Reviewer’s report

Title: Parental Depression and Child Conduct Problems: evaluation of parental service use and associated costs after attending the Incredible Years Basic Parenting Programme

Version: 1 Date: 15 October 2013

Reviewer: Lisa Gold

Reviewer’s report:

This paper presents useful additional data from the Welsh IY evaluation, focusing on the service use and depression data for parents.

The authors need to acknowledge the limitations of sample size more, repeat the limitation on lack of long-term control (that is in the 2009 BJP paper) and adjust their language around interpretation of effects, but otherwise this is an interesting addition to the literature. Having stated these concerns I trust the authors to consider them, so they are listed as minor revisions below.

I’ve listed everything else I noted down while reading the paper under Discretionary – these are all up to you.

Minor essential revisions

1) My main concern with this paper is the discrepancy between patterns of change in the data and statistical significance of those changes, in the context of a 2:1 randomisation procedure with a wait-list control. Now there are issues with wait-list controls and the lack of controls for the 12m and 18m data points which are dealt with in the limitations sections of other papers from this trial, but these should be re-stated in the limitations section here just so readers don’t have to go back and read the BMJ and BJP papers.

2) The one thing authors here need to revise is their language around how they interpret the changes in the data over time from the RCT (Results sections 1 and 2). Figure 2 shows a decline in mean BDI II in both groups and the last sentence of Results – 1 – tells us that there is no statistically significant difference between groups at baseline or 6m – but this means that there is no statistically significant effect on this measure in the RCT. This is important, because your discussion starts by telling us that the programme has a positive effect on parental depression. The BMJ papers show the reduction in intervention versus control to be statistically significant – you need to explain more in this paper why that finding is not replicated – probably due to reduced sample size, but also because some of the most depressed parents seem to be those not providing full data (not surprising…)

3) The problem with how Results – 1 – is currently presented is that before you tell us that there is no difference between RCT arms you provide two paragraphs detailing with lots of statistics a significant decline in intervention group and no
significant difference in control group – which conveys the exact opposite meaning. I think you need to be very careful here on all interpretations of statistical significance when the RCT had a 2:1 randomisation, meaning the numbers in the control arm are half those in the intervention arm and therefore stand less chance of showing any change as statistically significant. I would advise dropping all the statistical detail in these paragraphs (or moving these to a table instead) and just explaining the results a bit more in lay terms, especially for the difference between trial arms from baseline to six months. Then you can go on to describe what happens up to 18 months for the intervention arm.

4) For Results – 2 – I would first present your Figures 3 and 4 as side-by-side rather than stacked graphs as it’s hard to tell the difference between arms in a stacked graph. I would then be inclined to avoid describing the results as “decreased” and “increased” in the text and instead just direct the reader to look at Figures 3 and 4 and then tell us what is statistically significant (or not). By the way, the jump in the 18m data here makes me suspect you disproportionately kept in the high-users between 12m and 18m (i.e. those with low service use were more likely to drop out at 18m) – is that true?

5) All the above means the text in Discussion should change to account for 2:1 randomisation impact on sample size and chance of statistical significance. The “Changes in service use” paragraph also repeats the “decreased” and “increased” interpretations of the results section. I think that the discussion section should just admit to the limitations due to the sample size in each trial arm and move on to discussing the cut-point analysis, which does have a more clear result.

Discretionary Revisions

a) I'm not sure if this is analysis of a "subsample" - you actually have 3 more parents in here than the main economic evaluation paper had - I think you can say you are using the full data set, just then say you are using the complete case analysis (i.e. only those with all the data you need for this paper) - calling it a subsample seems to do you a disservice?

b) Just be a bit careful about the multiple citations of the papers from the one Welsh RCT - in Section 2 of Background it is not clear whether "Hutchings [23]" is a different analysis/dataset to "Gardner et al [24]"? To avoid confusion perhaps stick to just one reference per study to avoid this looking like two separate mediator analyses that independently came to the finding of positive mediation - which is contrary to previous literature, as you say.

c) This is related to (b) but in background section 3 and in discussion you reference "some research" or "previous studies" and include multiple references from the one Welsh IY RCT - again to avoid confusion (i.e., to make clear that this is one study not multiple) just use one reference.

d) The only typo I found is inconsistent use of brackets when citing multiple references at a time: " ( ....]." This is done in several parts of the text but first time is Background section 3 - just search and replace.

e) In Background section 3 please clarify whether the “increase” in children’s
service use at six months is relative to controls, or just relative to baseline.
f) In Background section 3 is the cited "total cost" of GBP8470 the average cost per child - if yes fine, if no then please give the average instead - readers can't divide by 42 all that easily...
g) The space devoted to the description of the foster carer study could be reduced - it seems a little at odds with your focus on (biological) parents - you are not claiming a link between depression in foster carers and CD, perhaps make the relevance of this study more clear to the reader, e.g. that for non-parental carers, where one is not so concerned about depression (perhaps), then service use mainly consists of social worker use and totals...
h) Also in Background section 3 I think the relevant statistic is just the level of expenditure (GBP5892bn in 2008/9) and not that it increased over one year, five years ago - just cite the total for the most recent year that you have (do you have something from this decade??)
i) Paper aims 1) just add words "the cut-off" at the end to be totally clear
j) Methods - sample - add I:C split of the original 153 so reader knows how many dropped out of each arm.
k) Methods - sample - state range of sessions attended (if you have this - I noticed it isn’t in the other published papers either)
l) Methods – measures – it looks odd to have mild depression as 14-19, moderate as 20-28 but then a score of “19 and over” as clinical – this would mean that someone scoring 19 is “mild” but over cut-off – is that right, or is the cut-off “over 19”? 
m) Methods – measures – not really sure you need to reference demographics – especially as PhD Theses are quite hard for open-access journal readers to access – could just say that the trial also collected basic socio-demographic and general health data on family members and leave it at that. Similarly in the procedures paragraph you can rephrase to say Researchers conducted home visits to complete measures at baseline and at six, twelve....
n) Methods – 4. Longitudinal analysis – why was this applied to frequencies and not total cost? Just looked a bit odd to leave out costs?
o) Results – In the first paragraph we don’t need to know 18 were female, we just need to know if drop-out was disproportionately by one gender/age group/marital status/etc. (because that might affect the change in measures between baseline, 6 and 12 months (when n=75 for intervention group) and 18 months). Similarly, for a general BMC HSR audience we don’t need the detail on K-S statistics, just say Tests for normality revealed that data was not normally distributed, therefore..
p) Results – 3 – in describing the cut-off results the inclusion of all the statistics on cut-off is very confusing and I’m not sure any of this is needed – just delete all the numbers from these Intervention group and Control group paragraphs and just comment on the table.
q) Results – 4. Longitudinal.. Can you rephrase this into more lay terms – e.g.
...ANOVA showed that parents scoring above .. had significantly higher frequency of service use over 24 months compared to those parents scoring below...

r) Discussion – BDI II paragraph – “below at baseline at six and eighteen” – should the second of these “at”s be a comma? This does change the meaning of the sentence so I wasn’t sure...

s) I would drop reference to study [19] from the limitations paragraph as you are explicitly making a point about controlled long-term follow-up and [19] does not have controls for the long-term follow-up

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests