PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Do parents recall and understand children's weight status information after BMI screening? A Randomised Controlled Trial.
AUTHORS	Dawson, Anna; Taylor, Rachael; Williams, Sheila; Taylor, Barry; Brown, Deirdre

VERSION 1 - REVIEW

REVIEWER	Inge Huybrechts International Agency for research on cancer Lyon, France
REVIEW RETURNED	17-Feb-2014

GENERAL COMMENTS	This manuscript investigates a new though important issue when aiming at obesity prevention in children. 1) you mention in the title and design that it is a randomised controlled trial but in the abstract it is not clear what the randomisation refers to. Only in the results section it is clear that 2 different feedback methods are being compared. It would also be interesting to know how randomisation was done. Please make sure that this information is better higlighted in the methods part of the abstract. 2) Also in methods section more details should be giving regarding the randomised controlled trial (e.g how was randomisation done, etc.). 3) It would be good to mention the n of both intervention groups in the participants section. 4) Duplicate measurements were performed, but what measurement was used in results (e.g. mean of the two ...)?

REVIEWER	Beatriz Sarriá Institute of Food Science, Technology and Nutrition (ICTAN-CSIC) Consejo Superior de Investigaciones Cientificas
REVIEW RETURNED	25-Feb-2014

GENERAL COMMENTS	The references are not in the correct format according to BMJ open: - the journal should not be in an italics - the last page is not indicated as it should be

	Please correct accordingly The manuscript is an interesting piece of work which addresses a relevant topic using a well-designed intervention. English style is excellent. Methods have been described in detail. The presentation of the tables and figures is correct, however references' format needs to be revised attending to the BMJ author instructions.

REVIEWER	David Gillespie Cardiff University, Cardiff, Wales, United Kingdom.
REVIEW RETURNED	22-Apr-2014

GENERAL COMMENTS	This is a well written paper of a randomised trial of two different methods of BMI/weight feedback nested within a larger trial. While I found the paper well written, I had some concerns over the following points, and suggest that the points/questions listed below would need addressing prior to any recommendation of publication taking place. 1. Were there any parents with multiple children screened / entered into the trial? If so, were they asked about multiple children? Their responses would probably be clustered and should be taken into account analytically if so. 2. Did the researchers collect reasons for the 20 parents declining interview? It would be useful to know this to assess the appropriateness of how missing data were accounted for in analyses. 3. What did the researchers do with the six who had difficulties with audio recording and one who brought booklet back? Strictly speaking, they should have been included in an intention to treat analysis (particularly the latter participant). 4. Ref 23 describes the larger RCT – In my opinion, this paper should be self-contained, and so I would like to see details of the sample size calculation and methods for randomisation within this paper. 5. From the protocol paper, the researchers estimated that they would need 200-225 participants in each group. Given that they were only able to recruit 271 participants in total (244 with analysable data), this study was clearly underpowered. What also concerns me from reading the protocol paper is that randomisation does not seem to be adequately described, the sample size calculation does not describe what this specific trial is powered to detect, and I cannot find any specific mention in the protocol paper of using this trial specifically for determining recall. As this is not mentioned (at all, as far as I can see) in the trial protocol, this appears to be a post-hoc analysis and this should be clearly specified in the paper. 6. Did the independent interviewer/s remain blind to treatment allocation? 7. Did the individuals transcribing the interviews remain blind to treatment allocation? 8. Were the coding rules and weights for scores defined a priori (i.e. before looking at any data)? If not completely (e.g. because it was an iterative process), to what extent were they defined a priori? 9. Were the outcome measures and weights derived from the qualitative interviews and author discussion validated in any way? 10. What was the primary outcome measure? How was this determined? 11. What did the distributions of the various "continuous" measures

	look like? What were the residuals from the regression models like? Were they checked? What was the interaction term in the mixed model and why was it included? Why was a p<0.2 threshold used to enter variables into a multivariate model? 12. I would like to see a table comparing the 27 not included with the remaining included, or at least description of effect sizes and confidence intervals in the text rather than just p-values. 13. What were the participants like at baseline? A table reporting this would be useful. 14. Is the mean an appropriate measure to use to report the number of items recalled (what was the distribution of responses like?) 15. How do the results adjusted and not adjusted for days between initial interview and recall interview differ? 16. Was 'time between interviews' included as a linear effect? Was the appropriateness of this checked? 17. Rather than using a p-value to make a statement, report effect sizes and confidence intervals. 18. On line 25 of page 13, one of the upper limits of the confidence intervals looks incorrect. Should it be 0.01 to 0.27 (instead of -0.27)? 19. Report absolute numbers along with proportions or percentages. 20. Why are the univariate findings described in more detail than the multivariate ones? Again, effect sizes and confidence intervals are more informative than p-values. 21. For the %s in the correct recall column of Table 4, it would also be useful to report the %s of the total n (not just the n who had recalled the items) 22. Report exact p-values in tables rather than stating that p < 0.05

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 (Inge Huybrechts)

1. You mention in the title and design that it is a randomised controlled trial but in the abstract it is not clear what the randomisation refers to. Only in the results section it is clear that 2 different feedback methods are being compared. It would also be interesting to know how randomisation was done. Please make sure that this information is better higlighted in the methods part of the abstract.

Our original methods state that the randomisation referred to feedback given either as best practice care or motivational interviewing (page 2, line 11). However, we have added "of different methods of feedback" to the design section of the abstract (page 2, line 6) to further highlight what the randomisation refers to. Further elaboration is not possible within the word constraints of the abstract.

2. Also in methods section more details should be giving regarding the randomised controlled trial (e.g how was randomisation done, etc.).

Additional information has been provided as requested (page 7, lines 11-12).

3. It would be good to mention the n of both intervention groups in the participants section.

The numbers have been added (page 7, line 10).

4. Duplicate measurements were performed, but what measurement was used in results (e.g. mean of the two ...)?

Additional information has been provided as requested (page 8, line 13).

Reviewer 2 (Beatriz Sarriá)
1. The references are not in the correct format according to BMJ open:
- the journal should not be in an italics
- the last page is not indicated as it should be

All references have been modified accordingly.

Reviewer 3 (David Gillespie)

1. Were there any parents with multiple children screened / entered into the trial? If so, were they asked about multiple children? Their responses would probably be clustered and should be taken into account analytically if so.

Siblings were eligible to enter the trial at the families request if the child also met the inclusion criteria and had been screened. In our overweight sample, one family enrolled 3 overweight children and 9 families enrolled 2 overweight children. The remaining 251 children did not have siblings in the study. All data have now been adjusted for clustering (page 12, lines 18-19). In essence this adjustment made little difference and did not alter any of our conclusions. Every change has been highlighted in red on the revised manuscript in the abstract and results sections.

2. Did the researchers collect reasons for the 20 parents declining interview? It would be useful to know this to assess the appropriateness of how missing data were accounted for in analyses.

Reasons were collected for non-participation and have now been incorporated in the revised manuscript (page 13, lines 9-14).

3. What did the researchers do with the six who had difficulties with audio recording and one who brought booklet back? Strictly speaking, they should have been included in an intention to treat analysis (particularly the latter participant).

There was nothing to indicate that these six participants differed in any way from the rest of the participants. All cases were due to malfunction of the audio equipment (we used multiple recorders) which occurred at different times on different days.

We undertook a modified intention to treat analysis which excluded 27 participants altogether as described. As very little data for the remaining 244 participants was missing (<1.5%) we have presented analyses for the available data, assuming the missing data is missing at random (MAR). A statement re missing data has now been included in the revised manuscript (page 13, lines 2-3).

It was not possible to include the last participant as they referred to their booklet each time they were asked a question, which does therefore not measure recall at all.

4. Ref 23 describes the larger RCT – In my opinion, this paper should be self-contained, and so I would like to see details of the sample size calculation and methods for randomisation within this paper.

The statistical power required for our study was based on recruiting a minimum of 250 participants into the intervention, as outlined in reference 24. We did indeed achieve our goal with final recruitment of 271 participants meaning that we are adequately powered to detect our main outcomes of interest from the two-year RCT. No sample size calculations were performed prior to analysis for this paper as it was a secondary data analysis. However, we did achieve significant differences in total recall score

(albeit small) between our groups of interest suggesting that this analysis is also sufficiently powered. We have added a sentence to the manuscript reflecting this (pages 12, lines 21-23 and page 13, lines 1-3). Additional detail on the methods for randomisation have been added (page 7, lines 11-12).

5. From the protocol paper, the researchers estimated that they would need 200-225 participants in each group. Given that they were only able to recruit 271 participants in total (244 with analysable data), this study was clearly underpowered. What also concerns me from reading the protocol paper is that randomisation does not seem to be adequately described, the sample size calculation does not describe what this specific trial is powered to detect, and I cannot find any specific mention in the protocol paper of using this trial specifically for determining recall. As this is not mentioned (at all, as far as I can see) in the trial protocol, this appears to be a post-hoc analysis and this should be clearly specified in the paper.

Please see response to point 4.

6. Did the independent interviewer/s remain blind to treatment allocation?

The interviewers were independent in that they were not the same interviewers who were involved in providing feedback to parents. At the end of the interview, the researchers invited each parent to participate in phase 2 of the larger study (the intervention). This was a separate randomisation procedure and thus interviewers knew which feedback condition parents had receieved as they needed to collect the next appropriate randomisation envelope in case the family decided to be involved in the intervention. To ensure consistency between interviewers, all participated in interview training prior to starting the follow-up interviews, and then followed a semi-structured interview schedule (as outlined in Table 1 of the revised manuscript).

7. Did the individuals transcribing the interviews remain blind to treatment allocation?

Yes, the text has been modified to reflect this (page 10, line 3).

8. Were the coding rules and weights for scores defined a priori (i.e. before looking at any data)? If not completely (e.g. because it was an iterative process), to what extent were they defined a priori?

The first phase of the coding was developed a priori from the interview schedule, which was designed and developed prior to the study, based on the information we expected to elicit. The second phase of the coding, involving the development of specific codes and weightings, were developed after the data had been collected and researchers became familiar with the categories of responses that parents gave. As no prior literature appears to exist that has examined recall after weight feedback, it was difficult to assign meaningful codes before the data were collected. We have provided an additional table if the journal would like to include this, perhaps as supplementary information. At this point though, our revised manuscript does not refer to this additional table but additional information has been provided in the text (page 10, lines 7-11).

9. Were the outcome measures and weights derived from the qualitative interviews and author discussion validated in any way?

These were not able to be validated in any way but were developed after consultation with the wider research team. As our manuscript highlights, similar results were obtained whether weighted or unweighted data were used.

10. What was the primary outcome measure? How was this determined?

The primary outcome was the total recall score which was calculated as described in Table 2.

11. What did the distributions of the various “continuous” measures look like? What were the residuals from the regression models like? Were they checked? What was the interaction term in the mixed model and why was it included? Why was a p<0.2 threshold used to enter variables into a multivariate model?

The distribution of the total recall score was indeed normally distributed as were the residuals for the regression analyses. We chose to use a cutoff of $p < 0.2$ to indicate which variables could be considered for the multivariate model because using a combination of variables can affect their association with the dependant variable, which means that variables may be statistically significant in a multivariate model but not statistically significant in a univariate model (Hosmer and Lemeshew, 1989). The interaction term in the mixed model was used to find out whether the type of imformation (lifestyle changes versus implications) was different in the MI and BPC groups (page 12, lines 13-14).

12. I would like to see a table comparing the 27 not included with the remaining included, or at least description of effect sizes and confidence intervals in the text rather than just p-values.

Baseline characteristics of the sample as a whole, as well as divided according to participants (n = 244) and non-participants (n = 27) has now been included (Table 3 of the revised manuscript). No differences were apparent between these groups. The results text has also been modified accordingly (page 13, lines 6-9).

13. What were the participants like at baseline? A table reporting this would be useful.

A table of baseline characteristics has now been added (Table 3 of the revised manuscript).

14. Is the mean an appropriate measure to use to report the number of items recalled (what was the distribution of responses like?)

The mean was an appropriate measure to report the total number of items recalled as it was normally distributed.

15. How do the results adjusted and not adjusted for days between initial interview and recall interview differ?

The univariate model presented in Table 6 shows the data unadjusted for days between feedback and the recall interview. Multivariate model 1 then adjusts for both days between the interviews and the feedback condition parents were allocated to, given both variables were weakly associated with recall score. This adjustment made no substantive changes to any of the estimates or confidence intervals.

16. Was ‘time between interviews’ included as a linear effect? Was the appropriateness of this checked?

When time between interviews was included as a linear effect, the residuals were normally distributed.

17. Rather than using a p-value to make a statement, report effect sizes and confidence intervals.

The majority of results presented in our manuscript already had confidence intervals included. However, we have added the requested data to the baseline comparison table (new Table 3) and to a few relevant places in the manuscript (page 13, line 19, page 15, lines 20-23 and page 16, lines 4-9).

18. On line 25 of page 13, one of the upper limits of the confidence intervals looks incorrect. Should it be 0.01 to 0.27 (instead of -0.27)?

We have corrected the error.

19. Report absolute numbers along with proportions or percentages.

These numbers are contained with the table (table 5) but we have added them to the manuscript as requested where appropriate (page 14, lines 22-23).

20. Why are the univariate findings described in more detail than the multivariate ones? Again, effect sizes and confidence intervals are more informative than p-values.

We have completely re-written the section regarding table 6 in line with the reviewer's comments (page, 15, lines 20-23 and page 16, lines 1-9).

21. For the %s in the correct recall column of Table 4, it would also be useful to report the %s of the total n (not just the n who had recalled the items).

The information has been added as requested, including changes to the text (page 15, lines 1-10).

22. Report exact p-values in tables rather than stating that $p < 0.05$

All tables only include exact p-values or confidence intervals as appropriate.

Thank you for the opportunity to submit our paper. We hope that the revised manuscript may be deemed suitable for publication in BMJ Open and we look forward to hearing from you in due course.