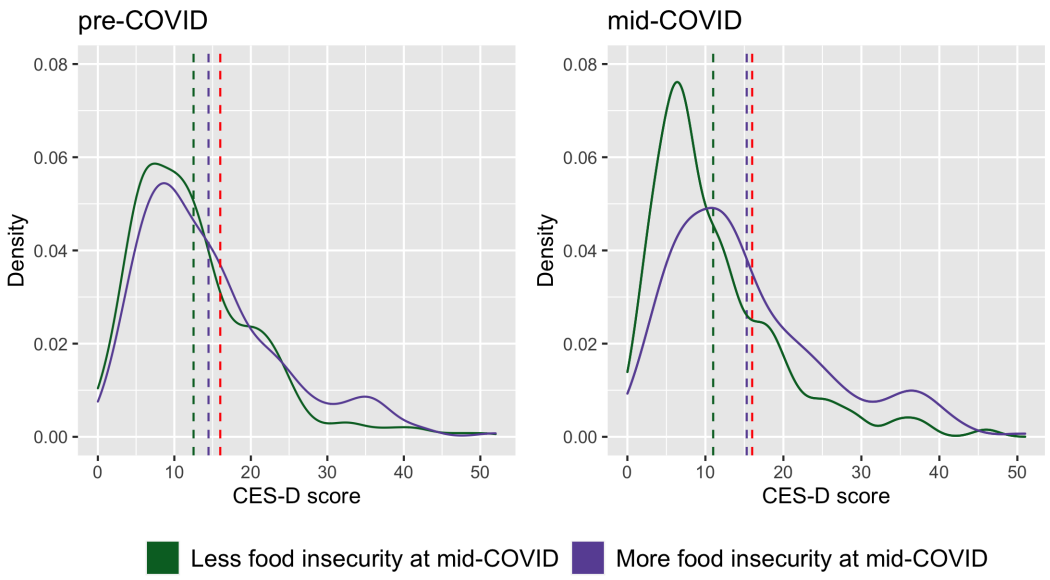
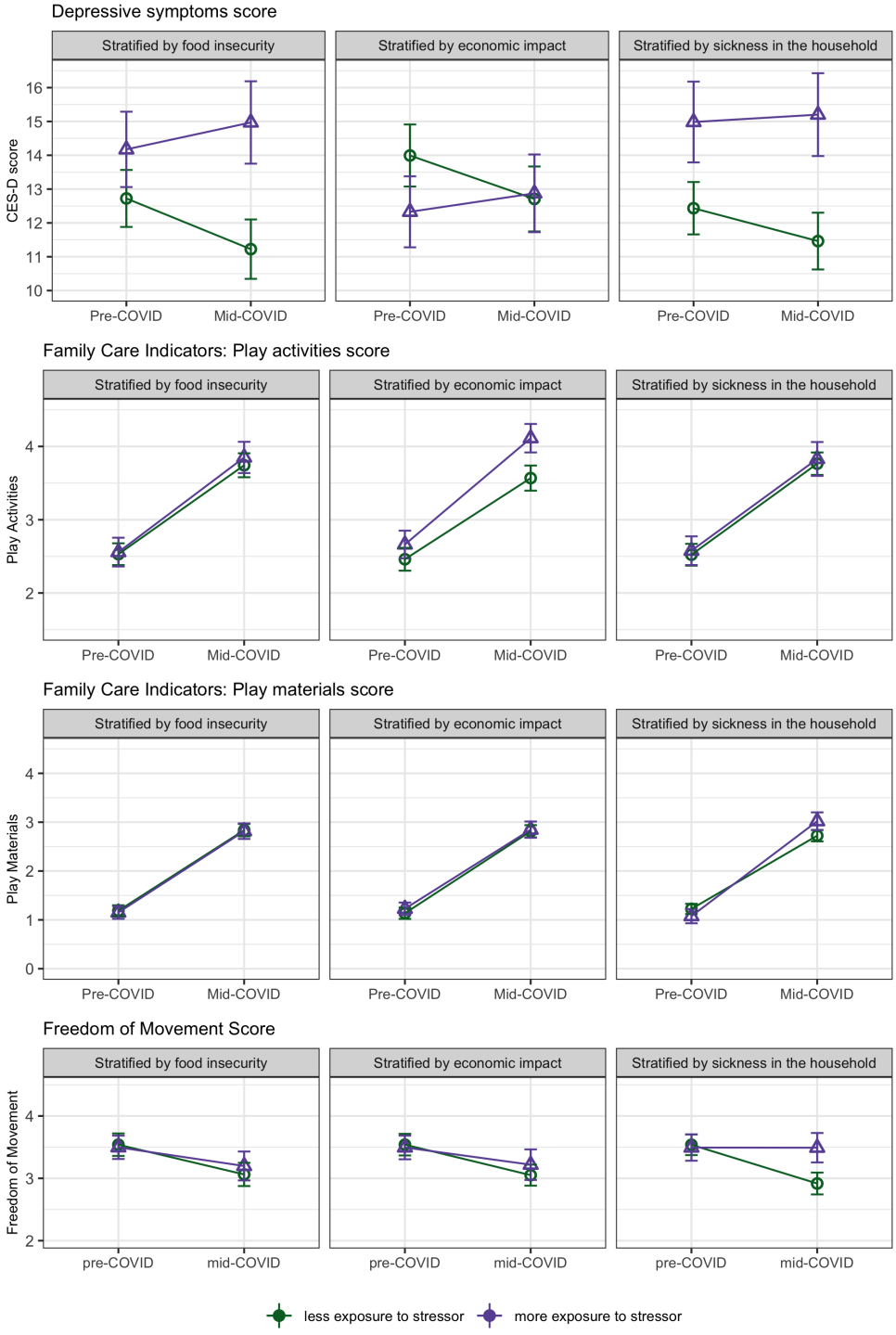


Figure 3.2: Kernel Density plot of CES-D scores comparing those who have more vs. less indicators of food insecurity at pre- and mid-COVID time assessments



Colored dashed lines correspond to mean CES-D scores in each group; Red dashed line indicates CES-D score of 16

Figure 3.3: Difference-in-differences plots for CES-D score, play activities, play materials, and freedom of movement outcomes over time, stratified by experiences at mid-COVID timepoint



(see next page for notes)

Figure 3.3: (notes from figure on previous page) Figure presents matrix of results for twelve difference-in-differences analyses whereby three different exposures during the mid-COVID time point (columns: food insecurity; economic impact; sickness in household) are related to four outcomes (rows: CES-D; play activities; play materials; freedom of movement) during the mid-COVID time period. "More exposed to food insecurity" = had two or more indicators of food insecurity. "More exposed to economic impact" = had both the primary earning member of the household lose their job and reduced household income since April 2020. "More exposed to sickness in the household" = had one or more household member experience COVID-like symptoms since April 2020.

Table 3.1: Demographic characteristics of the cohort during the pre-COVID assessment (n=517)

	% (n) or mean ± SD
Caregiver Characteristics	
Completed primary education (6+ years)	69% (369)
Currently pregnant	1% (6)
Muslim	98% (510)
Child Characteristics	
Age (in months)	8.55 ± 1.8
<6 months	7% (34)
6-12 months	93% (482)
12-18 months	<1% (4)
Female	51% (263)
Household Characteristics	
Household size	5.23 ± 1.80
Number of children <15 yrs under care	1.9 ± 0.80
Has cement floor	19% (98)
Has brick walls	23% (123)
Has electricity	99% (514)
Has refrigerator	22% (112)

Table 3.2: COVID-19 related responses in the cohort sample at the mid-COVID assessment

Since the shutdown due to COVID-19:	**Cohort (n=517)**	
	%	**n**
Household composition		
Any changes to number of people in HH	22%	114
Increased HH members (n=114)	97%	99
Interaction with child compared to prior to April 2020:		
Same interaction	53%	274
More interaction	41%	210
Less interaction	6%	33
Changes made due to COVID-19 (free response)		
More handwashing	66%	334
More cleaning of household and outdoor space	61%	316
Interacting less with people outside the household	19%	98
Restricted movement	16%	83
Wearing a mask while outside	15%	80
No changes made	16%	81
Food security		
Not able to buy essential food last 7 days	23%	122
Reason (select multiple): HH has less money (n=122)	95%	116
Reason (select multiple): items cost more (n=122)	14%	17
In the last 7 days, consuming less food when compared to the same time last year, due to COVID-19	44%	230
In the past 7 days, did you or your household members use any of the following to cover your household's basic needs? (select all)		
Reduce the number or size of meals for some HH members[1]	29%	152
Rely on less preferred and less expensive foods[1]	39%	202
Use cash or bank savings	40%	208
Take loan from someone else	34%	176
Borrow food or ask for help from a friend, relative, or neighbor	17%	87
Look for ways to earn additional money	15%	80
Sell assets	8%	39
Rely on government or NGO assistance	3%	16
No strategy used	20%	104
Two or more indicators of food insecurity[2]	41%	214
Economic impacts		
Change of employment for any household member since April	68%	351
No job or income source of the main earning member of HH	42%	219
Status of HH income since April 2020		
No HH income since April	10%	55
Reduced HH income since April	71%	366
Same HH income since April	15%	79
Increased HH income since April	3%	17
Both main earning member of HH lost job, and reduced HH income	40%	208

Health		
Any HH member sick with COVID-19 symptoms[2]	35%	181
Respondent sick with COVID-19 symptoms[2]	15%	77
Family member sick with COVID-19 symptoms[2]	27%	138

HH: Household; [1]fever, cough, cold, loss of taste/smell, shortness of breath, difficulty breathing, COVID-19; [2]Two or more of: In the past 7 days 1) reduced the number or size of meals or 2) relied on less preferred and less expensive foods to cover basic household needs, 3) not able to buy essential food items, 4) consumed less food compared to the same time last year, due to COVID-19

Table 3.3: Differences in child-development risk factors in the cohort comparing pre- and mid-COVID time points (n=517).

Outcome	**mean ± SD**		**Adjusted difference**
	Pre-COVID	**Mid-COVID**	**(95% CI)**
CES-D score	13.4 ± 8.7	12.8 ± 9.2	-0.55 (-1.30 to 0.21)
FCI Play activities	2.5 ± 1.51	3.8 ± 1.6	1.25 (1.08 to 1.42)
FCI Play materials	1.2 ± 1.1	2.8 ± 1.2	1.66 (1.52 to 1.79)
Freedom of movement score	3.5 ± 1.6	3.1 ± 1.8	-0.41 (-0.58 to -0.25)

FCI: Family Care Indicators, the play activities subscale is a sum score of the number of play activities that the caregiver participated in with the child in the previous three days (0-6); the play materials subscale is the number of different types of play materials the child has played with in the past 30 days (0-6); CES-D: Center for Epidemiologic Studies 20 question Depression questionnaire (0-60), higher scores indicate more depressive symptoms experienced; Freedom of movement score is a sum score with one point for attending each of the following four places in the last 6 months, and an additional point if that location was attended alone: the market, medical facility, outside the village, paternal home or the home of a friend or a relative (0-8); Adjusted differences are adjusted for village, child age category, maternal education, income, antenatal care, control over assets, household size, child sex, housing materials

Table 3.4: Pre-COVID characteristics of participants stratified by the food security, economic, and health domains at the mid-COVID assessment

	Food insecure			**Lost job and reduced income**			**Any household member sick**		
	Less (n=303)	**More (n=214)**	**p-value[5]**	**No (n=309)**	**Yes (n=208)**	**p-value[5]**	**No (n=336)**	**Yes (n=181)**	**p-value[5]**
Caregiver									
Completed primary education[1]	224 (74%)	134 (63%)	0.008	212 (69%)	146 (70%)	0.775	223 (66%)	135 (75%)	0.067
Currently pregnant	3 (1%)	3 (1%)	0.992	3 (1%)	3 (1%)	0.946	4 (1%)	2 (1%)	1
Children <15yrs	1.8 ± 0.8	2.0 ± 0.8	0.142	1.9 ± 0.8	1.8 ± 0.8	0.260	1.9 ± 0.8	1.8 ± 0.8	0.191
CES-D score	12.5 ± 8.2	14.5 ± 9.3	0.012	13.8 ± 9.0	12.7 ± 8.4	0.159	12.4 ± 8.3	14.9 ± 9.3	0.002
Freedom of movement[2]	3.7 ± 1.7	3.3 ± 1.5	0.036	3.6 ± 1.8	3.5 ± 1.5	0.492	3.5 ± 1.7	3.6 ± 1.5	0.558
Spend own money	191 (63%)	101 (47%)	<0.001	178 (58%)	114 (55%)	0.59	185 (55%)	107 (59%)	0.427
Caregiver MDD[3]	238 (79%)	150 (70%)	0.037	238 (77%)	150 (72%)	0.246	250 (74%)	138 (76%)	0.723
Child									
Age (months)	8.6 ± 1.9	8.5 ± 1.8	0.297	8.6 ± 1.8	8.5 ± 1.9	0.579	8.6 ± 1.8	8.4 ± 1.8	0.169
Female	144 (48%)	112 (52)	0.323	150 (49)	106 (51%)	0.653	176 (52%)	80 (44)	0.092
Child MDD[4]	34 (12%)	19 (10)	0.571	33 (11)	20 (10%)	0.885	40 (12%)	13 (7%)	0.126
FCI Play activities	2.6 ± 1.4	2.5 ± 1.6	0.496	2.5 ± 1.5	2.6 ± 1.5	0.29	2.5 ± 1.5	2.7 ± 1.5	0.153
FCI Play materials	1.3 ± 1.2	1.1 ± 1.1	0.044	1.2 ± 1.2	1.1 ± 1.1	0.707	1.2 ± 1.2	1.0 ± 1.1	0.062
1+ children's books	13 ± 4.3	5 ± 2.3	0.342	8 ± 2.6	10 ± 4.8	0.269	17 ± 5.1	1 ± 0.6	0.016
Household									
Household size	5.3 ± 1.9	5.1 ± 1.6	0.151	5.3 ± 1.9	5.1 ± 1.7	0.203	5.2 ± 1.7	5.2 ± 2.0	0.985
Has refrigerator	83 (27%)	29 (14%)	<0.001	79 (26%)	33 (16%)	0.012	74 (22%)	38 (21%)	0.874
Has television	162 (54%)	96 (45%)	0.066	154 (50%)	104 (50%)	1	171 (51%)	87 (48%)	0.602
Has brick walls	82 (27%)	39 (18%)	0.026	79 (26%)	42 (20%)	0.19	75 (22%)	46 (25%)	0.494
Has cement floor	69 (23%)	28 (13%)	0.008	71 (23%)	26 (13%)	0.004	61 (18%)	36 (20%)	0.716
Lowest income tertile	131 (43%)	128 (60%)	<0.001	149 (48%)	110 (53%)	0.059	166 (49%)	93 (51%)	0.538
Hygienic latrine	71 (23%)	41 (19%)	0.292	61 (20%)	51 (25%)	0.236	68 (20%)	44 (24%)	0.337
Handwashing station	65 (22%)	25 (18%)	0.006	57 (18%)	33 (16%)	0.522	63 (19%)	27 (15%)	0.33
Attendance									
1 or more sessions attended	165 (55%)	111 (52%)	0.623	179 (58%)	97 (47%)	0.015	180 (54%)	96 (53%)	0.981

(See next page for notes)

Table 3.4: (notes from table on previous page) All data mean ± SD or n (%); FCI: Family Care Indicators, the play activities subscale is a sum score of the number of play activities that the caregiver participated in with the child in the previous three days (0-6); the play materials subscale is the number of different types of play materials the child has played with in the past 30 days (0-6); CES-D: Center for Epidemiologic Studies 20 Question Depression questionnaire, scores range from 0-60, with higher scores indicating more depressive symptoms experienced; MDD: minimum dietary diversity; [1]6+ years of education; [2]Caregiver freedom of movement score (0-8); [3]Caregiver reported eating 5 or more food groups in the last 24 hours, out of the following 10 groups: grains, legumes, nuts and seeds, dairy products, animal flesh foods, eggs, vitamin A rich fruits and vegetables, other vitamin A rich fruits and vegetables, other vegetables, other fruits; [4]Children $>$ 6 months reported eating 5 or more food groups in the last 24 hours, out of the following groups: breast milk, grains, legumes, dairy products, animal flesh foods, eggs, vitamin A rich fruits and vegetables, other fruits and vegetables; [5]p-values are for comparisons between groups within domains with chi^2 tests for binary or categorical variables or t-tests for continuous variables

Table 3.5: Results from difference in differences analysis in the cohort group who are more and less affected on the food security, economic, and health domains at follow-up

	Exposure variable at the mid-COVID assessment					
	More food insecure		**Lost job and reduce income**		**Respondent or household members sick**	
Outcome	**DID estimate**	**p-value**	**DID estimate**	**p-value**	**DID estimate**	**p-value**
CES-D Score	2.29 (0.72 to 3.86)	0.004	1.83 (0.31 to 3.35)	0.02	1.19 (-0.47 to 2.84)	0.16
FCI Play activities	0.07 (-0.27 to 0.43)	0.66	0.34 (0.01 to 0.68)	0.05	0.01 (-0.35 to 0.37)	0.96
FCI Play materials	0.02 (-0.26 to 0.29)	0.91	-0.06 (-0.34 to 0.22)	0.68	0.44 (0.16 to 0.73)	0.002
Freedom of movement	0.17 (-0.17 to 0.53)	0.32	0.21 (-0.14 to 0.57)	0.24	0.62 (0.26 to 0.97)	0.001

DID estimate: Difference-in-differences estimates from a generalized estimating equation model accounting for repeated measures within participants, adjusted for village, child age category, maternal education, income, antenatal care, control over assets, household size, child sex, housing materials; FCI: Family Care Indicators, the play activities subscale is a sum score of the number of play activities that the caregiver participated in with the child in the previous three days (0-6); the play materials subscale is the number of different types of play materials the child has played with in the past 30 days (0-6) CES-D: Center for Epidemiologic Studies 20 question Depression questionnaire, scores range from 0-60, with higher scores indicating more depressive symptoms experienced; Freedom of movement score is a sum score with one point for attending each of the following four places in the last 6 months, and an additional point if that location was attended alone: the market, medical facility, outside the village, paternal home or the home of a friend or a relative (0-8)

Conclusion

The overall aim of this work is to inform the design of effective, sustainable, and scalable interventions that invest in local capacity to improve child development through policy implementation in low-resource settings. Through data collected in rural Bangladesh, this dissertation contributes to the literature on intervention effectiveness, considerations for measurement of interventions at scale, and the shift in the burden of risk factors for poor child development during the COVID-19 pandemic.

The first chapter examines the effects of a group-based multicomponent intervention in rural Bangladesh on child, caregiver and other outcomes. In the group-only arm, the intervention improved stimulating caregiving and child development in magnitudes similar to those measured after small-scale responsive stimulation interventions. The group-only intervention arm also improved caregiver mental health and caregivers' knowledge of lead as an invisible environmental toxin that could cause harm to their children. It also increased the presence of a handwashing station with soap and water. These results provide evidence that group-based multicomponent interventions are a viable option for testing at scale. A successful implementation of such interventions may provide more holistic impacts on children's early environments. Follow-up of the children enrolled in this trial mid-childhood would provide information on the sustainability of this multicomponent intervention.

The second chapter examines the correlation between the Bayley Scales of Infant Development-III and the Ages and Stages Questionnaire: Inventory (ASQ:I). The Bayley Scales of Infant Development-III is a direct measure of child development, whereas the ASQ:I is a caregiver report measure for early child development. The latter can more feasibly be implemented in large-scale evaluations as it requires less time, uses less financial and human resources than the direct assessment tool, and is moderately correlated with the Bayley Scales of Infant Development-III in children 4-27 months of age in rural Bangladesh. In addition to providing more evidence on which to base decisions of measurement tools, this analysis allows for a comparison of concurrent correlations across intervention arms. We found no consistent differences in concurrent correlations for participants in the intervention compared to those in the control group. This evidence provides support for the use of the ASQ:I in evaluating intervention effects.

The third chapter examines experiences during the COVID-19 pandemic and related shifts in the distribution of risk factors for poor child development. The primary risk factors examined are caregiver mental health and stimulating caregiving practices. Caregivers who experienced more food insecurity, more economic loss, and more sickness with COVID-like symptoms in the household had larger increases in depressive symptoms between the pre-COVID-19 and mid-COVID-19 time points, compared to those who had less exposure on each of these domains. There were no consistent effects across exposures for stimulating caregiving practices. An examination of the quality of stimulating caregiving could provide further insight into the magnitude of the effects on children in rural Bangladesh. This work highlights the potential impacts on child development outcomes that could be related to shifts in risk factors and provides evidence for the urgent need for financial assistance or other forms of economic empowerment in rural Bangladesh.

The global focus on early child development is shifting from small scale impact assessments to the design and evaluation of interventions implemented at a large scale to improve population-level child development. This dissertation has implications for the design and measurement of such large-scale interventions in LMIC settings. Follow-up research on (i) the scalability and sustainability of multicomponent interventions, and (ii) the longer-term effect of the COVID-19 pandemic on caregivers and children in rural Bangladesh will complement the work presented in this dissertation to inform the design of interventions to improve child development at scale. In the coming years, several key questions remain in the design and implementation of interventions to improve child development on a large scale, including: (1) How can effective early child development interventions be integrated into existing health systems and services?; (2) How can intervention quality and fidelity be maintained when these interventions are delivered at a large scale, over large geographic distances?; (3) Which large-scale interventions for early child development have the largest impact on child development and caregiver well-being while being the most cost-effective?; and (4) How can interventions that have demonstrated effects in the early years maintain their effectiveness as children age? The answers to these questions will contribute to the design of large-scale interventions that directly benefit children and families over the life-course.

Bibliography

1. Gertler, P. *et al.* Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. *Science* **344,** 998–1001 (2014).
2. Heckman, J., Pinto, R. & Savelyev, P. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review* **103,** 2052–2086 (2013).
3. Peet, E. D. *et al.* Early Childhood Development and Schooling Attainment: Longitudinal Evidence from British, Finnish and Philippine Birth Cohorts. *PLOS ONE* **10,** e0137219 (2015).
4. Bhutta, Z. A., Guerrant, R. L. & Nelson, C. A. Neurodevelopment, Nutrition, and Inflammation: The Evolving Global Child Health Landscape. *Pediatrics* **139,** S12–S22 (2017).
5. Fox, S. E., Levitt, P. & Nelson, C. A. How the Timing and Quality of Early Experiences Influence the Development of Brain Architecture. *Child Development* **81,** 28–40 (2010).
6. Black, M. M. *et al.* Early Childhood Development Coming of Age: Science through the Life Course. *The Lancet* **389,** 77–90 (2017).
7. UNICEF. *Nurturing Care for Early Childhood Development: A Framework for Helping Children Survive and Thrive to Transform Health and Human Potential* (World Health Organization, Geneva, 2018).
8. Britto, P. R. *et al.* Nurturing Care: Promoting Early Childhood Development. *The Lancet* **389,** 91–102 (2017).
9. Lu, C., Black, M. M. & Richter, L. M. Risk of Poor Development in Young Children in Low-Income and Middle-Income Countries: An Estimation and Analysis at the Global, Regional, and Country Level. *The Lancet Global Health* (2016).
10. McCoy, D. C. *et al.* Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modeling. *PLoS medicine* **13,** e1002034 (2016).
11. UN General Assembly. Convention on the Rights of the Child. *United Nations, Treaty Series* **1577,** 1–23 (1989).

12. UN General Assembly. *Transforming Our World: The 2030 Agenda for Sustainable Development* (2015).

13. Aboud, F. E. & Yousafzai, A. K. Global Health and Development in Early Childhood. *Annual Review of Psychology* **66,** 433–457 (2015).

14. Engle, P. L. *et al.* Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries. *The Lancet* **378,** 1339–1353 (2011).

15. Aboud, F. E. & Akhter, S. A Cluster-Randomized Evaluation of a Responsive Stimulation and Feeding Intervention in Bangladesh. *Pediatrics* **127,** e1191–e1197 (2011).

16. Tofail, F. *et al.* Psychosocial Stimulation Benefits Development in Nonanemic Children but Not in Anemic, Iron-Deficient Children. *The Journal of Nutrition* **143,** 885–893 (2013).

17. Aboud, F. E., Singla, D. R., Nahil, M. I. & Borisova, I. Effectiveness of a Parenting Program in Bangladesh to Address Early Childhood Health, Growth and Development. *Social Science & Medicine* **97,** 250–258 (2013).

18. Nahar B. *et al.* Effect of a Food Supplementation and Psychosocial Stimulation Trial for Severely Malnourished Children on the Level of Maternal Depressive Symptoms in Bangladesh. *Child: Care, Health and Development* **41,** 483–493 (2014).

19. Singla, D. R., Kumbakumba, E. & Aboud, F. E. 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 Global Health* **3,** e458–e469 (2015).

20. Fernald, L. C. H. *et al.* Promoting Child Development through Group-Based Parent Support within a Cash Transfer Program: Experimental Effects on Children's Outcomes. eng. *Developmental Psychology* **53,** 222–236 (2017).

21. Chang, S. M. *et al.* Integrating a Parenting Intervention With Routine Primary Health Care: A Cluster Randomized Trial. *Pediatrics,* peds.2015–0119 (2015).

22. Jeong, J., Pitchik, H. O. & Fink, G. Short-Term, Medium-Term and Long-Term Effects of Early Parenting Interventions in Low- and Middle-Income Countries: A Systematic Review. *BMJ Global Health* **6,** e004067 (2021).

23. Vazir, S. *et al.* Cluster-Randomized Trial on Complementary and Responsive Feeding Education to Caregivers Found Improved Dietary Intake, Growth and Development among Rural Indian Toddlers. *Maternal & Child Nutrition* **9,** 99–117 (2013).

24. Fernald, L. C., Prado, E., Kariger, P. & Raikes, A. A Toolkit for Measuring Early Childhood Development in Low and Middle-Income Countries (2017).

25. Rubio-Codina, M., Araujo, M. C., Attanasio, O., Muñoz, P. & Grantham-McGregor, S. Concurrent Validity and Feasibility of Short Tests Currently Used to Measure Early Childhood Development in Large Scale Studies. *PLoS One* **11,** e0160962 (2016).

26. Li, Y., Tang, L., Bai, Y., Zhao, S. & Shi, Y. Reliability and Validity of the Caregiver Reported Early Development Instruments (CREDI) in Impoverished Regions of China. *BMC Pediatrics* **20,** 475 (2020).
27. Yue, A. *et al.* Concurrent Validity of the Ages and Stages Questionnaire and the Bayley Scales of Infant Development III in China. *PLoS One* **14** (2019).
28. Clifford, J. *et al.* Examining the technical adequacy of the ages & stages questionnaires: INVENTORY. *Infants & Young Children* **31,** 310–325 (2018).
29. Milner, E. M., Fiorella, K. J., Mattah, B. J., Bukusi, E. & Fernald, L. C. H. Timing, Intensity, and Duration of Household Food Insecurity Are Associated with Early Childhood Development in Kenya. *Maternal & Child Nutrition* **14,** e12543 (2018).
30. Galasso, E., Weber, A. M., Stewart, C. P., Ratsifandrihamanana, L. & Fernald, L. C. H. Effects of Nutritional Supplementation and Home Visiting on Growth and Development in Young Children in Madagascar: A Cluster-Randomised Controlled Trial. *The Lancet Global Health* **7,** e1257–e1268 (2019).
31. Cash, R. & Patel, V. Has COVID-19 Subverted Global Health? *The Lancet* **395,** 1687–1688 (2020).
32. Yoshikawa, H. *et al.* Effects of the Global Coronavirus Disease-2019 Pandemic on Early Childhood Development: Short- and Long-Term Risks and Mitigating Program and Policy Actions. *The Journal of Pediatrics* **223,** 188–193 (2020).
33. Baranov, V., Bhalotra, S., Biroli, P. & Maselko, J. Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial. *American Economic Review* **110,** 824–859 (2020).
34. Tofail, F. *et al.* Effect of Water Quality, Sanitation, Hand Washing, and Nutritional Interventions on Child Development in Rural Bangladesh (WASH Benefits Bangladesh): A Cluster-Randomised Controlled Trial. *The Lancet Child & Adolescent Health* **2,** 255–268 (2018).
35. Prado, E. L. *et al.* Do Effects of Early Life Interventions on Linear Growth Correspond to Effects on Neurobehavioural Development? A Systematic Review and Meta-Analysis. *The Lancet Global Health* **7,** e1398–e1413 (2019).
36. Bellinger, D., Leviton, A., Waternaux, C., Needleman, H. & Rabinowitz, M. Longitudinal Analyses of Prenatal and Postnatal Lead Exposure and Early Cognitive Development. *New England Journal of Medicine* **316,** 1037–1043 (1987).
37. Forsyth, J. E. *et al.* Prevalence of Elevated Blood Lead Levels among Pregnant Women and Sources of Lead Exposure in Rural Bangladesh: A Case Control Study. *Environmental Research* **166,** 1–9 (2018).
38. Forsyth, J. E. *et al.* Turmeric Means "Yellow" in Bengali: Lead Chromate Pigments Added to Turmeric Threaten Public Health across Bangladesh. *Environmental Research,* 108722 (2019).

39. UNICEF and Pure Earth. *The Toxic Truth* (2020).
40. WHO. *Improving Early Childhood Development: WHO Guideline* (2020).
41. Luo, R., Emmers, D., Warrinnier, N., Rozelle, S. & Sylvia, S. Using Community Health Workers to Deliver a Scalable Integrated Parenting Program in Rural China: A Cluster-Randomized Controlled Trial. *Social Science & Medicine* **239,** 112545 (2019).
42. Radner, J. M., Ferrer, M. J. S., McMahon, D., Shankar, A. H. & Silver, K. L. Practical Considerations for Transitioning Early Childhood Interventions to Scale: Lessons from the Saving Brains Portfolio. *Annals of the New York Academy of Sciences* **1419,** 230–248 (2018).
43. Tripathy, P. *et al.* 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. *The Lancet* **375,** 1182–1192 (2010).
44. Andrew, A. *et al.* Impacts 2 Years after a Scalable Early Childhood Development Intervention to Increase Psychosocial Stimulation in the Home: A Follow-up of a Cluster Randomised Controlled Trial in Colombia. *PLOS Medicine* **15,** e1002556 (2018).
45. Akter, F. *et al.* Adaptation and Integration of Psychosocial Stimulation, Maternal Mental Health and Nutritional Interventions for Pregnant and Lactating Women in Rural Bangladesh. *International Journal of Environmental Research and Public Health* **17,** 6233 (2020).
46. Grantham-McGregor, S. & Smith, J. A. Extending the Jamaican early childhood development intervention. *Journal of applied research on children: Informing policy for children at risk* **7,** 4 (2016).
47. Hamadani, J. D. *et al.* Use of Family Care Indicators and Their Relationship with Child Development in Bangladesh. *Journal of Health, Population, and Nutrition* **28,** 23–33 (2010).
48. Bradley, R. H. & Caldwell, B. M. The HOME Inventory and Family Demographics. *Developmental Psychology* **20,** 315–320 (1984).
49. FAO and FHI 360. *Minimum Dietary Diversity for Women A Guide to Measurement* (Rome, Italy, 2016).
50. WHO. *Operational Guidance for Tracking Progress in Meeting Targets for 2025* (2017).
51. Ahern, J., Hubbard, A. & Galea, S. Estimating the Effects of Potential Public Health Interventions on Population Disease Burden: A Step-by-Step Illustration of Causal Inference Methods. *American Journal of Epidemiology* **169,** 1140–1147 (2009).
52. Rothman, K. J. No Adjustments Are Needed for Multiple Comparisons. *Epidemiology* **1,** 43–46 (1990).
53. Grembi, J. & McQuade, E. R. *riskCommunicator: G-Computation to Estimate Interpretable Epidemiological Effects* 2020.

54. Grantham-McGregor, S. *et al.* Group Sessions or Home Visits for Early Childhood Development in India: A Cluster RCT. *Pediatrics* (2020).

55. Luoto, J. E. *et al.* Group-Based Parenting Interventions to Promote Child Development in Rural Kenya: A Multi-Arm, Cluster-Randomised Community Effectiveness Trial. *The Lancet Global Health* **9,** e309–e319 (2021).

56. Herba, C. M., Glover, V., Ramchandani, P. G. & Rondon, M. B. Maternal Depression and Mental Health in Early Childhood: An Examination of Underlying Mechanisms in Low-Income and Middle-Income Countries. *The Lancet Psychiatry* **3,** 983–992 (2016).

57. Jeong, J., Pitchik, H. O. & Yousafzai, A. K. Stimulation Interventions and Parenting in Low- and Middle-Income Countries: A Meta-Analysis. *Pediatrics,* e20173510 (2018).

58. Yeasmin, F. *et al.* Exploration of Attendance, Active Participation, and Behavior Change in a Group-Based Responsive Stimulation, Maternal and Child Health, and Nutrition Intervention. *The American Journal of Tropical Medicine and Hygiene* **104,** 1586 (2021).

59. Hamadani, J. D. *et al.* Integrating an Early Childhood Development Programme into Bangladeshi Primary Health-Care Services: An Open-Label, Cluster-Randomised Controlled Trial. *The Lancet Global Health* **7,** e366–e375 (2019).

60. Yousafzai, A. K. *et al.* Effects of Responsive Stimulation and Nutrition Interventions on Children's Development and Growth at Age 4 Years in a Disadvantaged Population in Pakistan: A Longitudinal Follow-up of a Cluster-Randomised Factorial Effectiveness Trial. *The Lancet Global Health* **4,** e548–558 (2016).

61. Forsyth, J. E. *et al.* Sources of Blood Lead Exposure in Rural Bangladesh. *Environmental Science & Technology* **53,** 11429–11436 (2019).

62. WHO. *Thinking Healthy: A Manual for Psychosocial Management of Perinatal Depression* (2015).

63. Richter, L. *et al.* Early Childhood Development: An Imperative for Action and Measurement at Scale. *BMJ Global Health* **4,** e001302 (2019).

64. Bayley, N. *Bayley Scales of Infant and Toddler Development-Third Edition* (Harcourt Assessment, San Antonio, TX, 2006).

65. Pendergast, L. L. *et al.* Assessing Development across Cultures: Invariance of the Bayley-III Scales Across Seven International MAL-ED Sites. *School Psychology Quarterly* **33,** 604–614 (2018).

66. Small, J. W., Hix-Small, H., Vargas-Baron, E. & Marks, K. P. Comparative Use of the Ages and Stages Questionnaires in Low- and Middle-Income Countries. *Developmental Medicine & Child Neurology* **61,** 431–443 (2019).

67. Stewart, C. P. *et al.* Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on Child Development in Rural Kenya (WASH Benefits Kenya): A Cluster-Randomised Controlled Trial. *The Lancet Child & Adolescent Health* **2,** 269–280 (2018).

68. Fernald, L. C. H., Kariger, P., Hidrobo, M. & Gertler, P. J. Socioeconomic Gradients in Child Development in Very Young Children: Evidence from India, Indonesia, Peru, and Senegal. *Proceedings of the National Academy of Sciences* **109,** 17273–17280 (2012).

69. Xie, H. *et al.* Adapting and Validating a Developmental Assessment for Chinese Infants and Toddlers: The Ages & Stages Questionnaires: Inventory. *Infant Behavior and Development* **49,** 281–295 (2017).

70. Pitchik, H. O. *et al.* A Holistic Approach to Promoting Early Child Development: A Cluster Randomised Trial of a Group-Based, Multicomponent Intervention in Rural Bangladesh. *BMJ Global Health* **6,** e004307 (2021).

71. Lancaster, G. A. *et al.* Creation of the WHO Indicators of Infant and Young Child Development (IYCD): Metadata Synthesis across 10 Countries. *BMJ Global Health* **3,** e000747 (2018).

72. Walker, S. P. *et al.* Inequality in Early Childhood: Risk and Protective Factors for Early Child Development. *The Lancet* **378,** 1325–1338 (2011).

73. Evans, J. D. *Straightforward Statistics for the Behavioral Sciences.* ISBN: 0-534-23100-4 (Thomson Brooks/Cole Publishing Co, 1996).

74. Alderman, H., Friedman, J., Ganga, P., Kak, M. & Rubio-Codina, M. Assessing the Performance of the Caregiver Reported Early Development Instruments (CREDI) in Rural India. *Annals of the New York Academy of Sciences* (2020).

75. Schonhaut, L., Armijo, I., Schönstedt, M., Alvarez, J. & Cordero, M. Validity of the Ages and Stages Questionnaires in Term and Preterm Infants. *Pediatrics* **131,** e1468–e1474 (2013).

76. Roberton, T. *et al.* Early Estimates of the Indirect Effects of the COVID-19 Pandemic on Maternal and Child Mortality in Low-Income and Middle-Income Countries: A Modelling Study. *The Lancet Global Health* **8,** e901–e908 (2020).

77. Akseer, N., Kandru, G., Keats, E. C. & Bhutta, Z. A. COVID-19 Pandemic and Mitigation Strategies: Implications for Maternal and Child Health and Nutrition. *The American Journal of Clinical Nutrition* **112,** 251–256 (2020).

78. Laborde, D., Martin, W., Swinnen, J. & Vos, R. COVID-19 Risks to Global Food Security. *Science* **369,** 500–502 (2020).

79. Lund, C. *et al.* Poverty and Mental Disorders: Breaking the Cycle in Low-Income and Middle-Income Countries. *The Lancet* **378,** 1502–1514 (2011).

80. Black, M. M. *et al.* Depressive Symptoms among Rural Bangladeshi Mothers: Implications for Infant Development. *Journal of Child Psychology and Psychiatry* **48,** 764–772 (2007).
81. Stein, A. *et al.* Effects of Perinatal Mental Disorders on the Fetus and Child. *The Lancet* **384,** 1800–1819 (2014).
82. Esposito, G., Manian, N., Truzzi, A. & Bornstein, M. H. Response to Infant Cry in Clinically Depressed and Non-Depressed Mothers. *PloS One* **12,** e0169066 (2017).
83. Pratley, P. Associations between Quantitative Measures of Women's Empowerment and Access to Care and Health Status for Mothers and Their Children: A Systematic Review of Evidence from the Developing World. *Social Science & Medicine* **169,** 119–131 (2016).
84. Glennerster, R., Walsh, C. & Diaz-Martin, L. A practical guide to measuring women's and girls' empowerment in impact evaluations. *Gender Sector, Abdul Latif Jameel Poverty Action Lab* (2018).
85. Mahmud, S., Shah, N. M. & Becker, S. Measurement of Women's Empowerment in Rural Bangladesh. *World development* **40,** 610–619 (2012).
86. Schuler, S. R. & Hashemi, S. M. Credit Programs, Women's Empowerment, and Contraceptive Use in Rural Bangladesh. *Studies in Family Planning* **25,** 65–76 (1994).
87. Chowdhury, A. M. R. *et al.* The Bangladesh Paradox: Exceptional Health Achievement despite Economic Poverty. *The Lancet* **382,** 1734–1745 (2013).
88. Bangladesh Bureau of Statistics. *National Poverty Level of Bangladesh Based on Quarterly Estimates* (2016).
89. World Bank. *Bangladesh Country Profile* (2018).
90. Bangladesh Bureau of Statistics & UNICEF Bangladesh. *Progotir Pathey, Bangladesh Multiple Indicator Cluster Survey 2019, Survey Findings Report* (2019).
91. National Institute of Population Research and Training (NIPORT), Ministry of Health and Family Welfare & ICF. *Bangladesh Demographic and Health Survey 2017-18* (2020).
92. UN General Assembly. *Sustainable Development Knowledge Platform* (2015).
93. Wachs, T. D. in *Childhood Poverty: Multidisciplinary Approaches* (eds Boyden, J. & Bourdillon, M.) 148–165 (Palgrave Macmillan UK, London, 2012). ISBN: 978-0-230-36279-6.
94. Hamadani, J. D. *et al.* Immediate Impact of Stay-at-Home Orders to Control COVID-19 Transmission on Socioeconomic Conditions, Food Insecurity, Mental Health, and Intimate Partner Violence in Bangladeshi Women and Their Families: An Interrupted Time Series. *The Lancet Global Health* (2020).

95. Gelaye, B., Rondon, M., Araya, R. & Williams, M. A. Epidemiology of Maternal Depression, Risk Factors, and Child Outcomes in Low-Income and Middle-Income Countries. *The Lancet Psychiatry* **3,** 973–982 (2016).

96. Bangladesh Bureau of Statistics. *Bangladesh Population and Housing Census 2011* (2011).

97. Bangladesh Bureau of Statistics. *Bangladesh Population and Housing Census 2011: Community Report, Zila Pabna* (2013).

98. Lopez-Pena, P., Davis, C. A., Mobarak, A. M. & Raihan, S. Prevalence of COVID-19 Symptoms, Risk Factors, and Health Behaviors in Host and Refugee Communities in Cox's Bazar: A Representative Panel Study. *Bull World Health Organ* (2020).

99. Radloff, L. S. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* **1,** 385–401 (1977).

100. Biswas, T. K. & Kabir, M. Measuring Women's Empowerment: Indicators and Measurement Techniques. *Social Change* **34,** 64–77 (2004).

101. Lewinsohn, P. M., Seeley, J. R., Roberts, R. E. & Allen, N. B. Center for Epidemiologic Studies Depression Scale (CES-D) as a Screening Instrument for Depression among Community-Residing Older Adults. *Psychology and Aging* **12,** 277–287 (1997).

102. Das, S. *et al.* Acute Food Insecurity and Short-Term Coping Strategies of Urban and Rural Households of Bangladesh during the Lockdown Period of COVID-19 Pandemic of 2020: Report of a Cross-Sectional Survey. *BMJ Open* **10,** e043365 (2020).

103. Hessami, K., Romanelli, C., Chiurazzi, M. & Cozzolino, M. COVID-19 Pandemic and Maternal Mental Health: A Systematic Review and Meta-Analysis. *The Journal of Maternal-Fetal & Neonatal Medicine* **0,** 1–8 (2020).

104. Lund, C. *et al.* Poverty and Common Mental Disorders in Low and Middle Income Countries: A Systematic Review. *Social Science & Medicine (1982)* **71,** 517–528 (2010).

105. Tampubolon, G. & Hanandita, W. Poverty and Mental Health in Indonesia. *Social Science & Medicine* **106,** 20–27 (2014).

106. Goodman, J. H. Postpartum Depression Beyond the Early Postpartum Period. *Journal of Obstetric, Gynecologic, & Neonatal Nursing* **33,** 410–420 (2006).

107. Raihan, M. J. *et al.* Effect of Seasons on Household Food Insecurity in Bangladesh. *Food and Energy Security* **7,** e00136 (2018).

108. Ehsan, A. M. & Silva, M. J. D. Social Capital and Common Mental Disorder: A Systematic Review. *J Epidemiol Community Health* **69,** 1021–1028 (2015).

109. Kurstjens, S. & Wolke, D. Effects of Maternal Depression on Cognitive Development of Children Over the First 7 Years of Life. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* **42,** 623–636 (2001).

110. Liu, Y. *et al.* Maternal Depressive Symptoms and Early Childhood Cognitive Development: A Meta-Analysis. *Psychological Medicine* **47,** 680–689 (2017).

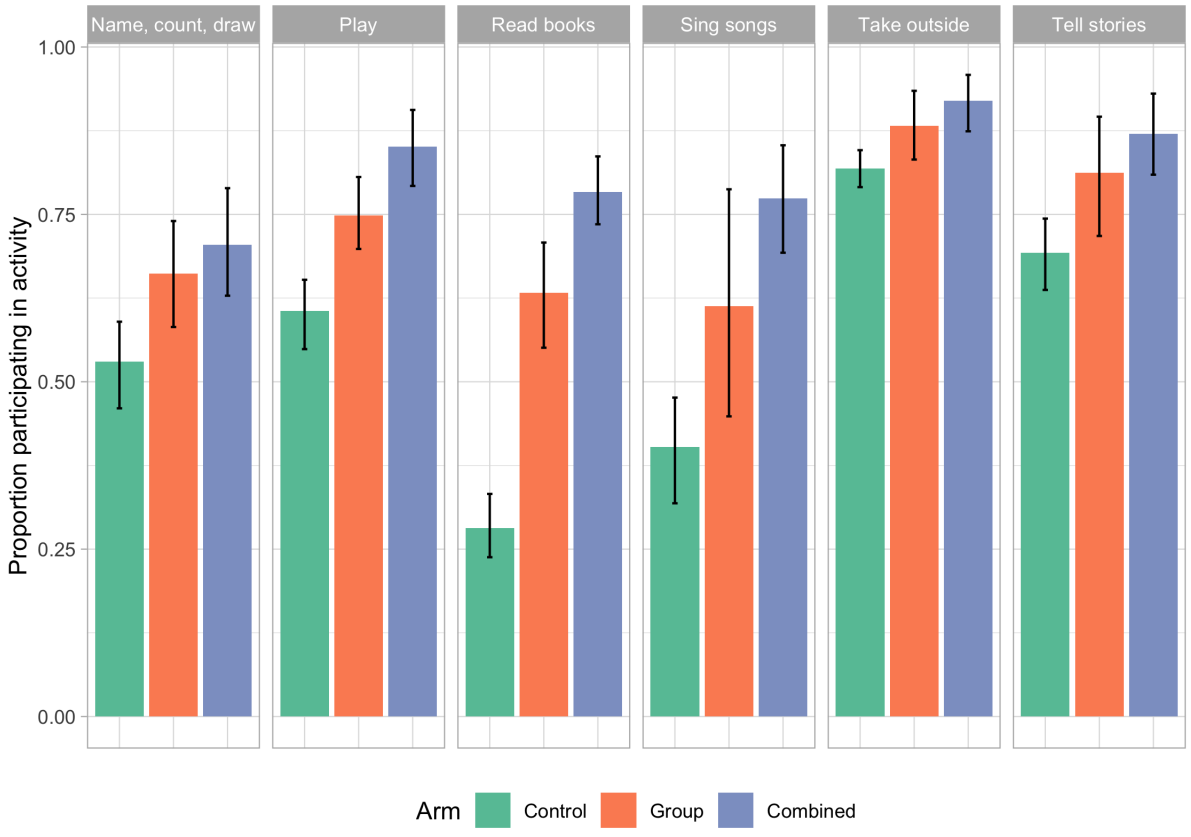
111. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2019* (2019).

112. United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision, Online Edition* (2018).

Appendix A

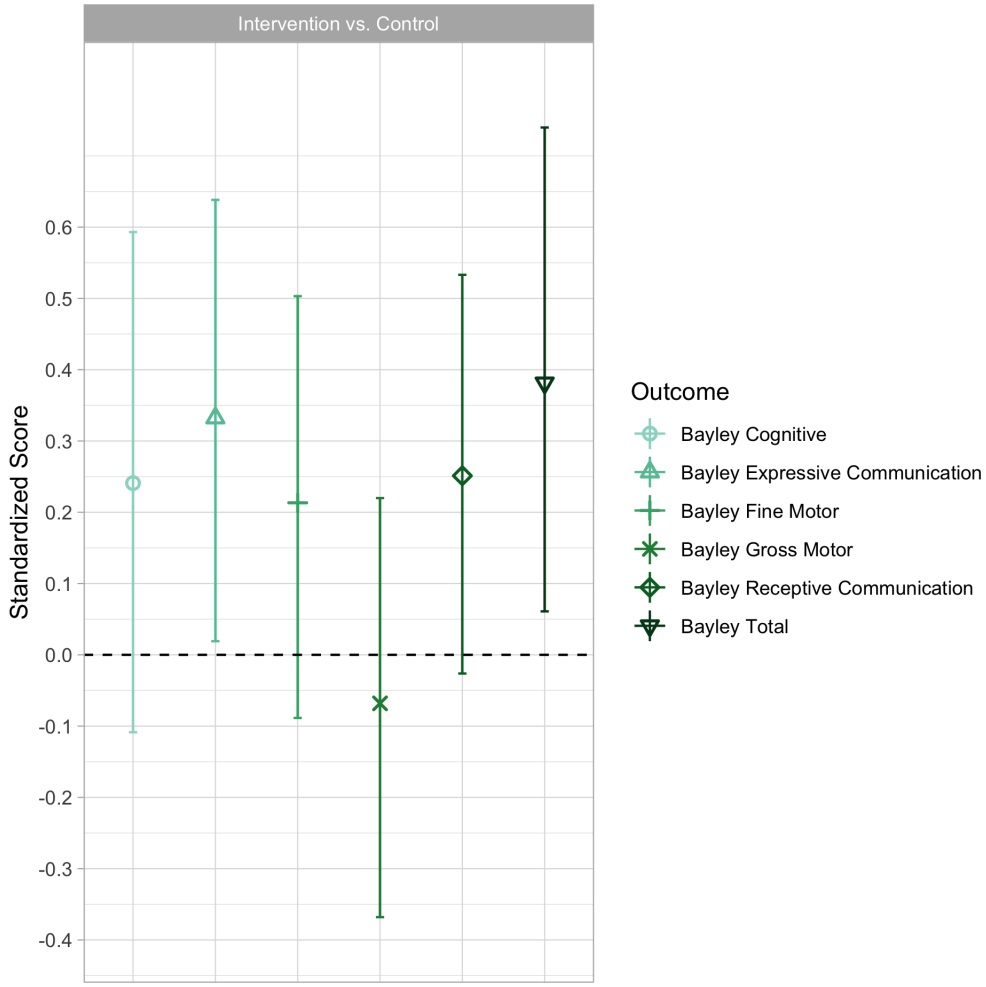
Appendix to Chapter 1

Figure A.1: Play activities in the home, by activity, at endline



Play activities in the home, by activity, at endline. Each of the six stimulation activities is caregiver reported participation in the previous three days

Figure A.2: Adjusted mean differences in age-standardized Bayley-III scores, any intervention vs. control



Adjusted mean differences in age-standardized Bayley-III scores, any intervention vs. control. All results are internally age-standardized to the control arm, and point estimates represent differences between any intervention group, and the control arm. Lines represent 95% confidence intervals n by domain, after removing outliers (± 4 SD) and individuals with missing data on each domain: Expressive Communication, $n = 250$; Receptive Communication, $n = 252$; Gross Motor, $n = 250$; Fine Motor, $n = 249$; Cognitive, $n = 249$; Total, $n = 245$ Potential covariates for inclusion in adjusted include interviewer, maternal and paternal education, child age, child sex, household income above the median, household wall material, presence of electricity in the home and the presence of household assets (wardrobe, table, chair, watch/clock, television, fridge, bicycle, sewing machine). For each outcome covariates were prescreened using a likelihood ratio test, and all covariates with $p < 0.20$ were included in adjusted analyses.

Table A.1: Exceptions to inclusion criteria for village selection

# of villages	Exception
1	Household ownership of 92.2% which was less than the mean -1.5*SD = 93.4%.
1	Village that had 193 households at the 2011 BBS census.
2	Village centers 1.15 km apart

For villages that were less than 1km apart we kept the village that was closer to the mean values on basic demographic variables. When three village centers were within 1 km of each other, we excluded the village in the middle, so that the remaining villages were approximately 2 km apart.

Table A.2: Brief description of intervention components by session (Group arm)

Session	Contents
G1	• Recap of community meeting (how WASH, nutrition, and child stimulation can help child development) • Importance of early stimulation • Introduction to thinking healthy • Interactions with baby
G2	• ANC • Importance of MMN/MNP/LNS • Maternal and Child nutrition, growth, and development • Interactions with baby • Talking while bathing baby • Distribution of micronutrients according to age/status
G3	• Sanitation, infection, growth and development • Water quality, infection, growth and development • Early care of neonate and cord care • Hand washing, infection, growth and development • Food hygiene, infection, growth and development • Interactions with baby • Talking while feeding the baby • Distribution of kangaroo care foot measuring stick to pregnant women
G4	• Pleasurable activities and reduced maternal workload • Physiological symptoms of pregnancy • Interactions with baby • Learning common nouns • Thinking healthy regarding “Mother’s health” • Distribution of micronutrients according to age/status
G5	• Maternal dietary diversity • Maternal weight gain • Interactions with baby • Talking while walking with baby • Lead information

G6	• Breastfeeding • Kangaroo Mother Care • Love and praise baby • Interact while breastfeeding • Interactions with baby • Learning action words • Thinking healthy regarding "Mother's relationship with baby" • Distribution of micronutrients according to age/status
G7	• Kangaroo mother care • Exclusive Breastfeeding • Breastfeeding challenges • Interactions with baby • Learning body parts
G8	• Safe disposal of human and animal feces • Constructing a hygienic latrine • Interactions with baby • Talking while dressing baby • Thinking healthy regarding "Mother's relation with people around her" • Distribution of micronutrients according to age/status
G9	• Use of child potty and potty training • Handwashing • Interactions with baby • Learning to follow directions
G10	• Child dietary diversity • Processed food • Responsive feeding • Interactions with baby • Touch and bring game • Thinking healthy regarding "Mother's health" • Distribution of micronutrients according to age/status
G11	• Handwashing, especially before feeding • Making soapy water • Interactions with baby • Talking while doing household activity • Sick child feeding
G12	• Child feeding frequency • Continued breast feeding • Responsive feeding • Interactions with baby • Funny game-act out activities • Thinking healthy regarding "Mother's relation with baby" • Distribution of micronutrients according to age/status

G13	• Water storage • Food storage • Exclusive breastfeeding • Continued breastfeeding • Interaction with baby • Finger game
G14	• Handwashing • Soapy water • Interaction with baby • Peek-a-boo game • Thinking healthy regarding "Mother's relationship with people around her" • Adverse health effect of arsenic on human health • Collect and drink water from arsenic free safe water source • Distribution of micronutrients according to age/status
G15	• Interaction with baby • Hand game • Sick child Feeding • Complementary feeding • Construct/upgrade and maintain child friendly toilet • Teach and encourage children elder than three years to use child friendly toilet
G16	• Integrated recommendation on food storage, hand washing, water storage • Exclusive breastfeeding • Continued breastfeeding • Maternal and child dietary diversity • 1000 days • Interaction with baby • Mirror game • Thinking healthy-real life experiences • Building collective efficacy avoid arsenic contamination • Distribution of micronutrients according to age/status
G17	• Healthy interaction process, centered around feeding – Complementary feeding, processed food, sick child feeding – Handwashing and responsive feeding – Healthy Thinking • Importance of early stimulation • Interaction with baby • Find-it-game
G18	• Healthy Community – Healthy Thinking – Whole community practices healthy behaviors • Recap of community meeting (Mention how WASH, nutrition, and child stimulation can help child development) • Interactions with baby • Learning action words

Table A.3: Observation items during the FCI interview

	Observation item
1.	(Mother/Guardian) Spontaneously spoke to child twice or more (excluding scolding)
2.	(Mother/Guardian) Responded verbally to child's speech or verbal bids for attention
3.	(Mother/Guardian) Provided toys or interesting activities for child
4.	(Mother/Guardian) Caressed, kissed, or hugged child at least once
5.	(Mother/Guardian) Kept child in view/could see child/looked at (him/her) often
6.	*(Mother/Guardian) Interfered with child's actions or restricted child from exploring more than 3 times
7.	Child's play environment is safe (no potentially dangerous health or structural hazards within a toddler's or infant's range)
8.	Reading material (newspapers, magazines, etc.) is present and visible
9.	Child and child's clothing appear clean
10.	*(Mother/Guardian) Slapped or spanked child one or more times
11.	*There is evidence that older children are handling the child inappropriately (i.e., handling roughly, hitting, etc.)

Observation items during the FCI interview. All questions have a binary yes/no response, and the total score is the sum of all items, with negative items reverse coded; Items responses are summed with starred items are reverse coded. Items are a subset of observation items from the Early Childhood HOME assessment that have been piloted, pretested, and previously used in Bangladesh, in addition to two additional items (questions 9 and 11).

Table A.4: Number of direct assessment items by ASQ:I domain

Domain	**Direct assessment items**	**Total items**
Communication	6	63
Gross Motor	11	65
Fine Motor	20	63
Problem Solving	9	68
Personal Social	4	50

Table A.5: Internally age-standardized child development outcomes at endline

	mean (SD)	
	Group	**Combined**
ASQi		
Communication	0.29 (0.87)	0.19 (0.96)
Fine Motor	0.26 (0.91)	0.15 (1.07)
Gross motor	0.24 (0.88)	-0.01 (1.11)
Problem-solving	0.13 (0.96)	0.19 (0.97)
Personal Social	0.27 (0.99)	0.24 (1.18)
Total	0.29 (0.92)	0.19 (1.09)
CDI		
Receptive	0.26 (0.93)	0.21 (0.91)
Expressive	0.25 (1.15)	0.16 (1.22)

Scores are internally age-standardized to the control arm: control arm means are between -0.02 and 0.02, and standard deviations are between 0.99-1.00, and not included in the table. n by outcome (excluding outliers±4 SD from the control arm mean): Communication $n = 566$; Fine Motor $n = 559$, Gross Motor $n = 563$; Problem-solving $n = 563$; Personal Social $n = 550$; Total $n = 532$; Receptive $n = 573$; Expressive (only children over 9 months of age included) $n = 498$

Appendix B

Appendix to Chapter 2

Figure B.1: Materials used during ASQ:I assessment

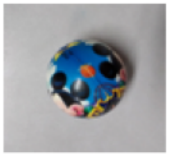
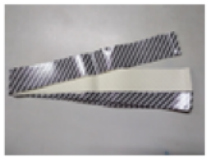
ASQI questionnaire forms, paper and pencil	Picture books 	Wooden blocks 	Shoe lace and beads 
Mirror 	Wooden push car 	Safety scissor 	Doll puzzle 
Large ball 	Small ball 	Stepping path 	Zipper shirt 

Table B.1: Number of direct assessment items by ASQ:I domain

Domain	**Direct assessment items**	**Total items**	**%**
Communication	6	63	10%
Gross Motor	11	65	17%
Fine Motor	20	63	32%
Problem Solving	9	68	13%
Personal Social	4	50	8%
Total	50	309	16%

Table B.2: Correlation between Bayley-III and ASQ:I assessments by child age and domain

ASQ:I domain	**Bayley-III domain**					
Control	**Cognitive**	**Receptive Language**	**Expressive Language**	**Fine Motor**	**Gross Motor**	**Total**
Problem Solving	**0.403 (0.18,0.59)**	0.335 (0.14,0.50)	0.297 (0.02,0.54)	0.198 (-0.06,0.41)	0.299 (0.04,0.54)	0.417 (0.16,0.62)
Communication	0.280 (0.11,0.45)	**0.381 (0.16,0.58)**	**0.228 (-0.05,0.49)**	0.192 (-0.02,0.41)	0.200 (-0.03,0.44)	0.324 (0.10,0.53)
Fine Motor	0.178 (-0.05,0.39)	0.254 (0.00,0.42)	0.257 (0.12,0.39)	**0.189 (-0.01,0.37)**	0.255 (0.00,0.52)	0.303 (0.09,0.49)
Gross Motor	0.168 (-0.07,0.38)	0.150 (-0.02,0.33)	0.197 (0.04,0.36)	0.078 (-0.08,0.24)	**0.619 (0.41,0.77)**	0.420 (0.21,0.61)
Personal Social	0.064 (-0.18,0.28)	0.370 (0.09,0.59)	0.215 (0.02,0.42)	0.243 (-0.01,0.46)	0.196 (-0.06,0.45)	0.291 (0.00,0.54)
Total	0.270 (0.17,0.39)	0.422 (0.19,0.59)	0.293 (0.08,0.50)	0.212 (0.01,0.41)	0.401 (0.14,0.66)	**0.453 (0.22,0.63)**
13-18 months						
Problem Solving	**0.332 (0.20,0.46)**	0.326 (-0.01,0.56)	0.260 (-0.02,0.49)	0.236 (-0.02,0.44)	0.078 (-0.15,0.32)	0.335 (0.13,0.49)
Communication	0.084 (-0.07,0.27)	**0.492 (0.22,0.65)**	**0.381 (0.06,0.63)**	0.193 (-0.15,0.45)	0.159 (-0.04,0.36)	0.311 (0.11,0.49)
Fine Motor	0.194 (0.04,0.36)	0.196 (-0.09,0.39)	0.243 (0.02,0.46)	**0.217 (0.00,0.40)**	0.110 (-0.17,0.33)	0.239 (0.04,0.40)
Gross Motor	0.076 (-0.12,0.26)	0.247 (-0.01,0.49)	0.101 (-0.04,0.24)	0.118 (-0.17,0.36)	**0.595 (0.48,0.72)**	0.384 (0.26,0.50)
Personal Social	0.140 (-0.04,0.38)	0.245 (-0.01,0.41)	0.206 (-0.01,0.38)	0.252 (0.01,0.42)	0.036 (-0.21,0.32)	0.212 (0.02,0.37)
Total	0.250 (0.08,0.43)	0.385 (0.06,0.59)	0.317 (0.04,0.53)	0.274 (0.01,0.41)	0.228 (0.01,0.48)	**0.391 (0.20,0.55)**
19-26 months						
Problem Solving	**0.216 (0.02,0.38)**	0.201 (-0.01,0.37)	0.259 (0.17,0.33)	0.419 (0.28,0.54)	0.343 (0.21,0.45)	0.286 (0.15,0.46)
Communication	0.254 (0.01,0.42)	**0.304 (0.07,0.50)**	**0.426 (0.22,0.59)**	0.268 (0.04,0.48)	0.321 (0.11,0.48)	0.305 (0.07,0.54)
Fine Motor	0.214 (0.09,0.31)	0.166 (-0.03,0.29)	0.173 (0.03,0.27)	**0.309 (0.19,0.44)**	0.255 (0.06,0.38)	0.155 (-0.05,0.33)
Gross Motor	0.255 (0.03,0.43)	0.179 (0.01,0.34)	0.107 (-0.09,0.25)	0.305 (0.15,0.45)	**0.385 (0.24,0.54)**	0.324 (0.14,0.48)
Personal Social	0.257 (0.02,0.47)	0.139 (-0.08,0.33)	0.146 (-0.01,0.29)	0.182 (-0.01,0.39)	0.157 (-0.08,0.35)	0.207 (0.01,0.41)
Total	0.315 (0.12,0.47)	0.186 (0.02,0.33)	0.269 (0.08,0.41)	0.336 (0.22,0.49)	0.365 (0.22,0.53)	**0.389 (0.23,0.54)**

Concordant domains across measures are represented in bold; Maximum n per age group: 4-12=73; 13-18=86; 19-26=85

Appendix C

Appendix to Chapter 3

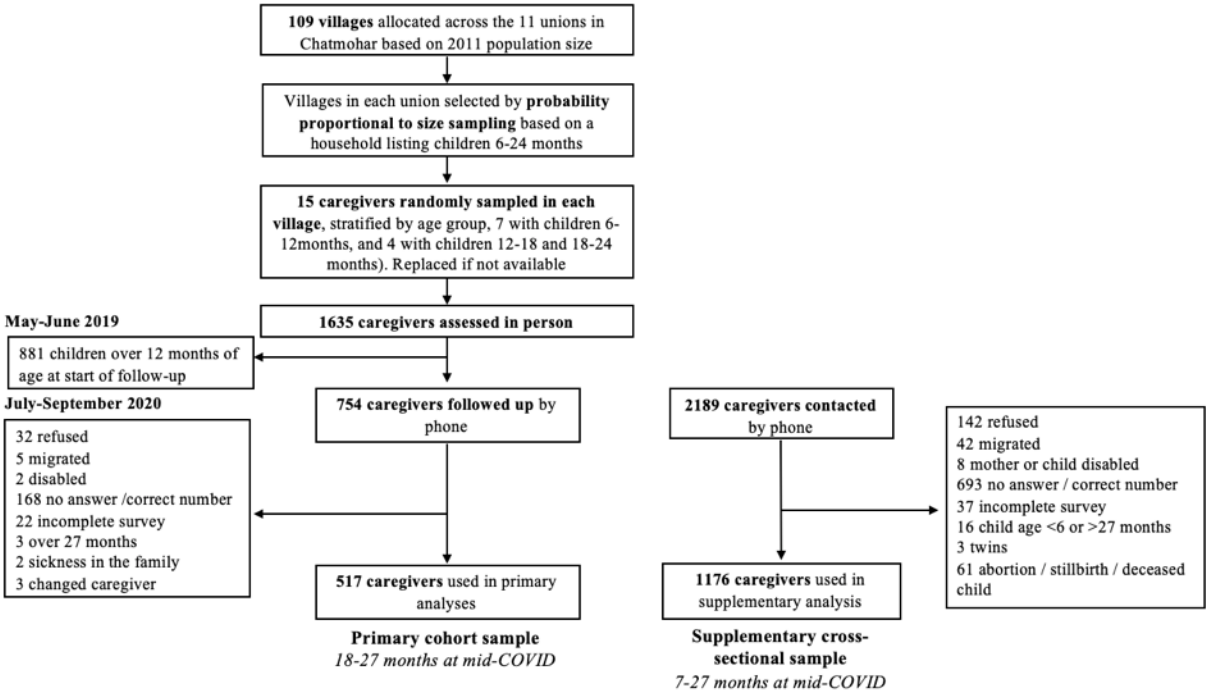
Description of multistage sampling

The stages of sampling were as follows: 1) 109 villages were allocated across the 11 rural unions of Chatmohar according to the proportion of the population that lived in each union according to the most recent national census, with 5–13 villages allocated to each union; 2) Using data from a household listing carried out by our study team, all villages with under 25 households with children 6–24 months of age were merged with the adjacent village so all villages had at least 25 eligible individuals; 3) Population proportional to size sampling was done within each union to select the specific villages to sample from; 4) In each sampled village 15 caregivers were selected using stratified random sampling with replacement (if participants refused or were unavailable) in the following three age ranges: 6–12 months ($n = 7$), 12–18 months ($n = 4$) and 18 to 24 months ($n = 4$).

Development of COVID-19 related questions

The research team first conducted qualitative interviews with 7 mothers of children 24-36 months of age in Chatmohar to better understand the ways in which the COVID-19 pandemic and subsequent mitigation strategies had affected their lives. We used parts of a survey previously done in Bangladesh to gather information on COVID-19 and added on some additional questions to address the specific concerns from the qualitative work. We then piloted this adapted survey with 15 caregivers of children 6–24 months to test for clarity and length (full survey in Table C.1).

Figure C.1: Study sample including cross-sectional mid-COVID sample



In the cross-sectional sample participants from additional villages were recruited for the cross-sectional assessment when all of the eligible participants in the original villages were attempted to be surveyed, and the desired sample size had not yet been met.

Figure C.2: Graphical depiction of samples used in primary and supplementary analyses.

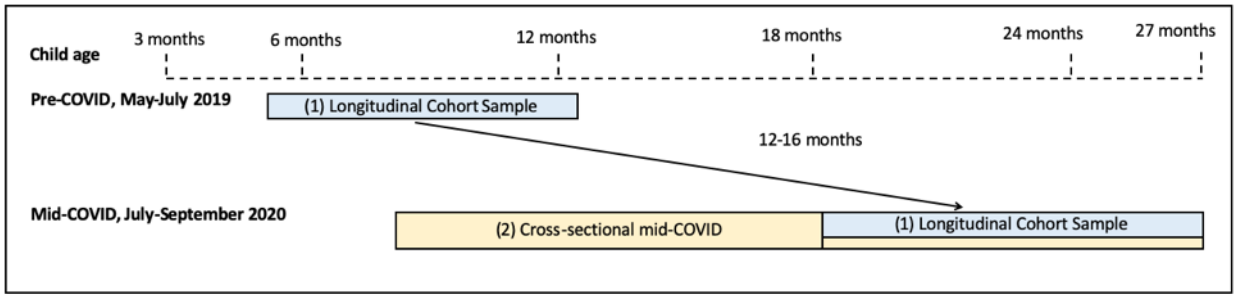


Table C.1: COVID-19 related questions

RINEW-G Module 23: Impact of COVID-19		
****Section A: Interaction and movement of the target child****		
q2301	Since the shutdown/lockdown due to COVID-19, have there been any changes in the number of people staying in your household (this number is the total number of household members who usually lived/stayed most of their time in that household)?	1. Yes 0. No
q2302	How has it changed? If the response is 4, then say to the mother, "I am sorry to hear that"	1. Increased; an adult or child has moved in 2. Increased; new baby member has been born 3. Decreased; an adult or child has left 4. Decreased; a household member has passed away
q2302a	How long an additional member (household members/relatives) is staying or stayed in your household? ***Record the duration in days***	
q2302b	Is the additional member (household member/relative) still staying in your household?	1. Yes 0. No
q2303	What was the cause of death? ***(Free response. Do not read options. Select all that apply)***	1. COVID-19 2. Fever 3. Cold 4. Difficulty in breathing 5. Accident/Injury 6. Stroke 77. Others (please specify) 99. Don't know
q2303_ot	Others (please specify)	77. Others (please specify)
q2304	Did the household member die at a health facility?	1. Yes 0. No
q2305	For the last 7 days, how much have you interacted (talking, playing, or spending time) with your child compared to time before lockdown due to COVID-19?	1. Same 2. More 3. Less

q2306	Why have you interacted (talking, playing, or spending time) with your child more/less? (Select all that apply)	1. Spending more time in the household due to restricted movement outside the house after lock down due to COVID-19 2. Child's sickness 3. Own sickness 4. More household chores 5. Less household chores 77. Others (please specify)
q2307	For the last 7 days, how much total time has your child interact with (talking, playing, or spending time) other household members compared to time before lockdown due to COVID-19?	1. Same 2. More 3. Less
q2308	Why other household members interacted (talking, playing, or spending time) with your child more/less?	1. Spending more time in the household due to restricted movement outside the house after lock down due to COVID-19 2. Child's sickness 3. More family members in the household since the lockdown due to COVID-19 4. Due to mother's sickness other family members spending more time with the child 77. Others (please specify)
q2309	What changes have been made in the household due to the COVID-19 pandemic? (Do not read the responses. Select all that apply)	0. No change due to COVID-19 1. More handwashing 2. More cleaning of household and outdoor space 3. Interacting less with people outside the household 4. Restricted movement 5. Wearing a mask when going outside 6. Children have stopped going to school / Madrasa 77. Others (please specify)
****Section B: Food security and household food consumption****		
q2311	Was your household able to buy essential food items over the past 7 days?	1. Yes 0. No

q2312	Which essential food items were you unable to obtain? ***Read all the options, select all that apply***	1. Water 2. Rice 3. Lentils 4. Oil 5. Fresh Fish 6. Chicken 7. Beef 8. Egg 9. Vegetables 77. Other (please specify)
q2313	Why were you unable to buy these items? ***Read out all options, Select all that apply***	1. Items were not available 2. Items were more expensive than usual 3. Items were the same cost but household had less money to spend 4. Markets/shops were closed 0. None of the above
q2314	In the last 7 days, are you consuming the same, more, or less food when compared to same time last year?	1. Same 2. Reduced 3. Increased
q2315	Why reduced?	1. Due to COVID-19 pandemic 2. Due to own sickness 77. Others (please specify)
****Section C: Health status of the household****		
q2321	Have you been sick since the lockdown due to COVID-19?	1. Yes 0. No
q2322	What kind of sickness? ***(Select all that apply)***	1. Fever 2. Cough 3. Cold 4. Loss of taste/smell 5. Shortness of breath or had difficulty in breathing 6. COVID-19 77. Others (please specify)
q2323	Did you go to a health facility for treatment?	1. Yes 0. No

q2324	Why didn't you go to the health facility for treatment? ***(Do not read out the options. Select all that apply)***	1. No facility was open 2. Facility was not officially closed but (we had heard) there was no staff 3. We were afraid of catching COVID-19 (or other illness) at the health facility 4. Too expensive 5. Sickness was possible to manage at home 77. Others (please specify)
q2325	Did any other members of your household fall sick since the lockdown due to COVID-19?	1. Yes 0. No
q2325a	Who fell sick?	1. Target child 2. Other child 3. Mother/Mother-in-law 4. Father/Father-in-law 5. Close Relative 77. Other (specify)
q2326	What kind of sickness? ***(Select all that apply)***	1. Fever 2. Cough 3. Cold 4. Loss of taste/smell 5. Shortness of breath or had difficulty in breathing 6. COVID-19 77. Others (please specify)
q2327	Did they go to a health facility for treatment?	1. Yes 0. No
q2328	Why didn't they go to the health facility for treatment? ***(Do not read out the options. Select all that apply)***	1. No facility was open 2. Facility was not officially closed but (we had heard) there was no staff 3. We were afraid of catching COVID-19 (or other illness) at the health facility 4. Too expensive 5. Sickness was possible to manage at home 77. Others (please specify)
q2329	Have there been any deaths in your household since the lockdown due to COVID-19 pandemic? ***If the response is "Yes" then please say the mother, "I am really sorry to hear that."***	1. Yes 0. No

q2329a	Who was it?	1. Target child 2. Other child 3. Mother/Mother-in-law 4. Father/Father-in-law 5. Close relative 77. Other (specify)
q2330	What was the cause of death? ***(Free response. Do not read options. Select all that apply)***	1. COVID-19 2. Fever 3. Cold 4. Difficulty in breathing 5. Accident/Injury 6. Stroke 77. Others (please specify) 99. Don't know
q2331	Did death occur at a health facility?	1. Yes 0. No
****Section D: Household economic status****		
q2341	Did the employment type of any of your household members change since the shutdown/lockdown due to COVID-19 started?	1. Yes 0. No
q2342	What type of changes happened? ***(Do not read out the options. Select all that apply)***	1. No job/income source of the main earning member of the household 2. Earning from a temporary source by the main earning member of the household 3. Main earning member of the household become a day laborer 4. Other members of the household started earning from a temporary source 5. Les s work/decreased salary for household members 77. Others (please specify)
q2343	What is the status of your household income since April 2020?	1. No income 2. Some income, but less than previous income 3. Same as previously 4. Earning more than previous

q2344	In the past 7 days, did you or your household members use any of the following to cover your household's basic needs? ***Readout all options, Select all that apply***	1. Look for ways to earn additional money (e.g., work more hours, do an occasional job, etc.) 2. Reduce the number or size of meals for some household members 3. Rely on less preferred and less expensive foods 4. Use cash or bank savings 5. Sell assets 6. Borrow food or ask for help from a friend or relative or neighbor 7. Rely on Government or NGO assistance 8. Donations 9. Taken loan from someone else 77. Others (please specify) 0. None of the above
****Section E: Other impacts****		
q2351	Now I would like to ask, if any other experience related to COVID-19 you want to share with me?	

Table C.2: Comparison of cohort sample assessed vs. lost to follow-up

	Sample, n (%) or mean ± SD	
	Assessed (n=520)	**Lost (n=234)**
Caregiver Characteristics		
Completed primary education (6+ years)	69% (369)	62% (144)
Currently pregnant	1% (6)	1% (3)
Number of children <15 yrs under care	1.9 ± 0.8	1.9 ± 0.80
Muslim	98% (510)	97% (227)
CES-D score (0-60)	13.4 ± 8.7	14.2 ± 9.4
Mobility score (0-8)	3.5 ± 1.6	3.6 (1.73)
Child Characteristics		
Age (in months)	8.6 ± 1.8	8.5 ± 1.9
Female	51% (263)	52% (122)
FCI Play activities subscale (0-6)	2.5 ± 1.5	2.4 ± 1.6
FCI Play materials subscale (0-6)	1.2 ± 1.1	1.1 ± 1.0
1+ children's book(s) present in home	3.7% (19)	1.7% (4)
Household Characteristics		
Household size	5.2 ± 1.8	5.3 ± 1.9
Has cement floor	19% (98)	15% (35)
Has brick walls	23% (123)	20% (49)
Has refrigerator	22% (112)	18% (41)

Table C.3: Difference-in-differences analyses stratified by sum of impact across food security, economic and health domains

Sum of impact across food security, economic and health	**N cohort**	**DID estimate**	**p-value**
0	147	Ref	
1	187	0.38 (-1.40 to 2.17)	0.673
2	135	1.56 (-0.51 to 3.60)	0.14
3	51	5.41 (2.58 to 8.24)	<0.001
2+	186	2.61 (0.71 to 4.50)	0.007

Table C.4: Sensitivity analyses with median regression for CES-D score outcome

Exposure variable:	**More food insecure**	**Lost job and reduce income**	**Respondent or household members sick**
Median CES-D	1.85 (0.62 to 3.08)	1.33 (0.05 to 2.60)	1.98 (1.13 to 2.83)

Table C.5: Difference-in-differences estimates for caregiver freedom of movement score, not including health clinic in the cohort group who are more and less affected on the food security, economic, and health domains at follow-up

Exposure variable:	**More food insecure**	**Lost Job and Reduce Income**	**Respondent or household members sick**
Freedom of movement not including health clinic (0-6)	0.15 (-0.15 to 0.44) p=0.30	0.14 (-0.16 to 0.44) p=0.36	0.41 (0.12 to 0.71) p=0.006

Difference-in-differences estimates from a generalized estimating equation model accounting for repeated measures within participants, adjusted for village, child age category, maternal education, income, antenatal care, control over assets, household size, child sex, housing materials; Freedom of movement not including health clinic score is a sum score with one point for attending each of the following four places in the last 6 months, and an additional point if that location was attended alone: the market, medical facility, outside the village, paternal home or the home of a friend or a relative (0–6)

Table C.6: Mid-COVID differences in child-development risk factors by exposure status in the full sample (N=1696)

Exposure variable:	**More food insecure**	**Lost job or reduce income**	**Respondent or household members sick**
CESD-Score	3.48 (2.66 to 4.30)	2.03 (1.24 to 2.82)	2.75 (1.81 to 3.68)
Play activities	0.21 (0.05 to 0.37)	0.38 (0.24 to 0.51)	0.19 (0.04 to 0.33)
Play materials	0.11 (-0.00 to 0.22)	-0.04 (-0.14 to 0.06)	0.23 (0.09 to 0.37)
Freedom of movement	0.06 (-0.08 to 0.21)	0.17 (-0.02 to 0.36)	0.58 (0.41 to 0.74)

Estimates are mean differences (95% CIs) from a generalized linear model with standard errors clustered by village. The comparison group for each column is those who did not experience the stated exposure at the mid-COVID time point. Differences are adjusted for child age (2 month categories), maternal education (category), income (tertile), antenatal care (4+ vs < 4), control over assets (binary), household size, child sex, housing materials (floor and roof). Estimates represent mean differences between respondents more and less affected on each domain during the mid-COVID assessment.