



Human Development
and Violence
Research Centre

**Statistical Analysis Plan for the Pelotas Parenting
Interventions for Aggression (PIA) trial**

Merryn Voysey

September, 2020

Trial registration: RBR-2kwfsk at the Brazilian Ministry of Health Register of Clinical Trials
(<http://www.ensaiosclinicos.gov.br/>)

Effective Date: 20 September 2020

Author: Merryn Voysey, University of Oxford

Reviewers/Approvers:

Joseph Murray, Federal University of Pelotas

Rafaela Martins, Federal University of Pelotas

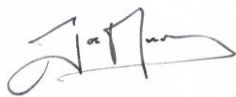
Signatures:



21 September 2020

Merryn Voysey

Date



15 September 2020

Joseph Murray

Date



15 September 2020

Rafaela Martins

Date

Version History

Version	Date	Details
1	15 September 2020	Approved version 1

Table of Contents

1	Introduction:	5
2	Hypotheses:	5
2.1	Primary Hypothesis	5
2.2	Secondary Hypotheses.....	5
3	Trial Methods:.....	5
3.1	Trial design	5
3.2	Randomisation	6
3.3	Sample size.....	6
3.4	Blinding	6
3.5	Interim analysis	6
3.6	Final analysis	6
3.7	Outcome assessments and other measures	6
	Table 1. Study Outcomes and Measures	7
4	Statistical principles	11
4.1	Significance level.....	11
4.2	Multiple comparisons	11
4.3	Analysis Populations	11
4.4	Outliers.....	11
5	Trial population.....	11
6	Efficacy Analysis	12
6.1	General analysis considerations.....	12
6.1.1	Descriptive statistics	12
6.1.2	Missing data	12
6.2	Primary analyses	12
6.3	Sensitivity analyses	13
	Table 2. Study Covariates for each Outcome.....	14
6.4	Mediator effects	17
6.5	Moderator effects	17
	Table 3. Mediators and Moderators for each Outcome.....	18
7	Changes from protocol defined statistical analysis	21
8	References	22

1 Introduction:

The study aims are to determine, via a three-arm randomised controlled trial, the impact of two group-based, parent training programmes implemented with mothers with young children. The two training programmes are *ACT: Raising Safe Kids*, which primarily aims to reduce harsh discipline and maltreatment against children, and a *Dialogic Book Sharing (DBS)* programme, which primarily aims to stimulate positive parent child interaction and child cognitive development. In particular, we aim to determine the impact of each training programme on child aggressive behaviour (primary outcome), and child cognitive development (secondary outcomes), and parenting practices (secondary outcomes).

2 Hypotheses:

2.1 Primary Hypothesis

1. Compared to a control group of children whose carers receive no additional intervention, children whose carers receive ACT or DBS will evidence significantly less aggressive behaviour.

2.2 Secondary Hypotheses

1. Compared to children of families who receive no intervention (the control group), the children in families receiving Dialogic Book-Sharing (DBS) will show less aggression at follow-up, and they will perform better on measures of language, executive function, attention, and empathy/emotion understanding; but parents will show more positive parenting, but parents will not show less harsh and abusive parenting.
2. Compared to children of control group families, the children of families receiving ACT will show less aggression at follow-up, and their parents will show less harsh and abusive parenting and less favourable attitudes about corporal punishment; but the children will not perform better on measures of language, executive function, attention, and empathy/emotion understanding.
3. For both the DBS and ACT groups, children and parents will show less stress at follow-up, and parents will show more positive parenting.

3 Trial Methods:

3.1 Trial design

The study is a three-arm randomized controlled trial (RCT) nested within an ongoing birth cohort study. The 2015 Pelotas Birth Cohort Study, has followed 4,275 children from birth. The cohort includes the entire population of children born in the city of Pelotas, southern Brazil, born in the calendar year 2015. Based on data collected on the cohort when children were aged 24-months, a sub-sample of mother-child pairs were recruited for the PIÁ trial, when children were between 2-3 years old. Interventions were delivered by the Pelotas municipal government staff in local educational facilities, under the supervision of the research team. Baseline, 4-week post-intervention, and 8-month follow-up assessments were conducted for the PIÁ trial.

3.2 Randomisation

Randomisation of mother-child pairs to control and the two intervention (ACT and DBS) conditions, was undertaken at the research centre immediately after baseline assessment, minimising for the following dichotomous variables: age of child (<3 years and ≥ 3 years), child sex (male and female), child aggression score at age 2 years (<4 and ≥ 4) harsh parental discipline score at age 2 years (<6 and ≥ 6). The probability that individuals entered into each of the three arms of the trial was 33.3%.

3.3 Sample size

The sample comprises mothers from the 2015 Pelotas Birth Cohort Study with children 2-3 years old, who were in the poorest 30% of the population, and who reported that their children showed average-high levels of child aggression at age 24-months.

Power calculation: With alpha set at 0.025 due to two pair-wise comparisons (i.e. between DBS and control, and between ACT and control), and beta at 0.20, each of the three arms in the trial requires a minimum of 104 participants (allowing for 10% attrition to the follow-up assessment) to detect a mid-range effect size of $d = 0.45$. Therefore a minimum of 312 participants were needed for the study. The actual number of participants recruited to the study was 369 with 369 (100%) followed at post-intervention, and 368 (99.7%) followed-up at 8 months post-intervention.

3.4 Blinding

To prevent assessment bias, assessments of children and caregivers have been carried out blind to group allocation, including explicitly asking participants not to reveal their allocation to the data collectors. All coding of video material has been made blind to allocation. The statistician will be provided with individual intervention allocations at the time of database lock, in the form of a categorical variable with three values (1= A, 2 = B, 3 = C) remaining blind to which of A, B, and C refers to the ACT group, the DBS, and the Control group. The statistician will then compare A-B, B-C, and C-A, in order to generate trial results comparing ACT-control and DBS-Control, while remaining blind to intervention status. After primary intention to treat analyses have been completed comparing outcomes in this way, the statistician will be unblinded in order to complete additional per protocol analyses and mediation and moderation analyses comparing ACT-Control, DBS-Control.

3.5 Interim analysis

There are no planned interim analyses.

3.6 Final analysis

The final analyses will be performed after all subjects have completed their follow-up visit, the data are coded, cleaned and the database locked. At the time of database lock the statistician will request to receive the individual level intervention information from the trial manager.

3.7 Outcome assessments and other measures

Table 1 shows the trial outcome constructs and measures.

Table 1. Study Outcomes and Measures

Outcome Domain	Outcome	Measures	Baseline (1 Week Prior to Intervention)	Post Intervention (1 Month Post Intervention)	Follow-Up (8 Months Post Intervention)	
Child aggression – combined measure (primary outcome)	Child aggression - combined measure	Combined score (mean of z-scores) of: • Child Behaviour Checklist – aggression subscale • ELDEQ study questionnaire aggression score	✓ p1zagg		✓ p3zagg	
		Child aggression - individual measures	Child Behaviour Checklist – aggression subscale	✓ p1cbclfinal		✓ p3cbclagress
			ELDEQ study questionnaire aggression score	✓ p1eldeqfinal		✓ p3eldeqfinal
			Filmed LabTab – aggression score	✓ p1labtab_manger		✓ p3labtab_manger
			Combined score (mean of z-scores) of: • Filmed ‘Don’t touch’ – child behaviour • Filmed ‘Clean Up’ – child behaviour	✓ p1zcbbehav		✓ p3zcbbehav
Interviewer rating – child defiance	✓ p1negativ		✓ p3negativ			
Child development (secondary outcomes)	Child language	Combined score (mean of z-scores) of: • Expressive language task • Receptive language task	✓ p1childlang		✓ p3childlang	
		Expressive language task	✓ p1tvesum		✓ p3tvesum	
		Receptive language task	✓ p1tvasum		✓ p3tvasum	
	Child attention and executive functions	Strengths and Difficulties Questionnaire – attention/hyperactivity subscale	✓ p1sdqfinal		✓ p3sdqfinal	
		Filmed Play Alone task – focus score	✓ p1playalone_mquality		✓ p3playalone_mquality	

Outcome Domain	Outcome	Measures	Baseline (1 Week Prior to Intervention)	Post Intervention (1 Month Post Intervention)	Follow-Up (8 Months Post Intervention)	
		Interviewer rating – child attention	✓ p1childatt		✓ p3childatt	
		Go no Go task from the Early Years Toolbox	✓ p1cgng		✓ p3cgng	
		Block Design task	✓ p1cblock		✓ p3block	
		Card Sort Task from the Early Years Toolbox	<i>Note: not measured at baseline</i>		✓ p3cs_switchacc	
	Child empathy- prosocial behaviour	Combined score (mean of z-scores) of: • Em-Que questionnaire – Emotion Contagion • Strengths and Difficulties Questionnaire - Prosocial behaviour score	✓ p1zempat		✓ p3zempat	
		Em-Que questionnaire – Emotion Contagion	✓ p1emqueemcont		✓ p3emqueemcont	
		Strengths and Difficulties Questionnaire - Prosocial behaviour score	✓ p1sdqprosocial		✓ p3sdqprosocial	
		Filmed Help Task	✓ p1helpt_score		✓ p3helpt_score	
		Denham’s puppet task	✓ p1puppetfinal		✓ p3puppetfinal	
		Dictator Game	<i>Note: not measured at baseline</i>		✓ p3altruism2	
		Theory of Mind	Triangle task	✓ p1ctrianaction		✓ p3ctrianaction
			Sally-Anne task	<i>Note: not measured at baseline</i>		✓ p3sallyfinal
	Parenting (secondary outcomes)	Positive parenting	Combined score (mean of z-scores) of: • Filmed Responsive Interactions – sensitivity • Filmed book-sharing - sensitivity • Filmed book-sharing task – reciprocity • Filmed Don’t touch – Guidance • Filmed Clean Up – Guidance	✓ p1zpparent	✓ p2zpparent	✓ p3zpparent

Outcome Domain	Outcome	Measures	Baseline (1 Week Prior to Intervention)	Post Intervention (1 Month Post Intervention)	Follow-Up (8 Months Post Intervention)	
		PAFAS - positive encouragement subscale	✓ p1pafasposen	✓ p2pafasposen	✓ p3pafasposen	
		PAFAS - parent-child relationship subscale	✓ p1pafasprelat	✓ p2pafasprelat	✓ p3pafasprelat	
		Filmed Book-sharing task - sensitivity	✓ p1bssensitivity	✓ p2bs4_m_sensitivity	✓ p3bssensitivity	
		Filmed Book-sharing task – reciprocity	✓ p1bsreciprocity	✓ p2bsreciprocity	✓ p3bsreciprocity	
		Filmed Responsive Interactions task - sensitivity	✓ p1respint_total	✓ p2respint_total	✓ p3respint_total	
		Filmed Don't Touch and Clean up tasks - mean of total guidance on two tasks	✓ p1totguid	✓ p2totguid	✓ p3totguid	
	Harsh parenting	PAFAS questionnaire - coercive subscale	✓ p1pafascoer	✓ p2pafascoer	✓ p3pafascoer	
		Filmed 'Don't touch' and 'Clean Up' Tasks – mean of coercion total scores on two tasks	✓ p1coercion	✓ p2coercion	✓ p3coercion	
	Attitudes about punishment	Attitudes about physical punishment	✓ p1aappfinal	✓ p2aappfinal		
	Maltreatment	Juvenile Victimization Questionnaire - maltreatment	✓ p1jvqanymalt		✓ p3jvqanymalt	
	Stress (secondary outcomes)	Maternal stress	Perceived Stress Scale	✓ p1pssfinal	✓ p2pssfinal	
			Pelotas Parenting Stress Index	✓ p1psifinal	✓ p2psifinal	
Maternal cortisol		3-month cortisol levels from hair samples	✓ p1mhairconc		✓ p3mhairconc	
Child cortisol		3-month cortisol levels from hair samples	✓ p1chairconc		✓ p3chairconc	

* Measures applied pre-baseline in context of 24-month assessment of birth cohort study in which the trial is nested.

The following additional data were collected prior to baseline, as part of cohort assessments:

- Maternal education measured in the perinatal assessment, and complemented with pre-natal data when perinatal was missing [abmateduc]
- Socioeconomic status measured when children were 24 months of age [fw24income]
- Frequency of reading-storytelling to child when children were 24 months of age [fw24reading]
- Child language when children were 24 months (to be used for missing language data at baseline)
 - Total language score: [fw24meanlanguage]
 - Expressive language score [fw24meanexpressilanguage]
 - Receptive language score [fw24meanreceptlanguage]

The following additional data were collected at baseline:

- Child age [p1cage, p1cagecat]
- Child sex [p1sex]
- Neighbourhood [p1area]
- Maternal ethnicity [p1mrace]
- Mother relationship status [p1partner]
- Time mother spends with child per week [p1stayweek]
- Child attendance at preschool [p1school]
- Child callous unemotional traits [p1icufinal]
- Intimate partner violence [p1vpifinalcat]
- Maternal depression [p1epdsfinal]
- Maternal problem drinking [p1auditfinalcat]
- Hair cortisol concentration confounders variables for children [p1cusecort, abmateduc, fw2cbmi, p1jvqmalt]
- Hair cortisol concentration confounders variables for mothers [p1musecort, abmateduc, fw24matimc]

The following additional data were also collected about intervention adherence:

- Mother completed the intervention (yes/no) [p3attend]
This was defined as attended 7+ out of 9 of the ACT intervention sessions and 6+ out of the 8 book sharing sessions in the DBS intervention group

The following additional data were also collected at post-intervention:

- Child language (combined receptive and expressive score) [p2zmeanlang], which will be analysed as a possible mediator of effects on the primary outcome.
- Child attention-hyperactivity (SDQ score) [p2sdqhiperac], which will be analysed as a possible mediator of effects on the primary outcome.

During the intervention phase, the number of sessions attended was recorded, and carer report of compliance and assessment of intervention quality and usefulness is also recorded at post treatment. Fidelity of facilitator implementation of the two parent-training programmes is also assessed by the research supervisors of each intervention.

4 Statistical principles

4.1 Significance level

All applicable statistical tests will be two-sided and will be performed using a 5% significance level. All confidence intervals presented will be 95% and two-sided.

4.2 Multiple comparisons

No formal adjustment for multiple testing will be made for the primary endpoints. The endpoints are associated with each other and an adjustment would over-correct (Schulz et al 2005).

All primary outcomes will be reported and interpreted together as specified in the primary hypothesis.

No formal correction will be made for multiple testing in the secondary and tertiary/exploratory analyses, but account will need to be taken in the interpretation where multiple statistical tests have been performed.

Interpretation of results will also take account of consistency across outcomes as well as clinical plausibility based on prior knowledge.

4.3 Analysis Populations

The intention- to- treat (ITT) population will be used for all analyses. This population includes all participants who were enrolled and randomised. Subjects will be analysed based on the groups to which they were randomized.

The per-protocol population will be used for a sensitivity analysis on the primary endpoints and key secondary outcomes. This population is a subset of the ITT population which excludes subjects who were allocated to but did not complete the ACT and DBS interventions. All ITT subjects in the control group will be included.

If the per-protocol population contains more than 95% of the ITT population then the sensitivity analysis will not be conducted.

4.4 Outliers

It is planned that all data will be included in the relevant analyses. However, if it is deemed necessary any subjects excluded from the summaries and/or statistical analyses will be documented along with the reason for exclusion in the report.

5 Trial population

The number of subjects who were enrolled, randomized and completed will be summarised and listed by intervention group. The number of subjects included in each population will also be presented by intervention group.

The consort diagram comprising the number of people screened/approached, eligible, randomised, received their allocated intervention, withdrawing/lost to the follow-up, will be produced in collaboration with the trial manager.

A listing of subjects removed from any population will be produced. Reports of compliance/attendance will also be listed and summarised.

6 Efficacy Analysis

6.1 General analysis considerations

For all analyses, the relevant assumptions will be checked. Alternative models may be used if necessary. If the normality assumption does not hold then a transformation such as log transformation, or an alternative distribution will be investigated. If there is no suitable transformation then a non-parametric testing method will be utilised.

Adjusted means and intervention group differences along with 95% confidence intervals and standard errors will be presented in a table, p-values will be presented for the difference between groups at post-intervention (where available) and follow-up. The primary comparisons between intervention and control group in child outcomes will be at follow-up.

6.1.1 Descriptive statistics

The primary endpoints, as well as the individual components, will be listed, and summarised by intervention group, and, where available, timepoint.

Demographic and baseline characteristics will be listed and summarised by intervention group. Categorical data will be summarised by numbers and percentages. Continuous data will be summarised by mean, SD, median, minimum and maximum values. The number and percent of missing data per outcome, group and time point will be summarised.

6.1.2 Missing data

For composite endpoints, missing data will be imputed on the individual measures and the composite will then be derived from the imputed components. If more than 1/3 of items on a questionnaire or measure are missing then the result will be considered missing for that person for that timepoint, otherwise prorated scores will be calculated and considered non-missing.

Due to technical difficulties language measures at baseline are missing for about one third of the participants. For the main analyses, multiple imputation will be used for these missing data, using previous measures of language measured at 24 months (pretrial cohort measure). In sensitivity analyses, this pretrial language measure will be used as a proxy for baseline language data for the whole cohort.

6.2 Primary analyses

Intervention effects will be assessed at post intervention and follow-up, and will be adjusted for: baseline scores, child age and sex, maternal education and depression, neighbourhood (as random effect), and additional, outcome-specific covariates shown in Table 2.

A mixed effects model will be fitted to the outcomes at post-intervention and at follow-up. Intervention, timepoint, intervention by timepoint interaction, covariates as listed above, and baseline values, will be fitted as fixed effects, neighbourhood and subject will be fitted as random effects, with repeated measures within a subject being accounted for. If the necessary assumptions of the models do not hold, suitable alternative models will be explored. Intention-to-treat analysis will be used to examine intervention effects.



6.3 Sensitivity analyses

Sensitivity analyses of the primary and main secondary endpoints will be conducted using the per-protocol population.

Table 2. Study Covariates for each Outcome

Outcome Domain	Outcome	Generic Study Covariates	Additional Covariates (measured at baseline, or pre-baseline from cohort)
Child aggression, combined measure (primary outcome)	Child aggression	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Child maltreatment • PAFAS coercive parenting • Combined child language • Child Go no Go Task • Child callous-unemotional traits
Child development (secondary outcomes)	Child language	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Child maltreatment • Child preschool attendance • Reading-storytelling at 24m • Combined positive parenting
	Child attention and executive functions	<ul style="list-style-type: none"> • Baseline measure (Card-Sort proxy = language) • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Child maltreatment • PAFAS coercive parenting • Combined positive parenting
	Child empathy-prosocial behaviour	<ul style="list-style-type: none"> • Baseline measure (Puppets proxy = language) • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Child maltreatment • PAFAS coercive parenting • Combined positive parenting

Outcome Domain	Outcome	Generic Study Covariates	Additional Covariates (measured at baseline, or pre-baseline from cohort)
	Child Theory of Mind	<ul style="list-style-type: none"> • Baseline measure (Sally-Anne proxy = Triangle Task) • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Child maltreatment • PAFAS coercive parenting • Combined positive parenting • Denham's puppet task • Child language • Older age sibling
	Positive parenting	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Maternal problem drinking • Reading-storytelling 24m [fw24reading] • Combined child aggression • Intimate partner violence • Pelotas Parenting Stress Index
Parenting (secondary outcomes)	Harsh parenting	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Maternal problem drinking • Combined child aggression • Intimate partner violence • Pelotas Parenting Stress Index • Attitudes about physical punishment • JVQ maltreatment
	Parental attitudes about corporal punishment	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Combined child aggression • Intimate partner violence • PAFAS coercive parenting
	Maltreatment	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Maternal problem drinking • Combined child aggression • Intimate partner violence • Pelotas Parenting Stress Index • Attitudes about physical punishment • PAFAS coercive parenting

Outcome Domain	Outcome	Generic Study Covariates	Additional Covariates (measured at baseline, or pre-baseline from cohort)
	Maternal stress	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Intimate partner violence • Combined child aggression
Stress (secondary outcomes)	Maternal cortisol	<ul style="list-style-type: none"> • Baseline measure • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • Intimate partner violence • Combined child aggression • Maternal steroids use • Maternal BMI
	Child cortisol	<ul style="list-style-type: none"> • Baseline measure • Child age • Child sex • Maternal education • Maternal depression • Neighbourhood 	<ul style="list-style-type: none"> • PAFAS coercive parenting • Child maltreatment • Child steroids use • Child BMI • Intimate partner violence

6.4 Mediator effects

To test the role of possible mediators of the intervention effects on child outcomes, we will use a multiple multilevel Mediator Model fitted using the `ml_mediation` package in Stata or similar. Table 3 shows the mediators that will be tested for each outcome.

For each outcome, a single outcome measure will be selected for the mediation analyses, contingent on the results of the primary analyses. Table 3 shows the principal outcome measure that will be considered for mediator (and moderator) analyses for each outcome. If effects are found ($p < .05$) in the primary analyses for the principal outcome measure shown in Table 3, then mediation analyses will be run on that principal outcome measure. However, if no main effect is found for the principal outcome measure, then the mediation analyses will be run on whichever other measure of that outcome has the largest significant effect.

Separate models will be fitted for each outcome and follow the main analysis modelling strategy. Mediators will be tested for both interventions, unless specified otherwise in Table 3.

6.5 Moderator effects

Table 3 shows the potential moderators of intervention impact that will be tested for each outcome. To assess the effect of potential moderators, the same model will be fitted as described in 6.2 with the addition of the moderator of interest and moderator*intervention interaction. Moderators will be tested for both interventions, unless specified otherwise in Table 3.

Table 3. Mediators and Moderators for each Outcome

Outcome Domain	Outcome	Principal Outcome Measure for Moderator-Mediator Analyses	Moderators (measured baseline, or pre-baseline from cohort)	Mediators (post intervention unless stated otherwise)
Child aggression, combined measure (primary outcome)	Child aggression - combined measure	Combined (mean) of z-scores from: Child Behaviour Checklist – aggression subscale ELDEQ study questionnaire aggression score	<ul style="list-style-type: none"> • Baseline measure of outcome • Child sex • Child age • Child callous-unemotional traits • Maternal education • Maternal depression • Intimate partner violence • PAFAS coercive parenting • Participation in PIM programme 	<ul style="list-style-type: none"> • PAFAS coercive parenting • Combined positive parenting • Combined child language (DBS only) • Child SDQ attention/ hyperactivity (DBS only) • Child cortisol at follow-up
Child development (secondary outcomes)	Child language (combined expressive and receptive)	Combined (mean) of z-scores from: Teste de Vocabulário Expressivo Teste de Vocabulário Receptivo	<ul style="list-style-type: none"> • Baseline measure of outcome • Child sex • Child age • Maternal education • Maternal depression • Intimate partner violence • Filmed book-sharing task – sensitivity (DBS only) 	<ul style="list-style-type: none"> • Filmed book-sharing task – sensitivity (DBS only) • Filmed book-sharing task – reciprocity (DBS only)
	Child attention	Strengths and Difficulties Questionnaire attention/hyperactivity subscale	<ul style="list-style-type: none"> • Baseline measure of outcome • Child sex • Child age • Maternal education • Maternal depression • Intimate partner violence • Filmed book-sharing task – sensitivity (DBS only) 	<ul style="list-style-type: none"> • Filmed book-sharing task – sensitivity (DBS only) • Filmed book-sharing task – reciprocity (DBS only)

Outcome Domain	Outcome	Principal Outcome Measure for Moderator-Mediator Analyses	Moderators (measured baseline, or pre-baseline from cohort)	Mediators (post intervention unless stated otherwise)
	Child executive functions	Block design score	<ul style="list-style-type: none"> • Baseline measure of outcome • Child sex • Child age • Maternal education • Maternal depression • Intimate partner violence 	<ul style="list-style-type: none"> • Combined average score of observed parenting
	Child empathy-prosocial behaviour	Combined (mean) of z-scores from: <ul style="list-style-type: none"> • Em-Que questionnaire Emotion Contagion • Strengths and Difficulties Questionnaire - Prosocial behaviour score 	<ul style="list-style-type: none"> • Baseline measure of outcome • Child sex • Child age • Maternal education • Maternal depression • Intimate partner violence 	<ul style="list-style-type: none"> • Combined average score of observed parenting
	Child Theory of Mind	Triangle task/ Sally-Anne task	<i>None to be tested</i>	<i>None to be tested</i>
	Positive parenting	Combined (mean) of z-scores from: <ul style="list-style-type: none"> • Filmed Responsive Interactions – sensitivity • Filmed book-sharing - sensitivity • Filmed book-sharing task – reciprocity • Filmed Don't touch - Guidance • Filmed Clean Up – Guidance 	<ul style="list-style-type: none"> • Baseline measure of outcome • Child sex • Child age • Maternal education • Maternal depression 	<i>None to be tested</i>
	Harsh parenting	PAFAS questionnaire - coercive subscale	<ul style="list-style-type: none"> • Baseline measure of outcome (ACT only) • Child sex (ACT only) • Child age (ACT only) • Maternal education (ACT only) • Maternal depression (ACT only) • Intimate partner violence (ACT only) 	<ul style="list-style-type: none"> • Attitudes about physical punishment (ACT only) • Parental stress (ACT only)
	Maltreatment	Juvenile Victimization Questionnaire (JVQ) – Maltreatment	<ul style="list-style-type: none"> • Baseline measure of outcome (ACT only) • Child sex (ACT only) • Child age (ACT only) • Maternal education (ACT only) • Maternal mental health (ACT only) • Intimate partner violence (ACT only) 	<ul style="list-style-type: none"> • Attitudes about physical punishment (ACT only) • Parental stress (ACT only)

Outcome Domain	Outcome	Principal Outcome Measure for Moderator-Mediator Analyses	Moderators (measured baseline, or pre-baseline from cohort)	Mediators (post intervention unless stated otherwise)
	Parental Attitudes about corporal punishment	Attitudes about physical punishment	<i>None to be tested</i>	<i>None to be tested</i>
Stress (secondary outcomes)	Maternal Stress	Perceived Stress Scale/ Pelotas Parenting Stress Index	<i>None to be tested</i>	<i>None to be tested</i>
	Maternal cortisol	3-month cortisol levels from hair samples	<ul style="list-style-type: none"> • Baseline measure of outcome • Maternal education • Maternal depression • Intimate partner violence 	<ul style="list-style-type: none"> • Pelotas Parenting Stress Index
	Child cortisol	3-month cortisol levels from hair samples	<ul style="list-style-type: none"> • Baseline measure of outcome • Child age • Child sex • Maternal education • Maternal depression • Intimate partner violence 	<ul style="list-style-type: none"> • Pelotas Parenting Stress Index • PAFAS coercive parenting • Combined positive parenting

7 Changes from protocol defined statistical analysis

Small corrections, adjustments and clarifications to the trial protocol have been made in this analysis plan.

First, the protocol specified that the primary outcome of the trial was child aggression, but it did not specify if all measures of child aggression would be reported as primary outcomes. The analysis plan defines one combined measure of child aggression as the primary outcome, and the individual measures of child aggression as secondary outcomes. (see Table 1).

Second, in some parts of the protocol text, some of the secondary outcomes were referred to as “main secondary outcomes and others as “additional secondary outcomes”. In the analysis plan all outcomes that are not “primary” are treated as secondary (see Table 1).

Third, some of the individual measures in the protocol have been changed:

1. For child attention, the Filmed Play Alone Task will be used as an additional measure and analysed as defined in this analysis plan (see Table 1).
2. The Card Sort Task measure was wrongly shown as a measure of child attention in the protocol, and will be analysed as a measure of child executive function (see Table 1).
3. Interviewers made short assessments of child behaviour after completing assessments. In the protocol it was planned to include a measure of self-control from these ratings, but the items were considered to better reflect child attention and child aggression. So, the analysis plan includes no interviewer rating of executive function, and includes instead additional interviewer ratings of child attention and aggression (see Table 1).
4. For child prosocial behaviour (called empathy/theory of mind in the protocol), the Strengths and Difficulties Questionnaire prosocial subscale score had not been specified in the protocol, but will be used as an additional measure of prosocial behaviour in the analyses (see Table 1).
5. The protocol stated that child protection service records would be measured as one measure of child maltreatment, but these records were not collected (because rates of official maltreatment in the population were discovered to be too low for analysis in this study), and so child protection service records will not be analysed.

Fourth, in the protocol it was stated that positive parenting would be measured at post intervention, and harsh parenting would be measured at follow-up, and hence they would be analysed accordingly. During the trial, parenting assessments were at both time points, and both positive and harsh parenting outcomes will be analysed at post-intervention and at follow-up, as defined in this analysis plan (see Table 1).

Fifth, the protocol stated that missing data would be handled using multiple imputation. However, the mixed effects models specified will handle the missing data without the requirement for formal imputation.

8 References

Cohen J, Cohen P, West SG, Aiken LS (2003). *Applied multiple regression and correlation for the behavioural sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

De Boer MR, Waterlander WE, Kuijper LDJ, Steenhaus IHM, Twisk JWR. Testing for baseline differences in randomized controlled trials: an unhealthy research behaviour that is hard to eradicate. *International Journal of Behavioural Nutrition and Physical Activity*. (2015)12:4.

Fenson L, Pethick S, Renda C, Cox JL, Dale PS, Reznick JS. Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*. 2000;21(1):95–115.

Hayes (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: a regression-model approach*.

Hedges LV. Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*. 2007;32(4):341–70.

National income deprivation affecting children Index (IDACI) (2015). <https://www.northamptonshireanalysis.co.uk/linked-data/id/indicator/IDACINationalDecile>.

Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. *Lancet*. 2005;365:1591–95.

Vally Z, Murray L, Tomlinson M, Cooper PJ. The impact of dialogic book-sharing on training on infant language and attention: a randomized controlled trial in a deprived South African community. *Journal of Child Psychology and Psychiatry*. 2015;56(8):865-73.

Dowdall N (2015). *The effectiveness of shared picture book reading interventions on child language outcomes: A systematic review and meta-analysis*. Master's Thesis, Oxford University.

Goldfeld S, Quach J, Nicholls R, Reilly S, Ukoumunne OC, Wake M. Four-year-old outcomes of a universal infant-toddler shared reading intervention: The Let's Read trial. *Archives of Pediatrics & Adolescent Medicine*. 2012;166(11):1045-52.