Author's response to reviews

Title: Tai Chi and Vestibular Rehabilitation Improve Vestibulopathic Gait via Different Neuromuscular Mechanisms: Preliminary Report

Authors:

Chris A McGibbon (cmcgibbon@partners.org)
David E Krebs (dkrebs@partners.org)
Parker W Stephen (swparker@partners.org)
Donna M Scarborough (dscarborough@partners.org)
Peter M Wayne (PWayne@nesa.edu)
Steven L Wolf (swolf@emory.edu)

Version: 3 Date: 19 October 2004

Author's response to reviews: see over
Response to Reviewers

Again, we thank the reviewers for their thoughtful critique and valuable insights. We have made rather extensive modifications to the manuscript to adequately address the reviewers concerns. The primary concern of all three reviewers was the statistical treatment. We have addressed this two ways: 1) we have reduced the number of variables and therefore, reduced the number of post-hoc comparisons made, and 2) we have used a Holms stepdown Bonferonni approach, as suggested by Dr. Ludbrook, to establish the alpha criteria for statistical significance. Please find below a point-by-point response to each concern raised by the reviewers, and a description of the changes we have made in response to those concerns.

Reviewer #1

2-3 As 19 subjects were omitted, data is presented concerning 36 subjects, not 53 subjects. Please revise the abstract so that it is not misleading.

Response: We have changed the opening line in the Abstract to describe 36 subjects instead of 53. However, to do this we also had to delete the two lines which followed the opening statement (describing the 19 of the 53 who were omitted from the study).

2-9. It would be clearer to say that measures of gait time-distance were improved in both groups following treatment. As it stands, it sounds as if measures did not improve till after treatment.

Response: This sentence has been edited as suggested (pg 2, line 8).

2-9. Please indicate here that the statistical criteria for significance were not conventional in this exploratory study. This must go in the abstract so that it is not misleading. An alternative, and much clearer approach, would be for you to say that you noted some trends that did not attain statistical significance due to the exploratory nature of this study.

Response: We agree, and have modified the text to be more accurate.

2-13. Conclusions. These conclusions do not follow from your data. Please defend or revise.

In particular, in 2-14 the term “better organized lower extremity neuromuscular patterns” embodies a qualitative value judgment (better), which seems out of place in the conclusion of a basic science paper. The results support the adjective “differently”. Also, in 2-15 the vague term “vigorous gait” seems out of place in the conclusion of a basic science article. Could you replace this term with something quantifiable? In 2-18, we do not see that these data suggest that TC confers benefits not attainable with more conventional interventions, because we don’t see that there is any proven benefit of TC vs. VC.

In the reviewers view, both groups of subjects walked more quickly after an exercise program. Kinematic measures suggested that they might have walked with a different organization, according to the training program. The VC subjects exhibited a pattern that may reflect optimization for head stability. The TC subjects may have optimized for something else, or have been less constrained, which resulted in the difference.

Response: We agree and have reworded the Conclusions section of the Abstract.
3-2. Suggestion – don’t use the “VSP” abbreviation – spell vestibulopathy out.

Response: We have replaced VSP with vestibulopathy throughout.

3-4. This sentence, the first one in the body of your manuscript, is awkward. What exactly do you mean by “typical role activities”?

3-7. This sentence is a little murky. Isn’t VR designed to compensate for VSR loss as well as VOR loss?

3-14. This sentence directly states that TC “balance the flow of … life force”. This is an inappropriate statement for a basic scientific paper. Please reword – you could say that TC is intended to balance flow of life force.

3-16. This sentence does not logically follow from the preceding sentence (about the “soft unfocussed gaze”). Please revise. You might wish to say that in VC, the explicit object of many of the exercises is to improve the stability of the eye in space, while in TC, emphasis is placed upon smooth and circular movement.

Response: The above referenced sentences have been edited to be clearer.

4-1. How do locomotor exercises enhance somatosensory input? Do you mean that they increased central weighting for somatosensory input? (this is a repeat of the same question in the first review).

Response: This line was deleted in an effort to make the paragraph shorter and more to the point (see following response).

4-4-7. Here you simply seem to be saying that both VR and TC should have positive carryover into gait because they both involve exercises done while standing. Could you shorten this paragraph?

Response: The above referenced lines have been edited to be clearer, and the paragraph shortened considerably.

4-16. Do you mean “outcome variables?” instead of “outcomes variables”? Is reference 19 (to yourselves) really needed in this sentence that says that you don’t know of anyone who has done this work before?

Response: The entire passage mentioning that these outcome variables have not been previously investigated has been deleted. We agree with the reviewer: the statement is not needed.

5-7. Your specific hypotheses do not follow from your rationale because you have not indicated how the locomotor activities of TC differ from VR. You are making a prediction about output without providing any information about input. Thus, they seem more likely to be a-posteriori than a-priori in nature. Please defend or revise.

Response: While we did describe briefly the nature of the two different interventions in the Introduction, the reviewer is correct that the “input” is not described. However, an adequate description of this “input” is best left for the Methods section which describes
the interventions. We have therefore edited the hypothesis rationale to reflect this (pg 4, lines 18-19).

5:9-10. Please use a more precise term than “directly related”. Do you mean “correlated” in the statistical sense?

**Response**: Yes, we were referring to positive correlations, and have edited the text accordingly.

6-1 – do you mean “had”?

**Response**: Yes, and edited accordingly.

6-5. What does the abbreviation “SVAR” mean? Assuming that it means something related to rotational testing, as patients with bilateral loss often have normal gain at 0.5 hz, the implication is that these bilateral patients were very severe. You might want to make this observation in the text.

**Response**: “SVAR”, or sinusoidal vertical axis rotation, has been replaced with “rotational testing”. Regarding the severity of bilateral loss, we have made this point in the methods (pg 5, lines 24-25) and discussion (pg 17, lines 10-11).

6-6. As 25% unilateral paresis is roughly the 5th percentile of for normal subjects, using a 30% unilateral reduction as a criterion for a unilateral loss means that your UL subjects may be nearly normal. Because your UL’s may have been nearly normal, and your bilaterals were very severe, the proportion of bilaterals between your two subject groups may have been critical.

**Response**: Although we tested for significant differences in group distribution of bilateral and unilateral vestibular deficit (there were none, see pg 11, lines 4-5), the reviewer makes an excellent point which we addressed above.

6-11. Should use past tense.

**Response**: Text modified accordingly.

7-13. During the 20 minutes of warm-up exercises, how many minutes were spent standing?

**Response**: Approximately 10-12 minutes of the 20 minute warm-up was performed while standing, and the remaining time while sitting. We did not mention this in the manuscript as it seemed a minor detail, though we now include it as requested (pg 7, line 6).

7-19 through 8-19. Please take out the irrelevant material about the intent of the exercises. This was also requested in the first review of this paper, and the material was not removed.

You describe a conventional vestibular rehabilitation program, and should shorten this section by referring to a textbook (such as Herdman’s). We do need to know what was done and how long was spent on it. How long did the subjects exercise? Did the stand for 70 minutes or did they spend some time sitting?

**Response**: We have removed some of the material regarding the details of the VR program, which is covered adequately in the manuscript Introduction section. We have
added clarification to the manuscript addressing the position of subjects during the exercises and the duration of the different treatment objectives. Subjects exercised for 60 minutes and an additional 10 minutes was made available for questions and answers with the instructor and assistance with individual progression of home exercises (see pg 7-8).

11-4 to 11. Here you appear to be testing you’re a-priori “general” hypothesis given on 4-20 through 21 that both groups would improve their time-distance measures. No correction is needed for the multiple t-tests.

Response: We agree.

11-12. You could leave out “when correcting for baseline differences”, as this is what “change” means. This construction is used in several other places, and could also be shortened there.

Response: Correcting for baseline refers to doing more than simply looking at changes. ANCOVA was used with the baseline value as a covariate, which effectively adjusts the means to have the same baseline value. We have made this clearer in the presentation of results by stating up-front, prior to presenting detailed results, that no significant between-groups differences were found for any variables (pg 11, lines 6-10).

11-15 and all later. Here you appear to be testing questions that I felt were a-posteriori (from 5:7-9), because no rationale was provided. Certainly very many t-tests were performed in which the investigators appear to have no expectations as to their results. In my view, you should have reduced level of “p” required for significance at this point, by dividing 0.05 by the number of a-posteriori comparisons (the Bonferroni correction) or something similar. On the web, there is a nice discussion of this problem at http://www2.sjsu.edu/faculty/gerstman/StatPrimer/anova-b.pdf As there are an immense number of comparisons, and the ‘p’ values are not terribly low, it seems unlikely that any of them would be significant using this conservative approach. The rewording in the manuscript would just be leaving out the word “significant” throughout.

Response: We did in fact provide a rationale for all our hypotheses. We have modified the statistical approach as suggested (see Methods, pg 10, lines 6-12).

11-25. Table 2 is neither interesting nor very informative. A graphical comparison would be more useful as one could both see the change as well as the variability. Another option would be to leave this table out. Same comments for table 3. These suggestions were made in the first review. These tables are not useful enough to keep. If you cannot put the data into a figure, please just delete them. Perhaps this journal has a way for you to put supplementary data on the web.

Response: We have eliminated Table 2 (as these variable are no longer considered in the analysis) and Table 3 (replaced with a bar-chart figure). Table 4 has also been eliminated (the data are simply included in the text).

13-16. Discussion: In general, the discussion is overly long.

Response: We have shortened the Discussion in this revised version of the manuscript.
13-26. Your hypothesis is about between-groups. Your results did not support a between groups difference. Therefore, your hypothesis was not “indirectly supported”, rather it was simply not supported. This comment was also made in the first review, and the authors did not adequately defend. Please revise.

Response: Edited as suggested (see pg 13, lines 9-13).

14:17-18. The statistics said that there was no significant difference between groups. By “clinically important” difference do you mean a trend?

Response: Yes. We now explain this more fully (see pg 14, lines 5-6).

14-24. This sentence strikes me as inappropriate as the goal for TC or VR is not to obtain a “more normal healthy coordinated pattern”, but rather is to optimize a number of interrelated variables – minimize energy expenditure, maximize speed, maximize safety, given their individual constraints in their sensorium, central processor, and motor plant. A “normal healthy coordinated pattern” of movement may be unsafe in a person with bilateral vestibular loss. In other words, the pattern observed in TC may not necessarily be good for them.

This idea is related to the topic brought up in the original review relevant to the original 9-11. The authors responded that reduced walking speed … are classical characteristics of impaired gait in people with … balance difficulties, and went on to argue that increases in these measures are signs of improved motor function. This argument misses the point that slowed gait (for example) in persons with imbalance may be an optimization weighting safety as noted before, and changes are not necessarily improvements. There is at least one study in the literature where increased activity from an exercise program was associated with increased falls.

Response: We agree with the reviewer, and have changed this passage accordingly (see pg 14, lines 12-13).

17-19. I disagree that the statistical approach used was reasonable for these multiple unplanned comparisons.

Response: Addressed above.

18-9. I disagree with this statement. Your BVH patients were extremely impaired, but your UVH patients could have been very nearly normal. It would help if the mean paresis and range was provided.

Response: As addressed above, we agree this should be highlighted in the discussion of limitations. Although we did address this in the prior revision, we now specifically state that the BVH patients were probably more disabled than the UVH patients (see pg 17, lines 10-11).

18-16. In the last sentence of this paper the idea of “spinal reflex compensations” is introduced. This idea was not mentioned in the discussion. This is not a good place to introduce a new idea.

Response: We agree, and have removed the reference to spinal reflexes.
Reviewer #2

The authors reject the possibility of performing an analysis based on last value carried forward on the basis that it will reduce power. However, this also implies that the authors are not being conservative in their analysis and therefore are significantly increasing their chances of making a type I error and therefore seeing significant results that do not really exist. The results would be much more convincing if the authors performed a conservative analysis and still observed the significant results. However, I suspect as they are unwilling to undertake such an analysis that the significant results would no longer be present. I would therefore like to see better information for the reader on the limitations of this study and the generalizability of the results.

Response: As detailed in our response to reviewers #1 and #3, we have completely reworked our statistical analysis, and take better care to identify the limitations of our data and conclusions.

The paper is still inconsistent in the way it presents P values. I strongly disagree that it is not important to quote the exact P value when a P value is not significant. A P value of 0.06 clearly gives a very different message to P=0.68. However, they would currently both be presented as P>0.05! I am also unclear how writing P=0.68 rather than P>0.05 will significantly increase the number of words in the paper. I am however, happy for P<0.001 to be used where the P values are very small.

Response: While we absolutely agree with the reviewer in principle, it is indeed more cumbersome to quote all exact p-values for all non-significant findings. If there are 10 variables, and only one is significant, this would require listing all 10 variables and corresponding p-values. Indeed, it is less cumbersome to state “X was significantly different between groups (p=.002), while all other variables were not different between groups (p>.05)” than to give the exact p-value for all non-significant values. There are some instances where this is not a problem, and in those cases we do give exact p-values.
Reviewer #3

1. On p. 10 (Statistical analysis) you should state what you mean by 'significant'. Eg. P =<0.05.

   **Response**: We agree, and now state explicitly our significance criteria (see pg 10, lines 6-12).

2. Much more seriously, you reject the opinion of both reviewers that you should hedge your statistical inferences by application of an adjustment for testing multiple hypotheses. I, in turn, reject your argument that this was only a "preliminary" or "pilot" study.

3. I have a particular concern about testing multiple hypotheses within the same study without adjustments to control the familywise Type I error-rate (see Ludbrook J, Clin Exp Pharmacol Physiol 1998; 25: 1032-1037). My favorite technique is the Ryan-Holm stepdown Bonferroni procedure.

4. Take, for instance, your Table 4. To take a very lenient view of families, there are 4 in this table: VR and TC for 'velocity peak'; VR and TC for 'velocity range'. The R-H adjustment tells me that, for 'velocity peak' only eccentric MEE for VR is 'significant', none for TC. And for 'velocity range', for VR, only 'eccentric hip MEE', 'total hip MEE', are 'significant'. And for TC, only 'eccentric hip MEE', 'total hip MEE', and 'total leg MEE'.

5. I do think that you have to do better. Try giving P (unadjusted) and P' (adjusted) values.

   **Response**: We have modified the statistical analysis as suggested (see pg 10, lines 6-12).

6. Your Figure 3. Your regression line should be based on ordinary least products regression analysis, not (as I suspect you did) ordinary least squares (see Ludbrook J. Clin Exp Pharmacol Physiol 1997; 24: 193-203).

   **Response**: We thank the reviewer for this suggestion, and had we been pursuing a predictive model relating the two measures (leg MEE and trunk velocity) we agree it would be more valid to use a Model II regression over a Model I regression. However, our intent was only to compute the correlation – the correlation is the same regardless of whether a Model I or Model II regression is used. Our intent in showing the regression plots was simply to demonstrate a positive relationship for the TC group and negative relationship for the VR group (the confidence boundaries were requested by a previous reviewer). Had we chosen not to show the plots, it is highly probable that the reviewer would not have requested we use a different regression approach. We have not shown, nor have we discussed, the regression model equation (which would be different for the two approaches), because for our purposes it is irrelevant. In summary, therefore, we have respectfully declined to use a Model II approach.