Author's response to reviews

Title: A Randomized Controlled Trial of the Effectiveness of Mechanical Traction for Sub-Groups of Patients with Low Back Pain: Study Methods and Rationale.

Authors:

Julie M Fritz (julie.fritz@hsc.utah.edu)
Anne Thackeray (A.Thackeray@m.cc.utah.edu)
John D Childs (john@texaspts.com)
Gerard P Brennan (Gerard.Brennan@imail.org)

Version: 2 Date: 2 April 2010

Author's response to reviews: see over
Response to Reviewer’s Report

Reviewer: Deidre Hurley

The following recommendations were made in this review. Our responses to these recommendations are also outlined:

Minor Essential Revisions

1. **Shorten the introduction and summary of previous work to identify key points and their relevance for the present trial**

   We have attempted to include topics that in our opinion are key points in establishing our approach in the present trial. We attempted to be more concise, but have not removed any topics from our introduction.

2. **State in the abstract and in methods (hypothesis) that you are testing for differences between groups.**

   The abstract states “... and will use linear mixed model analysis to compare treatment groups” The data analysis section of the methods description also indicates that comparisons are being made between treatment groups. We modified the wording of the hypotheses to clarify that comparisons will be made between the two treatment groups.

3. **In the methods clarify the proportion of subjects being stratified to each subgroup.**

   We are not certain exactly what the reviewer is requesting. Stratification describes a procedure used in the randomization process to assign subjects to treatment groups. Subjects are not stratified to a subgroup, they are stratified based on their sub-group status and then randomized to a treatment group. We added a statement that our previous study found a 50:50 split in sub-grouping status and we expect a similar distribution in this study.

Discretionary Revisions

1. **Clarify the rationale for exclusion of participants over 60 years of age.**

   This was done to exclude individuals with an increased likelihood of spinal stenosis, who may be contraindicated to receive extension exercises. We added clarification of this point in the manuscript.

2. **Provide a reference for clinical signs and nerve root compression in eligibility table.**
3. State the anticipated dropout rate at follow-up and how this is addressed in the sample size calculation.

*Please see the discussion of sample size estimates below.*

**Reviewer:** Martin Underwood

1. The origin of the statement that the clinically important difference in the OSW is not given. This clearly needs to be smaller than the clinically important change that many consider to be 10 points or a 30% improvement from baseline. My view is that quite small between group differences can be important at a population level. Thus, I would not reject a SMD of less than 0.47 as not being clinically important and would suggest that a larger study is needed. Considering that many consider Traction to be an ineffectual or possibly harmful treatment then a SMD of 0.47 seems rather optimistic.

We based our sample size justification on our preliminary work in which the effect size favoring the groups that received traction, in a more homogeneous group of subjects than has typically been used in previous studies was 0.47. This preliminary work is the basis for our optimism in the current protocol. In addition, we believe that the lessons learned from the preliminary study have helped us create a more effective traction protocol that may enhance the treatment effect in the current study. We agree that a smaller effect size may be clinically important, since the average effect size may obscure important effects examined at the individual patient level. We added a statement that our sample size of 120 subjects permits us to detect a 25% difference between groups for a dichotomized outcome measured at the individual patient level such as “success” based on dichotomizing the global rating of change.

2-3. When I calculated the sample size for a mean difference of 6 points in a measure with a SD of 14 with 80% power with significance of 5% I found that 87 participants were required in each group - 174 in total. This is nearly three times the size estimated in the paper. The effects of clustering have been recognised but I may not clear how the standard formulas for this has been applied. My experience is that ICC values for clustering are rarely >0.05 and that they may not be as important as we thought in individually randomised trials. Nevertheless, more information is needed on how various clustering factors have been considered (Site/Clinic/Therapist). The inflation factor proposed might be too generous.

We attempted to simplify and clarify the power that is achieved by our sample size of 120 subjects. Hopefully this will permit the reader to judge if this sample size will ultimately prove adequate or not. The calculation referenced by the reviewer is based on an unadjusted between-group comparison of means, which does not match the analysis that will be used in the study. Adjustment for baseline scores improves power, and we took this into account, adjusting the sample size based on formulas from Borm et al. As recommended we removed consideration of the inflation factor based on site and focused on the inflation based on the ability to detect an interaction effect, which is most relevant to the project. Hopefully this explanation will be more straightforward and permit potential readers to judge the adequacy of this sample size.
4. I am surprised that a sub-group effect that is larger than the main effect has been postulated. If the sub-group effect is greater than the main effect the consequence of this is that for the other sub-group that treatment will be detrimental. If one went for the more plausible, but still very optimistic effect size for the interaction, of 0.47 and sub-gotups were of equal size then a four-fold increase in sample size is needed. In any event for the subgroup effect of 0.71 specified the authors have estimated the increased sample size to be 1.4 but the authors have only multiplied their sample size by 1.2. Pedantically 102*1.2=122.4 not 120

Are presumptions relative to the size of the interaction effect are based on our findings from our preliminary study which did find that for subjects in the other sub-group, the use of traction resulted in a worse outcome than not using traction, as presumed by the reviewer. This is illustrated in figure 2 in the manuscript.

5. Sample size estimate is assuming no loss to follow-up. It would be wise to allow for up to 25% loss to follow-up at one year.

An advantage of the using a mixed model analysis is the imputation of missing data points that will preserve the sample size under intention-to-treat principles. We have outlined how our sample size of 120 can account for a presumed loss to follow-up of slightly more than 13%.

6. I note that the analysis for sub-group effects is the three way interaction time*subgroup*randomised group. I think that the two-way interaction of subgroup*randomised group is the more appropriate analysis. The hazard with the three-way interaction is that a false positive result might be obtained because, in most studies, change with time irrespective of treatment has the largest effect on outcome.

It seems that the risk would be a false negative if the concern is that time is the overwhelming factor determining outcomes, obscuring any effect of treatment or sub-group status. The same concern would seem to be relevant to the analysis for our primary aim that proposes to examine a group*time interaction without regard to sub-group status. Similar to the approach taken to explore the overall effects of the trial, taking time into account in analyzing the sub-group outcomes seems most appropriate to the hypotheses being examined in the study. Our hypotheses relate to the effects of treatment received and sub-grouping status on outcomes assessed over time.