Author’s response to reviews

Title: Lee Silverman Voice Treatment versus standard speech and language therapy versus control in Parkinson’s disease: a pilot randomised controlled trial (PD COMM pilot)

Authors:

Carl Clarke (carlclarke@nhs.net)
Catherine Sackley (k1453113@kcl.ac.uk)
Christina Smith (christina.smith@ucl.ac.uk)
Caroline Rick (c.e.rick@bham.ac.uk)
Marian Brady (M.Brady@gcal.ac.uk)
Natalie Ives (N.J.Ives@bham.ac.uk)
Smitaa Patel (s.a.patel@bham.ac.uk)
Rebecca Woolley (r.l.woolley.1@bham.ac.uk)
Francis Dowling (f.p.dowling@bham.ac.uk)
Ramilla Patel (em.east@parkinsons.org.uk)
Helen Roberts (H.C.Roberts@soton.ac.uk)
Susan Jowett (S.JOWETT@bham.ac.uk)
Keith Wheatley (k.wheatley@bham.ac.uk)
Debbie Kelly (debanhall1965@gmail.com)
Gina Sands (Gina.Sands@nottingham.ac.uk)

Version: 3 Date: 30 Nov 2017

Author’s response to reviews:

Thank you for the opportunity to respond to the comments made by the Reviewer regarding the above paper. We hope that the replies below and the changes made to the paper mean that the paper is now acceptable for publication.
We would like to remind the Editor that this was a paper reporting the results of a pilot trial to help design a pragmatic phase III trial of SLT in PD, which could be run and delivered within the setting of the NHS. The main PD COMM trial has been funded by the National Institute for Health Research (NIHR) HTA Programme, and has undergone peer review at various stages of the funding process. The data presented in the paper, and more, was provided to the HTA as part of the funding application.

Editor-in-Chief Comments

1. Add feasibility objectives into the abstract, and order the paper methods and results sections to address the 10 feasibility objectives given on page 4 more clearly. Objective 8 could be expanded also to include assessment of most suitable primary outcomes for main trial and obtain initial estimates for aiding sample size calculation.

Reply: In the abstract, we have added to the first sentence of the methods the main feasibility objectives of the trial, and changed the sentence on outcomes to mirror the objectives. The ten feasibility objectives on page 4 have been re-ordered to mirror the order in the results better. We have also re-ordered the data analysis section of the methods and linked the methods to the various Objectives, and similarly for the results sections of the paper, so the order of the feasibility objectives is the same throughout the paper. We have also added and renamed subheadings in the results section to match the objectives better and to aid the reader. We have made the suggested changes to Objective 8, and deleted objective 7 as this is how we assessed Objective 8.

2. Add references for all the measures listed on page 6 under outcomes and some comments/refs from own work on validity and reliability of use of each outcome on these patients.

Reply: We have added the references for all the outcomes listed on page 6. We have not undertaken any validity or reliability work on these outcomes as they are all validated and widely used tools, and this was beyond the scope of the PD COMM pilot trial. We have added a sentence to this effect to the methods.

3. Provide some extra explanation on consistency of measurement over time and sites e.g. training of assessors, arbitrarily chosen 4th reading etc. as per comments.

Reply: Regarding the arbitrarily chosen 4th reading. There was a lot of data to condense into this paper. We wished to present some data on the correlations, so we presented a cross section of
the data for illustrative purposes. The same message was observed across all the data. We have removed the footnote as we agree that this is confusing to the reader.

4. Provide some extra explanation on blinding of whom and when as per the response.

Reply: Information on who was blind is already in the methods on pages 5 and 7. Therapists and patients were not masked, however, assessors of the vocal assessments were all masked to treatment allocation. We have edited to text to make this clearer.

5. In the results data analysis section an exploratory analysis of differences between arms is not warranted/recommended except for those relating to sample size calculation effect sizes for the potential main outcomes. Therefore results in Table 3 should be presented for each group separately with CIs within group and then differences required for sample size presented in the text only. To see how the measures change over time please add extra time points to the table. A plot may be useful?

Reply: I’m sorry but we strongly disagree with parts of this comment.

We agree that any exploratory analysis of outcome data reported in the text of the paper should be restricted to outcomes in relation to the sample size calculation. We have now only presented data in the text for the VHI and PDQ-39 communication domains which were the two outcomes that were being considered for the primary outcome for the main trial, and have removed the carer outcome data. We do however agree that the results section reporting the outcome data could better reflect Objective 8, and we have changed the text to reflect this. We have taken text that was previously in the discussion to better reflect our assessment of outcome measures in the results section.

In Table 3, we have reported data on the other outcome measures collected in the pilot trial, and we think that it is important to report this data, so that the PD COMM pilot data can contribute to the evidence base and be included in any future meta-analyses in this area. As researchers we have an important role in ensuring that data is published (All Trials campaign) and available to add to the evidence base. The data for the other outcomes is included in Table 3 for this purpose, but we have made no reference to these data in the text. We feel that this is an appropriate way to ensure the data gets into the public domain.

We also feel strongly, for a number of reasons, that the data should be presented as a difference between groups, and that a within group comparison is not appropriate. First, a within group comparison is potentially misleading, as analysing change does not control for differences at baseline because of regression to the mean. In lay terms, if there are imbalances at baseline between the groups, this means some patients have more room for improvement, and if these
patients are all in the same group, this will make that group look better than it actually is. We then have to ask the question, is any change due to treatment or due to the fact they had more room to change? In a pilot study, due to small patient numbers the possibility of imbalances in the groups at baseline are higher. Secondly, the purpose of phase III trials is to assess for differences between groups, not to assess within group changes. We have undertaken an analysis similar to that which would be undertaken in the main trial as part of the framework for assessing the pilot trial, and to obtain some initial estimates to inform the sample size calculation. We refer the Reviewer and Editor to a paper by the eminent statistician Professor Doug Altman, which discusses the problem of analysing change data, and that analysis of covariance methods should be used; this is the analysis planned in the main trial (Vickers, Altman. Analysing controlled trials with baseline and follow-up measurements. BMJ, 2001; 323:1123-4). This pilot trial was not powered to detect differences between the groups and no hypothesis testing was planned, so we have only presented mean differences and 95% confidence intervals without statistical modelling. Thirdly, the analysis that we have presented is as per the Statistical Analysis Plan (SAP) for the trial, and it would be bad practice to analyse and present the data differently to that which we had specified up front in the SAP. I’m sure you are aware of Dr Ben Goldacre’s COMPARE initiative which is assessing reported trials and checking that they have been reported as per outlined in the protocol. We therefore do not think it is appropriate to present within group comparisons. However, to address this comment, we have expanded Table 3 to present the means at baseline and three months for each group, so readers can estimate the mean change within group for themselves should they wish to.

As suggested, we have added two plots showing the data for the VHI and PDQ-39 communication domains over the course of the trial.

6. Is there any useful information in per protocol results to consider as this pilot is looking at feasibility?

Reply: We are not sure we understand the reason or rationale for undertaking a per protocol analysis in a pilot trial and how this relates to an assessment of feasibility. We have reported data on compliance for the two SLT groups, which was one of the objectives of the trial. A per protocol analysis would be appropriate as a sensitivity analysis in a main trial where we are formally assessing differences between groups. Apologies, but we are not sure of the purpose of a per protocol analysis in the setting of a pilot, where we are not comparing the treatment groups.

7. Consider giving correlations in Table 2 for each time point to see how they change over time. Are no variables skewed as Pearson correlation used throughout?
Reply: We have added the correlations at baseline, so the correlations for the baseline and 3 month (primary outcome assessment point) data are available to the reader. They show the same take home message. No variables were skewed and Pearson’s correlation was appropriate to investigating a linear relationship.

8. Please include some limitations in the discussion to address more of the reviewers concerns - for example, regarding using the VH1 (VH1-10?) measure as the potential primary outcome for the main trial; differing lengths of treatment periods; confounding. Also some discussion on RCTs investigating LSVT LOUD as per reviewers/response comments and long list of publications. On page 10 please comment on the 73% in the SLT arm completing treatment.

Reply: It is difficult to know how to address this comment, as a lot of the limitations expressed by the reviewer are personal comments and the opinion of the reviewer. They also seem to be limitations they think will be relevant in relation to the full trial, which is currently underway, so we don’t yet know if these limitations hold. If we show a benefit for LSVT and SLT on the VHI, then will the criticism of this as our choice of primary outcome remain? We feel the discussion of limitations in this paper should relate to the pilot trial that we are reporting, and not in reference the full-scale trial which we have not completed. We have however added to the discussion regarding the 73% in the SLT arm completing treatment by three months, and noted that it will be important to monitor this within the main trial.

Reviewer #1:

Thank you for your additional comments and continued discussion. From the author's comments, and from looking at the manuscript, it appears that the extensive recommendations for revision provided in the previous two reviews were not incorporated, and that the authors do not agree that it is necessary to add additional information to the discussion to address the topics/concerns raised. Therefore, unfortunately the decision to reject the article for publication still remains.

As the authors have stated, it appears that a circular discussion has been ongoing, thus, the comments here will be provided as an overall summary, rather than a point by point discussion.

My comments in the previous two cycles of reviews, citing the importance of including additional information in the paper, have been based on over 17 years of vast experience with patients with Parkinson disease (PD) as well as the literature on PD, and my experience as a part of a research team focused on voice and speech research with people with PD. This team has been funded by the National Institutes of Health for 20 years. This team has included career statisticians as well as internationally recognized researchers. The recommendations provided
have been in regards to the considerations learned from this team when conducting voice/speech research with this PD population.

Based on the experience our team has had with multiple publications, when writing an article it is the responsibility of authors and reviewers to describe any possible confounds to a study, and clearly state all methods in order to ensure that the reader is provided with the proper tools and information in which to accurately ascertain and interpret the data.

The key elements raised previously are listed below:

1. The issue of decreased sensory awareness and the impact this can have on the self-perception of voice and speech for people with PD has been extensively described in the literature (Sapir et al., 2011; Ramig et al., 2011; Ho et al., 2000; Kwan and Whitehill, 2011; Mollaei et al., 2013; Arnold et al., 2013; Kompoliti, 2000; Sapir 2014; Liu 2012; Houde, et al., 2004; Cucci et al., 2010) and in my previous review comments.

This is a fundamental problem in PD and has been shown to limit an individual's ability to self-rate their speech/voice. It is common for people with PD to deny they have a problem with speech/voice, when objective data and professional assessment demonstrate they do have a problem with speech/voice.

In light of this, why then would it be concluded that a self-rating scale (especially without reference or discussion of these issues in the paper) be chosen as the single primary outcome variable for a larger study?

As stated by Coster (2013): "The best design and most rigorously executed procedures cannot make up for a poorly chosen measure. Important knowledge about the impact of the intervention may be lost because the selected measure was unable to capture it or, even worse, distorted the true results".

Information regarding the confounds of using a measure of self-perception as a primary outcome variable should have been included in the discussion.

Furthermore, a correction to the author's comments, the Jacobson article does not report reliability or validity data on PD specifically. Thus, while the Jacobson article has published reliability data, this is not specifically with people with PD. Therefore, and especially in light of the issues with self-perception in this population (which makes application of perceptual measures different than with other populations), reliability results from the Jacobson study cannot be generalized to the PD population.

A single primary outcome variable with no validity or reliability data on the population of study places the outcomes of this research in question.
The current study would have been an opportunity to evaluate intra-subject reliability of VHI in PD, but this was not done.

The VHI has indeed been included in research publications with people with PD led by the Ramig team. When it has been a part of the Ramig led studies, the VHI was not used as a primary outcome measure, research results also included quantifiable acoustic measures, the VHI was given two times at the time points of measurement, and finally, information regarding the issues of a sensory mismatch in PD and the potential influence of these issues of self-perception on interpretation of the results was provided. All of which are points that were raised previously by this reviewer, and should have been provided by the authors, as information important for interpretation.

Please be clear, that the recommendations made by this reviewer are not points raised because of post treatment results that might go one way or another, these same cautions would be raised if the VHI was being used only as a descriptive baseline measure.

Reply: We are sorry that the reviewer does not think that we have addressed all the points made previously. We appreciate than previous trials in SLT have used therapist-rated outcomes as the primary outcome. These trials showing possible benefit for SLT in people with PD have provided the important spring board for the PD COMM trial. However, these outcomes are not feasible for a large scale clinical trial recruiting nearly 550 people with PD; the trial would have become prohibitively expensive. The main PD COMM trial was a commissioned trial by the National Institute for Health Research (NIHR) HTA programme (the main funder of clinical research in the UK), with the brief stating the following with regard to outcome measures:

Primary outcomes: Speech intelligibility and effectiveness of communication.

Secondary outcomes: Vocal loudness; patient satisfaction; adherence and compliance; quality of life; cost-effectiveness; adverse effects; effect on carers.

The NIHR are very clear that they want trials with patient-centric outcome measures, and promote the use of patient-reported outcomes. They want the patient to feel the benefit and have improved quality of life. This provides part of the rationale for choosing the VHI as the primary outcome. As although we have seen in previous trials that SLT results in improvements in therapist-rated outcomes, how and does this translate back to how the patient feels?

Within the pilot, we undertook a variety of work to determine an appropriate outcome measure for a large-scale phase III trial. This included assessing correlations between therapist-rated and patient-reported outcomes, which we found to not be well correlated, and also talking to people with PD as to what they thought and felt was important to them. The output from this was that
the VHI and the PDQ-39 were considered appropriate outcome measures by both us the researchers and our patient group. The PDQ-39 is a well-validated and widely used tool in trials in PD, and the VHI has been used in trials by Lorig et al. We need large-scale evidence to assess whether SLT provides benefit to people with PD, and to do this we need to undertake large scale pragmatic trials with patient-centric outcome measures that can easily assessed and completed.

2. To make the reader aware of the impact that including data from a subject who didn't actually receive a particular treatment can have on the results. The reader who may not be familiar with ITT needs to be provided with an explanation that while this statistical measure was applied, X number of people didn't actually receive LSVT LOUD because they didn't complete the full 16 sessions. Readers may not be aware that LSVT LOUD treatment is not just considered LSVT LOUD because of the exercises that are conducted, but it is a combination of the exercises that are completed with the intensity of a specific protocol (4 times a week for 4 weeks, one hour sessions). If this is not discussed, it would be similar to using ITT to report on the results of X number of people who were randomized to a swallow exercise group, but then not providing an explanation that X individuals who were included in the final results never actually received the specified swallow treatments. So, the point again is that an explanation needed to be provided to elucidate the reader and help them understand the results, regardless of if the results supported or didn't support the hypothesis.

Reply: Whilst we agree with this point regarding the use of ITT analyses and people’s understanding of this analysis method, this was a pilot trial. We have not performed any hypothesis testing, and the analysis is purely a descriptive analysis to inform the sample size calculation for the main trial. A per protocol analysis is part of the sensitivity analyses included in the Statistical Analysis Plan for the main PD COMM trial currently underway. We have added an explanation of the ITT analysis to the methods.

3. Finally, as stated previously, reporting on post baseline correlations for the entire sample of individuals with PD is not valid because it is confounded by treatment group effects. For post measures, when stratifying by group, correlations between outcome measures for an entire sample are meaningless.

It would only be correct to do these entire sample correlations for baseline assessments, which would not yet be confounded by treatment group. Thus, these results are misleading and should not be included as currently reported. In order to accurately represent the results, the correlations should have been reported by group.
Thus, while the research question is an important one, for all of these reasons stated above, unfortunately, in its current form, conclusions cannot be drawn and the manuscript should not be published.

Reply: To address this, in Table 2, we have added the correlations for the baseline data alongside the data at 3 months. The data are very similar, so we feel our conclusions still stand, and we hope this satisfies your comment above.

We are sorry that you feel so strongly about this manuscript not being published. We have re-framed the paper, and we hope that the paper now addresses the majority of your comments. We would like to remind you that this was a pilot study to inform the design of a full trial, the PD COMM trial, which has been funded by the NIHR HTA after extensive peer review, and which is currently open to recruitment.