Reviewer's report

Title: The effect of rhythmic cued motor imagery on walking in people with multiple sclerosis: a randomised controlled pilot trial

Version: 1 Date: 7 May 2015

Reviewer: Toby Prevost

Reviewer's report:

Design = randomised controlled pilot trial
n = 30
3 arms all having 17 minutes of motor imagery, 6 times per week, for 4 weeks plus
arm 1 music (A)
arm 2 metronome cues (B)
arm 3 controls (C)

Aim - feasibility of a RC pilot T

Primary outcomes = walking speed (Timed 25-Foot Walk, T25FW) and walking distance (6-Minute Walk Test, 6MWT)
Feasibility outcomes = recruitment rates, attrition and adverse events

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.

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?

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.

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.

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.

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.

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

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.

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 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.

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.

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

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.

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.

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

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

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

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.

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

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?

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.

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".

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.

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.

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.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests