

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol for a cluster randomized controlled trial evaluating a parenting with home visitation program to prevent physical and emotional abuse of children in Indonesia: The Families First Program
AUTHORS	Ruiz-Casares, Monica; Lilley, Sarah; Thombs, Brett D; Platt, Robert W; Scott, Susan; Isdijoso, Widjajanti; Hermanus, Emmy; Andrina, Michelle; Mayo, Nancy

VERSION 1 – REVIEW

REVIEWER	Philip Wilson University of Aberdeen Scotland I have been involved as a principal investigator or co-investigator in several randomised trials of a number of parenting interventions, but not Families First. My University has received payment for my involvement in parenting evaluations, but I have not received any personal remuneration for this work.
REVIEW RETURNED	11-Feb-2018

GENERAL COMMENTS	<p>This is a very important study, and in general the proposal as described appears to be adequate for getting an answer to the research questions posed. I do however have two significant concerns, one about the power calculation and the other about inadequate description of the process evaluation and qualitative research. In both cases it would be useful if the authors could refer to the 2008 MRC Framework for the Evaluation of Complex Interventions, and the 2015 MRC guidance for process evaluations of complex intervention trials.</p> <p>My concern about the sample size calculation is that an intraclass correlation coefficient of 0.02 is assumed, and no evidence is presented for this assumption. It is very likely that the figure will be higher than this and the study power consequently lower than described. An independent data monitoring committee (referred to in the SPIRIT checklist but not actually mentioned on page 17) could potentially examine the baseline data, calculate the ICC and advise the research team on whether additional villages should be recruited</p>
-------------------------	---

	<p>at the outset.</p> <p>The process evaluation is very poorly described. The logic model could be used as a framework to help the authors determine which mediator and moderator analyses might be conducted. What process data will be collected? Completion rates of sessions (total as well as by demographic characteristics)? Parent satisfaction? Fidelity of delivery? Etc. Furthermore the qualitative interviews and focus groups are mentioned (mainly in the flow diagram) but no detail is given. The reader needs to be able to understand the point of the qualitative evaluation. Who will be interviewed? How many interviews will be required? What topics might need to be covered? What method of analysis will be used?</p> <p>A good process evaluation of a groundbreaking trial like this is essential whether the results are positive or negative. In either cases the process evaluation will inform commissioners and practitioners about what contextual and operational factors will be most likely to lead to successful implementation in other settings.</p> <p>More minor points:</p> <ul style="list-style-type: none"> - there should be some discussion of how likely self-report questionnaires are to give honest answers about child maltreatment - readers will be interested to know about 'village leaders'. Are they easy to identify? - the timeline of the study should be described. - The method by which baseline covariates will be used to adjust outcomes is not specified clearly. Further detail would be useful.
--	---

REVIEWER	Jamie M. Lachman Department of Social Policy and Intervention, University of Oxford, United Kingdom
REVIEW RETURNED	07-Apr-2018

GENERAL COMMENTS	<p>Overall</p> <p>This manuscript provides a very important contribution to the science regarding the testing of parenting programs in low- and middle-income countries. The proposed protocol demonstrates how researchers can conduct rigorous evaluations of the effectiveness of parenting programs on the reduction of violence against children in low resource areas where there is limited evidence. It also addresses key policy issues regarding the global prevention of violence against children and the need for evidence-based interventions that meet this need. The manuscript is well-written and organized, and closely adheres to the SPIRIT guidelines for protocols of intervention trials. It also meets the high standard for scientific excellence of the BMJ Open. After addressing the specific comments discussed below from each section, this reviewer is happy to recommend the manuscript for publication.</p> <p>Abstract</p> <ol style="list-style-type: none"> 1. Please specify that villages in the intervention group will also receive standard community health and social services as stated in the body of the manuscript. 2. Even though the manuscript is a protocol and thus does not have results to report, please include information regarding the potential
-------------------------	---

	<p>scientific contribution and policy implications of the study regardless of whether results are positive or negative.</p> <p>Strengths and Limitations of the Study</p> <ol style="list-style-type: none">1. Please be more explicit about the last limitation referring to potential biases as a result of social desirability due to the reliance of the self-report outcome measurements.2. Please also be more explicit regarding the lack of blinding, although this may be possible at post-test. <p>Introduction</p> <ol style="list-style-type: none">1. The introduction provides a clear overview of the background of violence against children in LMICs, the role of positive parenting and effective discipline in reducing risks of child maltreatment, and the extent of evidence of effectiveness of parenting programs in LMICs, and the East Asia Pacific region in particular. The section on risk factors for violence against children could include more research that is specific to Indonesia, or if lacking, the region. Similarly, please include any evidence on the effect of child maltreatment that is particular to Indonesia.2. It would be helpful to have more discussion regarding the current evidence of effectiveness of parenting programs in Indonesia. For instance, there was a recent randomized controlled trial of a parenting program in Indonesia that might provide a useful context and justification for additional research in the country and region (see Sumargi, A., Sofronoff, K., and Morawska, A. (2014). A randomized-controlled trial of the Triple P Positive Parenting Program Seminar Series with Indonesian parents. (Unpublished doctoral thesis). University of Queensland).3. Please also provide more information regarding any existing policy frameworks or initiatives in Indonesia that might facilitate the eventual take up of the program should the study show positive results.4. Please include any existing evidence of the effectiveness of the program or its parent program, Positive Discipline in Everyday Parenting. Is there a citation for the pilot and/or adaptation study of the program that was conducted in 2015?5. Please also include a specific hypothesis, or hypotheses that are connected to the primary objective of the study at the end of this section (as recommended by the SPIRIT guidelines). <p>Methods</p> <ol style="list-style-type: none">1. Study design: The study design using a cluster randomized controlled trial with a 1:1 allocation ratio is justified given the potential for contamination and rivalries within villages.2. Communities: Please explain the reasoning behind selecting the Cianjur District in West Java. Was this due to the NGO or government preference?3. Participants: It would be helpful to define what the investigators mean by requiring participants to "have never engaged in another parenting program," in terms of what constitutes a parenting program. Would caregivers who attended a one-day workshop or once-off lecture on parenting be ineligible? The eligibility criterion
--	--

	<p>requiring parents to read Bahasa may limit recruitment to only more highly-educated families who are literate. This may exclude those who are more at risk in using physical and verbal abuse as discipline strategies given the strong associations between poverty and maltreatment. Please also justify the age range of selecting families with children from birth to seven years of age. Lastly, it is concerning that the investigators have already resigned themselves to recruiting a sample that is predominantly women, especially since male caregivers often have more authority and control in Indonesian families. Even though male recruitment is an issue that studies on parenting interventions face throughout the world, it is advisable that investigators explore strategies to recruit men into the study, or at least to include them in the delivery of the intervention.</p> <p>4. Randomization and allocation concealment: It is understood that the study will include stratified randomization procedures to a) select 20 clusters out of eligible cluster, b) to allocate selected clusters to intervention and control groups, and c) select participating families from within selected clusters. Although the authors adequately explain the allocation to intervention and control groups, more information is required to describe how the stratified random selection of 20 clusters out of the 38 eligible clusters will be conducted, and the selection of participating families. What methods will be used and how will selected villages/families be informed. How many urban and rural villages are part of the original sample of 38 eligible clusters (i.e., what proportion of each)? Regarding the selection of families within villages, what measures will be used to assure that selection does not create tensions or inter-family conflict due to some families being excluded from the study?</p> <p>5. Intervention and control sites sections should be included in the Randomization and Allocation Concealment section.</p> <p>6. Intervention sites: The authors state that 12 caregivers will participate in each group session. Does this mean that there will be three parent groups per village? If so, please state explicitly.</p> <p>7. Control sites: It is premature to state that participants in the control arm will be offered the intervention, especially given that there is no evidence of intervention effectiveness and results may show either harmful or null effects, or mixed results. Are there any alternatives that might be offered?</p> <p>8. Program delivery and training: More detail is needed here. Please provide details on the adaptation of the Families First Program from the original PDEP curriculum. How was it adapted and what were some of the key adaptation to the program? Is it already being delivered in Indonesia? What is the theory of change and underlying principles regarding how it is hypothesized to reduce violence against children? More details are also required regarding the program content, delivery process, and structure. Does it use group discussions or a more didactic approach? How are community facilitators recruited and what qualifications are necessary prior to training? Do they live in the community? Who are the mentors and what qualifications, training, experience do they have? How long will the supervision sessions be each week, and what is the nature of support provided by program developers? There is also a grammatical error in the third sentence of section, "The home visits will use visual materials and role plays to promote positive...."</p> <p>9. Measurement strategy and measures</p> <p>a. The decision to use outcome indicators instead of complete</p>
--	--

	<p>measures is an unusual one. Please provide more justification and citations of the approach used to create these indicators.</p> <p>b. It is unclear how meeting with local authorities and community councilors will increase sample retention. Please provide more explanation.</p> <p>c. Instead of randomly selecting one child per family, it may be more effective to have the families self-select the included child based on the child whose behavior is most problematic to the caregiver since this child might have an elevated risk of being maltreated.</p> <p>10. Outcomes</p> <p>a. The age range of the children (0-7) is also concerning given the challenges in measuring child behavior and parenting for the children under the age of 2 years. For example, the Alabama Parenting Questionnaire is normally used for parents of children ages 6-18 years (although there also exists a version for younger children). Likewise, the Parenting of Young Children Scale was developed for children ages 2 to 9, and the Strengths and Difficulties Questionnaire does not measure child behavior for those under the age of 2 years. Please explain how you will account for potential measurement issues given that the study is focused on families with children ages 0-7 years.</p> <p>b. Please provide examples of items for the explanatory and exploratory outcomes, as well as minimum and maximum scores.</p> <p>c. Please report whether the measures have been used previously in Indonesia, and if so, please report their psychometric properties, including reliability and validity.</p> <p>11. Blinding: While it may be impossible to conduct a double-blind study in which the participants are not aware of their allocation status, this reviewer disagrees with the investigators' assertion that it will not be possible to blind investigators and data assessors to allocation status. There are plenty of examples in which those directly involved in data collection and analysis can be blinded in studies on parenting interventions (e.g., see Lachman JM, Cluver L, Ward CL, Hutchings J, Wessels I, Mlotshwa S, et al. Randomized controlled trial of a parenting program to reduce the risk of child maltreatment in South Africa. <i>Child Abuse and Neglect</i> 2017; 72:338-351). At the very least, it is recommended that the investigators implement measures that will minimize the non-blinding of those involved in data collection and analyses. They should also report any instances of compromised blinding.</p> <p>12. Qualitative and process evaluation assessments: Please provide more detail on the satisfaction questionnaire in terms of number of items, an example of an item, and citation if available.</p> <p>13. Statistical analysis and sample size calculation: Please provide information about how the intra-class correlation was calculated.</p> <p>14. Types of analyses</p> <p>a. The manuscript states that analyses will be conducted "once the intervention period has ended." Do the investigators mean once the data collection has been completed? If not, please explain.</p> <p>b. Please explain the reasons behind nesting individuals within families when assessments are only going to be one parent per family.</p> <p>c. It is commendable that the authors have elected to use an intention-to-treat approach in which all persons who complete baseline assessments will be analyzed at post-test and follow-up regardless whether they participated in the additional assessments</p>
--	--

	<p>or attended the program (for the intervention group villages). However, the strategy using Last Value Carried Forward to account for missing data is susceptible to detection bias and not considered a robust approach. Instead, this reviewer recommends using either multiple imputation or sensitivity analyses to account for the missing data at each time point.</p> <p>d. Please provide information regarding power calculations for gender-based and other subgroup analyses.</p> <p>e. It is commendable that the authors will adhere to the CONSORT guidelines for reporting on randomized controlled trials (including the recommendations to publish and register trial protocols!).</p> <p>f. Please provide more detail regarding the qualitative analysis approach in terms of type of analysis and coding strategy.</p> <p>15. Data collection and data management sections should be placed before the analysis section. The manuscript is also missing a section specifically focused on data collection. Please include this section and describe in greater detail the selection and training of data collectors, as well as the development, translation, piloting, and administering of questionnaires.</p> <p>16. Ethics and Dissemination</p> <p>a. ClinicalTrials.gov registry should be placed at the beginning of the methods section.</p> <p>b. Please provide more information regarding the referral protocols in terms of possible disclosure of harm to either parents or children.</p> <p>c. Please provide more details about the protocol in the possibility that the investigators find negative effects of the intervention on the wellbeing of adults or children either during data collection, intervention implementation, and data analyses.</p> <p>Discussion</p> <p>This section requires a more detailed discussion regarding the potential strengths and limitations of the study design and its potential policy implications, contribution to science and practice, and ultimate benefit for the welfare of children in Indonesia, especially since it is the first study of its kind to be implemented using a cluster randomized controlled trial to test the effectiveness of a parenting intervention in the region.</p> <p>Competing Interests Statement</p> <p>The authors should state more explicitly potential conflicts of interest since Save the Children, the developer and implementer of the Families First Program and Positive Discipline in Everyday Practice, is a major donor to the study and one of the authors (Lilley) is a representative from the organization.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 (Philip Wilson, University of Aberdeen, Scotland)

3. This is a very important study, and in general the proposal as described appears to be adequate for getting an answer to the research questions posed. I do however have two significant concerns, one

about the power calculation and the other about inadequate description of the process evaluation and qualitative research. In both cases it would be useful if the authors could refer to the 2008 MRC Framework for the Evaluation of Complex Interventions, and the 2015 MRC guidance for process evaluations of complex intervention trials.

Thank you for the positive assessment of our study. The two documents indicated have been consulted and used as references where appropriate.

4. My concern about the sample size calculation is that an intraclass correlation coefficient of 0.02 is assumed, and no evidence is presented for this assumption. It is very likely that the figure will be higher than this and the study power consequently lower than described. An independent data monitoring committee (referred to in the SPIRIT checklist but not actually mentioned on page 17) could potentially examine the baseline data calculate the ICC and advise the research team on whether additional villages should be recruited at the outset.

We based the ICC on work previously done by R Platt, the statistician on this project, on a clustered-randomized trial in Belarus [1]. He found this same small magnitude of correlation among outcomes in clinic clusters. A review of ICC values in public health also have values of this magnitude [2]. This value was set a priori. Relevant references have now been provided on p. 19 to justify this decision.

5. The process evaluation is very poorly described. The logic model could be used as a framework to help the authors determine which mediator and moderator analyses might be conducted. What process data will be collected? Completion rates of sessions (total as well as by demographic characteristics)? Parent satisfaction? Fidelity of delivery? Etc. Furthermore the qualitative interviews and focus groups are mentioned (mainly in the flow diagram) but no detail is given. The reader needs to be able to understand the point of the qualitative evaluation. Who will be interviewed? How many interviews will be required? What topics might need to be covered? What method of analysis will be used? A good process evaluation of a groundbreaking trial like this is essential whether the results are positive or negative. In either cases the process evaluation will inform commissioners and practitioners about what contextual and operational factors will be most likely to lead to successful implementation in other settings.

We thank the reviewer for these important points. Figure 2 is a theoretical model for how we hypothesize the intervention impacts on physical and emotional punishment. The variables in the pathway from intervention to outcome (punishment) are considered mediators of the effect of intervention on punishment. The contextual factors are considered to be effect modifiers or confounders. We will use structural equation modeling (SEM) to test whether this model fits the data. In other contexts, such as an observational study, variables like parental mental health could be mediator variables but not in this context as the intervention was not specific for this construct. As we have a strong theoretical model, we will not use statistical approaches to determine the role of the variables in the causal pathway, we will test this model. One of our investigators has considerable experience with SEM (see Mayo N structural equation modeling in pubmed).

The sections "Process evaluation and qualitative assessments" (p. 16-18) as well as "Types of analysis" (p. 19) have been expanded within the limited space provided and now provide more detailed information on these study components.

More minor points

6. There should be some discussion of how likely self-report questionnaires are to give honest answers about child maltreatment

Enumerators are carefully trained and supervised not to pass on judgement when administering the questionnaire. Although some level of violent discipline is normative in the study area (and therefore socially accepted), we are aware that some bias may emerge due to social desirability, particularly following the intervention (for respondents in intervention group as well as for controls as a result of repeated administration of measures).

7. Readers will be interested to know about 'village leaders'. Are they easy to identify?

Village leaders are elected government officials who play a key role in the implementation of any local intervention. Without their approval and buy in at the local level, it would be very difficult to proceed with the intervention, even more so since half of the villages are placed in the waiting list as a result of randomization. In terms of data management, analysis, and reporting, attention will be paid to protecting confidentiality, should data be collected from any of them during key informant interviews.

8. The timeline of the study should be described.

The sequence of activities as well as duration of each phase is specified in the study flow chart (Figure 1). More details about the progress of the study are provided in the "Timeline" section at the end of the manuscript (p. 24) and in the Letter to the Editor.

9. The method by which baseline covariates will be used to adjust outcomes is not specified clearly. Further detail would be useful.

Adjustment for covariates will be done by fitting a multi-variate model to include any variables that are not balanced by the randomization (p. 19).

Reviewer 2 (Jamie M. Lachman, Department of Social Policy and Intervention, University of Oxford, United Kingdom)

Overall

10. Overall, This manuscript provides a very important contribution to the science regarding the testing of parenting programs in low- and middle-income countries. The proposed protocol demonstrates how researchers can conduct rigorous evaluations of the effectiveness of parenting programs on the reduction of violence against children in low resource areas where there is limited evidence. It also addresses key policy issues regarding the global prevention of violence against children and the need for evidence-based interventions that meet this need. The manuscript is well-written and organized, and closely adheres to the SPIRIT guidelines for protocols of intervention trials. It also meets the high standard for scientific excellence of the BMJ Open. After addressing the specific comments discussed below from each section, this reviewer is happy to recommend the manuscript for publication.

Thank you very much for this encouraging assessment of our study and the thorough review of our manuscript.

Abstract

11. Please specify that villages in the intervention group will also receive standard community health and social services as stated in the body of the manuscript.

The abstract now specifies that the intervention group will also receive standard community health and social services (p. 2).

12. Even though the manuscript is a protocol and thus does not have results to report, please include information regarding the potential scientific contribution and policy implications of the study regardless of whether results are positive or negative.

A reference to the potential contribution of this study beyond the results on this particular program outcomes is briefly mentioned in the abstract (p. 3).

Strengths and Limitations of the Study

13. Please be more explicit about the last limitation referring to potential biases as a result of social desirability due to the reliance of the self-report outcome measurements.

This limitation has been more clearly indicated on p. 4.

14. Please also be more explicit regarding the lack of blinding, although this may be possible at post-test.

This limitation has been more clearly indicated on p. 4.

Introduction

15. The introduction provides a clear overview of the background of violence against children in LMICs, the role of positive parenting and effective discipline in reducing risks of child maltreatment, and the extent of evidence of effectiveness of parenting programs in LMICs, and the East Asia Pacific region in particular. The section on risk factors for violence against children could include more research that is specific to Indonesia, or if lacking, the region. Similarly, please include any evidence on the effect of child maltreatment that is particular to Indonesia.

Additional references re. risk factors and effects within the country and/or region have been added to both sections (pp. 4-5).

16. It would be helpful to have more discussion regarding the current evidence of effectiveness of parenting programs in Indonesia. For instance, there was a recent randomized controlled trial of a parenting program in Indonesia that might provide a useful context and justification for additional research in the country and region (see Sumargi, A., Sofronoff, K., and Morawska, A. (2014). A randomized-controlled trial of the Triple P Positive Parenting Program Seminar Series with Indonesian parents. (Unpublished doctoral thesis). University of Queensland).

Thank you for bringing this work to our attention. We have incorporated it into this section to further support the need for further work on parenting in this country and region (p. 6).

17. Please also provide more information regarding any existing policy frameworks or initiatives in Indonesia that might facilitate the eventual take up of the program should the study show positive results.

This is now mentioned on p. 7.

18. Please include any existing evidence of the effectiveness of the program or its parent program Positive Discipline in Everyday Parenting. Is there a citation for the pilot and/or adaptation study of the program that was conducted in 2015?

References to existing evidence of effectiveness (pretest-posttest evaluation of the parent program in Canada) and relevance are included on p.6. Unfortunately, the results from the pilot have not been published yet those were shared with our evaluation team and used by SC and the program developers to refine the local program before the trial.

19. Please also include a specific hypothesis, or hypotheses that are connected to the primary objective of the study at the end of this section (as recommended by the SPIRIT guidelines).

Specific hypotheses have now been included on p. 7.

Methods

Study design

20. The study design using a cluster randomized controlled trial with a 1:1 allocation ratio is justified given the potential for contamination and rivalries within villages.

We concur with the reviewer that this design responds adequately to contextual circumstances.

Communities

21. Please explain the reasoning behind selecting the Cianjur District in West Java. Was this due to the NGO or government preference?

Cianjur has been a priority area for intervention for SC given important child protection concerns such as it being a major sending area of child trafficking, child prostitution, women migrant workers, early child marriage as well as high incidence of divorce, and "contract marriage". Cianjur is also the first prone area of disaster in Indonesia, and its population has generally low education status and high of unemployment. This is now mentioned on p. 8.

Participants

22. It would be helpful to define what the investigators mean by requiring participants to "have never engaged in another parenting program," in terms of what constitutes a parenting program. Would caregivers who attended a one day workshop or once off lecture on parenting be ineligible?

Parenting programs are not common in this setting. The only common activity by Posyandu (integrated government health services) is rather locally referred to as "socialization" and is focused on child health and development rather than parenting. Nonetheless, as a safeguard, caregivers are asked whether they had participated in any parenting program, regardless of length.

23. The eligibility criterion requiring parents to read Bahasa may limit recruitment to only more highly educated families who are literate. This may exclude those who are more at risk in using physical and verbal abuse as discipline strategies given the strong associations between poverty and maltreatment.

The reviewer raises a good point re. literacy and we will keep this in mind when interpreting results. Only a basic reading level will be required so that participants can provide written consent; the information on the informed consent form will be verbally summarized for participants and they will have opportunities to ask questions in the local language.

24. Please also justify the age range of selecting families with children from birth to seven years of age.

The focus on children aged 0-7 years fits within the larger program of the implementing agency that focuses on early childhood education and considers 0-7 years as basic foundation of parenting and also character building of children. Positive discipline is important to practice since

delivering the baby. Government program on parenting in Indonesia does not yet use a positive discipline approach, rather it puts more emphases on health and early stimulation. Providing positive discipline in this early age will provide comprehensive parenting program link with other government program on health and early stimulation. SC also aims to reduce institutionalization of children, which often occurs at the time they reach school-going age. One sentence has been added on p. 8 in this regard.

25. Lastly it is concerning that the investigators have already resigned themselves to recruiting a sample that is predominantly women especially since male caregivers often have more authority and control in Indonesian families. Even though male recruitment is an issue that studies on parenting interventions face throughout the world it is advisable that investigators explore strategies to recruit men into the study or at least to include them in the delivery of the intervention.

We concur with the reviewer that more work is needed with fathers and male authority figures in efforts to curtail violence against children. In fact, the implementation and evaluation teams discussed at length the issue of involving fathers in the program and the trial. Challenges sustaining the participation of fathers due to work and cultural barriers were identified in this setting. In the end, a decision was made to exclude fathers in this run of the program for several reasons: this would require significant additional human and financial resources (including for training of facilitators and delivery of the program), further extend the timeline (difficulty in recruiting male facilitators), and likely create challenges in the implementation of the program (e.g., adapting schedules to male/female availability), and the pilot program had been delivered with female caregivers only. Given the focus on younger children, who are more often under the purview of females, this decision seemed feasible and otherwise justified. Nonetheless, fathers are invited to participate in the home visits. This decision was shared with the local Advisory Committee, who approved of the design. Limited resources for implementation of the program and evaluation set boundaries to the selection of participants yet the implementing agency is aware of this issue and may consider involving males in the future. The perspectives of key male stakeholders (village authorities, program staff, etc) will be captured through the qualitative assessment.

Randomization and allocation concealment

26. It is understood that the study will include stratified randomization procedures to a) select 20 clusters out of eligible cluster b) to allocate selected clusters to intervention and control groups and c) select participating families from within selected clusters. Although the authors adequately explain the allocation to intervention and control groups more information is required to describe how the stratified random selection of 20 clusters out of the 38 eligible clusters will be conducted and the selection of participating families. What methods will be used and how will selected villages/families be informed.

As the reviewer rightly points out, the allocation of clusters to intervention/control groups and the selection of families within each cluster from the initial list of potential participants is done following strict randomization procedures as explained on p. 9. However, as it was described in the section "Procedures", the selection of 20 clusters out of 38 is done between implementers and evaluators by carefully analyzing all the inclusion criteria rather than randomly from a list of clusters. Further details are now provided regarding the selection and allocation of villages and families on pp. 9-11.

27. How many urban and rural villages are part of the original sample of 38 eligible clusters (i.e. what proportion of each)?

There are 65% rural and 35% urban villages in the 47 villages in the four sub-districts in our sample. In reviewing background files for this information we realized that the 38 village count excluded one of the subdistricts; the correct total of villages is 47. This has now been corrected in the manuscript. This has been clarified on p. 9.

28. Regarding the selection of families within villages, what measures will be used to assure that selection does not create tensions or inter family conflict due to some families being excluded from the study?

Enumerators are trained to respond to inquiries by community members and to explain the exploratory nature of this round of the program (i.e., we want to find out how well it works so that decisions can be made about its future). Attention will be paid to keep good communication with village authorities and service providers (cadres) so that they are also able to convey the same message.

Intervention and control sites sections

29. Intervention and control sites sections should be included in the Randomization and Allocation Concealment section.

This change has been made (p. 10).

Intervention sites

30. The authors state that 12 caregivers will participate in each group session. Does this mean that there will be three parent groups per village? If so, please state explicitly.

That is exactly right. This has now been made explicit on p. 10.

Control sites

31. It is premature to state that participants in the control arm will be offered the intervention especially given that there is no evidence of intervention effectiveness and results may show either harmful or null effects or mixed results. Are there any alternatives that might be offered?

Although the program is offered to all participants when obtaining informed consent, we will indeed share results of the trial with the implementing agency in advance to the delivery of the program to the control group. An alternative has not been offered but we appreciate this comment and we will discuss with the implementing agency and the IRB, as needed.

Program delivery and training

32. More detail is needed here. Please provide details on the adaptation of the Families First Program from the original PDEP curriculum. How was it adapted and what were some of the key adaptation to the program? Is it already being delivered in Indonesia? What is the theory of change and underlying principles regarding how it is hypothesized to reduce violence against children? More details are also required regarding the program content delivery process and structure. Does it use group discussions or a more didactic approach? How are community facilitators recruited and what qualifications are necessary prior to training? Do they live in the community? Who are the mentors and what qualifications training experience do they have? How long will the supervision sessions be each week and what is the nature of support provided by program developers? There is also a grammatical error in the third sentence of section "The home visits will use visual materials and role plays to promote positive...."

This section now provides more information about the intervention being evaluated, including bibliographic reference (p. 10-12). The grammatical error has now been corrected.

Measurement strategy and measures

33. The decision to use outcome indicators instead of complete measures is an unusual one. Please provide more justification and citations of the approach used to create these indicators.

The measures collectively contain more than 180, items many of which are redundant, ambiguous, and idiomatic, and it was shown by our pretesting in the field that this volume of items was

borderline feasible. In addition, it is well known that summing ordinal items to create total scores is not optimal as the values assigned to categories are not actual numbers but labels and do not represent quantities; summing also assumes each contributes equally to the total score and this is also unlikely. Our measurement approach is to use the items from the measures to form constructs; some of these constructs will be formative constructs (summarized by single indicators or a count) and some will be reflective. For constructs that fit a reflective conceptual model, Rasch Analysis will be used to test the extent to which the items fit a uni-dimensional hierarchical model with interval-like properties such that the total score can be used in mathematical transformations. There is no room in this protocol paper to present details on this approach but there is a vast literature on modern psychometric approaches such as Rasch analysis or Item response theory (IRT). Items that are crucial to the measurement approach but that do not fit the Rasch model can be used as single indicators. E.g. "How do you rate your health? (Excellent very good, good, fair, poor) is one of the most widely used single indicators of health. A brief explanation and several references [3-5] have been added to the protocol on p. 13.

34. It is unclear how meeting with local authorities and community councilors will increase sample retention. Please provide more explanation.

Thank you for the opportunity to further clarify this point. Since several months will pass between measures and local authorities (including community volunteers) play a key role in communicating with residents, our plan is to visit the communities shortly before the survey team arrives to remind them of the ongoing evaluation (this is particularly crucial in control villages, as no further contact with those villages will occur in between measures) and answer any questions they may have in this regard. Local authorities will be, in turn, able to inform participants that the team will be in the community in the days to follow. We have reworded the sentence and added another one for further clarity (p. 13).

35. Instead of randomly selecting one child per family, it may be more effective to have the families self select the included child based on the child whose behavior is most problematic to the caregiver since this child might have an elevated risk of being maltreated.

By randomly selecting an index child we want to gain protection from any bias that may result from the selection of children for a range of other reasons (e.g., a caregiver may select an orphan or a non- biological child regardless of the problematic nature of their behavior). This could also result in a skewed distribution of age groups.

Outcomes

36. The age range of the children (0-7) is also concerning given the challenges in measuring child behavior and parenting for the children under the age of 2 years. For example, the Alabama Parenting Questionnaire is normally used for parents of children ages 6 -18 years (although there also exists a version for younger children). Likewise, the Parenting of Young Children Scale was developed for children ages 2 to 9, and the Strengths and Difficulties Questionnaire does not measure child behavior for those under the age of 2 years. Please explain how you will account for potential measurement issues given that the study is focused on families with children ages 0- 7 years.

We agree with the reviewer that measuring child behavior and parenting for very young children is challenging. Thus, some non-primary outcomes will not be measured for children under the age of 2 years (e.g., child social and emotional wellbeing) and some outcomes have slightly different set of items. Because inadequate understanding of child development and children's rights can trigger negative disciplinary reactions in parents of children of all ages and good parenting practices are best learned early (in fact, the PDEP program is aimed at parents of children of all ages), we did not want to exclude parents of infants from the program and the study. Instead, we did our best to select measures that had been validated in LAMICs with caregivers of young children. The local team validated the appropriateness of the measures for different age groups during the pilot testing of the questionnaire. Even if certain items were originally developed for somewhat older children (e.g., Speak calmly with your child when you were upset with him or

her?) if it was deemed reasonable for a younger child, the question was retained. Else, skip logic was applied for younger children. As can be seen in the Appendix, in some cases, some items are reserved for certain age groups (e.g., “Your child comes home from school more than an hour past the time you expect him/her” is only asked to 5-7 year-olds). We will ensure adequacy of selected items for a range of age groups when selecting indicators.

37. Please provide examples of items for the explanatory and exploratory outcomes as well as minimum and maximum scores.

All the items and scales used are provided in the Appendix. Given the current length of the manuscript, we have referred the readers to the Appendix but can provide examples in this section if the Editor so requests.

38. Please report whether the measures have been used previously in Indonesia and if so please report their psychometric properties including reliability and validity.

There is a lack of validated measures to study parenting by Indonesian parents. Although some measures we use have previously been used in Indonesia (e.g., SDQ, MICS), their psychometric properties are not reported in these studies. In other cases, the psychometric characteristics of the measure have been studied in another South-East Asian country. Whenever the measure has been used in Indonesia or the psychometric properties are studied in the region, we have added their bibliographic references (pp.13-15), yet we have opted for not reporting psychometric properties that are not available specifically for Indonesia. Note that the questionnaire was piloted and refined prior to start of trial.

Blinding

39. While it may be impossible to conduct a double blind study in which the participants are not aware of their allocation status, this reviewer disagrees with the investigators assertion that it will not be possible to blind investigators and data assessors to allocation status. There are plenty of examples in which those directly involved in data collection and analysis can be blinded in studies on parenting interventions (e.g., see Lachman JM, Cluver L, Ward CL, Hutchings J, Wessels I, Mlotshwa S, et al. Randomized controlled trial of a parenting program to reduce the risk of child maltreatment in South Africa. *Child Abuse and Neglect* 2017; 72:338-351). At the very least, it is recommended that the investigators implement measures that will minimize the non blinding of those involved in data collection and analyses. They should also report any instances of compromised blinding.

Although we agree with the reviewer that blinding researchers conducting self-report interviews as to allocation would be best, the feasibility of so doing was carefully discussed with our local partners in Indonesia and realized that this would be very difficult if not impossible to guarantee given normative patterns of communication and the fact that although no interviewer was hired from the participating communities, many originated in this District due to language skills selection criteria (fluency in Sundanese required from interviewers). Instead, special attention was paid to discussing researcher bias during all the trainings that preceded data collection as well as frequent reminders and direct observation of interviews (for quality check) by field supervisors throughout data collection. Emphasis was placed in the need for interviewers not to pass judgement, comment on the program or what they had heard about it with any participants, and report any instances of compromised blinding to their supervisor and in the ‘Notes’ section at the end of each interview. We did not insist on analysing the data blind as the data analysis is overseen by a group of experienced statisticians and epidemiologists who will validate that the raw data and the results of the modeling are concordant.

Qualitative and process evaluation assessments

40. Please provide more detail on the satisfaction questionnaire in terms of number of items an example of an item and citation if available.

The program satisfaction questionnaires for caregivers, facilitators, and mentors have now been described on p. 16-17.

Statistical analysis and sample size calculation

41. Please provide information about how the intra class correlation was calculated.

Please see response to Q4 above.

Types of analyses

42. The manuscript states that analyses will be conducted "once the intervention period has ended." Do the investigators mean once the data collection has been completed? If not please explain.

That is correct. Statistical analyses of outcomes will be conducted once data collection has been completed. This has been more clearly stated on p. 19.

43. Please explain the reasons behind nesting individuals within families when assessments are only going to be one parent per family.

We thank the Reviewer for pointing this out. This was an oversight from an earlier draft of our proposal. We dealt with this by randomly selecting one child per family. This has now been corrected in the manuscript.

44. It is commendable that the authors have elected to use an intention to treat approach in which all persons who complete baseline assessments will be analyzed at post test and follow up regardless whether they participated in the additional assessments or attended the program (for the intervention group villages). However, the strategy using Last Value Carried Forward to account for missing data is susceptible to detection bias and not considered a robust approach. Instead, this reviewer recommends using either multiple imputation or sensitivity analyses to account for the missing data at each time point.

We have pondered this consideration and decided to use multiple imputation. This has been changed on p. 20.

45. Please provide information regarding power calculations for gender based and other subgroup analyses.

Only planned subgroup analyses are considered in sample size calculations. We will conduct gender-based analysis for exploration, not for inference.

46. It is commendable that the authors will adhere to the CONSORT guidelines for reporting on randomized controlled trials (including the recommendations to publish and register trial protocols!).

Despite some staff and timing challenges faced last year, the team is committed to sharing learning from this study with decision-makers and the wider scientific community.

47. Please provide more detail regarding the qualitative analysis approach in terms of type of analysis and coding strategy.

Further information on the qualitative analysis can now be found on pp. 20.

Data collection and data management sections

48. Data collection and data management sections should be placed before the analysis section. The manuscript is also missing a section specifically focused on data collection. Please include this section and describe in greater detail the selection and training of data collectors as well as the development translation piloting and administering of questionnaires.

The (new) Data collection section and the Data management section now precede the analysis sections (pp. 17-18).

Ethics and Dissemination

49. ClinicalTrials.gov registry should be placed at the beginning of the methods section.

A new sub-heading on “Trial Registration” has now been created in the Methods section (p. 7).

50. Please provide more information regarding the referral protocols in terms of possible disclosure of harm to either parents or children.

All team members in contact with families will undergo special training on the Save the Children Safeguarding policy and procedures which clearly detail the steps to follow should any abuse or harm be observed or suspected. We have provided a reference to this policy on p. 21.

51. Please provide more details about the protocol in the possibility that the investigators find negative effects of the intervention on the wellbeing of adults or children either during data collection intervention implementation and data analyses.

Prior administration of this program in Indonesia and the pilot did not show evidence of any harm. Nonetheless, should any harm be identified, investigators and data collectors are familiar with the SC Safeguarding policy, which applies as approved by the Ethics Boards involved, and the PI and partner NGO will be alerted (p. 21) and action taken accordingly.

Discussion

52. This section requires a more detailed discussion regarding the potential strengths and limitations of the study design and its potential policy implications contribution to science and practice and ultimate benefit for the welfare of children in Indonesia especially since it is the first study of its kind to be implemented using a cluster randomized controlled trial to test the effectiveness of a parenting intervention in the region.

This section has been reworked to highlight potential contributions of this trial (p. 23). Given space constraints, however, readers are referred to the Strengths and Limitations section for more detailed information on those particular points.

Competing Interests Statement

53. The authors should state more explicitly potential conflicts of interest since Save the Children the developer and implementer of the Families First Program and Positive Discipline in Everyday Practice is a major donor to the study and one of the authors (Lilley) is a representative from the organization.

The conflict of interest statement has been reworded for further clarity re. the role of Save the Children in the study (p. 30).

Works Cited

1. Kramer, M., et al., *Effect of prolonged and exclusive breast feeding on risk of allergy and asthma: cluster randomised trial*. BMJ : British Medical Journal, 2007. **335**(7624): p. 815.
2. Murray, D.M. and J.L. Blitstein, *Methods to reduce the impact of intraclass correlation in group-randomized trials*. Evaluation Review, 2003. **27**: p. 79-103.
3. Grimby, G., A. Tennant, and L. Tesio, *The use of raw scores from ordinal scales: time to end malpractice?* . J Rehabil Med, 2012. **44**(2): p. 97-8.
4. Bond, T. and C. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 3rd ed. 2015, New York: Routledge.

5. Tennant, A. and P. Conaghan, *The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper?* . Arthritis Rheum, 2007. **57**(8): p. 1358-63

VERSION 2 – REVIEW

REVIEWER	Jamie M. Lachman University of Oxford, United Kingdom
REVIEW RETURNED	27-Jul-2018
GENERAL COMMENTS	I commend the authors for adequately addressing the concerns raised in the previous review within the word limitations of the manuscript. It is essential that study protocols for research in low- and middle-income countries are published and widely accessible in order to expand our knowledge base in low-resource settings. I believe that the manuscript should be accepted for publication.

