Author's response to reviews

Title: Rhythmic cued motor imagery and walking in people with multiple sclerosis: a randomised controlled feasibility study

Authors:

Barbara Seebacher (b.seebacher@aon.at)
Raija Kuisma (r.kuisma@brighton.ac.uk)
Angela Glynn (a.glynn@brighton.ac.uk)
Thomas Berger (thomas.berger@i-med.ac.at)

Version: 2 Date: 28 May 2015

Author's response to reviews: see over
The effect of rhythmic cued motor imagery on walking in people with multiple sclerosis: a randomised controlled pilot trial Barbara Seebacher, Raija Kuisma, Angela Glynn and Thomas Berger

Dear Pilot and Feasibility Studies Editorial Team, dear Dr Walters,

Thank you very much for allowing us to send you our revised manuscript and cover letter with our point-by-point response to the reviewers’ comments. Please find enclosed the revised version of our manuscript MS: 2345835161659644: “Rhythmic cued motor imagery and walking in people with multiple sclerosis: a randomised controlled feasibility study” which we resubmit for consideration as research article in PILOT AND FEASIBILITY STUDIES. All changes to the manuscript are highlighted in red text.

Reviewer 1
Dear Dr Annabel Allison,

Thank you for your helpful comments on our manuscript. Please see our response as follows.

In general the authors place too much emphasis on hypothesis testing, when it is likely to be
severely underpowered for detecting a meaningful difference. Instead the authors should focus on the feasibility and acceptability of a full RCT, and on deriving estimates of parameters used in the sample size calculation for the main study.

Statistical Analysis
1. The statistical analyses should be mainly descriptive in a pilot study, so I suggest removing the information on hypothesis testing and accompanying P-values and present summary statistics only. Any differences between groups (e.g. gender at baseline) may still be highlighted.

   We agree with you to descriptively show the results in our pilot study. Therefore, we removed the information on hypothesis testing and accompanying p-values and presented summary statistics only. We still highlighted the differences between groups, including gender at baseline.

2. The authors should also state the feasibility measures that have been recorded in the statistical analysis section (e.g. recruitment rate, retention rate) and the feasibility criteria.

   We have now clearly stated the feasibility measures that have been recorded in the statistical analysis section (recruitment rate, duration of recruitment, retention rate, safety, adverse events, compliance and acceptability of the interventions) and the feasibility criteria.

Results
1. In the results section, paragraph 1, the authors state that the ‘larger study will be feasible’ before the sample size for the full RCT has been presented. It is therefore unclear to the reader as to whether the full RCT is feasible. Maybe reorder the information in the results section.

   In the results section, we reordered the information to step-by-step show the reader that the full-scale RCT is feasible. Then, we presented the implications for the main study including the sample size calculation.

2. Table 1 - testing for differences in baseline variables does not need to be performed in an RCT so please remove the P-values. Differences in a baseline variable are only important if the variable is thought to be related to the outcome, in which case the analysis is adjusted for this. For age/EDSS, please present as either continuous or categorical variables as opposed to both.
We removed the p-values from Table 1 as we do not consider the differences in baseline variables to be related to the outcome. We presented age as continuous variable and EDSS as categorical variable.

3. Table 2 - again remove the P-values.

We removed the p-values from Table 2.

4. I suggest adding a subheading for the sample size calculation for the main study, as it is one of the key aims of the trial. I would like to see more detail in this section including any parameters used for the calculation, e.g. clinically relevant difference/standard deviation. Is there a reference for the a priori sample size? Also, it is not clear whether the sample size achieves 90% power for both primary endpoints - please add this information.

We added a subheading for the sample size calculation for the main study and additional subheadings for all other implications for the main study (recruitment, randomisation, allocation concealment, and blinding). In the sample size section we presented detailed information on parameters used for the calculation, based on the effect size within the 95% confidence interval of this study, plausibly-achievable effect size against the literature of similar interventions in this population and people with stroke as well as clinically relevant changes (as suggested by reviewer 2). The a priori sample size was estimated in the study protocol and approved by the ethics committees of the Universities Brighton and Innsbruck. We added this information to the manuscript. The current sample size calculation is based on 80% power for both (former: primary) secondary walking outcomes; we also added this information to the manuscript.

5. I am not sure that Figure 2 really adds anything to the paper since most of the information is already presented in Tables 1 and 2. The only additional information are the individual patient scores, however, the y-axis scales do not make it possible to find out the exact values. I suggest removing this.

We agree with your suggestion and removed Figure 2.

Minor essential revisions

1. Throughout the paper remove the +/- sign before standard deviation throughout and present results as mean (sd).

Following your advice and suggestions from reviewer 2, Professor Toby Prevost, throughout the paper we presented the results as mean (95% CI).
2. (Abstract, paragraph 2) I would state that the control group are receiving usual care. In the abstract, we added that all participants received usual care as this was the case. We also added this information to the CONSORT flow diagram.

3. (Abstract, paragraph 3) Highlight that the results show mean (sd). We highlighted that the results show mean (95% CI). Please see your first minor comment and our response as well.

4. (Methods, Primary outcomes, paragraph 2) Remove ‘Timed’ from the first sentence. We did accordingly.

5. (Discussion, paragraph 8) Change ‘…to enable to calculate…’ to ‘…to enable us to calculate…’
   We changed the sentence to “The study was underpowered because of the small sample size, but this was a pilot study to enable us to calculate the required sample size,…”

   (Discussion, paragraph 9) Suggest reword of ‘Third, a single experienced…’ (e.g. ‘Another possible limitation of the study is the use of a single physiotherapist to give instruction to participants and undertake all measurements’). Remove ‘on’ from ‘…could not influence on participant allocation…’
   We did accordingly.

   (Results, paragraph 10) Change the last sentence to ‘The stratification will help ensure that the intervention and control groups are similar with respect to patient characteristics’.
   We did accordingly.

Whilst the CONSORT diagram is useful to the reviewers, I do not think it should be included as supplementary material if the paper is published since page numbers will not appear in the final publication.

   We did not include the CONSORT checklist as additional/supplementary material (but included the flow diagram as Figure 1); thank you for your advice.

Reviewer 2
Dear Professor Toby Prevost,
Thank you for your helpful comments on our manuscript. Please see our response as follows.

**Major**

1. Abstract - as it is a feasibility study, it is not powered to do tests to detect intervention efficacy effects and so the results for these ought to be given as 95% confidence intervals rather than p-values for testing hypotheses.

   According to your suggestions, we presented the results as mean (with 95% confidence interval).

2. Abstract - line 46: the aims of the study surround feasibility outcomes. This is given very little proportion of the abstract results. Secondly, is it also possible to report on fatigue which might be a barrier for the main trial?

   We rewrote the abstract and focussed mainly on the feasibility outcomes. Secondly, we also reported on fatigue which we have assessed using the Modified Fatigue Impact Scale.

3. Abstract lines 49-49: it is rather 'dangerous' to say that the intervention is effective. The conclusion I think ought to be on the aims of feasibility. e.g. see line 150.

   We agree with your comment. In the abstract and throughout the manuscript we deleted the statements on efficacy and reported the feasibility results. Please see also our response to reviewer 1, Dr Annabel Allison.

4. Line 174: More details need to be provided for the randomisation (such as the process) because it is unclear how the plan to achieve 10 per arm could be known to be achieved if the method were simple randomisation as stated. Line 186 instead suggests that this is a form of restricted randomisation rather than simple randomisation. For the main trial it would be worth considering devising better concealment. For example towards the second half of the study one treatment arm can have (nearly) run out already, and knowledge of this can lead to potential recruitment bias where the recruiter consciously or unconsciously recruits patients for whom it is (pretty well) known will go into certain groups using loose eligibility criteria. This point is relevant to lines 485-486 where is would be worth blinding the researcher in the main trial.

   We agree with you that restricted randomisation as opposed to simple randomisation was used and we have renamed this in the text. However, we had used NHS based
information for our decision to refer to simple randomisation (please see [http://www.nbt.nhs.uk/sites/default/files/filedepot/incoming/Randomisation_in_clinical_trials.pdf](http://www.nbt.nhs.uk/sites/default/files/filedepot/incoming/Randomisation_in_clinical_trials.pdf)). We provided more details for the randomisation process in the manuscript. Please let us know should you regard the information being nonsufficient. We devised better concealment for the main study which is described in the results section “Allocation concealment” (lines 496-504). We referred to blinding in the main trial in the results section “Blinding” (lines 506-514).

5. Line 246: I think that the title and the text in line 247 refer to primary outcomes intended for the main trial rather than the primary outcomes of this feasibility trial. In which case, it would be worth re-labelling these so that the feasibility outcomes become the prominent outcomes of this feasibility trial. Really, the feasibility outcomes ARE the primary outcomes of this trial.

This comment was very helpful to us because we became aware of the fact that the feasibility outcomes are the primary outcomes of our study. Thus, we changed the title to “Rhythmic cued motor imagery and walking in people with multiple sclerosis: a randomised controlled feasibility study”. Further, in the methods section, we have re-labeled the outcomes so that the feasibility outcomes including fatigue became the primary outcomes, and the walking outcomes became the secondary outcomes. Please see also our response to reviewer 1.

6. Line 278-279: As above, I think the focus ought to be on using these methods to provide 95% confidence intervals (rather than p<0.05) for measures which are not intended to address efficacy hypotheses which are to be evaluated in the main trial. Lines 284-292 would work in the main trial but not be sensitive enough to work in 10 per arm. More could be put in on what would be estimated with 95% confidence intervals on the feasibility side of the outcomes. Estimating the recruitment rate per month would seem important, as an initial piece of information to find out the number of centres needed for the main trial. Also, the proportion of patients in this period who were eligible (189 out of 471 with 95% CI), and the proportion of these who consented to take part in order to assess the proportion of centre patients likely to participate. Also an exact 95% confidence interval for the 100% retention which can then assist in justifying the choice of 10% dropout allowance in the main trial, along with dropout seen in other similar but larger studied.

Instead of hypotheses testing and efficacy statistics, descriptive statistics (mean with 95% confidence intervals) was used. Medians (range) were reported for ordinal
(fatigue, EDSS) data, and raw count (number, %) was reported for nominal data. Feasibility outcomes were estimated according to your advice with 95% confidence intervals, such as the recruitment rates (per month), retention rate, eligibility rate, and consent rate. The latter were used to assess the proportion of eligible participants likely to consent. An exact 95% confidence interval was estimated for the 100% retention to justify a (newly calculated) 12% dropout allowance, along with dropout in other similar but larger studies. Please see methods and results sections.

7. Lines 374-375: the results on improvement here might be better than what is currently in the abstract on potential early efficacy.

Following your suggestions, we changed these lines in the abstract to “The mean improvement in walking speed in both groups A and B was -0.9 (95% CI -1.3, -0.5) seconds, and mean worsening in group C was 0.4 (95% CI -0.3, 1.1) seconds. The mean improvement in walking distance in group A was 68.1 (95% CI 51.4, 84.7) metres and in group B 92.9 (95% CI 55.2, 130.5) metres, and mean worsening in group C was -9.4 (95% CI -35.6, 16.9) metres.”

8. Line 386: “true difference”: it is important to provide the effect size that is to be detected. This might be a difference between arms in mean of the outcome. It needs to be a plausibly-achievable choice of effect size (within the 95% confidence interval of this feasibility trial) and plausibly-achievable effect size against the literature of similar interventions in this population. It also needs to not to be too large that one would then be underpowered to find a smaller useful difference. This will need clinical judgment as to how many metres is useful in practical terms for patients.

We provided the effect sizes for the differences in means of the walking outcomes (Cohen’s d). A plausibly-achievable effect size was chosen (within the 95% confidence interval of our study) and this effect size was plausible-achievable against the literature of similar interventions in this population as available, and in people with stroke. According to our clinical judgement, the percentage of participants with a clinically relevant improvement at 20% and above was used for further effect size estimation. This effect size estimation was rather conservative in order to not being underpowered in the main study.

Discretionary

9. Abstract lines 36 and in Results: the use of +/- makes it ambiguous as to what number that follows this represents. Recommend using mean(sd) or indicate se,
this this is what it is. Though better still to summarise results with 95% CI. Although the meaning of +/- is mentioned lower down, the abstract needs to be unambiguous when it is read stand-alone.

We used mean (95% CI) in the abstract and throughout the article.

10. Background line 90 and surrounding paragraphs: it is useful to have the background references and general findings, but it would be better to have an idea of the sizes of the effects found in these papers for the various interventions, especially if they used the same or similar outcome measures being feasibility assessed here for the main trial. Were they from similar populations and severity of MS.

We enclosed the effect sizes of the improvements found in the referenced papers, if they were stated. Unfortunately, many articles did not report effect sizes. In addition, for description of the studies in people with MS, we added the EDSS or a severity description (e.g. mild, moderate, severe MS) to enable comparisons to be made by the reader.

11. Line 145: was acceptability of selected nature of music, and tone pitches assessed in the 30 feasibility participants?

The acceptability of the music melody and beat and the tone pitches of the metronome was assessed narratively during the weekly phone calls and at the follow-up assessments. We included this information in the manuscript (in the primary outcomes section, lines 267-270, and in the results section: safety and acceptability, lines 392-397)

12. Line 173: I’m not sure that recruitment using unselected consecutive Sampling should be also described as convenience sampling, as this may put a lower light on what you did.

Thank you for this comment. We removed “convenience sampling”.

13. Line 243: some description if usual care might be useful as the readers may be interested to know how the context may apply to their situation. Was it all 3 arms that had this usual care or just the control group; if all 3 then the consort flow diagram could have ‘usual care’ in each of the 3 arms. It may be worth discussing which, if any, of the arms had more attention and time spent on them.

All participants had usual care, and we changed the manuscript and CONSORT flow diagram accordingly. Please find a description of usual care and time spent on
participants in the “intervention” section, lines 255-259. In paragraph one of the study limitations we discussed the issue of attention and time spent on participants of the three groups, lines 628-637.

14. **Line 322-323**: were these walking aids used at both timepoints for these individuals or at which timepoint only?

The walking aids were used at the baseline and follow-up measures to allow for comparisons. We have changed the relevant sentences. (please see lines 307-309 and 432-434)

15. **Lines 335-336**: "most participants" - this could be more specific.

We provided more detailed information and changed parts of the “safety and acceptability” chapter to: “All participants in group A reported that they liked the music melodies, particularly because different music styles were used, and that the beats supported their motor imagery of walking. Five out of 10 participants in group B regarded the metronome cues being helpful to keep the tempo of their imagined steps, 2/10 concerned them neutral, and 3/10 found them boring. All participants in group B regarded the variety of tone pitches being acceptable. All but one participant (in group B) reported that the verbal cueing of the researcher facilitated their attention with the motor imagery and helped them to stay with the beat. No other participants reported problems, and 8 out of 10 participants in group A and 4 out of 10 in group B reported the intervention as being pleasurable.”

16. **Line 380**: if a test needs to be used, then please check this is an exact test.

We removed the p-value as it relates to significance testing (referring to your comments above and comments by reviewer 1).

17. **Line 396**: The main trial will have a primary outcome that is a change in value from baseline to follow-up. So it may be worth considering which baseline factors are most predictive of change in primary outcome. This generally is the baseline of the primary outcome itself in many trial. It may be relevant to consider stratifying the randomisation on one (or two) predictive factors such as this, rather than use the method of blocked randomisation.

We had proposed to do what you suggest, but obviously our wording was somewhat unclear. All three baseline factors age, gender and disability (EDSS) are predictive of a change in the primary outcomes (walking speed, walking distance) of the main trial. Typically, people with MS at a higher age also have more severe disability, but this is
not always the case. Additionally, males usually develop a more severe disability over time, even if they are younger. Therefore, we intended to stratify the randomisation on age, gender and disability, with 8 strata resulting. We did not describe the strata in the article, but might do so if you would suggest that. We would prefer stratified blocked randomisation with permuted blocks to stratified randomisation because our trial will be too small to ensure balance in groups (Lachin, Matts et al. 1988, Rosenberger and Lachin 2002, Kang, Ragan et al. 2008), and during recruitment, we will not know the number and baseline characteristics of all participants in advance (Lachin, Matts et al. 1988, Suresh 2011). Moreover, permuted blocked randomisation would help to force balance in the numbers of participants allocated to the intervention and control groups (Matts and Lachin 1988, Rosenberger and Lachin 2002). We chose the small block sizes of three to be able to fill the blocks (Rosenberger and Lachin 2002). With allocation concealment, allocation bias should be avoided.

18. Table 1 - it would help readers to have the names of the Groups as well as the A, B, C.
   We changed both tables accordingly.

19. Was the feasibility study designed to assess whether one of the intervention arms should be dropped e.g. over-fatigue in the combined arm?
   Yes, we also assessed fatigue using the Modified Fatigue Impact Scale. Following your advice, we included this information in the feasibility outcomes (lines 266, 270 and 274-282), statistical analysis (line 333) results (lines 355 and 412-420), discussion (lines 522-525 and 563-567) and conclusion (line 670) sections.

20. Table 1 - p-values are never supposed to be used in comparison of baseline in randomised trials because they test the hypothesis that the randomisation is fair generator; and it is know that this is the case. The impact of an imbalance depends also on the correlation it has with the outcome.
   We completely agree with your comment. Therefore, and also in response to reviewer 1, we removed the p-values from Table 1. Due to the limited number of participants, it was not possible to evaluate the influence of gender imbalances on the walking outcomes with appropriate statistical power.

21. Table 2 - the results presented are significant for change from baseline to follow-up but not for comparison at follow-up. This lessens, to me, the impact of the results presented in the abstract, and increases the case not to put much at all in the
abstract about "significance".

As above, we agree with your comment. Thus, we removed the p-values from Table 2 and the abstract, and we presented the results using summary statistics.

22. Given that there may be only 2 timepoints, it would be more transparent to show the results of change from baseline without interaction between group and timepoints. This would then naturally lead to providing 95% CI for change scores.

According to your suggestion, we showed the results of change from baseline to follow-up without interaction between group and time, providing 95% CI for change scores.

23. Table 2 - participants with clinically significant improvement: it would be useful to present exact 95% confidence intervals within each of the 3 groups for this (3/10, 3/10, 0/10) because these indicate the current potential of what might occur in the main trial for these 3 arms.

We enclosed exact 95% confidence intervals for participants with clinically significant improvement within each of the 3 groups; please see Table 2 and “results”: “walking speed” and “walking distance”.

24. Lines 467-471 - this limitation is at the bottom of the article but the readers of abstracts will not catch it. This is why I think it is important that the stand-alone abstract needs to have hardly any emphasis on the efficacy, and that the methods section ought not to be going beyond confidence intervals into tests and a long way beyond to sphericity/normality assumption testing on 10 per arm.

Thank you for this suggestion. We rewrote the abstract and removed any information on efficacy testing and assumptions for Two-Way Mixed Design (Factorial) ANOVA from the methods section and remaining sections of the manuscript. Instead, we included descriptive statistics as appropriate. Additionally, we kept the (slightly changed; according to suggestions from reviewer 1) limitation “The study was underpowered because of the small sample size, but this was a pilot study to enable us to calculate the required sample size. The results are to be considered preliminary and they cannot be generalised yet to a larger population of people with MS. A subsequent well-powered main study is being conducted.” (lines 637-641).

Thank you again for your helpful comments which supported us to improve our manuscript.

Kind regards,
Barbara Seebacher