Author’s response to reviews

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

Authors:

Dr 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@nessa.edu)
Steven L Wolfe (swolf@emory.edu)

Version: 2 Date: 4 May 2004

PDF covering letter
May 4, 2004

Editor, *BMC Neuroscience*
BioMed Central Ltd,
Middlesex House,
34-42 Cleveland Street,
London W1T 4LB, UK

Re: Manuscript # 2993057262879920

Dear Colleagues:

Please accept for publication in *BMC Neurology* our Revised Research Article entitled “Tai Chi and Vestibular Rehabilitation Improve Vestibulopathic Gait via Different Neuromuscular Mechanisms: Preliminary Report”. Each author has read and concurs with the content of the final manuscript.

Please find attached our responses to the reviewers.

We look forward to your correspondence.

Sincerely yours,

Chris A. McGibbon, PhD

CAM: hpd
Responses to Reviewers

We thank the reviewers for their very thorough and thoughtful critique of our study. Please find below a detailed response to each concern raised by the reviewers.

Reviewer #1

Major Compulsory Revisions

3-24 through 4-2. It would be helpful to define VR and TC more precisely in the introduction. It is difficult to keep track of what you are talking about – does VR just consist of gaze stabilization exercise, or does it incorporate whole body movements such as TC. If it does, what is so different about VR and TC, operationally?

Response: We have added more detail in the Introduction so that the differences between VR and TC interventions for balance impairment are clearer (pg 3, lines 15-19 and pg 4, lines 5-7).

4-15 to 16. Please clarify. When you use the word “trend” do you mean it isn’t a statistically significant effect? If there is an effect, is it more prominent in persons with impairment? The term “functional limitation” is very vague.

Response: There is a statistically significant decline in ankle power, and increase in hip power, with aging in healthy elders, and even more so in elders with functional limitations (so yes, it is often more prominent). The studies cited support this statement. We have eliminated the word “trend”, however, to avoid confusion. “Functional limitation” simply denotes a stage in the disability process. We now use “mobility impairments” to be more specific (pg 5, line 1).

4-18. I find this hard to follow (non sequitur). What is the definition of a “improvement in lower extremity motor control”. I don’t see how the previous two sentences lead one to this hypothesis, other than in the most vague of all connections.

Response: We agree the statement was somewhat obtuse. We now clarify by qualifying the statement with what the aim of such an improvement would be – to increase propulsion and trunk stability (pg 5, lines 3-4).

5-9. I count 53 subjects here, but only 36 were studied. Please confine yourself to the subjects that completed your protocol.

Response: As requested, we now limit the gender, diagnosis (UVH vs BVH) and demographics to the 36 subjects whose data were analyzed (pg 5, lines 21-23 and pg 6, lines 9-13).

5-10. What were your criteria for UVH and BVH. Did all UVH patients have calorics? What were the results (range). Did all BVH patients have calorics and/or rotary testing? What did it take to call a patient BVH? It is not enough to refer to a previous paper (10).
Response: All patients, both UVH and BVH underwent caloric and rotary testing. We describe in detail the specific symptomatic and objective testing to diagnose each subject (pg 5, lines 24-26 and pg 6, lines 1-9).

5-17. Dropping out for “health related issues” is too vague. Did they drop out because they got tired of the program, because they had something serious happen (die perhaps)?

Response: Drop outs were primarily due to a subject having a new medical condition arising (unrelated to the study) that limited participation in the study, or the inability to attend treatments due to the sudden onset of illness of a subject’s family member. Some additional details have been described in the manuscript to clarify (pg 5, lines 18-20).

5-18. Please indicate that the informed consent process was reviewed by your local IRB, if true.

Response: We now explicitly state that the protocol was approved by the IRB (pg 6 lines 15-16).

6-4. and most of paragraph. This is too vague. We don’t need to know about the instructor’s experience or objects. We do need to know about what was done and for how long.

Response: The length of the both interventions was 10 weeks, and stated in the opening paragraph of the Intervention section (see pg 6, lines 21-22). We have deleted the statement regarding the instructor’s experience and have added more detail about the TC intervention, including more information on the content and duration of each of the components (pg 7, lines 10-18).

6-13. A little more detail about the warm-up exercises would be helpful. How much time was spent upright during the class?

Response: More detail regarding the content of the warm-up exercises is now provided, including the amount of time devoted to upright vs. seated exercises (pg 7, lines 12-18)

6-15 and most of paragraph. This is too vague. We don’t need to know the objectives, intentions or theoretical background of the VR program. We do need to know what exactly was done and for how long.

Response: As indicated all 10 weekly sessions were 70 minutes long, for both groups (see pg 6, lines 21-22). We have added more details to the section describing the VR intervention (pg 8, lines 2-16). The VR interventions are also described in detail in references 4, 10 and 29.

7-20. Please move this sentence earlier, with the description of subjects. A problem is that there were 8 BVH in the TC group vs 5 BVH in the VR group. It is well known that balance in UVH patients can be near normal. With BVH patients, it depends on their degree of deficit (unfortunately we do were not [sic] provided this information).
Response: This information was moved as requested (to pg 5, lines 21-23), and diagnosis information also provided (see responses above). In fact there is little relationship between patients’ degree of balance impairment and vestibular function, as detailed in reference 10.

If the BVH patients were severe, and the UVH patients were nearly normal in their balance, as most are, differences between the two groups might be reasonably attributed to the subject composition, as there are nearly twice as many BVH in the TC group.

Response: As now described in detail, to be included all patients had gait instability for which they sought treatment, and patients who "were nearly normal in their balance were excluded from the study. Moreover, there were no significant differences in proportion of BVH and UVH (or gender) in the two treatment groups (see now in Results section: pg 11, lines 5-7). However, this difference might have contributed, nonetheless, to baseline differences between groups. It should be noted that we statistically corrected for baseline differences in the treatment effect comparisons (by ANCOVA), and therefore any differences should not have greatly influenced the outcomes. We do however discuss this now as a potential limitation of the study (pg 18, line 5-9).

7-25 to 8-6. Which of these time-distance measures measure stability and why? It might help if you defined “stability”, in a mathematical way, and then said why the measured values quantify “stability”.

Response: Unfortunately, this semantic question plagues all "balance" studies, and upon reflection the authors agree the word “stability” may not have been the most appropriate descriptor. Step width and gait speed, for example, are often called measures of dynamic gait stability, since they indicate the degree to which the patient is willing to translate their center of mass over their base of support. We therefore have chosen to use “Measures of dynamic gait function”; we now leave out the “and stability” descriptor (pg 9, line 13); in addition, our project has both objective and systematic indicators of inner ear dysfunction, unlike many (especially elderly) studies of balance therapy.

8-19 paragraph. Please comment about whether the measures you have chosen to reflect “trunk stability”, in some way are related to “balance” or “overall stability”.

Response: This is a difficult question to answer, as it depends quite substantially on what one means by “balance” or “overall stability”. If one considers CG sway in standing (such as Posturography) as a measure of balance or overall stability, then it is unlikely there is close relationship between the two. This is primarily because the control strategies needed to maintain standing are probably quite different from those needed to maintain stability while ambulating (Evans & Krebs, Otolaryngol Head Neck Surg 1999;120:164-173). However, if one considers dynamic control of the whole-body CG during walking as a measure of balance or overall stability, then it is likely that trunk CG kinematics reflect those more general quantities. The rationale for using kinematic measures of trunk stability, instead of overall stability, however, was to enable us to examine the relationship between lower extremity neuromuscular function (using the mechanical energy analysis as such a measure) and the kinematics of the upper body, as a mass to be controlled apart from the legs. We now include this last sentence in the trunk stability description section (pg 10, lines 10-14).
8-26. There seem to be an immense number of paired t-tests. Was the value chosen for significance corrected using the Bonferroni or similar method.

Response: As the title of the study states, this is preliminary or pilot study of the effects of two different interventions on a relatively stable sample of patients with vestibular hypofunction. In keeping with the nature of a preliminary study, we were concerned with seeking variables that hold promise for acting as primary outcomes in a larger, full-scale, study. This intent underscored the rationale for the study, as described in the Introduction. However, we have reworked the Introduction to be more clear in this regard (pg 4, lines 11-19). As such, we have relaxed the statistical constraints one would normally apply to a full-scale intervention study, and do not apply a Bonferroni correction to the paired-samples t-tests. We discuss the implications of this in the limitations paragraph of the Discussion (pg 17, lines 16-20).

Minor Essential Revisions

3-8. The last sentence needs to be reworded because of the unclear referent – gaze stability vs both gaze stability and balance retraining.

Response: VR uses both gaze and balance retraining exercises to help the patient improve the patient’s gait function. We believe the confused wording is remedied by deleting the word “stability” after gaze (pg 3, line 8).

3-14. Surely there are more references than [15] whose main distinguishing feature seems to be that it is written by one of the authors.

3-15. Surely there are more references than [17] whose main distinguishing feature seems to be that it is written by one of the authors.

Response: Additional references have been added (pg 3, lines 15 and 20).

3-23. Do the locomotor exercises intended to heighten somatosensory input actually do this? How do exercises increase sensory input?

Response: We now provide more detail on how sensory inputs are systematically manipulated in the interventions (pg 8, lines 2-16).

4-6. Is should be was.
4-12. is should be was.
Please use past-tense when you discuss things taking place in the past (ie. 6-25)

Response: Above tense issues have been fixed (as well as others, throughout the manuscript).

9-11. Why are greater values for stance duration and length, “improvements”? Perhaps they were already optimal when they started and any changes were detrimental. Again, there are a lot of t-tests here, with no suggestion of a Bonferroni correction.
Response: There is solid epidemiological evidence that reduced walking speed and reduced duration of single limb support, for example, are classical characteristics of impaired gait in people with motor or balance deficiencies. It stands to reason therefore that increases observed in these measures post-intervention are signs of improved motor function. We do recognize, however, that there are limits to these assumptions, and acknowledge this in the discussion of limitations (pg 18, lines 2-3).

9-17. Again, a lot of t-tests without a correction. One wonders if any of these changes are significant if one included a Bonferonni correction. If not, they need not be described in this much detail – it would suffice to say that there were no significant changes. Wouldn’t it be good to increase concentric MEE, but bad to increase eccentric MEE? Here I am speaking from an efficiency standpoint. Wouldn’t you expect practice to make people more efficient?

Response: The reviewer poses two questions. The first, regarding multiple t-tests, we have already addressed above; we do recognize the limits of interpretation of the findings because of this, and have justified our approach based on the exploratory nature of this study. The second question is directed more at the interpretation of the meaning of changes in MEE. Whether an increase in concentric MEE is good or not depends on which joint. Our, and others, observations suggest that increased ankle MEE is good, while increased hip eccentric MEE may not be good, as the latter appears to occur as a compensatory mechanism of underlying neuromuscular deficits. Practice may indeed make people more efficient, but mechanical energy measures of the lower extremities are probably not good measures of overall efficiency of movement, except in very controlled experiments (such as cycle ergometry).

10-3 through 10-14. Again I suspect due to multiple t-tests, none of these changes are significant.

Response: Already addressed above.

This paragraph seems to be the key one differentiating VC [sic] and TC. In figure 1, the error bars are “95% confidence intervals” (of what?). SEM would make it a little easier to decode. More clarity in the writing would helpful to determine what you expected and what you found here.

Remainder of results section – same comments regarding lack of clarity in writing and neglect of Bonferonni corrections.

Response: The error bars were 95% confidence intervals on the mean. We have improved the clarity of the writing in all sections, particularly the section on correlations: we have summarized the correlations in a table instead of inserting the information as text. We believe this presents a clearer picture of the important difference between the two groups – that correlations for the VR group are all negative, versus all positive for the TC group (pg 13, lines 2-10).

12 6-8. Please explain in more detail how your first hypothesis comparing the two groups was “indirectly supported” by data which showed no between-groups differences. As previously mentioned, a rationale for this hypothesis is also missing.

Response: The first hypothesis was supported only indirectly by examining the within-groups changes of the two groups. We now clarify this (pg 13, line 26 to pg 14, line 1). The rationale for
this hypothesis was described in detail in the Introduction, and has now been reworded to be more clear (pg 4, lines 20-26 and pg 5, lines 1-6).

12 10-11. Again, a rationale for this hypothesis is missing. When using multiple t-tests, without the Bonferonni correction, it is possible for seemingly significant associations to appear purely by chance.

Response: As above, the rationale for the hypotheses has been reworded to be more clear (pg 4, lines 20-26 and pg 5, lines 1-6).

13-19. What exactly were the warm up exercises and how long did subjects spend on them?

Response: See response above (to 6-13).

13-25. How does increase ankle ROM improve ankle MEE? Wouldn’t ankle strength be the more direct variable?

Response: There is no direct mathematical calculation to link ankle strength with ankle power during a specific activity (such as terminal stance phase of gait). Thus, ankle strength is a more indirect influence. The more direct influences are ankle rotation speed and ankle torque – these two quantities are multiplied to arrive at power. Increases in either variable can increase the resulting power (assuming the other does not decrease proportionally). An increased ROM means that the ankle is rotating through a greater range, and therefore probably faster. We now include both (ROM and torque) as potential explanations (pg 15, line 18).

14 11-18. This all seems very tenuous and speculative.

Response: We have bolstered our argument by adding more detail and citations (pg 16, lines 3-13).

As a general comment, the discussion is too long and too speculative.

Response: We have tightened up the discussion.

Figure 1. Please clarify what these bars are. What is the 95% confidence interval for (?) Data vs mean ?) Figure 2. Same comment.

Response: We have clarified this in the Figure legends.

Tables:
Table 1 – could be condensed and inserted as text.
Table 2 – numerical format is not very instructive. Might eliminate entirely in favor of figures.
Table 3 – numerical values are again not very instructive. Might eliminate entirely in favor of figures.
**Response:** We feel that Table 1 is needed, as including the values in the text would only serve to make the Results section more difficult to read. For Tables 2 and 3, figures (such as bar charts) do not work well because of the difference in magnitude of the MEE for different modes and joints. This would require a separate bar chart for each joint, at minimum. Our experience leads us to the conclusion that a table is the best way to present this type of data.

Discretionary Revisions

8-7 to 8-18. Would it be correct to say that improved motor control, or more precisely, more efficient movement, would be reflected by a higher ratio of concentric MEE to total MEE?

**Response:** It stands to reason that this is probably true. However, as discussed above, we have purposefully stayed away from the issue of “efficiency”, as our measures of mechanical energy expenditures at joints during gait are only a component of the energy expenditure of the body, and may not themselves provide meaningful estimates of metabolic efficiency.

9-8. One wonders if the two groups were different in vestibular function.

**Response:** There were no baseline differences in VOR or gait function

14-1,2 This seems very tenuous, especially given the numerous t-tests. How do we know that the change in MEE are optimizing?

**Response:** Based several prior studies in our lab on mechanical energy analysis of patients with motor deficits, we have identified what we believe is a pattern of maladaptive gait. This provides a relatively solid rationale for what to expect when motor function improves, such as with therapy.

Reviewer # 2

Major Compulsory Revisions

I am concerned about the sample size of 53 people. There is no justification for the sample size. I suspect that 53 people is really too small to provide any meaningful results.

**Response:** The overall study was funded by NIH/NCAM as a high risk-high impact R21 exploratory grant. The sample size was determined as the minimum number of subjects needed to show a treatment effect in various ADL outcomes (such as gait speed, etc.). We have identified the sample size as being a significant limitation to the generalizability of the study (pg 17, lines 16-20), but disagree in general that small samples cannot provide meaningful results. In fact, this is the first report of which we are aware showing that a VR/TC comparison does generate meaningful results.
From reading the paper my understanding is that the analysis of the underlying mechanisms is really a post-hoc analysis. I am concerned that undertaking a post-hoc analysis on such a small number of patients where the analysis is not prespecified is highly likely to lead to false positive results.

**Response:** The reviewer is correct that in some ways our study was a post-hoc (or perhaps more accurately, a retrospective) analysis of the mechanisms underlying changes in motor function following balance rehab. Our intention, in keeping with the expressed purpose of the R21 funding mechanism, is to develop theories and future directions for the project on a larger scale. Our current study represents such an attempt, and we feel that our analysis is based on a solid rationale and our findings meaningful enough to warrant a larger-scale study to analyze these mechanisms in more detail. Indeed, our title identifies the work as a “Preliminary Study”.

For a small study there are a large number of dropouts. These dropouts may have a significant impact on the analysis. The influence of the dropouts on the analysis needs to be considered. The usual method would be to undertake an intention to treat analysis (with possibly a last value carried forward approach). However, I accept that this may not be possible under these circumstances. An approach considering differences in baseline factors between the completers and non-completers could be considered.

**Response:** We implemented the standard approach of controlling for baseline by using analysis of covariance (ANCOVA) for all group comparisons (pg 10, lines 17-18). Regarding dropouts, the "last value forward" method brings with it a further reduction in statistical power, imputing as it does "no change" from baseline among the missing data points.

The authors present a large number of correlations in the results section and a couple of these are further explored using scatter charts. The first scatter chart shows a significant correlation of -.536. However, visual inspection of the chart shows that this level of correlation is being strongly influenced by one point and that this point is strikingly different from all the others. If this point was removed then the correlation would be very different and I suspect non-significant. I am therefore rather dubious of the worth of the correlations in the paper. Please can the authors provide some indications of the robustness of these calculations. Ie Calculating confidence intervals for the correlations possible using Jack-knifing.

**Response:** We now include 95% confidence intervals on the scatter plot regression lines (see Figure 3). Regarding multiple correlations, pilot projects almost always risk insufficient statistical power. Even with 29 subjects, however, a correlation of r=.5, with p=.05 generates power of .81, which is clearly sufficient for the present data. Again, we refer to the stated objectives of our paper as an exploratory study of the mechanisms of neuromuscular adaptation in balance rehab with two very different therapeutic approaches (pg 4, lines 11-19).

**Minor Essential Revisions**

The authors are inconsistent in the way that they use P values. On some occasions they give the exact values where on others they only specify that it is larger than a certain amount (usually 0.05). The authors should alter the paper such that the P actual values for the P values are given. This will provide the reader with information about the degree that individual tests are significant or non-significant and not rely on the .05 dichotomy.

**Response:** The only time a non-exact value is quoted (such as p>.05) is when a group of variables are being described as non-significant. Given that these variables are non-significant, it does not
seem justified to increase the amount of text in the results to give individual non-significant p-
values for variables determined to be unimportant in the analysis. When individual variables are
described as being non-significant, however, we do show the actual p-value, as this takes no
additional space. We hope this explanation will satisfy the reviewer, as we would prefer not to
make the Results section more cluttered with numbers than it already is.

Some of the information represented as means and SDs are clearly skewed. Eg. page 5 In 11: 2.94 years
(+/− 2.73 years). This information needs to be presented in a statistically valid manner.

Response: We believe the identified case above (time post-onset of vestibulopathy) is the only
case where there is indeed a skewed distribution. We have changed this to a range and mean,
instead of mean and standard deviation (pg 6, line 10).

Discretionary Revisions

None.