Author’s response to reviews

Title: Triple P - Positive Parenting Program for parents of preterm born preschoolers: A randomized, clinical trial

Authors:

Renske Schappin (r.schappin@umcutrecht.nl)
Lex Wijnroks (a.wijnroks@uu.nl)
Monica Uniken Venema (m.unikenvenema@umcutrecht.nl)
Barbara Wijnberg-Williams (b.j.wijnberg@isala.nl)
Ravian Veenstra (r.r.veenstra@isala.nl)
Corine Koopman-Esseboom (c.koopman@umcutrecht.nl)
Susanne Mulder-De Tollenaer (s.m.mulder@isala.nl)
Ingeborg Van der Tweel (i.vandertweel@umcutrecht.nl)
Marian Jongmans (m.j.jongmans@uu.nl)

Version: 2 Date: 13 March 2013

Author’s response to reviews: see over
Author’s response to reviewers

Title: Brief parenting intervention for parents of NICU graduates: A randomized, clinical trial of Primary Care Triple P

Authors:
R Schappin: r.schappin@umcutrecht.nl
L Wijnroks: a.wijnroks@uu.nl
M Uniken Venema: m.unikenvenema@umcutrecht.nl
B Wijnberg-Williams: b.j.wijnberg@isala.nl
R Veenstra: r.r.veenstra@isala.nl
C Koopman-Esseboom: c.koopman@umcutrecht.nl
S Mulder-de Tollenaer: s.m.mulder@isala.nl
I Van der Tweel: i.vandertweel@umcutrecht.nl
M Jongmans: m.j.jongmans@uu.nl

Date: March 12th, 2013
Dear editor,

Thank you for the extensive review and the opportunity to clarify some issues with our manuscript. Each comment of the reviewer is in bold text, followed by our response and text changed in the manuscript in italics.

Reviewer: Jane Barlow

1. I am unclear what the logic was for testing a programme with unproven evidence of effectiveness with a group of children with specialist needs in addition to their emotional and behavioural problems. Why wasn’t a version of Triple P used that had already been proven effective? The consequence of this is that the current study does not show whether the absence of effectiveness is a result of the programme or the study group;

At the time the study was designed (the year 2007), most evidence for the Triple P program looked very promising but none of the versions of Triple P could be labelled as ‘evidence based’. Given this promising perspective, several (major) local health authorities in the Netherlands were on the brink of implementing Triple P. Indeed, Triple P is currently the most widely used (systematic) parenting program in the Netherlands.

Since there were (and still are) no specialized parenting programs for parents of NICU graduates at toddler age, our expectation was that these parents would be referred to Triple P once Triple P was implemented in the Netherlands. Because of the specific parenting problems these parents reported, we also expected that they would be referred to Primary Care Triple P (level 3), a short parenting intervention. The target population of Primary Care Triple P is: “Parents with specific concerns about their child’s behaviour or development that require active skills training” for “children with mild to moderate emotional and behavioral problems”. From our clinical experience, this fitted the needs of parents of relatively healthy (i.e. without serious medical problems) preterm NICU graduates at toddler age, since these parents mainly reported common toddler problems (dinner-time issues, sleeping problems, temper-tantrums).

The goal of our trial was to investigate the effectiveness of Primary Care Triple P specifically for the population mentioned above, and our trial was designed accordingly. At the time of the study design, the discussion about the effectiveness of Triple P was not expected since the intervention looked very promising; several studies conducted in other countries were able to demonstrate its effectiveness. Nonetheless, at that time, no study was published on the effectiveness of Triple P in a sample of parents of NICU graduates. Therefore, it must be stated that our study was never designed to draw conclusions about the general effect of Primary Care Triple P (i.e. beyond parents of NICU graduates). We agree that this complicates the interpretation of the working mechanism of our results. Unfortunately, this problem cannot be resolved.

The available evidence for the effectiveness of Primary Care Triple P at the time our trial was designed is now made explicit in the second paragraph of our introduction: “Furthermore, at the time our trial was designed, several studies had demonstrated the effectiveness of Primary Care Triple P in non-clinical populations”.

2. Related to the above point, most of the parenting programmes that have evidence of effectiveness are delivered over a period of 12 weeks, not 4.
As mentioned above, at the time our study was designed, Primary Care Triple P looked promising and seemed to fit the type of parenting problems encountered by our target population. We considered more intensive or longer interventions such as Standard Triple P to be disproportionate compared to the problems parents reported to us. Furthermore, in the stepped care system of Triple P it is custom to start with the briefest possible but of course appropriate intervention, and when this brief intervention proves to be not effective, move on to a more intensive intervention.

We have made our choice for a brief intervention more explicit in the second paragraph of our introduction: “A brief version of Triple P was chosen because it fitted the problems reported by parents during regular clinical follow-up.” Furthermore, we discussed the possibility that a longer intervention could have been more effective in the second paragraph of our discussion: “This also suggests that it is possible that although Primary Care Triple P was not effective, more intensive versions of Triple P that consist of more sessions, such as Standard Triple P would have been effective.”

3. I am bemused by the introduction of other forms of treatment post-intervention. This potentially renders the follow-up results to be completely meaningless.

Although we agree that our follow-up cannot be interpreted in terms of intervention effects, we do not agree that it is completely meaningless. The follow-up represents what would normally occur in clinical practice: parents who are satisfied with their outcome do not need extra care, whilst parents who are not satisfied receive additional treatment. We wanted our follow up to reflect this practice. This means that our follow-up can only be interpreted in terms of trends in problem behaviour, and not in terms of intervention effects. We have carefully re-read our manuscript to ensure that we convey this message clearly.

On a practical note, it was deemed unethical by the institutional review board and therefore not allowed to withhold intervention in the control group for more than 6 months, and to withhold additional intervention if necessary in the intervention group for more than 6 months.

We now emphasize that our follow-up results should be interpreted with caution under the subheading ‘12-month follow-up outcomes’ in our results section: “Due to the possibility of (additional) psychological treatment in both the intervention and control group after the trial endpoint at 6 months, allocation was no longer random during follow-up and results should be interpreted with care.”

4. In terms of the justification for the study, it is not clear from the background section why children who were born preterm are different in terms of the development of emotional and behavioural problems, or what the implications of this might be in terms of the provision of parenting support. There is an interesting literature on mechanisms involved with this group of children, which is touched on briefly (reference 6 and 7) but not adequately addressed.

Thank you for this suggestion. We have tried to keep our introduction as brief as possible, which may have lead to a simplification of this issue. We have now discussed it more elaborately in the first paragraph of our introduction: “Transactional theories on the development of behavior problems in preterm-born children suggest that the interplay between parents’ preexisting personality and family factors, prenatal experiences, and emotional distress during the NICU period, alters parents’ perception of their child and their parenting style [7]. Parents may inadequately perceive their child as extra vulnerable even up to 6 years after birth [8, 9], and employ a parenting style that is characterized by
overprotection and inconsistent discipline [10]. In combination with the neurological vulnerability of the child, these altered parenting practices may negatively impact the behavior of the child [11]."

5. The background literature also does not address well the literature on existing interventions with is rather jumbled in terms of the distinction between preventive and specialist interventions, and the point at which they are offered. It would make much more sense to address the literature on other interventions that have been provided to this population of children at the same time point as the proposed intervention.

The problem is that as far as we are aware, there are no specific parenting interventions for parents of NICU graduates at toddler age. We would welcome suggestions on the existence of such interventions.

6. I’m not clear why it makes sense to include term infants who suffered from asphyxiation…did they spend extended periods of time in NICUs?

Infants born at term with perinatal asphyxia are the second largest patient group in the NICU (after preterm infants) in the two centers that participated in this trial and other similar centers in the Netherlands. They spend prolonged periods in the NICU and resemble preterm born children in terms of the prevalence of behaviour problems, which is approximately 20% in both patient groups. This is now stated more specific in our introduction: “These children are two major patient groups in the NICU, and have a prevalence of behavior problems of 20%, versus approximately 10% in healthy term-born children.” (Van Handel M, Swaab H, De Vries LS, Jongmans MJ: Behavioral outcome in children with a history of neonatal encephalopathy following perinatal asphyxia. J Ped Psychol 2010, 35:286-295)

7. Other methodological concerns are as follows:

a) The failure to register this trial with one of the two organisations that maintain trial registers;

Our trial is registered at the Netherlands National Trial Register, which is a WHO and ICMJE approved primary registry. This is confirmed by the editor.

b) The absence of a concealed and independent randomization process;

We discussed these issues with our methodologist, and the randomization process used in our study is common practice in intervention research. Because of the nature of our intervention, open allocation to either intervention or control group was not possible. In order to use open allocation, we would have had to design a placebo intervention, but it was deemed impossible to design a placebo intervention that would resemble parent training but would have no effect at all on parenting practices.

The randomization of our study was completely traceable. We used computer-generated random permuted blocks of 6. This creates a numbered list of random assignments to either the intervention or control group (e.g. 1=control, 2=control, 3=intervention, etc.). Depending on the order in which they consented to participation in the study, the first author matched the participants to this numbered list. For example, the first participant to consent was in the control group (number 1), the second was also in the control group (number 2), the third was in the intervention group (number 3), etc.
c) The power calculations do not appear to have been undertaken with reference to existing studies in terms of the size of change that might be expected; the study is small in terms of numbers and one explanation for the findings is that the study is underpowered;

At the time our study was designed, there were no existing comparable studies. We therefore decided that we wanted to be able to show a reduction of at least 5 points on the CBCL Total scale t-score in the intervention group, assuming the control group to remain stable. We did not think that smaller reductions in behaviour problems would be clinically relevant.

Our exact calculation was:
Effect size Cohen’s D = difference between groups of at least five points / preliminary standard deviation = 4 / 7.1 = 0.67
With an alpha of .05 and a power of .80, a total sample size of 58 is needed to find this value of Cohen’s D. We expected 10% attrition, so 58 / 0.9 is a total sample size of 64.

8. The discussion does not address why this intervention did not prove effective with this group of children, and this may reflect the confounding that is present in the study design.

Thank you for this comment. Although we do not agree that there is confounding in our study design (see our response to question 1), we have more clearly addressed why Primary Triple P may not have been effective specifically for NICU graduates.

Our discussion of this issue can be found in the second paragraph of our discussion:
“Our finding that Primary Care Triple P is not effective in reducing problem behavior in NICU graduates could be due to the specific characteristics of our population. After all, the study was specifically designed to investigate the effectiveness of Primary Care Triple P in NICU graduates, assuming this version of Triple P to be effective for healthy term-born children. Given the specific parenting practices and parenting problems of parents of preschool-aged NICU graduates, such as inconsistent discipline which may lead to temper tantrums, a brief parenting program like Primary Care Triple P seemed indicated. However, this may have overlooked that a significant amount of attention should be paid to the emotional mechanisms behind these parents’ parenting practices. Feelings of guilt on behalf of the mother, lingering thoughts about the NICU period, and perceptions of vulnerability may need to be properly addressed before practical solutions find ground.”

Reviewer: Stephen Scott

1. It would benefit from explicit reporting of CONSORT criteria, they are nearly all met but a few are missing and they should comply with this.

The reviewer has a good point. Somehow, we forgot to include the CONSORT checklist as supplementary material. Item 18 (multiplicity) and 19 (adverse events) of the CONSORT checklist were not applicable to our trial, and were therefore not reported in the manuscript.

2. The training and skills of the interventionists needs to be better reported, we are told they are licensed in triple P, does that mean they passed the exit examination after initial training, could they state how frequently they had supervision and whether they videotaped sessions? This is important as there is an increasing literature showing that
successful outcomes are not just an issue of fidelity to the manual, but the skill of the practitioner also is important.

We agree that this is an important point and have now elaborated on this in our manuscript. The training and skills are now described in the method section of our manuscript, under the subheading ‘intervention’:

“The Primary Care Triple P training was provided by one experienced social worker, two registered healthcare psychologists, and two registered clinical psychologists. They received three-and-a-half-days of training in Primary Care and Standard Triple P and passed an individual examination and accreditation test to become licensed Triple P practitioners. Peer supervision between these professionals and the first author conducted at least once a month assured adherence to the intervention protocol.”

3. The discussion section needs to consider the explanations for the lack of effect more systematically. The authors are right that it fits in with the nonrandomised trials suggesting that primary care triple P is insufficient to bring about change. However, there are a number of possibilities.
a. The first and most obvious one is that the triple P programme is not of the highest quality, thus a recent trial where both triple P and incredible years were used showed the former did not work whereas the latter did.
b. Secondly, do they believe that it was due to the characteristics of the children that were less responsive to change in their environment than children who had not had burst distress or been premature? This seems an unlikely explanation, since parenting did not change.
c. Thirdly, is for sessions enough for bringing about substantial parenting change? Some would argue that you need 10 or 12 sessions to achieve this.

We thank you for these valuable suggestions and have incorporated them in the second and third paragraph of our discussion:

“Our finding that Primary Care Triple P is not effective in reducing problem behavior in NICU graduates could be due to the specific characteristics of our population. After all, the study was specifically designed to investigate the effectiveness of Primary Care Triple P in NICU graduates, assuming this version of Triple P to be effective for healthy term-born children. Given the specific parenting practices and parenting problems of parents of preschool-aged NICU graduates, such as inconsistent discipline which may lead to temper tantrums, a brief parenting program like Primary Care Triple P seemed indicated. However, this may have overlooked that a significant amount of attention should be paid to the emotional mechanisms behind these parents’ parenting practices. Feelings of guilt on behalf of the mother, lingering thoughts about the NICU period, and perceptions of vulnerability may need to be properly addressed before practical solutions find ground. This also suggests that it is possible that although Primary Care Triple P was not effective, more intensive versions of Triple P that consist of more sessions, such as Standard Triple P could have been effective.

Another possible explanation for our findings is a general lack of effectiveness of the Triple P program. Recently published independent research on Primary Care Triple P and Triple P in general questions the effectiveness of Triple P and suggests that it may not be as effective as was reported a few years ago. Currently, there are in total 5 published peer-reviewed publications evaluating Primary Care Triple P, of which 3 studies included a control group, but none included clinical populations [16, 17, 29-31]. The largest study employed a quasi-experimental design and found no significant differences between Primary
Care Triple P and care-as-usual in terms of parenting stress, parenting practices, and family functioning [29]. Measures of child behavior were not included in this study. Another quasi-experimental study that compared Primary Care Triple P to care-as-usual did find significantly higher levels of parental competence and more positive parenting in the Primary Care Triple P group, but no significant differences between groups in terms of child emotional and behavioral problems [30]. The only study that used a wait-list control group in a quasi-experimental design was a developer-led study that was published before the start of our trial, and concluded significant differences existed between groups in terms of measures of child emotional and behavioral problems. The Primary Care Triple P group also had less dysfunctional parenting styles and less parental anxiety and depression. Nonetheless, groups did not differ with regards to observations of parent child interactions [16]. Compared with the two independent studies, the developer-led study reported the most positive results on the effect of Primary Care Triple P. This is in line with findings of a recent meta-analysis on Triple P in general, that independent researchers generally report smaller or non-existent effects compared to those seen in developer-led studies [32].”

4. The findings are negative and are somewhat over optimistically reported in terms of triple P being as good as a waiting-list, it would be more honest to say that the intervention had no effect.

We wholeheartedly agree on this. Because we had problems in the past with other reviewers, we became too cautious in our conclusions. We now have changed our conclusions in: “This randomized clinical trial demonstrates that the brief parenting intervention Primary Care Triple P is not effective in reducing parent reported child emotional and behavioral problems, in families with a preterm-born child or a term-born child with perinatal asphyxia at preschool age.”

Reviewer: Frances Gardner

1. This is a clear and well-written report on what appears to be a well-conducted trial. However, it should include a CONSORT checklist to show that it has conformed with best reporting standards.

The reviewer has a good point. We apologize for the fact that we forgot to include the CONSORT checklist as supplementary material and have now included it. Item 18 (multiplicity) and 19 (adverse events) of the CONSORT checklist were not applicable to our trial, and were therefore not reported in the manuscript.

2. It should be made much clearer in the title, abstract and conclusions that this is a very brief intervention, and that the findings do not necessarily generalise to longer versions of Triple P, or to other parenting programs.

We agree with this and have now tried to express that Primary Care Triple P is a brief intervention throughout the manuscript. Furthermore, we have made amendments in the mentioned sections of the manuscript:

Title: “Brief parenting intervention for parents of NICU graduates: A randomized, clinical trial of Primary Care Triple P”

Abstract: “Our objective was to investigate whether a regular, brief parenting intervention, Primary Care Triple P, is effective in decreasing emotional and behavioral problems in
preterm-born or asphyxiated term-born preschoolers.” … “Primary Care Triple P, a brief parenting intervention, is not effective in reducing child emotional and behavioral problems in preterm-born children or term-born children with perinatal asphyxia.”

Conclusions: “This randomized clinical trial demonstrates that the brief parenting intervention Primary Care Triple P is not effective in reducing parent reported child emotional and behavioral problems, in families with a preterm-born child or a term-born child with perinatal asphyxia at preschool age.”

3. The rationale for using parenting interventions with this group is made clear, but why a brief one? In the discussion, the authors state that most of the other 5 studies of this brief Triple P primary care either found it didn’t work, or were designed so it was hard to draw conclusions. So what is the rationale for using it with this group? This should be made clear in the introduction - for example, was it warranted because of weak methods in the other trials, or because they were conducted by the developer, and independent replication may be worthwhile? Moreover, if the results of prior trials were not robust with this brief program, then why power the trial using a one-sided significance level? Here I’d repeat the wise words of Iain Chalmers, that all trials should be justified with reference to the lesson from a (new or recent existing) systematic review.

Unfortunately, most of the studies on (Primary Care) Triple P were conducted after the start of our trial. At the time we designed our trial, the effects of Primary Care Triple P looked promising. Please also see our response to comment 1 of prof. dr. Barlow.

The sole reason for using a one-sided significance level was that we had a one-sided hypothesis: we expected problem behaviour to decrease, not increase.

4. Furthermore, if this intervention is of unclear effectiveness with non-preterm born children, then there seems little to be gained by speculating about why the intervention was ineffective for preterm born children per se. (top of p15, and then twice on p 16-17).

This is true, but as our response to your previous question shows, Triple P was not of unclear effectiveness at the time our study was designed. Therefore, our study was never designed to draw conclusions about the general effect of Triple P level 3. The result of this is that the only conclusion that can legitimately be drawn from our trial is that Primary Care Triple P is not effective for NICU graduates. Therefore, we feel that we also need to discuss the possibility that characteristics of parents of this specific group of children contributed to not finding an effect.

5. In my opinion it goes MUCH BEYOND THE DATA to state that the experience of being in the trial, for this preterm group, “may have beneficial effects in its own right for parents in both the Triple P intervention and the wait-list control group”. This is stated in different ways 3 times and should be removed. It is surely illogical to make such a statement on the basis of (modest) improvements in both groups, in the absence of a third comparison group that had no such attention. After all, if we could make such inferences without a comparison group, then we would not need to bother with randomised trials! I appreciate that it is not very practical to make such a comparison. Luckily we don’t need to, as there are 25 year of trials and systematic reviews showing superiority of many parenting interventions over the mere filling in of questionnaires and talking to someone.
Thank you for this comment. We had this debate in our research group several times over the last years, and in hindsight the inclusion of a ‘post-test only’ group would have facilitated the interpretation of our results. Although we agree that we should not repeatedly state that attention in its own right could have had beneficial effects, we do think that a Hawthorne effect could be one of several possible explanations for the decrease in behaviour problems in both groups. Therefore, we have now only mentioned this effect once in the fourth paragraph of our discussion: “Another explanation that can not be excluded is a Hawthorne effect, in which the decrease in problems reflects the effect of participating in research.”

Furthermore, we added not including a post-test only control group to the limitations of our study (discussion, fifth paragraph): “Third, we did not include a post-test only control group. Including a post-test only group would have facilitated the interpretation of our findings that problem behavior decreases in both the intervention and control group, for example by estimating the magnitude of a possible Hawthorne effect.”

6. Clinical implications: arguably it is implausible, given what we know about parenting and the development of children’s behaviour problems, both in the general population and in those born preterm, and given the programme theories underlying these interventions, that they would be effective for kids identified as having behavioural problems, but not for those who are so identified, but who are also preterm. After all, the children included in trials of parenting interventions for conduct problems are a very heterogeneous group anyway, many of whom will have neonatal and other biological and social risk factors. Although I do not know of any studies that directly test the question of whether neonatal risk is an effect modifier, analyses of other subgroups that may include a higher percentage of children with neonatal risk factors (eg boys, high ADHD scores; very low-income families; Jones et al 2008; Gardner et al, 2009; 2010), do not suggest any differential (diminished) effects. Most good parenting programmes are flexible and at least somewhat tailored with respect to parent and child needs (Gardner et al, 2009; 2010), and so for many reasons we can be optimistic that evidence based parenting interventions will work with preterm born children. The paper should draw on this wider literature on risk factors, effect modifiers and applicability of parenting programmes, in order to make a brief but more nuanced discussion of the implications of the trial for this moderately high risk population. (In terms of conduct problems, the long term outcomes may not be so bad for these very premature & very low birthweight kids, Gardner et al 2004; Hack et al 2002).

Although we acknowledge that trials on parenting interventions may include children with neonatal risk factors, we do not agree that general parenting interventions may be as effective for term born children without medical complications as for children in need of neonatal intensive care. This is the reason we conducted our study. In the first paragraph of our introduction, we have now made more explicit why children born preterm or term-born children who suffered from perinatal asphyxia are different in terms of the development of emotional and behaviour problems, as well as parenting styles: “These children are two major patient groups in the NICU, and have a prevalence of behavior problems of 20%, versus approximately 10% in healthy term-born children [3, 5]. Furthermore, having a child in the NICU burdens parenting, not only during the acute phase of illness but also in the years thereafter [6]. Transactional theories on the development of behavior problems in preterm-born children suggest that the interplay between parents’ preexisting personality and family factors, prenatal experiences, and emotional distress during the NICU period, alters parents’ perception of their child and their parenting style [7]. Parents may inadequately perceive their child as extra vulnerable even up to 6 years after birth [8, 9], and employ a parenting
style that is characterized by overprotection and inconsistent discipline [10]. In combination with the neurological vulnerability of the child, these altered parenting practices may negatively impact the behavior of the child [11].”

7. Given point #3, is the trial underpowered?

At the time our study was designed, there were no existing comparable studies and only few studies on Primary Care Triple P. We therefore decided that we wanted to be able to show a reduction of at least 5 points on the CBCL Total scale t-score in the intervention group, assuming the control group to remain stable. We did not think that smaller reductions in behaviour problems would be clinically relevant.

Our exact calculation was:
Effect size Cohen’s D = difference between groups of at least five points / preliminary standard deviation = 4 / 7.1 = 0.67

With an alpha of .05 and a power of .80, a total sample size of 58 is needed to find this value of Cohen’s D. We expected 10% attrition, so 58 / 0.9 is a total sample size of 64.