Author's response to reviews

Title: A comparison of three parenting programmes for parents of children with behavioural difficulties when implemented in a large scale effectiveness study in community settings: The Parenting Early Intervention Pathfinder (PEIP)

Authors:

Geoff Lindsay (geoff.lindsay@warwick.ac.uk)
Steve Strand (steve.strand@warwick.ac.uk)
Hilton Davis (hilton.davis@kcl.ac.uk)

Version: 3 Date: 25 October 2011

Author's response to reviews: see over
Dear Jim
I would like to thank the Associate Editor and the two reviewers for the comments on our paper. You do not include the Associate Editor’s name so I cannot make the comments direct.

My apologies for misunderstanding the mechanics of the submission process. As I recall I did submit the title, authors etc in separate fields and removed these from the manuscript file. Thanks for clarifying the process. I confirm that the resubmission manuscript file should indeed include the title page with content as specified in your email and an abstract in the journal style.

We thank the reviewers for their comments and I confirm that we have addressed each point in both reviews and detail these responses below, indicating locations of the changes made as appropriate. There is one point which concerns editorial policy which we are happy to follow.

Please note also that we have corrected a typo in Table 2 where the WEMWBS mean post-test score for Triple P should be 50.1 (not 59.1 as previously shown).

Finally, please note that the 3rd author is now emeritus professor. I have therefore included his academic address but his home address is also available.

Our responses to the reviewers are given in order below.

Reviewer 1: Clarice Martins

We thank Clarice Martins for these comments. We present the issue raised in bold and then our response in each case, with additional text highlighted in yellow as appropriate. Please note that page numbers have changed.

Major compulsory revisions

I am concerned about a methodological question: in the analysis, SFSC revealed lower efficacy in almost all the evaluated parameters. Then, in page 14, authors explained that this results should be due to the aims and content of the program which are much more broader than the other two programs. If the SFSC has different aims, why to compare it with the other two that have the same aim?

The three parenting programmes were selected by the UK government. The government selected these three as they all had the same basic aims.
   i) All aimed to improve parenting skills (including behavioural strategies) and parents’ sense of competence as a parent; and
   ii) A primary focus of all three programmes was on improving children’s behaviour, as a result of having improved these parenting skills. In addition, all had evidence of efficacy with respect to both aims, improving parenting and child behaviour. Consequently we selected measures that were well established in
research in this field to examine these domains. However, each programme had some different additional foci, the SFSC in particular. We therefore draw attention to this in our interpretation of the findings of relative improvements between the three programmes as it is possible to argue that the relatively lower level of improvement for SFSC could (not should as stated in the reviewer’s comments) be related to the additional content. It is important to stress therefore that all three were effective, as demonstrated by individual programmes’ results, but it is the relative improvements that are the focus of this part of the discussion.

To clarify, we have amended the text in two places by adding the following:

p.9: The effectiveness of the parenting programmes was measured by the following primary outcome measures; there were no secondary measures. The instruments were selected to measure the domains expected to show improvements following attendance at the programme and common to all three programmes.

p.16-17:

All three had been selected by the UK government as appropriate programmes to improve parenting skills and reduce children’s behavioural difficulties; hence a three way comparison using measures selected to show improvements in the primary domains common to all three programmes is a reasonable analysis.

We then go on, as before, to propose possible explanations for the relatively lower levels of improvement for SFSC and point out the need for further research to test these out.

Discretionary Revisions

1. I recommend to briefly explain what kind of activities are done during the intervention process (page 7);

We have added information on p. 6-8 to describe each programme individually and then provide information common to all three programmes on p.9, namely:

Implementation was representative of the way these three programmes are typically conducted. The three programmes described above were delivered to groups of parents typically comprising about 10 parents. The sessions were conducted by people (referred to as facilitators) specially trained for the purpose. Group sessions were guided by the programme manual to optimise fidelity and comprised watching programme videos (DVDs) created as stimulus material for specific teaching points, group discussions and role play. Parents had their own personal handbook and carried out homework between sessions.

Reviewer 2: José Cazuza Farias Júnior

We thank José Cazuza Farias Júnior for these comments. We present the issue raised in bold and then our response in each case, with additional text highlighted in yellow as appropriate. Please note that page numbers have changed.
Major compulsory revisions

Introduction

1. The introduction was very long and there are many results. There are many results of several studies. It is necessary to synthesize more information.

The introduction was longer than is the norm in other papers for the journal – almost 3½ pages compared with 2 in a sample of other papers examined. We have reduced its length as requested to 2½ pages. We have limited the Background where possible to studies that review literature, e.g. systematic reviews. The Background addresses seven main issues that are pertinent to the rationale of the study plus one paragraph that describes the policy background for the large scale intervention funded by the UK government. Each element is important but the overall length has been reduced and information has been synthesized.

2. The hypothesis/aim/objective of the study did not was clearly described.

We recognise that the original text was not as clear as it should have been. We have specified two distinct aims of the study more clearly (p. 5).

3. Authors should present the state of the art and briefly highlight the knowledge gaps and so this study will help fill those gaps.

As noted above, we have set out seven main themes in the Introduction, indicating both current knowledge and the issues to address where knowledge is limited. These comprise as background to the study: p.3. prevalence of behaviour problems; p. 3 issues concerning the delivery of parenting programmes; p3-4 evidence for the efficacy of parenting programmes to reduce child behaviour problems; p.4 limitations of efficacy evidence; p.4 the importance for public health policy of evidence of effectiveness in community settings and when programmes are rolled out on a large scale; p.4 challenges to effectiveness studies; and p.5 the importance of comparative studies of the effectiveness of different parenting programmes.

Methods

4. The intervention programs were not adequately described. What was developed in each program?

We recognize that more information to describe the programmes would be helpful. Each of these three programmes is described in previously published literature and has a published manual and we provide references to this material. We have added descriptions of each programme, taking into account that all three had common aims:

Incredible Years (p.6-7)

The focus of the programme is the enhancement of effective, positive parenting, so as to enable children's development and education and to manage behavioural problems where necessary. However, there is also a strong concern with parents' adaptation more generally so that they are better able to deal with their own problems and relationships.
Triple P (p. 7)

The evaluation within this study was concerned with the implementation of courses for groups of parents with the focus on parental management of child behaviour and reduction of parental stress. Core principles in the courses included enabling parents to provide a safe and interesting environment for their children, a positive learning environment and assertive discipline, while maintaining realistic expectations and taking care of themselves as parents.

Strengthening Families Strengthening Communities (p.7-8)

As with the other programmes it was implemented in a group format with the central concern being the development of effective parenting skills and the reduction of child behavioural difficulties. However, there are broader themes involved in the course which include cultural and spiritual dimensions, enhancing relationships, rites of passage and community involvement.

5. The main outcomes to be measured did not clearly described in the methods section.

The parent mental well-being, parenting style and child behavior measures are described in the methods section (p9-11). We describe the nature of each scale, the more specific domains each addresses, their origins with references to authors, reliability (Cronbach’s alpha) and validity – again, with references to research.

6. The intervention in each group was not clearly described. This is very important in interventions studies.

As noted above, these three programmes comprise highly detailed interventions which are described in programme handbooks. However, we agree that more information regarding implementation is necessary for the paper and have added the following to p.9, stressing that the implementation followed the required protocol as set out in the manuals and fidelity was supported by training of all group facilitators.

Implementation was representative of the way these three programmes are typically conducted. The three programmes described above were delivered to groups of parents typically comprising about 10 parents. The sessions were conducted by people (referred to as facilitators) specially trained for the purpose. Group sessions were guided by the programme manual to optimise fidelity and comprised watching programme videos (DVDs) created as stimulus material for specific teaching points, group discussions and role play. Parents had their own personal handbook and carried out homework between sessions.

Facilitators (total 1100), from a range of backgrounds including social workers, psychologists and health visitors, were recruited by each LA and trained by the relevant programme provider over 3 to 5 days, according to the programme’s usual training requirements. Supervision and follow-up checks on implementation fidelity were made subsequent to initial training in accord with each programme’s normal practice. This included observation by experienced staff from the programmes as well as local senior
facilitators. The parenting programmes were delivered in a variety of settings including community centres, schools, clinics, and the premises of voluntary bodies: this range is representative of the locations typically used for these programmes.

7. The authors should be to talk about this point: Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?

Yes, as noted above in 6.

8. The authors did not define which primary and secondary outcomes that would be used to evaluate the effectiveness of the intervention.

We only had primary outcomes. This point is clarified as follows on p.9.

The effectiveness of the parenting programmes was measured by the following primary outcome measures; there were no secondary outcomes to be measured. The instruments were selected to measure the domains expected to show improvements following attendance at the programme and common to all three programmes. The outcome measures are described in the methods section, as described above.

9. The description of the data analyze be very long and hard. There are many explanations of the analysis performed. This is unnecessary. Authors should be limited to inform the analysis used for each objective of the study.

We consider that the description of our analytic strategy an important component of the paper. This section (P11-13) discusses competing Intention to Treat and Per Protocol approaches to analysis and our rationale for a Per Protocol analysis; evaluates the issue of sample attrition and establishes the justification for focusing on the sample with both pre-test and post-test scores; and finally gives the rationale for the ANOVA of change scores rather than ANCOVA. These are all important issues which we feel need to be clearly outlined and explained.

10. The probability values not been reported for the main outcomes

We provide $p$ values on p. 12 where results of statistical tests are presented in the text but not in a table. The $p$ values for the ANOVAs comparing the three programme groups are reported in Tables 2 & 3. The specific $p$ values for the subsequent multiple comparisons are not reported in the text but are indicated as *= $p$<.05; **= $p$<.01; ***= $p$<.001. This we believe gives the required information while saving space in already crowded tables. However, we recognize that full reporting of $p$ values is good practice and so have now added to the main text the $p$ values that apply to the comparisons between programmes. To add clarity we have added a final sentence to the Statistical Analysis section (p.13) which specifies that we set alpha at .05.
Bonferroni post hoc tests are reported where significant (alpha level of .05 for all statistical tests).

Discussion

We discuss the two issues raised in reverse order

11. There was no information about the losses of follow up, and whether the groups were homogeneous at baseline.

We have explicitly discussed the question of attrition and loss at follow up initially on p. 11-12 and again in the Discussion (p. 17). We argue that the level of attrition was similar to that typically found in trials and distinguish true attrition by ‘drop out’ by parents from administrative loss of data (p.17). We also discuss whether there are any significant differences in the demographics and pre-test scores of those who responded at post-course versus those who did not respond. We discuss this on p 11-12 and show there was no difference on 13 of the 14 measures and a very small difference on the remaining measure (see also point 12). We have added a specific reference to this point to the Discussion (p. 16):

Furthermore, the parents for whom we have post-course data are comparable on 13 of the 14 measures, indicating that the results are generalisable to the population from which the sample providing pre- to post-group comparison data was taken.

This supports the validity of focusing on the sample responding at post-test.

With respect to whether the groups were homogenous at baseline, the data on the demographics for each group are clearly presented in Table 1 and data on baseline (pre-course) scores in Tables 2 and 3. We discuss this issue of heterogeneity in demographics and pre-course scores in the Statistical Analyses section on page 12-13. It is because there is some variability between groups on demographics and pre-test (baseline) scores that we use ANOVA to make comparisons between programmes, after adjusting for demographics (parent gender, education and income and child gender and age) and why we express the dependent variables as change scores (from pre-test to post-test) to control for baseline score. It is precisely because there is some variation between groups with regard to baseline scores that we use ANOVA rather than ANCOVA (because the assumption in ANCOVA of equal pre-test means is violated, see p13).

12. It’s necessary discuss whether of the findings of the study they may be generalized to the population from which the study subjects were derived.

We have described the sample selection and discussed the limitations in the Discussion. We also show that the findings are comparable to those from efficacy trials in the UK and the studies in a meta-analysis of 55 studies (p.16). We show that the parents for
whom we have both pre- and post-course data are comparable in almost all respects, on our measures and demographics, to those parents for whom no post-course data are available (see point 11 above). We go on to consider limitations and, as we note in the third limitation (p. 18), there were no data on those parents who were not recruited at all by the local authorities and who might have refused to participate in these programmes.

Therefore our analysis indicates that these results can be generalized to parents who undertake these three parenting programmes. We have made this point clearer by two means: adding a new final sentence to para 1 on p.17 (see point 11 above for this addition) and by starting a new paragraph for limitation 3 in this discussion. (p. 18)

**Minor Essential Revisions**

**Title**

13. In the title of the article, it missing the information about that the program would change

The title has been changed to refer to change as indicated by *improving.*

14. The intervention programs were not adequately described. What was developed in each program?

Duplicate point – see 4 above

15. It is not necessary to present results of studies that talk about the effectiveness of these programs in the methods.

It is important in our view to demonstrate that the three parenting programmes have evidence for their *efficacy*; hence this is included in the description of the programmes. Given that they have evidence for *efficacy,* our study examines whether they are *effective* in this large scale roll out

16. The authors should be to talk about this point: were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?

Duplicate – see 7 above

17. The power of the sample has not been shown.

The sample size was not determined by power calculations so we have not presented these. However if we assumed two-tailed t-tests (not predicting which programme, if any, might be more successful than others), an effect size of around 0.3, set alpha=0.05, and require the minimum power 0.80, we would need 176 participants in each group. The size of the sample we have (n=1,121 at follow up with smallest group n=314) substantially exceeds the minimum sample size needed to detect an effect if one exists. (NB the minimum sample size we would need to establish ‘real’ changes in pre-test vs. post-test scores within each programme would be substantially smaller; using one-tailed tests on repeated measures with the same parameters we would need samples of only
30 within each programme. The actual power we have here, taking the average observed effect size of 0.6 and the smallest achieved sample size of 314 is a massive 1.0!)

18. Its need shower how the main outcome measures used.

All the outcome measures comprised questionnaires, as described in the Outcome Measures section, and were completed by the parents, at the start and end of their course (pre- and post-course). This is described on p.11 in Procedure

19. The authors must be showed the instruments accurate (valid and reliable) for the primary and secondary outcomes.

As stated above, the reliability of all of the parent mental well-being, parenting style and child behaviour measures (Cronbach’s alpha) are described in the Outcome Measures section. We also provide references to studies supporting their validity. We have added the following to p. 10:

Evidence for validity is well established through many studies which have used these scales, in the form of positive correlations with direct observation of parenting behaviour (e.g. [28], [32]).

We also discuss the issue of the validity of self-report measures in the Discussion on p.18, and include a reference to studies that have found improvements in both direct behavioural measures as well as standardized questionnaires (e.g. Lundahl et al, 2006 – reference [40]).

20. There are many explanations of the analysis performed. This is unnecessary. Authors should be limited to inform the analysis used for each objective of the study.

Duplicate question, already addressed.

**Discretionary Revisions**

21. It’s need to show p values of statistic applied for hypothesis tests.

Duplicate question, already addressed.

END