Group-based parenting interventions to promote child development in rural Kenya: a multi-arm, cluster-randomised community effectiveness trial

Jill E Luoto, Italo Lopez Garcia, Frances E Aboud, Daisy R Singla, Lia C H Fernald, Helen O Pitchik, Uzaib Y Saya, Ronald Otieno, Edith Alu

Summary

Background Early childhood development (ECD) programmes can help address early disadvantages for the 43% of children younger than 5 years in low-income and middle-income countries who have compromised development. We aimed to test the effectiveness of two group-based delivery models for an integrated ECD responsive stimulation and nutrition education intervention using Kenya's network of community health volunteers.

Methods We implemented a multi-arm, cluster-randomised community effectiveness trial in three rural subcounties across 60 villages (clusters) in western Kenya. Eligible participants were mothers or female primary caregivers aged 15 years or older with children aged 6–24 months at enrolment. If married or in established relationships, fathers or male caregivers aged 18 years or older were also eligible. Villages were randomly assigned (1:1:1) to one of three groups: group-only delivery with 16 fortnightly sessions; mixed delivery combining 12 group sessions with four home visits; and a comparison group. Villages in the intervention groups were randomly assigned (1:1) to invite or not invite fathers and male caregivers to participate. Households were surveyed at baseline and immediately post-intervention. Assessors were masked. Primary outcomes were child cognitive and language development (score on the Bayley Scales of Infant Development third edition), socioemotional development (score on the Wolke scale), and parental stimulation (Home Observation for Measurement of the Environment inventory). Analysis was by intention to treat. This trial is registered with ClinicalTrials.gov, NCT03548558.

Findings Between Oct 1 and Nov 12, 2018, 1152 mother–child dyads were enrolled and randomly assigned (n=376 group-only intervention, n=400 mixed-delivery intervention, n=376 comparison group). At the 11-month endline survey (Aug 5–Oct 31, 2019), 1070 households were assessed for the primary outcomes (n=346 group only, n=373 mixed delivery, n=351 comparison). Children in group-only villages had higher cognitive (effect size 0·52 SD [95% CI 0·21–0·83]), receptive language (0·42 SD [0·08–0·77]), and socioemotional scores (0·23 SD [0·03–0·44]) than children in comparison villages at endline. Children in mixed-delivery villages had higher cognitive (0·34 SD [0·05–0·62]) and socioemotional scores (0·22 SD [0·05–0·38]) than children in comparison villages; there was no difference in language scores. Parental stimulation also improved for group-only (0·80 SD [0·49–1·11]) and mixed-delivery villages (0·77 SD [0·49–1·05]) compared with the villages in the comparison group. Including fathers in the intervention had no measurable effect on any of the primary outcomes.

Interpretation Parenting interventions delivered by trained community health volunteers in mother–child groups can effectively promote child development in low-resource settings and have great potential for scalability.

Funding Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health.

Lancet Glob Health 2021; 9: e309–19

Published Online December 17, 2020 https://doi.org/10.1016/S2214-109X(20)30469-1

See Comment page e229

RAND Corporation, Santa Monica, CA, USA (J E Luoto PhD, I Lopez Garcia PhD, U Y Saya MSc); **Department of Psychology, McGill University, Montreal, QC, Canada** (Prof F E Aboud PhD); **Department of Psychiatry, Sinai Health and University of Toronto, Toronto, ON, Canada** (D R Singla PhD); **Division of Community Health Sciences** (Prof L C H Fernald PhD) **and Division of Epidemiology** (H O Pitchik MSc), **School of Public Health, University of California, Berkeley, Berkeley, CA, USA; Safe Water and AIDS Project, Kisumu, Kenya** (R Otieno, E Alu)

Correspondence to: Dr Jill E Luoto, RAND Corporation, Santa Monica, CA 90401, USA **jluoto@rand.org**

Introduction

Approximately 250 million (43%) children younger than 5 years in low-income and middle-income countries (LMICs) will not reach their full developmental potential due to poverty, nutritional deficiencies, or inadequate psychosocial stimulation.[1] Early childhood development (ECD) interventions that integrate responsive parenting and early learning with nutrition education have effectively improved short-term child development outcomes in numerous LMIC settings.[1–3] Early childhood can be a cost-effective period to address developmental outcomes, because early investments have the potential to improve adult human capital.[4] Despite increasing evidence for the effectiveness of ECD interventions to improve short-term child developmental outcomes, key questions remain about how to make these programmes scalable,[5] particularly in rural, low-income settings.[2,6]

Two primary methods for delivering ECD interventions are individual home visits and group-based meetings in a primary care or community setting.[3] Individual home visits can offer personalised feedback, support, and problem solving to overcome personal and family

Research in context

Evidence before this study

We reviewed two systematic reviews published in 2015 and 2017 and searched the Global Health (Ovid) database on Dec 31, 2019, for publications since Jan 1, 2015. We focused on studies done in low-income and middle-income countries (LMICs) with children younger than 3 years and that featured psychosocial stimulation as a primary component. Our search terms included: "child development", "responsive care", "responsive play", "parenting behavior", "parenting", "psychosocial stimulation", "early childhood", and "intervention". We found eight studies in addition to the 40 previously included in the meta-analyses, which overall show effect sizes for cognitive and language development in the order of 0·40. Of the most recent eight studies, five yielded significant child outcomes, and three were non-significant. In sum, although effect sizes for psychosocial stimulation parenting programmes are in the moderate range, there is still the question of why some are more effective than others. Recommended features of successful programmes included having a structured curriculum, with opportunities for parental practice and feedback on the new behaviours, as well as sufficient frequency of sessions. However, most ECD interventions in LMIC settings feature a home visiting model of delivery, which can be prohibitively expensive to scale. Studies of ECD programmes that feature a group-based delivery model have been predominantly efficacy studies and often feature smaller group sizes (two to eight mothers), with supplemental home visits that sometimes outnumber the group sessions, drawing into question the effectiveness of group-only delivery and larger group sessions. Although some studies allowed fathers to participate in the interventions, none had explicitly tested the effects of including fathers.

Added value of this study

Our study directly compared a purely group-based model versus a mixed-delivery model that combines home visits with group sessions in a community effectiveness trial in a rural LMIC setting. We also tested the effectiveness of a group-based ECD intervention using community health workers in an African country and attempted to rigorously test the added value of inviting fathers into the intervention. We tested the effectiveness of both delivery models and found that group visits alone were sufficient to improve child cognitive, receptive language, and socioemotional outcomes by magnitudes similar to or larger than previous efficacy studies. Our results show that a group-based intervention is at least as effective as a mixed-delivery model that combines home visits, but is less burdensome to deliver. Our use of existing paraprofessional community-based health workers as delivery agents and relatively large group sizes (median attendance of 13 mothers) makes our model potentially scalable to other settings in which potential demand vastly outpaces supply of these programmes.

Implications of all the available evidence

ECD interventions can be delivered effectively in large village group settings by existing paraprofessional community agents without loss of effects compared with a more personalised (and thus more expensive and time-consuming) delivery model featuring home visits. The added benefits of involving fathers in these programmes is unclear and more research is needed to understand how to successfully engage fathers into ECD programmes in low-resource settings.

barriers to behaviour change, but can be expensive to implement at scale in LMIC settings.[7] Group-based models offer potential economies of scale, might modify group norms for child-care, and can provide mothers with increased peer support.[8] However, group meetings might be comparatively weak in overcoming personal barriers to behaviour change.[9] A mixed-delivery model that combines group sessions with a small number of personalised home visits has been hypothesised to balance the cost savings and peer support of groups with the benefits of personalised attention and feedback from home visits, but the added value of home visits to a purely group-based intervention is unknown.[9] ECD programmes have primarily focused on mothers and children, although involving fathers has also been suggested to be beneficial for child and family wellbeing.[10]

The aim of this study was to assess the effectiveness of two potentially scalable models of delivery for an integrated parenting intervention to improve child developmental outcomes among families with young children in rural Kenya. Because individual home visits in dispersed rural settings such as those found in Kenya would be prohibitively time-consuming and expensive for scaling, we tested a model based on group meetings versus a mixed-delivery model that combined a small number of home visits with group meetings. We also aimed to assess the added value of explicitly inviting fathers into the intervention across both delivery models.

Methods

Study design and participants

In collaboration with a Kenyan non-governmental organisation, Safe Water and AIDS Project (SWAP), we did a multi-arm, cluster-randomised community effectiveness trial in the subcounties of East Rachuonyo, South Rachuonyo, and Sabatia in western Kenya. These predominantly rural areas are characterised by high rates of poverty, child mortality, and stunting (31–34%).[11] Sabatia's population is predominantly from the Luhya tribe and speaks Luhya and Swahili. The populations of South Rachuonyo and East Rachuonyo are predominantly Luo and speak Luo. No other ECD programmes or

interventions existed in these areas at the time of the study. The majority of villagers are subsistence farmers or unskilled informal workers.

Eligible participants within villages were mothers or other female primary caregivers aged 15 years or older with a child aged 6–24 months without signs of severe mental or physical impairment. If married or in established relationships, fathers or male caregivers aged 18 years or older were also eligible to participate. We identified eligible mother–child dyads using a census run by SWAP immediately before baseline data collection. All participants (female and male caregivers) provided written informed consent at the time of data collection before randomisation. Ethics approval was obtained from the ethics committee at Maseno University in Kisumu, Kenya, and the RAND Corporation. The study protocol has been published elsewhere.[12]

Randomisation and masking

Villages comprised clusters and were the unit of randomisation to minimise the risk of contamination within villages, and to align with the territory of community health volunteers.

We first listed all villages in the three subcounties estimated to have at least 20 households with eligible children; next, using a computer-generated random number in Stata and stratifying by subcounty, 60 villages were randomly sampled for inclusion as long as they maintained a minimum distance (1·5–2·5 km) from all other sampled villages. A team of trained enumerators did a census within the villages and a baseline survey on the final sample of eligible households. Using a random number generator in Stata, two authors (JEL and ILG) randomly assigned villages in a 1:1:1 ratio to one of three study groups, stratified by subcounty. One group featured a group-only delivery model with fortnightly sessions; the second group featured a mixed-delivery model combining 12 group sessions with four home visits; and the third group served as a comparison group and received no intervention.

Among the villages assigned to one of the two intervention groups, half were randomly assigned (1:1) to invite only mothers and children (mothers-only group); the remaining villages also invited fathers to the 16 sessions (father group). The geographical location of the sampled villages and their randomisation status are shown in the appendix (p 3).

Due to the nature of the intervention, masking of participants and community health volunteers who delivered the intervention was not possible. Data collection teams worked independently from the community health volunteers, and all efforts were made to keep them as masked to group status as possible. Data analysis was blinded. Interviews of enumerators at the end of data collection showed that they were unaware of details about the intervention and treatment allocation. For quality assurance, subcounty supervisors randomly supervised 20% of child and mother interviews.

Procedures

Each village had a predesignated community health volunteer who was invited to deliver the programme in intervention villages; none refused. Community health volunteers are part-time volunteers and members of their communities tasked with improving community health and linking individuals to primary health-care services. The research project paid community health volunteers a monthly stipend for their duties, according to local policy.

SWAP introduced the project to village leaders, secured local approvals, and managed local implementation. Community health volunteers delivered 16 fortnightly sessions to both intervention groups between mid-November, 2018, and mid-July, 2019. The responsive stimulation and nutrition education intervention was named Msingi Bora (Good Foundation in Swahili). It was based on a structured curriculum adapted from previous successful parenting trials in LMICs and expanded to include more activities around responsive play and talk with children.[8,13] Six sessions were piloted in April–June, 2018, in six villages not included in the main trial. The finalised curriculum included session-specific activities and materials, with Luo or Swahili and English manuals for each community health volunteer. The Msingi Bora curriculum focused on five key practices: responsive play, responsive communication, hygiene, nutrition, and love and respect in the family. The sessions emphasised parents learning new practices with their child, spouse, and peers through demonstration and coached practice, group-based problem solving, and peer support. Sessions took place in local community centres or churches. More details of the intervention are in the appendix (p 1).

Mothers and children were invited to attend all 16 sessions, and received a small gift for attendance (eg, small bar of soap [US$0·15]). Every fourth session served as a review session, for which households receiving the group-only intervention continued with group meetings and households receiving the mixed-delivery intervention received individual home visits (appendix p 4). Community health volunteers in mixed-delivery villages visited each participant household during the same week that a group review session was held in group-only villages. During these home visits, the community health volunteers delivered review messages identical to those in the group reviews, but the focus was tailored to that family. In the villages where fathers were invited to participate in the intervention, they were invited to all 16 sessions; 12 of which were for both mothers and fathers, and four of which were separate sessions by sex, including the first two, as a way to try to encourage their participation (appendix p 4). The father-only sessions emphasised topics such as practising respectful

See Online for appendix

communication, father involvement in child-care, and emotional support between spouses. Similar topics were covered in the four corresponding mother–child sessions so that the curriculum was identical across mothers, regardless of the intervention group. Households in villages assigned to the comparison group did not receive any interventions other than information about child feeding during the baseline survey.

Community health volunteers in villages assigned to an intervention group received 8 days of intensive training in November, 2018, covering sessions 1–8, and another 8 days in April, 2019, covering sessions 9–16. Monthly 1-day refresher trainings were done in each subcounty for that month's sessions. The 40 community health volunteers who were trained included ten men and 30 women, with a mean age of 44 years (SD 11·02; range 26–69), 11 years of education (minimum 7 years; SD 1·74), and 9 years of community health volunteer experience (SD 5·98; range 1–24). The initial training was led by FEA, ILG, JEL, HOP, and DRS; all subsequent trainings were led by SWAP staff in a train-the-trainers model.

For more on **SurveyCTO** see https://www.surveycto.com

Sessions were monitored by trained SWAP supervisors who rated community health volunteers on skills such as facilitating discussion, coaching parents, answering questions, as well as overall session quality and engagement. Community health volunteers were provided with supervisor feedback immediately after each session. Attendance sheets were also completed at each session. Qualitative interviews of community health volunteers (n=11), participant mothers (n=8), fathers (n=3), and SWAP supervisory staff (n=6), selected to cover the districts equally, were done at the end of the intervention and used to assist in the interpretation of quantitative data.

15 local enumerators with a minimum of 2 years' post-secondary education were recruited and given 8 days of intensive training on the assessment measures before baseline in September, 2018. The strongest three were promoted to a supervisory role, and the other 12 collected data, four in each subcounty. For the endline survey, six of the 12 were retrained for 15 days on child assessments. The other six enumerators were retrained for four days on the mother and father measures. All questionnaires and child assessments were administered in Luo or Swahili as appropriate, and were translated and back-translated using standard methods.

Child language and cognition were assessed using the Bayley Scales of Infant Development third edition,[14] a commonly used direct child assessment previously adapted and validated in many African countries.[15–17] We adopted previous adaptations to the Bayley to make it appropriate for our context, including by modifying the picture and stimulus booklets with attire and objects that would be familiar to rural Kenyan children.[8] Socioemotional development was assessed using an adapted Wolke scale[18] that required observational ratings on a 1–5 scale of children's behaviour during testing along seven dimensions: approach, emotional tone, gross motor activity, cooperation, vocalisation, emotional security, and exploration. Parental stimulation practices were assessed at baseline with the Family Care Indicators,[19] a caregiver-reported measure of 12 stimulating materials and activities, and at endline with the Home Observation for Measurement of the Environment (HOME) inventory, a 45-item measure using mother-report and observation.[20] We found good concurrent validity of the raw Bayley cognitive and receptive language scores with age (cognitive scores r=0·842, receptive language scores r=0·768; p<0·0001) and with HOME scores (cognitive scores r=0·235, receptive language scores r=0·204; p<0·0001). Inter-rater reliabilities during field testing were high to moderate (cognition n=46, κ=0·81, p<0·0001; receptive language n=43, κ=0·88, p<0·0001; expressive language n=48, κ=0·71, p<0·0001; socioemotional n=49, κ=0·49, p=0·05; HOME κ=0·59, measured with Cohen and Conger's κ).

Data were collected using SurveyCTO on Android tablets for all outcomes.

Outcomes

The primary outcomes were child development (standardised Bayley scores for receptive and expressive language and cognition, and raw Wolke score for socioemotional development) and parent stimulation (HOME) measured at endline. Secondary outcomes, also measured at endline, were child stunting based on length-for-age, measured using the standard WHO Multicenter Growth Reference Study;[21] child dietary diversity, measured using a 0–7 scale on which parents report the categories of food eaten by the child in the past 24 h following WHO recommendations for child feeding;[22] and self-reported household food security using the Household Hunger Scale.[23] Prespecified exploratory outcomes measured at endline included self-reported measures of maternal mental health, knowledge, fathers' scores on the six behavioural items of the Family Care Indicators, and other potential mediators of behavioural change and are listed in full in the appendix (p 5).

Statistical analysis

Our sample size was calculated on the basis of the primary outcome of child cognitive development assessed with the Bayley score, which has a usual mean of 10 (SD 1·5).[14] We assumed 75% session attendance among households invited to the intervention, 15% attrition, and an intraclass correlation coefficient of 0·07 within villages. In a side-by-side comparison between any two study groups, a sample of 400 children per group was estimated to provide 80% power to detect an increase in Bayley score of 0·30 SD. All analyses clustered SEs at the level of villages, which is the unit of randomisation. Significance was defined as p<0·05.

Analyses were done with Bayley scaled scores, as well as Bayley raw scores that were internally age-standardised in 2-month bands relative to the comparison group.

This approach was adopted because the official scaled scores show substantial age gradients in our sample, as is common in LMICs.[24] For ease of interpretation, we adopted the same standardisation procedure in all child and parent measures.

Analyses of all outcomes were by intention to treat among the final endline sample. Random assignment of villages to each study group allowed us to estimate intention-to-treat effects for all outcomes by doing a series of pairwise comparisons across intervention groups. All reported results include pre-planned adjustments for child sex, birth order, maternal education, household wealth, corresponding baseline outcomes (if available), and sub-county strata to account for sampling design.[12] Adjustments used multivariate linear regressions for continuous outcomes and logistic regression for stunting, which is binary. To account for imperfect attendance of parents at sessions, we also estimated the average treatment effect on the treated using a two-stage least squares regression approach and the recorded attendance data as our measure of compliance. We corrected for multiple hypothesis testing using the Romano-Wolf estimator.[25]

To understand whether our interventions are more effective for some households than others, we tested heterogeneous intervention effects in prespecified analyses by children's age, sex, and baseline assessments, as well as maternal education. We did a mediation analysis using structural equation modelling to explore the role of induced parental behavioural change in observed effects on child outcomes.[26] All analyses were done using Stata version 16. This trial is registered at ClinicalTrials.gov, NCT03548558.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

During August, 2018, 1265 mother–child dyads were identified in the census across the 60 included villages, and 1152 (91%) dyads were successfully enrolled into the

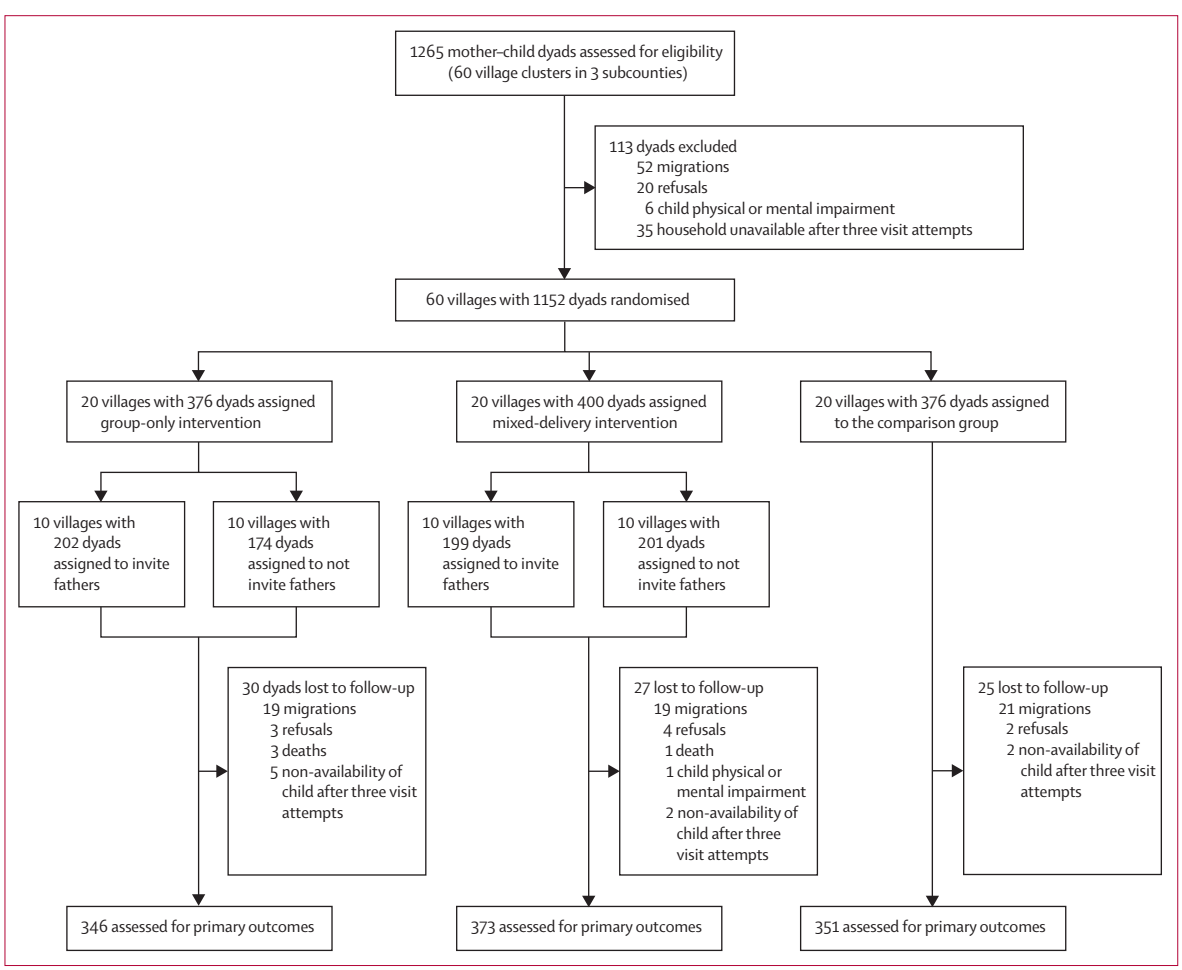

Figure 1: **Trial profile**

	Comparison group (n=376)	Group-only model (n=376)	Mixed-delivery model (n=400)	Fathers not invited (n=374)	Fathers invited (n=402)
Household characteristics					
Wealth index from assets					
Quintile 1	70 (19%)	74 (20%)	87 (22%)	86 (23%)	75 (19%)
Quintile 2	86 (23%)	69 (18%)	75 (19%)	74 (20%)	70 (17%)
Quintile 3	83 (22%)	78 (21%)	70 (17%)	65 (17%)	83 (21%)
Quintile 4	71 (19%)	73 (19%)	86 (21%)	67 (18%)	92 (23%)
Quintile 5	66 (18%)	82 (22%)	82 (20%)	82 (22%)	82 (20%)
Father in household	229 (61%)	228 (61%)	256 (64%)	210 (56%)	274 (68%)
Household size	5·6 (1·9)	5·4 (2·2)	5·8 (2·8)	5·6 (2·0)	5·7 (2·9)
Mother's education, years	8·9 (2·6)	8·8 (2·8)	8·8 (2·8)	8·8 (2·7)	8·8 (2·9)
Mother's age, years	29·0 (9·2)	27·8 (9·7)	28·0 (8·5)	28·6 (9·1)	27·3 (9·0)
Father's education, years	9·4 (2·9)	9·7 (3·0)	9·4 (2·9)	9·3 (3·0)	9·7 (2·9)
Child characteristics					
Child's age at recruitment, months	14·2 (4·7)	13·8 (5·0)	14·4 (4·7)	13·9 (5·0)	14·3 (4·7)
Girls	192 (51%)	178 (47%)	203 (51%)	178 (48%)	203 (50%)
Birth order	3·1 (1·7)	2·9 (1·7)	2·9 (1·7)	2·9 (1·7)	2·9 (1·7)
Child stunting status	44 (12%)	44 (12%)	58 (15%)	52 (14%)	50 (12%)
Child unwell in past 2 weeks	114 (30%)	117 (31%)	149 (37%)	134 (36%)	132 (33%)
Child Bayley scale score					
Cognitive score, 0–19	9·5 (2·3)	9·3 (2·2)	9·5 (2·3)	9·1 (2·2)	9·7 (2·2)
Receptive language score, 0–19	9·7 (2·3)	9·3 (2·2)	9·6 (2·1)	9·2 (1·9)	9·7 (2·3)
Parental measures					
Maternal Family Care Indicator score, 0–12	5·0 (2·0)	4·7 (2·1)	4·6 (1·9)	4·6 (2·0)	4·8 (2·0)
Father's Family Care Indicator behavioural score, 0–6*	3·1 (1·5)	3·2 (1·4)	3·3 (1·4)	3·2 (1·5)	3·4 (1·4)
Child dietary diversity, 0–7	3·2 (1·2)	3·1 (1·2)	3·0 (1·1)	3·1 (1·2)	3·1 (1·2)
Maternal depression, 0–40†	10·2 (7·1)	10·3 (6·1)	9·8 (7·1)	11·0 (6·9)	9·2 (6·3)
Mother's relationship support scale, 0–30	25·2 (3·5)	25·0 (3·7)	25·3 (3·4)	24·9 (3·7)	25·4 (3·4)
Mother's social support, 0–20	10·0 (3·1)	9·5 (3·0)	9·9 (2·8)	9·5 (3·0)	10·0 (2·8)
Mother's knowledge of infant development, 0–12	4·4 (2·1)	4·8 (2·0)	4·7 (1·9)	4·9 (2·1)	4·6 (1·8)

Data are n (%) or mean (SD). SDs are clustered at the village level. Bayley scores are scaled scores age-standardised (range 0–19) using the publisher's manual. *Family Care Indicator scores for fathers at baseline have n=162 for the comparison group, n=160 for the group-only group, n=190 for the mixed-delivery group, n=193 for father villages, and n=157 for mother-only villages. †Center for Epidemiologic Studies Depression Scale scores at baseline are on a 0–40 scale due to an error in scoring at baseline and are not comparable with scores at endline or to outside samples. Child length-for-age at baseline used Seca mobile measuring mats (model 210; Seca; Hamburg, Germany); enumerators measured the child three times and calculated the mean; all measures were converted to length-for-age Z scores following WHO recommendations and stunting was defined as <2 SD below the mean.

Table 1: **Baseline characteristics of the intention-to-treat population**

study between Oct 1 and Nov 12, 2018. 20 villages with 376 enrolled dyads were randomly assigned to the group-only intervention, 20 villages with 400 enrolled dyads were randomly assigned to the mixed-delivery intervention, and 20 villages with 376 enrolled dyads were randomly assigned to the comparison group (figure 1). Each village had a mean of 19 children enrolled (SD 2·74; range 14–26). The endline survey was done approximately 11 months after enrolment (between Aug 5 and Oct 31, 2019) when children were aged 16–34 months. 1070 (93%) of the enrolled children were followed up and included in the primary analysis (figure 1). There were no differences in attrition across study groups (30 [8%] in the group-only group, 27 [7%] in the mixed-delivery group, and 25 [7%] in the comparison group; p value of joint F-test 0·76). Households lost to follow-up are similar to retained households in sociodemographic characteristics, but attrition was more likely in the two Luo subcounties (appendix p 6).

Child and household characteristics, as well as all child and parental measures, were balanced across the groups at baseline, with the exception that villages inviting fathers had higher rates of fathers being present (table 1). Rates of stunting were low at baseline, probably due to the young mean age of children at enrolment (14 months; SD 4·79).

Monitoring data on session implementation showed that the intervention was delivered as intended, with 16 fortnightly sessions per intervention village. 496 (89%) of 560 mother group sessions were monitored by a SWAP supervisor, and 122 (9%) of 1411 home visits

	Group-only model (n=346)	Mixed-delivery model (n=373)	Comparison group (n=351)	Group-only model *vs* comparison group		Mixed-delivery model *vs* comparison group		Group-only model *vs* mixed-delivery model
				Effect size (95% CI)*	p value	Effect size (95% CI)*	p value	p value
Child primary outcomes								
Standardised Bayley cognitive score, 0–19†	9·45 (1·73)	9·10 (1·44)	8·74 (1·48)	0·52 (0·21 to 0·83)	0·0013	0·34 (0·05 to 0·62)	0·021	0·20
Standardised Bayley receptive language score, 0–19†	10·56 (2·23)	10·12 (1·78)	9·75 (1·75)	0·42 (0·08 to 0·77)	0·017	0·20 (–0·11 to 0·52)	0·20	0·17
Standardised Bayley expressive language score, 0–19†	8·92 (1·71)	8·64 (1·61)	8·85 (1·74)	0·08 (–0·19 to 0·34)	0·56	–0·09 (–0·33 to 0·15)	0·45	0·19
Raw Wolke socioemotional score, 7–35	25·69 (5·42)	25·77 (5·19)	24·80 (4·94)	0·23 (0·03 to 0·44)	0·024	0·22 (0·05 to 0·38)	0·011	0·85
Child secondary outcome								
Child stunting status‡	80 (24%)	107 (29%)	77 (22%)	0·02 (–0·06 to 0·09)	0·68	0·06 (–0·02 to 0·13)	0·13	0·18

Data are mean (SD) or n (%) except where otherwise stated. Results based on a final sample of n=1070 at endline with the exception of child stunting, for which n=334 group-only model, n=364 mixed-delivery model, and n=343 comparison group. *Effect sizes and p values obtained from regression-adjusted intention-to-treat estimates for each outcome measured at endline using internal age-standardisation to the comparison group. Adjustments were prespecified in our study protocol and include child's age, household wealth, maternal education, child sex, birth order, the corresponding outcomes at baseline (if measured), and subcounty fixed effects (the strata). Unadjusted estimates are statistically equal to the adjusted results presented. Standard errors are clustered at the village level. Robust 95% CIs in parentheses. All statistically significant effects tested individually are also robust to multiple hypotheses testing at the 5% level using the Romano-Wolf estimator. †Bayley scaled scores (age-standardised). ‡Effect sizes cited are marginal effects based on logistic regression-adjusted estimates for a binary outcome.

Table 2: **Child primary and secondary outcomes in the intention-to-treat population**

were supervised. The median number of sessions attended was 13 (IQR 8–15), with a median group size of 13 (10–15) mothers at a group meeting. Attendance was higher for the mixed-delivery intervention (mean attendance of 11·9 [74%] of 16 sessions) than for the group-only intervention (10·3 [64%] of 16 sessions), with the bulk of this difference stemming from higher attendance in the four review sessions, which were home visits for the mixed-delivery group (mean attendance of 3·53 [88%] of 4 home visit review sessions *vs* mean attendance of 2·58 [65%] of 4 group review sessions, p<0·0001). Attendance among fathers was substantially lower overall than among mothers: fathers attended a median of one of 16 sessions (IQR 0–3) and 168 (52%) of 328 fathers ever present in the household attended at least one session. Sessions lasted a median of 90 min (IQR 70–110). Community health volunteers were reviewed by supervisors to be "very good" or "excellent" in how well they coached parents for 325 (67%) of 487 sessions, and in how well they facilitated discussions for 296 (60%) of 496 sessions.

Children in group-only villages showed significant improvements in cognitive, receptive language, and socioemotional scores compared with the comparison group (table 2). Children in mixed-delivery villages showed smaller but still significant improvements in cognitive and socioemotional scores, but not in receptive language scores. There was no improvement in expressive language in either intervention group.

Mothers in both group-only and mixed-delivery villages showed significantly better HOME scores at endline than mothers in the comparison group (table 3). We observed significant improvements in the secondary outcome of child's dietary diversity in group-only villages compared with the comparison group, but not in mixed-delivery villages. There was no improvement in child stunting status in either intervention group compared with the comparison group (table 2).

Improvements were seen in mediator measures of maternal knowledge of child development and maternal recall of intervention messages under both intervention models compared with the comparison group. The intervention did not significantly improve measures of maternal wellbeing under either delivery model (table 3). Inviting fathers to the sessions did not have an effect on child and maternal outcomes (appendix p 7).

All significant results remained significant after accounting for multiple hypothesis testing at the 5% level or above. Adjusting for covariates did not affect the magnitude or statistical significance of estimated effects, but increased precision. Results were not sensitive to the internal standardisation.

Higher levels of compliance with the intervention resulted in greater cognitive gains under both delivery models, with signs of increasing returns by number of sessions attended in analyses of the effect of treatment on the treated (figure 2). For example, in the group-only intervention group, increasing from attending at least one to at least 12 sessions improved cognition by 0·36 SD (95% CI 0·54–0·90). Similarly positive dose–response relationships were observed for all other outcomes with significant effects in the intention-to-treat analyses (appendix p 8). When we focus on attendance at only the four review sessions that differed between the two intervention groups (mixed-delivery villages had home visits for these sessions), the added benefits for child cognition of attending each additional review session increased steeply for groups, but not for home visits (appendix p 9). For example, children who attended all four group review sessions had cognitive

	Group-only model (n=346)	Mixed-delivery model (n=373)	Comparison group (n=351)	Group-only model *vs* comparison group		Mixed-delivery model *vs* comparison group		Group-only model *vs* mixed-delivery model
				Effect size (95% CI)*	p value	Effect size (95% CI)*	p value	p value
Parent primary outcome								
HOME score, 0–45	30·59 (5·55)	30·45 (5·53)	27·48 (4·19)	0·80 (0·49 to 1·11)	<0·0001	0·77 (0·49 to 1·05)	<0·0001	0·83
Other parental behaviours (secondary outcomes)								
Dietary diversity, 0–7	4·28 (1·12)	4·09 (1·11)	4·06 (1·17)	0·23 (0·01 to 0·45)	0·038	0·04 (–0·17 to 0·25)	0·70	0·044
Household hunger scale score, 0–6	0·48 (0·77)	0·50 (0·79)	0·61 (0·85)	–0·16 (–0·32 to 0·01)	0·064	–0·10 (–0·27 to 0·06)	0·21	0·52
Parental wellbeing and knowledge (exploratory outcomes)								
Father's Family Care Indicator score, 0–6	4·27 (1·69)	4·33 (1·72)	4·20 (1·70)	0·08 (–0·10 to 0·26)	0·38	0·03 (–0·15 to 0·22)	0·72	0·65
Depressive symptoms, 0–60	14·41 (8·67)	15·62 (9·28)	14·74 (9·28)	–0·01 (–0·22 to 0·20)	0·90	0·18 (–0·01 to 0·37)	0·067	0·092
Stress, 0–105	15·98 (12·41)	15·23 (12·60)	14·98 (13·35)	0·10 (–0·11 to 0·30)	0·35	0·07 (–0·12 to 0·26)	0·45	0·78
Self-efficacy, 13–65	39·95 (3·98)	40·33 (3·97)	39·84 (4·15)	0·05 (–0·15 to 0·25)	0·59	0·11 (–0·14 to 0·35)	0·39	0·65
Relationship support scale, 0–30	25·58 (4·32)	25·32 (4·45)	25·24 (4·83)	0·21 (–0·01 to 0·42)	0·057	–0·01 (–0·24 to 0·23)	0·94	0·042
Social support, 0–18	10·12 (2·00)	10·12 (2·30)	10·21 (2·17)	0·07 (–0·09 to 0·23)	0·37	–0·00 (–0·20 to 0·20)	1·000	0·47
Beliefs about childrearing, 0–20	13·51 (3·91)	13·85 (3·79)	12·95 (3·71)	0·11 (–0·11 to 0·34)	0·32	0·25 (0·01 to 0·48)	0·039	0·21
Knowledge of infant and child development, 0–48	23·13 (6·72)	22·93 (5·85)	21·84 (6·64)	0·22 (0·06 to 0·37)	0·0062	0·16 (0·01 to 0·30)	0·032	0·31
Recall intervention messages, 0–15	6·77 (4·01)	6·88 (3·75)	2·40 (3·21)	1·50 (1·12 to 1·87)	<0·0001	1·46 (1·14 to 1·79)	<0·0001	0·83

Results based on a final sample of N=1070 at endline, except for father's Family Care Indicator score with N=789 (group-only model n=270; mixed-delivery model n=268; comparison group n=251). HOME=Home Observation for Measurement of the Environment. *Effect sizes and p values obtained from regression-adjusted intention-to-treat estimates for a given outcome measured at endline using internal age-standardisation to the comparison group. Adjustments were prespecified in our study protocol and include child's age, household wealth, maternal education, child sex, birth order, the outcome at baseline (if measured), and subcounty fixed effects (the strata). Unadjusted estimates are statistically equal to adjusted results presented. Standard errors are clustered at the village level. Robust 95% CIs in parentheses. All statistically significant effects tested individually are also robust to multiple hypotheses testing at the 5% level using the Romano-Wolf estimator.

***Table 3*: Parental primary, secondary, and exploratory (mediator) outcomes at endline**

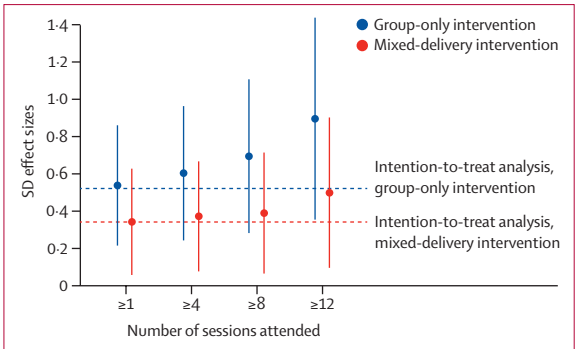

***Figure 2*: Effects of treatment on the treated for child cognition**
The figure shows mean estimated effects of treatment on the treated with 95% CIs under both delivery models. Compliance to the intervention is defined based on the number (out of 16) of sessions attended. CIs get larger as the number of compliers decreases. For comparison, the estimated effects in the intention-to-treat population under both delivery models are included using horizontal dashed lines (data shown in table 2).

scores 1·26 SD (95% CI 0·48–2·05) higher than children in the comparison group, whereas children who attended all four home visit review sessions had 0·41 SD (0·07–0·74) higher cognitive scores (p=0·044 on the difference between group and home visit reviews). Monitoring data show that supervisors generally rated group review sessions higher than home visits in many aspects of delivery quality (appendix p 10). In qualitative interviews, mothers commented that both they and their children liked interacting with others in their group, benefited from questions and problems raised by other mothers, and learned how to play games and talk with their child by watching others.

We found no evidence of differential effects in prespecified subgroup analyses by maternal education, child age, or baseline outcomes (appendix pp 11–12), although there is suggestive evidence of greater effects on primary outcomes for children with less educated mothers. Effects on child outcomes and parental stimulation practices were systematically larger and occasionally statistically significant for girls versus boys (appendix p 12). A mediation analysis using structural equation modelling showed that stimulation, dietary diversity, maternal knowledge, and recall of intervention messages together explain up to 33% of the effects on cognition and 58% of the effects on receptive language in the group-only intervention group (appendix p 13).

Discussion

In this cluster-randomised trial, we tested the effectiveness of two potentially scalable group-based delivery models for an early childhood development intervention in rural Kenya. We found that both the group-only and mixed-delivery models significantly improved child cognitive and socioemotional development, as well as maternal stimulation practices. The group-only model also showed significant improvements in child receptive

language, although the mixed-delivery model did not. Expressive language was not significantly improved in either intervention group. Effect sizes for the group-only model were generally larger than for the mixed-delivery model, although the difference was statistically significant only for dietary diversity. However, we can nearly reject our initial hypothesis that the mixed-delivery intervention would provide larger benefits than the group-only intervention for cognition (p=0·099) and receptive language (p=0·084) based on one-sided tests derived from the two-sided p values in table 2. We interpret this as suggestive evidence that the group-only model was more effective in our setting.

Children who attended more sessions showed greater benefits from both delivery models. Maternal stimulation practices similarly improved with attendance. These positive dose–response relationships have important implications for the optimal design and intensity of ECD programmes. Tests of heterogeneity also suggest that our intervention might have greater benefits for girls than for boys.

Effect sizes from this community-based effectiveness trial are similar to or larger than in previous studies using either intensive home visits, group interventions, or a combination thereof,[9,24] yet are generally smaller than two effectiveness trials from Asia (from 2014 and 2019).[13,27] However, those two studies featured more intensive interventions over a period of 1–2 years instead of the 8 months of our programme, and differed in other crucial ways as well. For example, the 2-year trial in Pakistan had much more emphasis on home visits, and the 25-session, year-long trial in Bangladesh worked with small groups of two mothers and targeted children who were stunted at baseline (for whom ECD interventions can have particularly large effects).[28] Our results show that an exclusively group-based ECD intervention for all age-eligible children in an LMIC (who are not receiving conditional cash transfer benefits) can have positive effects on cognitive, language, and socio-emotional development. Moreover, our delivery format of relatively large groups of women makes our results particularly promising to inform ECD policy on the implementation of scalable and potentially cost-effective programmes in similar contexts.[5] Our use of existing community health workers to deliver an intervention that improves child development in a rural low-income setting directly addresses the top two questions (whether child development packages focusing on nurturing care and parent support can improve child cognitive development in rural low-income settings and whether community health workers or paraprofessionals can be trained to deliver ECD interventions effectively) generated by a panel of worldwide ECD experts in a recent ranking of priority questions for the field.[6]

Our intervention brought about improvements in parental stimulation practices under both delivery models. Maternal knowledge of child development also increased, yet our intervention did not improve measures of maternal wellbeing, contrary to our expectations and programme emphasis on topics related to maternal mental health. Both stimulation and maternal knowledge might have affected how the intervention improved child outcomes. However, our preliminary mediation analysis does not correct for the endogenous nature of parental investments nor potential measurement error in outcomes, which future work will do. If parents continue to practise the newly learned behaviours at home and responsively adapt them as children age, the effects of our intervention might be sustainable. Yet we recognise that this is a central challenge of ECD interventions worldwide, in which immediate effects have dissipated over time.[29]

Contrary to our hypotheses, group meetings consistently outperformed a mixed-delivery model in key outcomes such as child cognition and receptive language, although differences were generally not significant with two-sided tests. Nonetheless, this is important for considerations of cost-effectiveness because group sessions were also less burdensome to deliver: a time-use survey (internal implementation data from our study that will be used in a cost-effectiveness analysis to be reported later) showed that community health volunteers doing home visits spent an average of 27 additional h per session travelling to and doing home visits relative to group sessions. This opportunity cost of delivery agents' time would need to be accounted for in any plans for scaling the programme.

The review sessions were the only sessions that differed between the two intervention groups. Our particular model of mixed delivery differed from other interventions that had mostly home visits and a few group sessions[9] in that our home visits were infrequent and, as review sessions, introduced little new material. One possible explanation for why children seem to benefit more from the group review sessions might have to do with peer learning. Learning, or imitation of behaviours of similarly aged peers, is noted under specific conditions that are similar to our four review sessions, namely, when the behaviour is familiar and being consolidated, and when peers are familiar from past experience.[30] In our group review sessions, children were consolidating the conversation and play behaviours learnt in previous sessions. Thus the group, but not home visit, review sessions might have enhanced learning through the cognitive and social functions of peer imitation. More opportunities for peer learning at group review meetings could explain the divergence in child outcomes across delivery models, especially as there were larger returns from attendance at successive group review sessions compared with home visits. These results also suggest that having periodic reviews of ECD programme content for participants might help to consolidate learning.

Our study has many strengths related both to the intervention and evaluation of its effects. The intervention was implemented by a local organisation whose staff were also trained to train delivery agents. It used

paraprofessional community health workers as delivery agents, was low cost, and showed generally high rates of uptake among mothers in the intervention groups despite large group sizes. These factors make it a strong contender for future replication and scaling in similar low-resource settings. Strengths of our experimental evaluation included data collectors being masked to group assignment, low attrition with no differential attrition by study group, and the collection of child outcomes based on direct assessments, not maternal report. We also estimated causal effects on participants (average treatment effect on the treated) to show the key role of compliance in the intervention, we corrected for multiple hypothesis testing, and we explored mediating pathways for how the intervention affects children's development.

Despite these strengths, a key limitation of our study is that outcomes were assessed immediately following the end of the intervention, constraining our ability to know the longer-term effects of the intervention. Another limitation is the use of subjective measures of maternal wellbeing and parenting practices. Although these measures are often used to evaluate intervention effects in low-resource settings, they might be subject to self-report biases. We also cannot say what the effects would have been in a pure home visiting delivery model. Finally, we had hoped that our study would inform whether involving fathers in these programmes delivers additional benefits, but low father attendance prevented us from succeeding. Future work should focus on learning ways to engage fathers into an ECD programme, as well as re-examine child and household effects in later years to understand the sustainability of measured effect over time.

Our results show that an integrated child development intervention featuring exclusively group sessions delivered by paraprofessional community health workers to large groups of mothers can benefit multiple child outcomes and parental behaviours. We find no added benefit from the substitution of personalised home visits for group sessions. Our results represent a promising avenue for scaling similar interventions in low-resource rural settings where many children and families could potentially benefit from ECD programming.

Contributors

JEL and ILG secured the funding, designed the study, did the statistical analyses, verified the underlying data, and wrote the first draft. HOP and UYS oversaw data collection at different points. FEA, DRS, JEL, ILG, and HOP trained the trainers from the Safe Water and AIDS Project in the intervention content. RO and EA became lead trainers for community health volunteers and managed the study's implementation throughout. FEA trained data collection teams, with assistance from JEL, ILG, and UYS. FEA, DRS, and LCHF reviewed and provided critical input to study design and conceptualisation. All authors reviewed and provided input to the final draft. JEL had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

The de-identified datasets generated during the study along with the statistical plan and analytic code will be available from the corresponding author on reasonable request at the end of the 5-year project after all planned manuscripts have been accepted for publication. We will make the data without identifiers available to users only under a data-sharing agreement that provides for: (1) a commitment to using the data only for research purposes and not to make the attempt of identifying any individual participant; (2) a commitment to securing the data in case there are still some sensitive variables after the identifiers have been removed, by using appropriate computer technology; (3) a commitment to destroying or returning the data after analyses are complete; and (4) a commitment to not publish any information that is not treated at the aggregate level so that no specific characteristics can be linked to small communities.

Acknowledgments

We thank Alie Eleveld, Alex Mwaki, Ursula Liona, Fredrick Mseveni Nyadol, Caroline Odhiambo, George Simbiri, Arieda Joab Ochieng, the community health volunteers, and the entire team at the Safe Water and AIDS Project in Kenya for invaluable support in the carrying out of this study. We thank the County Health Management Teams and Sub-County Health Management Teams. We thank Marilyn N Ahun, Rebecca Zhu, and Hilda Chavala for invaluable research assistance in the field. This work is supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development with award number R01HD090045.

References

1 Black MM, Walker SP, Fernald LCH, et al. Early childhood development coming of age: science through the life course. *Lancet* 2017; **389:** 77–90.

2 Richter LM, Daelmans B, Lombardi J, et al. Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *Lancet* 2017; **389:** 103–18.

3 Britto PR, Lye SJ, Proulx K, et al. Nurturing care: promoting early childhood development. *Lancet* 2017; **389:** 91–102.

4 Gertler P, Heckman J, Pinto R, et al. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science* 2014; **344:** 998–1001.

5 Britto PR, Singh M, Dua T, Kaur R, Yousafzai AK. What implementation evidence matters: scaling-up nurturing interventions that promote early childhood development. *Ann N Y Acad Sci* 2018; **1419:** 5–16.

6 Tomlinson M, Darmstadt GL, Yousafzai AK, et al. Global research priorities to accelerate programming to improve early childhood development in the sustainable development era: a CHNRI exercise. *J Glob Health* 2019; **9:** 020703.

7 Horton S, Black MM. Chapter 24. Identifying an essential package for early child development: economic analysis. In: Bundy D, Nd S, Horton S, et al, eds. Child and adolescent health and development, 3rd edn. Washington DC: The International Bank for Reconstruction and Development, 2017: 343–53.

8 Singla DR, Kumbakumba E, Aboud FE. Effects of a parenting intervention to address maternal psychological wellbeing and child development and growth in rural Uganda: a community-based, cluster randomised trial. *Lancet Glob Health* 2015; **3:** e458–69.

9 Aboud FE, Yousafzai AK. Global health and development in early childhood. *Annu Rev Psychol* 2015; **66:** 433–57.

10 Jeong J, McCoy DC, Yousafzai AK, Salhi C, Fink G. Paternal stimulation and early child development in low- and middle-income countries. *Pediatrics* 2016; **138:** e20161357.

11 The Demographic and Health Surveys Program. Kenya demographic and health survey 2008–2009 final report. https://www.dhsprogram.com/publications/publication-fr229-dhs-final-reports.cfm (accessed Aug 3, 2016).

12 Luoto JE, Lopez Garcia I, Aboud FE, Fernald LCH, Singla DR. Testing means to scale early childhood development interventions in rural Kenya: the Msingi Bora cluster randomized controlled trial study design and protocol. *BMC Public Health* 2019; **19:** 259.

13 Yousafzai AK, Rasheed MA, Rizvi A, Armstrong R, Bhutta ZA. Effect of integrated responsive stimulation and nutrition interventions in the Lady Health Worker programme in Pakistan on child development, growth, and health outcomes: a cluster-randomised factorial effectiveness trial. *Lancet* 2014; **384:** 1282–93.

14 Bayley N. Bayley Scales of Infant Development and Toddler Development: technical manual. Bloomington, MN: Pearson, 2006.

15 Rademeyer V, Jacklin L. A study to evaluate the performance of black South African urban infants on the Bayley Scales of Infant Development III. *South African J Child Heal* 2013; **7:** 54–59.

16 Hanlon C, Medhin G, Worku B, et al. Adapting the Bayley Scales of Infant and Toddler Development in Ethiopia: evaluation of reliability and validity. *Child Care Health Dev* 2016; **42:** 699–708.

17 Pendergast LL, Schaefer BA, Murray-Kolb LE, et al. Assessing development across cultures: invariance of the Bayley-III Scales across seven international MAL-ED sites. *Sch Psychol Q* 2018; **33:** 604–14.

18 Wolke D, Skuse D, Mathisen B. Behavioral style in failure-to-thrive infants: a preliminary communication. *J Pediatr Psychol* 1990; **15:** 237–54.

19 Hamadani JD, Tofail F, Hilaly A, Huda SN, Engle P, Grantham-McGregor SM. Use of Family Care Indicators and their relationship with child development in Bangladesh. *J Health Popul Nutr* 2010; **28:** 23–33.

20 Bradley RH, Caldwell BM. The HOME inventory and family demographics. *Dev Psychol* 1984; **20:** 315–20.

21 de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. *Food Nutr Bull* 2004; **25** (suppl)**:** S15–26.

22 WHO, UNICEF. Global strategy for infant and young child feeding. Geneva: World Health Organization, 2003.

23 Ballard T, Coates J, Swindale A, Deitchler M. Household Hunger Scale: indicator definition and measurement guide. Washington, DC: Food and Nutrition Technical Assistance II Project, 2011.

24 Attanasio OP, Fernández C, Fitzsimons EOA, Grantham-McGregor SM, Meghir C, Rubio-Codina M. Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. *BMJ* 2014; **349:** g5785.

25 Romano JP, Wolf M. Stepwise multiple testing as formalized data snooping. *Econometrica* 2005; **73:** 1237–82.

26 MacKinnon DP, Pirlott AG. Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Pers Soc Psychol Rev* 2015; **19:** 30–43.

27 Hamadani JD, Mehrin SF, Tofail F, et al. Integrating an early childhood development programme into Bangladeshi primary health-care services: an open-label, cluster-randomised controlled trial. *Lancet Glob Health* 2019; **7:** e366–75.

28 Grantham-McGregor SM, Powell CA, Walker SP, Himes JH. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *Lancet* 1991; **338:** 1–5.

29 Andrew A, Attanasio O, Fitzsimons E, Grantham-McGregor S, Meghir C, Rubio-Codina M. 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 Med* 2018; **15:** e1002556.

30 Zmyj N, Seehagen S. The role of a model's age for young children's imitation: a research review. *Infant Child Dev* 2013; **22:** 622–41.