Author’s response to reviews

Title: A Randomized Controlled Trial comparing Parent Child Interaction Therapy - Toddler, Circle of Security– ParentingTM and Waitlist controls in the treatment of disruptive behaviors for children aged 14-24 months: Study Protocol

Authors:

Jane Kohlhoff (Jane.Kohlhoff@unsw.edu.au)
Sara Cibralic (s.cibralic@unsw.edu.au)
Nancy Wallace (nancy.wallace@unsw.edu.au)
Susan Morgan (Susan.Morgan3@health.nsw.gov.au)
Cathy McMahon (cathy.mcmahon@mq.edu.au)
Erinn Hawkins (erinn.hawkins@gmail.com)
Valsamma Eapen (v.eapen@unsw.edu.au)
Nancy Briggs (nancy.briggs@unsw.edu.au)
Anna Huber (ahuber@familiesinmind.com.au)
Cheryl McNeil (cmcneil@wvu.edu)

Version: 2 Date: 28 Jul 2020

Author’s response to reviews:

Reviewer 1

Comment #1. In general, the outcome measures chosen appear well-justified given the theoretical framework for both interventions as well as the theoretical framework of the study. The only exception may be Parental Emotion Regulation (p. 20). It is not clear from the theoretical framework for the study as well as the theoretical basis for the interventions how these interventions would directly affect parental emotion regulation. Therefore, should the efficacy of the interventions be assessed on the basis of this outcome? This is all the more important given the large ratio of outcome measures against sample size.

Response: Thank you for raising this point. We note, however, that one of the specific aims of the PCIT-T intervention is improving parental emotion regulation ability. The program does this using a number of targeted intervention techniques including “check in” sheets and discussion with the parent at the start of every coaching session (which include questions related to emotion
regulation), and incorporation of an “Adult CARES” module, which teaches the parent a series of steps/techniques that can be used to promote their own emotional regulation. We would like to draw the reviewer’s attention to the following paragraph (page 9), which provides an outline of the adult CARES module:

“Throughout both phases, to enhance parents’ ability to effectively implement the CDI-T and PDI-T skills, parents are also encouraged to develop their own emotion regulation skills through application of a parallel adult-focused C-A-R-E-S model [Coming in close to your own feelings, Assisting yourself to manage (e.g., deep breathing, progressive muscle relaxation), Reassuring yourself (cognitive challenging), Labeling Emotions experienced, Soothing yourself (e.g., self care)].”

To provide information about the emphasis in the “check ins” on parental emotion regulation, we have also now added the following sentence to the same paragraph: “All sessions begin with a brief “check in” conversation in which the therapist and parent discuss and problem-solve any issues around home practice, and also reflect on the parent’s experiences using the PRIDE and CARES skills (for both themselves and the child) during the week.” (page 9)

To further articulate this we have now added in the following sentence to the introduction: “The program aims …. to improve the parent’s own emotional regulation skills, ….” (page 8)

Comment #2. In general, the methods are well described and document. One detail that was missing on page 16 is the number of therapists involved in offering COS-P. This detail is important given the potential for therapist effects and the opportunity for assessing such effects.

Response: Thank you for raising this important point. We have now added the following sentence to the section about the COS-P groups: “In total, approximately 6 different therapists will be involved as facilitators for the COS-P groups.” (page 19)

Comment #3a. Linear mixed model repeated measures analysis of variance for the three time conditions (p. 23) is generally appropriate for testing effects of interventions, but important details are missing to assess how the authors deal with some of the intricacies of their approach.

– For example, the WLC-group is not included in T3. So are the authors planning a test of their hypotheses on the basis of T1-T2 and a test of the hypothesis of superiority of PCIT-T over COS-P on the basis of T1-T3?

Response: Yes, given that there will be no non-treated control group for the T3 time point, we will test the major hypotheses at T2 and then will test a second hypothesis that PCIT-T remains superior to COS-P at the T3 follow-up. We have now articulated this more explicitly in the Introduction with extra details about the second hypothesis, as follows: “Second, we expect that for the two active treatment conditions, this gradient effect will be maintained at a 4-month post-treatment follow-up, whereby parent-child dyads that receive PCIT-T will show superior outcomes on all variables compared to parent-child dyads where the parent received COS-P.” (page 13)

To further clarify this point, we have also added the following section to the analysis section: “Group by time interactions will be examined in fixed effects models, with planned comparisons
constructed to examine the differences between the three groups at T2, and between the PCIT-T and COS-P groups at T3.” (page 26)

Comment #3b – Will a separate analysis be conducted for the outcomes that were not included in T2 (i.e., the SSP)?

Response: Yes, a separate analysis will be conducted for the SSP, both because it is the only categorical measure in the study and also because it is the only measure that will not be administered at T2. This has been described in the text on page 27, now with an added note about the fact that the SSP will not be administered at T2, as follows: “To examine changes in categorical SSP classifications from T1 to T3 (SSP not administered at T2), generalized linear mixed models with binomial distribution and logit link will be conducted with insecure attachment (0/1) and disorganized attachment (0/1) as dependent variables.”

Comment 3c – What are the decision rules if the results from these different hypothesis tests are (partially) conflicting?

Response: We have not specified decision rules to define the performance of PCIT-T over the other groups. However, a consistent pattern of better outcomes across multiple measures and clinically interesting change will support the idea that PCIT-T is a good, perhaps preferable option for these children.

Comment #3d – Furthermore, it is unclear how the hypothesis of a gradient effect would be tested in the T1-T2 analysis of WLC, COS-P, and PCIT-T. Are the authors using planned contrasts of just testing whether the within (time) x between (group) interaction factor is significant?

Response: Yes, to test the hypothesis of a gradient effect from T1 to T2, we will examine the group by time interaction in a fixed effects model, followed by planned comparisons. We will construct comparisons that examine the differences amongst all groups at T2, and for the PCIT-T and COS-P groups at T3. This is now more clearly articulated in the manuscript, with the addition of the following sentence: “The group by time interaction will be examined in the fixed effects model, and planned comparisons will also be constructed to examine the differences between all groups at T2, and for the PCIT-T and COS-P groups at T3.” (page 26)

Comment #3e - The authors plan to use Cohen's criteria for interpreting the clinical significance of their findings. However, what are the arguments for using this benchmark? And how is the benchmark related to effect size expectations? Most research on psychosocial interventions suggests that effect sizes for differences between treatment and waitlist-controls might be inflated due to the nocebo effect for being on a waitlist. In contrast, the difference in effects between two active treatments is likely to be much smaller than for treatment versus no treatment. Therefore, it seems reasonable to use different benchmarks for both these types of effects.

Response: We have chosen to use the Cohen’s criteria (.2 small, .5 medium and .8 large) for interpreting the clinical significance of our findings because it is commonly used and well accepted in intervention research, which will allow us to compare our results to other studies in the area (Durlack, 2009). While we acknowledge that researchers should never blindly use the
benchmarks suggested by Cohen (1988), we believe that for this study, they will provide a useful
indicator of the size of observed effects, and effort will be made to interpret all effect sizes in light
of clinical relevance.

We will use the Cohen’s criteria to interpret the effect size differences on study measures from T1
to T2 (for each of the three groups), and from T1 to T3 (for the PCIT-T and COS-P groups). In
addition, we will use the Cohen’s criteria to interpret the size of the differences between groups at
T1 and T2, and between the PCIT-T and COS-P groups at T3.

Thank you for the interesting point about the possibility of inflated differences between the
treatment and waitlist groups due to a nosebo effect, and of smaller differences between the two
active treatment groups. We agree that this is a likely scenario for our study, and we will keep
these issues in mind when interpreting and discussing observed effect sizes.

Comment 3f - Related to the previous point, it is unclear what the basis is for the effect size
expectation on p. 24 and it is also not clear whether this refers to the absolute efficacy (treatment
versus WLC) or relative efficacy (PCIT-T versus COS-P).

Response: With regards to the basis for our effect size expectation, we have added the following
details to the manuscript: “The effect size f is defined as $f = \frac{\sigma_m}{\sigma}$, where $\sigma_m$ is the SD of the effect
of interest and $\sigma$ is the common error SD. Cohen (1963) suggested that $f$ of 0.25 is a moderate
effect. This effect size was chosen because (a) it is a reasonable and achievable statistical effect,
and (b) it is substantively and clinically meaningful for the primary outcome measures.” (page 27)

As a point of clarification, when we state an expectation of finding a moderate effect size
difference, we are referring to the size of the differences between the three groups, at two time
points (T1 and T2). Thus, we believe that we are referring to both absolute efficacy (i.e., PCIT-T
vs WLC and COS-P vs WLC) and relative efficacy (i.e., PCIT-T vs COS-P).

Comment #3g- More details are needed on the power calculations, given the mixed linear models
that will be employed and the Bonferroni adjustments for family-wise error rate. Also, it is unclear
whether the power calculations refer to difference between any intervention and waitlist (which is
limited to 2 measurement occasions) or difference between the two active interventions (which has
3 measurement occasions but would also have to be based on a much lower effect size).

Response: Power was calculated for the group x time interaction between the three groups, at time
points T1 and T2. Then, assuming the sample size obtained in this first set of calculations, the
power to detect the relevant mean differences at T2 and T3 (adjusting alpha for multiple
comparisons) was calculated.

Comment 3h - The sentence about the post-hoc testing is quite dense and needs fuller explanation,
especially given that the alpha was divided by 4 to adjust for multiple comparisons, but the study
includes 12 outcome variables. The study plan and power analysis might benefit from a
consideration of primary hypotheses and secondary hypotheses, with the primary hypotheses
including the outcomes that would at least have to show significant effects for the intervention to
be favored over alternatives. The tests for these outcomes would not need to be Bonferroni-
corrected, because the whole 'family' of tests falls under the same error rate (.05) of the decision rule. For any secondary hypotheses with less strict decision rules, Bonferroni-style controls of error rate may still be applied. Structuring the hypotheses and outcomes like this would allow for clearer study outcomes while preserving statistical power.

Response: The power calculation was performed on the basis of one primary outcome measure: the externalizing subscale of the CBCL. The adjustment for multiple comparisons in the sample size calculation was thus made for this one primary outcome variable. It was also done using the very conservative Bonferroni correction to ensure that Type I error was minimized for this primary outcome. In the actual analysis, Holm’s stepdown Bonferroni procedure will be used.

We feel that it was appropriate to use the CBCL externalizing subscale score as the primary outcome variable in the power calculation because ultimately, the reason that children will be referred to the trial is because the parent is concerned about the challenging behaviour that the child is displaying. All of the other variables were chosen for inclusion in the study due to a belief that they are, in some way, contributors to child externalising behaviour. Thus, change on any of the study variables will be interpreted as suggestive that the PCIT-T intervention should be favoured over the alternatives. Therefore, apart from viewing the CBCL as prominent due to the abovementioned reasons, we do not feel that there is an obvious order for the variables in terms of importance. For this reason, after careful consideration, we have chosen to not re-structure the hypotheses into primary and secondary hypotheses as per the reviewer’s suggestion. Instead, we have communicated clearly in the manuscript that the sample size calculation was conducted using the CBCL externalizing subscale score (page 26). We have also provided reasons for this decision, as follows: “The CBCL externalizing subscale score was used in the power calculation because of all the measures in the study (a), it is the one that corresponds most directly to the presenting problem for most participants; and (b) it was used in a previous study conducted at the same site and with a similar study sample, allowing us to obtain realistic estimates of mean CBCL externalizing subscale scores for T1, and for the PCIT-T group at T2.” (page 27)

In response to the reviewer’s comment that the sentence about post hoc testing was dense, we have broken this sentence into two sentences, and have added some additional clarifying details. The section now reads as follows: “Post hoc testing of differences between the three groups at T2, and a PCIT-T versus COS-P difference at T3, show that n=30 in each group would give 80% power to detect a mean difference between groups of at least 8.0, with SD of 9. And alpha of .05/4, to correct for multiple comparisons for all group comparisons at T2 (PCIT-T versus COS-P, PCIT-T versus Waitlist, COS-P versus Waitlist) and T3 (PCIT-T versus COS-P).” (pages 27-28)

Comment #4. Regarding the Conflict of Interest statement, I wondered whether the authors have also considered their authorship of the adapted PCIT protocol (Kohlhoff and Morgan) as presenting a possible reason for at least intellectual interest that readers may be made aware of? This also regards the affiliation of several of the authors with the clinical center that is offering the intervention. Although this information is transparently within the manuscript itself, it may be important to review the journal policy on declaring information that has the potential for being perceived as presenting a possible conflict of interest.
Response: This is a good point. We have now amended our conflict of interest statement so that it reads as follows: “Authors JK, NW, SM and CM acknowledge that they are co-developers of the PCIT-T adaptation, and co-authors of the PCIT-T treatment manual (42)”

Reviewer 2

Comment #1. Authors do a good job of describing the ways in which PCIT-T differs from traditional PCIT, although additional details would be appreciated, especially as regards discipline techniques.

Response: We have now added some additional details about the specifics of PCIT-T, so that it can be differentiated from standard PCIT. This includes specific details about the discipline (or as referred to in the paper, the “limit setting” technique), as follows: “In the CDI-T phase the parent is also taught to use “under-reaction and re-direction” in response to disruptive child behaviour judged to be attention seeking (e.g., throwing toys) and to use a brief, developmentally appropriate limit-setting sequence in response to child aggression. The limit-setting sequence involves the parent coming close to the child, taking his or her hands, and, with direct eye contact, saying, “no hurting”. The parent then looks away for 3 seconds, and then looks back at the child and says, “remember, gentle hands” before rotating the child towards the toys and using PRIDE skills (and CARES skills, where needed) to re-engage the child in positive play.” (page 9)

We have also added some extra details to the description of the program on page 9-10, as follows: “Between PCIT-T sessions, parents are encouraged to practice the skills for 5 minutes on a daily basis. All sessions begin with a brief “check in” conversation in which the therapist and parent discuss and problem-solve any issues around home practice, and reflect on the parent’s experiences using the PRIDE and CARES skills (for both themselves and the child) during the week. Following the “check in”, the therapist observes the parent and child for 5 minutes (without providing coaching), and in this time codes the parent’s use of the PCIT-T skills (e.g., PRIDE skills, Don’t skills, CARES skills). This allows the therapist to choose a focus area for the session and to track progress across sessions. The PCIT-T program is mastery-based, meaning that parent-child dyads progress from the CDI-T phase to the PDI-T phase when they show satisfactory use of the CDI-T skills in the 5 minute observational period (i.e., they are observed to use 10 labeled praises, 10 behavioral descriptions and 10 reflections, and to adequately implement the CARES skills when needed). Graduation from the PDI-T phase takes place when the parent can competently implement the guided compliance teaching sequence. For most dyads, when PCIT-T is delivered twice weekly (30-40 mins per session), CDI-T and PDI-T mastery is typically achieved after approximately 10-12 sessions.”

Comment #2. The authors do an excellent job describing the methodology used, and propose appropriate measures and assessment protocols. Additional detail on study eligibility would be helpful, however. The authors refer to a complex screening procedure conducted by KTC services that would determine if the client is deemed appropriate for services, but do not specify how that judgement is made in regard to clinical presentation. The two questions used to screen for more
specific study eligibility seem very mild and may lead to a sample with very low levels of symptoms; however, if other symptom screening is taking place before that step it would be useful to know that.

Response: We acknowledge that it is possible that our approach to symptom screening may yield to a sample with low levels of symptoms. To address this possibility, it will be a requirement that the family has been referred to the clinical service from a health professional. We have now added in some extra details to the text about study eligibility (e.g., requirement for a health professional referral), and an explicit statement that no other screening questions or measures will be used to assess eligibility.

The section now reads as follows:

“To be included in the study, the family must have a child aged 14-24 months, a referral from a health professional, and the parent must give a positive response to one or both of two screening questions (‘Do you have concerns about your child’s behavior?’ and/or ‘Do you have difficulties managing your child’s behavior?’). No other screening questions or measures will be used to assess eligibility, however families will be excluded from the study if there is evidence (in the information obtained in the referral or intake assessment) of severe parental depression with suicidality or other serious mental health conditions causing significant impairment in cognition or behaviors (e.g., psychosis) or if they are not sufficiently proficient in speaking English to complete study measures and protocols.” (page 15)

Comment #3. The measures proposed appear sound and well-supported in regard to reliability and validity, however, there is some concern about the reliability and validity of the CBCL with the younger end of the age range of these subjects.

Response: We acknowledge as a weakness of the study the fact that the measure of child externalizing behavior, the CBCL, is not validated for children at the younger end of the study age range. We draw the reviewer’s attention to previous studies that have successfully used the CBCL in this age range (these are mentioned in the Measures section, page 24), but we have also now added mention of this issue in the study limitations section, as follows: “Finally, while the parent-report measure of child externalizing problems, the CBCL, has been used successfully in previous studies with children aged as young as 12 months (79), it has not been validated for children aged less than 18 months and so there may be issues with the measure for children in the study aged 14-17 months.” (page 29)

Comment #4. In addition, the number of pre-treatment assessments required is substantial and there may be significant attrition related to the assessment burden, or delays due to difficulty fitting so many assessments into a single week. Similarly, the appointment schedule of 2x per week also seems ambitious, and it would be helpful to see a more thorough discussion of how delays due to no-shows, illness, and other types of cancellations will be handled.

Response: We agree that the number of pre-treatment assessments and the requirement for the PCIT-T group to attend 2 times per week for treatment, are ambitious. No shows and cancelled sessions due to illness, holidays and other life events are expected. We will be somewhat flexible
regarding the pre-treatment sessions (e.g., if required, the pre-treatment sessions will take place over a 2 week period instead of 1 week period). In a series of preliminary analyses, we plan to test for differences in (i) the total treatment time, and (ii) total number of weeks between T1 and T2. If any of these variables are found to significantly impact results, we will control for these variables in the final analysis.

Reviewer 3

Comment #1. Page 6, Line 36- Can you give a couple of examples of what "daily hassles" are?
Response: We have now included two examples: money problems or troubles at work.

Comment #2. Page 7, Line 4- If there is room, it would be helpful to have a sentence or two very briefly describing each of these interventions.
Response: In response to this advice, we have added in a brief description of each of the interventions prior to describing evidence regarding efficacy, as follows:

Page 6: “VIPP-SD (30) involves review of video-taped parent-child interaction sessions in the home and aims to increase parental sensitivity and sensitive parental discipline, defined as the ability to take into account the child’s perspective and signals when discipline is required.”

Page 6: “ABC is based around active coaching of the parent from a therapist in ten 1-hour home-based sessions, with coaching focused specifically on enhancing nurturance of the child, following the child’s lead with delight, and reducing frightening/intrusive behavior.”

Page 7: “COS-I comprises a mix of individual and group components spread over 20 weeks, and aims to increase caregiver sensitivity and responsiveness to child cues, empathy for the child by supporting parental reflective functioning, recognition and understanding of child attachment cues, and awareness of the impact of the caregiver’s own attachment history on caregiving.”

Comment #3. Page 8, Line 46- confusing as to whether it is mastery-based or time-limited- which skills are quantified? What is average length of treatment?
Response: To make this clearer, we have now added the following explanation to the text:
“The PCIT-T program is mastery-based, meaning that parent-child dyads progress from the CDI-T phase to the PDI-T phase when they show satisfactory use of the CDI-T skills in the 5 minute observational period (i.e., they are observed to use 10 labeled praises, 10 behavioral descriptions and 10 reflections, and to adequately implement the CARES skills when needed). Graduation from the PDI-T phase takes place when the parent can competently implement the guided compliance teaching sequence. For most dyads, when PCIT-T is delivered twice weekly (30-40 mins per session), CDI-T and PDI-T mastery is typically achieved after approximately 10-12 sessions.” (pages 9-10)
Comment #4. Page 9, Line 14- Starting with "Kohlhoff and Morgan (45) conducted a retrospective..." awkward sentence, hard to understand.

Response: We have now re-worded this sentence, as follows: “Kohlhoff and Morgan (45) retrospectively reviewed outcomes of the first 29 cases treated with this early version of PCIT-T and showed that toddlers who received PCIT-T showed statistically and clinically significant improvements in child behavior and that their parents reported feeling less depressed.” (page 10)

Comment #5. Page 11, line 14- after "specify," "an" should be "a"

Response: Thank you for picking up on this typo. It has now been corrected.

Comment #6. Why is future tense used throughout methods?

Response: Future tense is used because this a protocol for a study not yet completed. We believe that use of future tense in published study protocols is common practice.