Author's response to reviews

Title: Preferred intensity exercise for adolescents receiving treatment for depression: A pragmatic randomised controlled trial

Authors:

Tim Carter (timothy.carter@nottingham.ac.uk)
Boliang Guo (guo.boliang@nottingham.ac.uk)
David Turner (David.A.Turner@uea.ac.uk)
Iaonnis Morres (iomorres@pe.uth.gr)
Elizabeth Khalil (Elizabeth.khalil@nottingham.ac.uk)
Emily Brighton (emilybrighton@hotmail.co.uk)
Patrick Callaghan (Patrick.callaghan@nottingham.ac.uk)

Version: 2
Date: 11 August 2015

Author's response to reviews: see over
Dear editor and reviewers

Thank you for your consideration and feedback, it has substantially improved the paper. Please find a point by point response for each reviewer’s comments below. I have included the corresponding reviewer comment above each response.

Referee 1

I suggest that the title reflects the PICO ‘criteria’ for clarity: Aerobic exercise versus treatment as usual for adolescents with major depression. A Randomized clinical trial. Just my opinion.
Response: We have decided to keep the title as it is – especially considering the participants were not formally diagnosed with MDD. Thank you for the suggestion.

Please specify that the primary outcome was depression score post-intervention.
Please specify type of exercise (aerobic exercise
Response: This has been done

I disagree with the authors on the conclusion. I think the conclusion should primarily state that there was no effect of exercise post-intervention on depressive symptoms. The positive finding was a secondary outcome, borderline significant and not related to physical activity otherwise or quality of life – i.e. could therefore be a type 1 error.
Response: The conclusion now includes a statement that the findings should be taken with caution due to the small sample size. Further points have been included in the discussion regarding the potential for type 1 error.

The references chosen to argue that exercise is beneficial is adults are peculiar. Why not site systematic reviews of RCT. E.g. Cochrane. These also suggest that exercise not beneficial in studies with low risk of bias, which could explain the findings in the current study.
Response: Reference to the mentioned Cochrane review has been included

The authors cite and discuss results from longitudinal non-controlled studies, which states that lack of exercise is associated to high risk of depression. No causality can be claimed here. I suggest that the authors focus on results from RCTs.
Response: Results from RCTs have now been focussed on

No plausible mechanism for an antidepressant effect is offered?
Response: A brief overview of antidepressant theories of exercise has been added

Please describe the TAU intervention in more detail. ‘Psychological intervention’ is a very broad and in a professional context meaningless term.
Response: I have included a separate section outlining TAU – I have also pointed the reader towards Table 2 where more detail is given regarding the specific treatments the participants were engaged in.

Was any form of exercise capacity measured before and after to see whether they actually increased their fitness levels? And why not?
Response: No this was not undertaken as it was not the purpose of the study, we did not anticipate that fitness would be the mechanism by which exercise may alleviate depression. Furthermore, the duration of the intervention did not allow time for fitness levels to be substantially changed. We also
felt it to be very intrusive and as we did not recruit an exercising sample (see baseline LTEQ scores) we did not want to discourage participation with the study.

It is stated that the trial statistician generated the allocation sequence — but how?
Computer generated?
Response: I have now stated that it was computer generated

Primary outcome: Why choose a new depression scale? As I understand this is a self-reported instrument and therefore the primary outcome is not blinded?
Response: It was felt appropriate to use the CDI-2 as there were a number of changes from the original based on updated clinical perspectives. For instance, items on hypersomnia and hyperphasia in youth and the age-relevant depressive symptom of dysfunction in thinking and concentration were included. Furthermore, some items were revised from the CDI due to the discrepancy between respondents’ interpretation and initial intent [1]. We have also included more psychometric information on validity in the paper.
Response: The outcome assessors who were present when the young people filled them out were blinded. Considering the young people typically required some assistance at times from assessors it was important that they were blinded to treatment condition so as to not unduly influence the outcome.

Sample size description is not sufficient. Please report how many points on the CDI which is expected from exercise and cite a relevant paper. Citing Cohen is insufficient. It was meant for effect of teaching and not exercise!
Please state the expected SD post-intervention.
Response: We have cited the Cochrane Review in this area as this established a moderate effect size for exercise in a general population - this was our best estimate and was the basis of our sample size calculation. Sample size calculation using effect size is common practice in psychology. Furthermore, as there were no previous studies that had compared exercise to TAU in depressed adolescents we had no mean differences to cite and use in a power calculation.

It is stated that ITT analysis is used. According to the tables this is not the case. Please define ITT analysis. Please report results using all patients regardless of compliance and follow-up and use appropriate statistics than can handle missing data.
Response: We had previously conduced imputed and observed case analysis for comparison. We have now included the ITT analysis and amended the tables accordingly.

Figure 1 is problematic. A lot of patients appear to be excluded from analysis. Post-hoc exclusion of patients for failing to fulfill inclusion criteria is also problematic. This patient should be included.
Response: This patient has now been included. Also, in error, Figure 1 was showing how many people were missing at each time not missing from analysis. All people who were randomised were included in the analysis. Figure 1 has been amended as the analysis presented is ITT and therefore all participants are included in the analysis.

‘A non-significant effect was observed’ – I disagree. It should clearly be stated that no effect was found post-intervention.
Response -This has been changed

The ‘significant’ effect at 6 months should be discussed in terms of the risk of type 1 error, bias (non blinded outcome assessment), and lack of appropriate handling of missing data known to cause exaggeration of treatment effects.
Response: The MLM model we used is identical to the method used for a trial published in The Lancet [2]. Using MLM to analyze trial data with repeated measures is common practice. Missing data have been imputed through multiple imputations using Realcom software. The person who was excluded has also now been included in the analysis. We feel that we have not inappropriately handled the six month data.

While the results from an unpublished study is quite interesting it is inappropriate to refer to these data in the discussion. Response: As this publication has now been accepted for publication it remains in the study and is referred to it as ‘in press’

Referee 2

• Minor Essential Revisions
1) Line 264 describing the participant excluded from analyses: Why was this participant not excluded from participating prior to randomization if they did not meet the cut point for depression? The choice to remove this participant’s data from the control group post-hoc needs to be clarified to reduce risk of bias. Response: This has been addressed, as this person has now been included in the analysis.

2) Lines 50-51: Change “Preferred intensity exercise in addition to treatment as usual” to something along the lines of “Preferred intensity exercise combined with treatment as usual / Treatment as usual with an additional pref. intensity exercise program” Currently reads awkwardly/unclear that it is a combined treatment or a list.

3) Line 60: remove “for this group”

4) Line 61: “Alongside this” awkward/colloquial, suggest using “Additionally”

5) Line 68: Add a comma after “national”

6) Line 116: Explain the slash – was stretching either at the start or end, or is it both?

7) Paragraph at Line 142 about heart rate measurement could be simplified. Use active voice for describing how measures were taken, i.e.: “To measure heart rate, participants placed two fingers on separate panels of the faceplate and held for 5 seconds... The exercise therapist could also visually verify heart rate from the watches.”

8) State in the recruitment process what kind of compensation was offered to study participants, or whether no compensation was offered.

9) Line 174: The acronym (SNOSE) is never used again, suggest removing. Response: Points 2-9 have been amended as suggested.

10) Paragraph at Line 196 regarding EuroQol measures has several concerns: A) Keep the short form of the EQ-5D-5L consistent (hyphens vary slightly between uses. Response: This is done

B) Description of the EQ5D-5L scoring is unclear. Does the scoring algorithm result in a range of possible scores from 1 to -0.59, or is that from your sample. In either case, how can a score be less than 0 if that is considered equivalent to death? Also, in what way is this considered equivalent to death?
Response: The methods used to score the EQ5D-3L have been fully described in the literature. (e.g. Dolan et al, 1996). There is also evidence that death is not the worse health state for many respondents (Patrick et al 1994). We merely reported the range of possible values obtainable in a commonly used HRQOL measure. A discussion of these issues, and the meaning/validity of negative scores was considered beyond the scope of the current manuscript.

C) Line 205-206 describing the relationship to “the 3 level questionnaire” is non sequitur. Why is this relevant? I’m assuming this is the EQ-5D-3L (apparent only after looking up the EQ-5D-5L) – and that the 3L is what was previously shown to be related to depression.
Response: Yes, that is correct. The 3L has been shown to correlate with changes in depression symptoms in a variety of studies. The 5L questionnaire was based closely on the 3L. The 5L scoring system was based on that of the 3L. There is currently not a body of literature supporting the 5L in depression. The evidence for the 3L was therefore used as the best available indicator that the use of this instrument was reasonable. However, we accept that this is speculative and will have to remain so until a body of literature using the EQ5D-5L is available.

11) Paragraph at Line 337 is not explicitly connected to the discussion. Connect with the subsequent paragraph and/or describe how the present study builds on it.
Response: This has been done

12) Line 453: Insert “capable OF substantial”
Response: This has been changed

13) Line 493: Heart rate should be reported as a % of maximum heart rate (using age based maximum estimate) or % heart rate reserve, as this allows for better comparison to guidelines for exercise intensity.
Response: This has been changed as advised

14) Line 532: Insert “with respect to INTERPRETTING the post-intervention results”
Response: This has been changed as advised

• Discretionary Revisions
15) Suggest adding a brief description to the introduction about why “preferred intensity” was selected
Response: This has been added

Referee 3

Major compulsory revisions
1. The trials register entry https://clinicaltrials.gov/ct2/show/NCT01474837 suggests that the trial intervention was exercise plus motivational interviewing. There is no mention of the latter in the description of the intervention. Although the authors do state that two project staff members “exercised and interacted with participants in all sessions”. This needs clarification.
Response: This changed from the trial register as we wanted to isolate exercise participation as the mechanism of change as far as possible (we wanted to reduce the potential for additional psychosocial therapeutic impact from facilitators). The reason the staff exercised with participants was to reduce the potential negative effect (pressure/judgement) that may occur with staff members watching the participants.
2. In the description of the analysis, the authors suggest that the primary model for the primary outcome included an interaction between treatment by time. However in the reporting of the results, the authors only present the findings for outcomes post-intervention and at 6 months. This is inconsistent and requires clarification. Was there any evidence of a differing effect of time over time? Given the numbers, I suspect that any such formal test of interaction would be substantially underpowered.

Response: During exploratory data analysis, we plotted the outcome against time by group and noticed the fit was not parallel which suggested we should include the interaction term in the modelling to reflect the nature of the data. The reviewer is right that the coefficient of interaction term is not statistically significant, however, the outcome of treatment effectiveness is the difference of change scores at each time point, to analyse this data we used a reliable and acceptable method (See Tyrer et al, 2013, Lancet paper).

3. In the results section, the authors talk in terms of findings that were or were not ‘statistically significant’. As the authors themselves acknowledge, they did not recruit the number as specified in the original sample size calculation in order to be adequately powered to detect the pre-specified difference. It is more helpful, in general, to talk in terms of the strength of evidence based on p values and use confidence intervals to describe whether the results exclude a meaningful difference between groups (see:226 Sterne, BMJ 2001;322). This needs to be addressed through this section.

Response: Some discussion regarding confidence intervals has been added into the discussion section.

4. In the discussion, the authors suggest that a lack of power explains the null finding post-intervention (6 weeks) and that the treatment effect at 6 months is a novel finding. However, there is substantial and, importantly, differential attrition. 43% of those in the intervention group and 58% of those in the TAU group are lost to follow-up at 6 months. Therefore I have concerns about the potential for bias in the results presented. The authors conduct a number of comparisons to test for differences in drop-out between groups and in terms of baseline characteristics that predict missingness, and finding no significant differences conclude that there is no bias as the missing mechanism is MAR. [The latter is incorrect, if there are truly no factors that predict the missingness then the mechanism is MCAR rather than MAR.] However, given the small sample, basing the conclusions regarding patterns of missingness on statistical significance may be misleading. Given the substantial amount of missing data, particularly at the 6 month point, the authors need to conduct additional sensitivity analyses to determine the potential for bias. Without doing so, I believe that there is a substantial risk that the authors may be drawing a conclusion regarding a long-term effect of the intervention that is not justified.

Response: We have removed this sentence regarding MAR from the paragraph. As per what Carpenter said, ‘We cannot tell from the data at hand whether the missing observations are MCAR, NMAR or MAR (although we can distinguish between MCAR and MAR)’ [3].

We acknowledge that more sensitivity analysis can always be conducted to present more evidence for the result’s robustness. However, the Multilevel modelling we presented took into account missing values automatically in order to give sensible results. We also imputed the missingness using a
Multiple Imputation procedure. As the 6 month measure is not the primary outcome, we made the decision to not conduct further sensitivity analysis.

Minor essential revisions

5. There is no mention of the CONSORT guidelines in the reporting of this trial and it would be helpful to ensure that the manuscript fully aligns with the required reporting standards for RCTs.
   Response: A statement assuring this has been added

6. The sample size calculation indicates that the target sample size was 158. However, only 87 individuals were randomised. There is no mention of the challenges that recruitment posed in this trial and brief details would be helpful.
   Response: This has been added to the limitations section

7. In discussing the literature in adults, the authors should reference the latest version of the Cochrane review of exercise for depression http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD004366.pub6/abstract;jsessionid=2852D -
   Response: This has been included

8. I struggled with the term ‘preferred intensity’. I think it needs to be clearer how this was agreed. There is no indication of how long each exercise session was. This needs adding.
   Response: More information on the exercise intervention and preferred intensity has been included to address these issues.

9. The authors state that intervention “engagement” was measured using a scale that measured exertion and heart rate readings. I would consider “engagement” to be attendance at sessions and better quantified as the number of sessions attended. Exertion and heart rate are linked with how hard the person is exercising and the level of intensity of exercise. The authors also describe how the participants were asked to report exertion at intervals throughout the exercise session but it is not clear on how many occasions this happened and whether this was consistent across sessions.
   Response: We have included the frequency (3 time points at each session). We have also amalgamated heart rate, RPE and attendance under ‘engagement’ as ‘exertion’ is a measure of participants specifically engaging with the exercises themselves.

10. Under eligibility criteria, 3rd line, typo ‘Imventory’
   Response: This has been changed

11. Under ‘randomisation and allocation concealment’, further information is required as to who gained informed consent for trial participation – was this the referring clinician or a member of the research team?
   Response: This information has been included

12. In the description of the LTEQ, it would be helpful to have a description of what the categories ‘active’, ‘moderately active’ and ‘insufficiently active’ equate to in relation to national guidelines of PA levels for children.
   Response: This information is not available as how these link to national guidelines is not expressed explicitly in the LTEQ.
13. The total number of participants assessed for eligibility is given as 128 – is this the same as the number of referrals that was received by the team? I wondered whether there were any individuals who were referred but who did not agree to be screened for eligibility.

Response: No this is unlikely to be the same as the total number asked and we have included a sentence stating this.

14. The description of the average number of sessions attended should include those who did not attend any sessions.

Response: This has been included

15. The presentation of data in the Table 3 requires attention. The means for the intervention and TAU groups are presented for the various outcomes. The ‘difference in reduction’ is incorrect. This is a (between-group) difference in mean scores. For the EQ-5D-5L, the authors present the median (IQR) values but then present what appears to be a difference in means. This is inconsistent and needs to be addressed. There should only be one p values for the comparison of LTEQ scores between the groups (rather than for each level).

Response: The results were the ‘difference in reduction’, ie. The difference in change between groups from baseline to follow up [4] – we have changed the wording to reflect this.

Referee 4

- Major Compulsory Revisions

1) Limitations of the work are clearly stated; however, many of them appear throughout the discussion section and should be added to the Limitations sections, along with some additional limitations as noted below.

Response: I have added to the limitations section.

The authors discussed lack of statistical power as one of the primary contributors to the lack of significance at post-intervention and do explicitly state this in the Limitations section. However, the Conclusion that effects may not occur until six months is misleading given the fact that the study was underpowered

Response: I have included a sentence that states this should be taken with caution due to sample size.

b. The selection of 6 weeks for treatment duration is low compared to other studies of treatments for depression, both pharmacological and nonpharmacological, with many of the studies examining exercise as a treatment for depression in adults having a duration of at least 8 weeks, and some 12 weeks. This may be a primary
factor in the post-intervention results being non-significant. The authors acknowledge this as a potential limitation in the Discussion.

c. Although it is worthwhile that the study had limited eligibility criteria in order to be highly generalizable, the inclusion of participants who were moderately active and active at baseline is a limitation, and another possible contributor to the non-significant finding at 6 weeks. If these participants were already exercising at baseline, it is unclear how the intervention provided by the study compared to existing levels of activity. Did participants continue with their existing activities while in the study? This should also be acknowledged as a potential limitation.

Response: Participants were not encouraged to stop other activities. This has been added a potential limitation

Additionally, it is interesting that there were fewer inactive and more moderately active participants in the TAU group at week 6, suggesting that there was change in activity level within that group.

Response: As there were no statistically significant differences on the LTEQ we have decided not to include a discussion on this.

2) Greater details of the exercise intervention should be reported. Was all exercise aerobic, and what types of exercise were performed? How many minutes of exercise? These factors may be helpful in understanding the results as well.

Response: This has been included

- Minor Essential Revisions

3) Because the study is evaluating exercise as an adjunctive treatment, it may be more appropriately referred to as an augmentation study throughout the manuscript and abstract.

Response: We have taken this under consideration but feel it is appropriate to remain as it is considering it is clearly outlined that the intervention is delivered alongside TAU.

4) Both the CDI-2 and EQ-5D-5L are relatively new measures. As such, it would be quite helpful to provide some psychometrics to orient the reader to the available information on validity and reliability of these measures, and particularly how they compare to other commonly utilized measures (e.g., Children's Depression Rating Scale-Revised).

Response: We have now included psychometric data on the outcomes
