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: 1 Date: 20 May 2017

Author’s response to reviews:
Thank you for the opportunity to review the article, "Lee Silverman Voice Treatment versus standard speech and language therapy versus control in Parkinson's disease: a pilot randomised controlled trial (PD COMM pilot)".

This is an important topic for discussion, as in order to provide and receive the best rehabilitative care, it is essential for therapists and patients to have information that can guide the decisions they make regarding selection of therapeutic methods that have been proven to be effective via research.

This paper is intended to be a pilot report of the feasibility of implementation of a larger study, and based on information collected in this pilot study, to determine which types of outcome measures to include in a larger study. In order to determine the feasibility for implementation of a larger study, and to determine which outcome measures should be used in a future study, it is essential that the pilot study be conducted in a way that controls for possible confounds and implements methods that will result in reliable and valid data.

While this study has demonstrated the feasibility of this type of a project from a recruitment standpoint, the information currently provided regarding implementation, and the methods for assessment and data collection, are not reported in a way to determine the validity and reliability of the data.

Thank you for your comments but we disagree with this statement, the study was multi centre and the data completeness was excellent, it was independently analysed and the researchers were blind to allocation. It has to be emphasised that this is not an efficacy study but a feasibility study.

Additionally, the current study has note-worthy potential for confounds. It is not possible to determine from the information presented whether or not the items that will be described as confounds in this review, and the methods that are lacking were actually applied, but just not reported, or if they were not included in the original study design. Thus, if it is a case of the former (information is available but wasn't reported), then it would be important that significant and substantial revisions are made to the paper to include this information. If the comments cannot be addressed and information cannot be added to affirm the fidelity of treatment and validity/reliability of information provided, then these would be significant design flaws, the data would not be able to be interpreted as reliable or valid, and it would not be recommended that the study be published.
I will start this review with a summary of the critical design issues that would make one question the data as valid and reliable, and then provide a more detailed review of specific line items in a sequential order following this:

* While a variety of dependent variables were assessed, the focus in the discussion section regarding future studies was on the VHI. This discussion did not mention or consider the confounds of this type of perceptual scale when considering the sensory awareness issues that are present in individuals with PD, and the impact these issues can have on perceptual ratings of voice. Also, conclusions cannot be drawn from the reported correlations that the VHI would be the "best" measure to use, without information regarding reliability, the time point for the correlation comparison, and separation of groups. See item 30 for more details regarding this issue.

The groups were independently randomised with concealed allocation. Variations in sensory awareness issues would be managed by randomisation, there was no systematic bias in stage of disease or other variables. However, to repeat, the study was not designed to look at efficacy.

* It is mentioned that there was a blinded assessor in addition to therapist reported outcomes, more information is required to clarify who collected the data, methods for collection, and how bias was avoided in the collection of this data. See item 4 for more details regarding this issue.

The therapist are unblinded to treatment allocation in order to provide treatment and so unblinded when filling out the treatment logs as they would need to know which logs to fill out. The listeners were blinded to treatment allocation when listening to the AIDS data. The therapist will have been blinded when doing the AIDS, cookie theft picture, rainbow passage and vocal loudness assessments.

* Reliability of the outcome measures is not reported, if reliability data was not collected there is no measurement stability and the data cannot be considered valid. See item 18 for more details regarding this issue.

We did not further assess reliability for the outcome measures as part of the feasibility study. We referenced 2 publications.


* Information is lacking regarding whether or not the clinicians who delivered LSVT LOUD were certified in LSVT LOUD; and if they were LSVT LOUD certified, what type of training was incorporated to ensure that all therapists were implementing LSVT LOUD treatment according to protocol. See item 12 for more details regarding this issue.

Yes, clinicians delivering LSVT were all LSVT certified with appropriate top up training and they delivered it as clinical therapists within the NHS health system. As this study was comparing treatments as delivered within the NHS no additional training was given to therapists in either arm.

* Conclusions cannot be drawn regarding the correlations of patient reported outcomes as compared to the therapists measures, as all of the subjects were grouped together, and it is not stated whether these correlations were made from 3 month, 6 month or 12 month data. See items 23 and 30 for more details regarding this issue.

All correlations were done at 3 months to compare patients reported outcomes to therapist reported outcomes. This was to determine which outcomes would be taken forward to the larger trial. However, we are unclear of this point, as correlation is a way of ascertaining whether or not there is a linear relationship between two variables. So grouping the subjects would not affect the correlations. Variables at a time point would only be plotted against other variables at the same time point.

* It is not possible to draw conclusions regarding the "best" outcomes measures to include in future trials without also analyzing the data that were collected at 6 and 12 months as well. See items 32 and 34 for more details regarding this issue.

In the paper only report outcomes at 3 months, as this was a trial of feasibility to ensure we could collect this data at 3 months, the proposed primary end point.

* It appears that data was included in the LSVT LOUD arm for individuals who did not complete the full protocol and in the no treatment arm for individuals who may have started to receive treatment before the end of the 12 month trial, this would invalidate data for both groups. See items 19, 21, 22 and 31 for more details regarding this issue.
An intention to treat analysis was used. The primary end point was 3 months, but any drop outs or cross overs would be included. However, this was a feasibility trial with clear end points.

Please see below for specific comments regarding critical issues organized in sequential order:

1. Abstract (page 2, line 3) and Background:

The comment is made that there is little evidence for LSVT.

Over the past 20 years, three RCTs (funded by the National Institutes of Health in the US) have been conducted on the efficacy of LSVT LOUD, resulting in multiple publications. The two initial RCTs in PD compared LSVT LOUD to an alternative respiratory treatment and to untreated control groups (PD and Healthy Control). Outcomes of the trials demonstrated that LSVT LOUD produced significant, immediate and long-term improvements, and the magnitude of these changes surpassed those in the control treatment group. Additionally, changes accompanying LSVT LOUD significantly exceeded those observed in the speech of untreated individuals with PD over time (Ramig et al., 1995; Ramig et al., 1996 and Ramig, et al., 2001a, 2001b.). For both of these RCTs, improvements from LSVT LOUD were maintained for the duration of the follow-up (6 to 24 months post-treatment), with within treatment effect sizes ranging from .85-2.93. (Ramig et al, 1995, 1996, 2001a, b; Sapir et al., 2011).

This was our conclusion following our Cochrane review where only one single centre trial ‘Ramig 2001’ was controlled with a no intervention group, it had 29 participants, however the trial had a reasonably high risk of bias due to issues of selection, randomisation and allocation concealment.

Preliminary data from the third RCT are also demonstrating treatment effects that are in line with these previous RCTs.

We would be very grateful for the details of this publication.

Finally, the evidence behind LSVT LOUD has been documented by additional research groups, separate from the work done by Ramig et al. (e.g. Theodoros et al., 2016; Sale et al., 2015; Sauvageau et al., 2015; Whitehall et al., 2011; Constantinescu et al., 2011; Wenke et al., 2010).

2. Page 2 line 4: Now that there is a physical therapy arm of LSVT® as well, LSVT® is referred to as LSVT LOUD®.

Thank you- corrected
3. Page 3, line 14: LSVT LOUD is being implemented into the clinical world globally (over 25,000 clinicians have been trained in 73 countries). It is rare that a research protocol gets implemented into the clinical world. LSVT Global is a socially responsible commercial enterprise committed to providing clinicians globally with access to LSVT LOUD training. The goal is to maintain treatment fidelity of LSVT LOUD so the same treatment delivered in the research studies is available to patients and clinicians globally. LSVT LOUD is an evidence based treatment protocol, based on years of research, it is not a commercial programme.

Noted- amended

4. Page 4 Line 16: This says that vocal assessments were performed by blinded assessors, but later it says speech measures were based on therapist reported outcomes. This is confusing. Please define who collected the voice and speech information at baseline, 3, 6 and 12 months. If this person was a therapist, how was bias avoided? Also, if the person collecting the data was a therapist it should be noted that even if a therapist doesn't provide verbal cues, their presence serves as a cue to the treatment techniques that were incorporated in therapy. Thus, data collected by a treating therapist would be confounded. Were any cues regarding loudness, articulation, breathing, etc. given by the person collecting the data?

Thank you we will clarify. AIDS words and sentences, vocal loudness, rainbow passage and cookie theft were conducted by blinded therapists. It is possible that therapists encouraged participants during the performance of the vocal tests which may have enhanced the results, however the correlations between vocal loudness (arbitrary 4th test) and patient reported measures of intelligibility, communication and disease specific quality of life were very low. The clinical relevance of such bias would be questionable. Analysis of recordings was completed by blinded assessors at a separate location.

5. Page 4 line 22-23: This statement is not entirely correct as vocal strain on its own is not a contra-indication for LSVT LOUD. In some cases, vocal strain is a result of poor compensatory mechanisms to produce a louder voice, and these poor compensations can be overcome by retraining proper respiratory laryngeal coordination and proper voice production techniques via LSVT LOUD (Countryman et al., 1997; Smith et al., 1995). LSVT LOUD can also be utilized with people who have had previous laryngeal surgeries. In order to create a homogeneous group for purposes of the study, it is valid to exclude people based on the criteria of previous laryngeal surgeries, however the exclusion would not be because they are not "appropriate" but rather to eliminate possible variables that might influence outcomes.
The exclusion criteria included laryngeal pathology and surgery were used following the published trial, we accept that there may be individual variation however as methods of differentiating participants accurately are not published we followed the earlier guidance.

6. Page 5 line 5: What was included in the vocal assessments and who conducted these? Was this also done by the blinded assessor or by the treating therapist? Were any cues given for the voice and speech tasks (e.g. cues to speak or produce voice/speech in a particular way?) This is important to know to be clear that there was no cueing influences or bias on the assessment.

Blinded assessors conducted the assessment and the analysis was independent.

7. Page 5 line 20: LSVT LOUD Treatment sessions should be 60 minutes in length, not 50-60 minutes.

This was an evaluation of treatment as delivered in the NHS, there is no implementation data to demonstrate this is any different from the implementation of LSVT Loud® in other countries.

8. Page 5, Line 23 - 24: This is not accurate. The pitch exercises are done by starting at a modal pitch and going up, and starting at a modal pitch and going down. They are not low to high and high to low glides.

Corrected

9. Page 5 line 25: This is not entirely accurate. The purpose of these exercises is to provide a foundation of vocal effort and loudness for translation into functional speech.

Corrected

10. Page 6, line 2: The description of LSVT LOUD treatment is not complete. The treatment is not just focused on motor practice, but also focuses on retraining the sensory system to enable individuals to learn to internally produce and feel comfortable with the effort and loudness required to achieve a normal loudness voice. This is an essential component of the treatment. It would be important to confirm that this aspect was included in LSVT LOUD treatment delivery.
Noted we have expanded the description.

11. Page 6 line 11-12: If data was collected out to 12 months, why were individuals in the no treatment group allowed to receive treatment at 6 months. This would invalidate the 12 month assessments for this no treatment group.

Again this was a feasibility trial not an efficacy trial. This approach was specified in the protocol as our patient advisors wanted the participants to have the option of treatment after 6 months. We have changed this to 12 months in the main trial.

12. Page 6 line 13: What type of training was provided to the clinicians in both the SLT and LSVT LOUD groups?

Trial related training of research processes i.e. how to collect and record data for the study. Amended in text.

Was LSVT LOUD administered by SLTs who have participated in the LSVT LOUD Training course, passed the certification exam and hold current certification to deliver LSVT LOUD?

YES- amended.

This is essential to ensure fidelity of treatment delivery, and this information should be included in the article. If the clinicians were not certified to delivery LSVT LOUD, this is a significant confound for the study, as the treatment would not have been LSVT LOUD and no conclusions could be drawn from the results.

All the therapists were certified with appropriate refresher course, we have added to the text.

How was it ensured that LSVT LOUD was delivered according to protocol?

This was not assessed beyond the number and duration of sessions. As discussed above this is a study of how these therapies are delivered within the health care system. There is no routine checking of therapist performance following LSVT Loud training routinely within the NHS and we followed that protocol. Is there in other countries?

What type of information was included on the intervention record forms for both the LSVT LOUD and SLT groups?
Primarily dose, length of session, number of sessions per week, number of weeks, home practice reports.

This is important information to establish the fidelity of treatment delivery for both groups.

Additionally, for a larger study, it will be important to collect information from the therapists in the traditional SLT group that will allow analysis of which interventions and which frequency of delivery used in the group were most effective.

Yes, we are collecting this but expect wide variation as the provision of therapy is regionally variable and is unavailable in some.

13. Page 6 line 23: I would question the use of the PDQ -39 total score for this study, as so few of the questions are directly related to communication.

The PDQ 39 was used as a disease specific quality of life measure, it is routinely used in studies of PD interventions. It has a communication domain which we report.

14. Page 6 line 25: More information needs to be included regarding how the measures collected by the therapists were obtained.

Please see response to ‘4’ above.

The therapist giving the intervention filled in the intervention log at each visit

15. Page 7, line 1: Were any cues given during collection of loudness data? How was loudness data collected? This is important to know to be clear that there was no cueing influence or bias.

The blinded therapists administered the tests which were independently analysed, but it is possible the therapists encouraged participants during testing however, as this is not an efficacy study we are unclear how this would bias results.

16. Page 7 line 1: It says data were collected before, at 3, 6 and 12 months. It appears that part of this data was collected by the treating therapist. How was bias avoided for the data collected by the treating therapist? A treating therapist is an automatic cue for the therapeutic strategy being assessed. What methods were utilized to ensure that the treating therapist did not provide cues during the data collection?
All QOL pen and paper measures that were completed by participant or carers was done in clinic before randomisation with a nurse or clinician. Then the 3, 6 and 12 months were done by post.

The intervention logs were completed by the treating therapist at each visit.

17. Page 7 line 6-7: Why was vocal strain/abuse identified as a possible risk? If behavioral/voice treatments (including LSVT LOUD) are delivered correctly, they are designed to decrease vocal strain/abuse, not increase it.

Yes, the therapists identified vocal strain as a risk if therapy was not delivered correctly therefore we included as an adverse event. None were reported.

18. Page 7 line 10: What type of reliability measures were in place to ascertain reliability of the data collected at each time point? Due to the variability in performance that occurs as a part of PD, it is important when collecting objective and perceptual data that more than one measurement is collected at each time point to establish reliability of results.

We were testing between group data from randomised groups, we demonstrated balance between groups and we have no reason to expect systematic bias, however the sample size is small and that is why we are conducting a Phase iii trial. We used various questionnaires to collect data at baseline, 3, 6 and 12 months which are all validated tools. These include VHI, PDQ-39, V-RQoL, Living with Dysarthria, EQ-5D and carer QOL.

Test, retest reliability needs to be included for dependent variables, in order to ensure that the dependent variables are stable, otherwise the validity of the data cannot be established.

We do not routinely conduct further validation of measures.

What type of measures were in place to combat fatigue (which could result in decreased accuracy) from the individuals filling out so many forms. Were the forms given all at one time, were instructions given for taking breaks between filling out forms?

The burden of form filling was assessed by the patient support group who considered it acceptable. The completion rates were assessed to ascertain acceptability. Again, randomisation would account for variations in fatigue levels, however form completion rates and return rates
were excellent which would imply that there was no burden on form filling. Response rate was >90% across all time points.

19. Page 7 line 11: It is noted that because this is a feasibility study it was not appropriate to make definitive comparisons and that the outcomes were described descriptively. If this was the reason that more stringent statistical analyses were not reported, then it would be important to just focus this report on feasibility of recruitment and participation and not include outcome results. If outcome results are to be considered, then factors such as data collection by non-treating therapists, collecting and reporting repeat measures at each time point, training of therapists, fidelity of treatment delivery, inclusion only of those individuals who fully completed protocols, and strong statistical methods and analyses appropriate to multiple treatment groups and RCT need to be incorporated. Otherwise, validity of the data as currently reported cannot be established and conclusions regarding feasibility and which outcome measures might be most appropriate for future trials cannot be made.

Form completion rates and return rates were very good which would imply that there was no burden on form filling. Response rate was >90% across all time points. In addition, the purpose of collecting these outcomes was so we could determine our primary outcome and end-point in the main trial and use data from the pilot trial to perform a sample size calculation.

(As first author and Chief Investigator I would like to clarify that the statisticians working within Birmingham Clinical Trials Unit are appropriately qualified, career statisticians).

For further information on feasibility trial methods please see;


20. Page 8 line 8: How did the trial ensure that if an individual was randomized to the LSVT LOUD arm, that there would be a LSVT LOUD certified clinician available to provide the treatment. Was this a part of the decision making of centres for inclusion? This would be an important point for consideration of feasibility of implementation.

Yes it was. Recruitment can be paused of no therapist available.
21. Page 8 line 22: A median of 16 sessions. All individuals in the LSVT LOUD arm should have had 16 sessions, or they did not receive LSVT LOUD. Data for individuals should not be included in a LSVT LOUD arm if they have not had 16 sessions. Including individuals in the LSVT LOUD arm who did not complete 16 sessions would be similar to including individuals in this arm who received a completely different treatment. As the treatment is not LSVT LOUD unless all 16 sessions are completed. It is not clear if some data from the 7 subjects described in the following paragraph were included in the LSVT LOUD group data or not. If this data was included the results are not valid.

LSVT treatment was intended to be 16 sessions, however as delivered in a health care setting not all sessions are delivered, for example this can be because of illness of therapist, person with PD, carer or person transporting them or simply the participant did not want to continue in the trial. As discussed above, the study is a feasibility study assessing therapy as delivered in the NHS and because an intention to treat analysis was used these people are included.

22. Page 8 line 26: It appears from the Figure that forms were still collected later in the study from the 4 individuals that only did 1-3 sessions of LSVT LOUD. If this is the fact, that is not valid data to include, as these 4 did not complete the intervention. Any assessments they filled out would not be an accurate reflection of their opinion of the treatment, as they did not actually receive the full treatment.

Again, this is a true reflection of how care is delivered within a health care system and we were assessing feasibility for a phase III study. Outcome measures were utilised not the participants’ opinion of the treatment. We were not assessing efficacy but we needed to look at all patients data as this most accurately inform the full trial which is why we used intention to treat analysis (ITT). Data was analysed on an ITT basis. Even if participants did not comply with intervention they still returned data so it was analysed.

If this data was included due to an intention to treat analysis the following cons to intention to treat should be considered. Intention to treat (ITT) does not allow for an accurate assessment of treatment efficacy, and when comparing two different treatment groups, it may falsely cause results from the two groups to appear similar. Thus, for a more accurate depiction of results, ITT alone should not be utilized in this type of a feasibility study.

For an accurate assessment of groups, an as treated or per protocol analysis could be considered for use instead.

We were not assessing efficacy with this study. It is a feasibility for a phase III study of clinical and cost effectiveness.
23. Page 9, line 7 and Table 2: Is this data from 3 month, 6 month or 12 month?

It is not valid to combine correlations for all groups together, since there were different treatment targets for each. For example, if you are investigating the correlation of loudness levels with other measures, this should only be used with treatments that focused on loudness. If loudness was not a treatment target (e.g. as in an augmentative communication treatment), this is not a valid correlation, and skews the relationship for other comparisons when everything is lumped together.

We disagree with this statement. For example, if loudness were correlated well with a patient reported measurement, then high loudness should correspond to a high/low score in this measurement. So if the patient were in the control arm they may have reduced loudness and a corresponding high/low score in the measurement. The more data we have the better we can look at the assess relationship.

24. Page 9, line 12: Analyses of Outcome Measures: It would be important to report the within group 3, 6 and 12 month difference scores as compared to pre, in addition to the group comparisons. Some groups may have shown greater within group change.

We were conducting a randomised study of group comparisons to compare interventions, within group analysis is not appropriate.

25. Page 9, line 15: Was the PDQ Communication Score obtained from questions 34-36?

Yes it was, this is the standard and validated way of calculating score for this domain.

26. Page 9, line 21: Carer information. How many carer's were in each group? It would be important to report this information separately, as the carer's role and time commitment might vary largely depending upon the group, and their perceptions of treatment type, dosage or no treatment will significantly influence their ratings.

We have added the number per group to the text. (13 LSVT, 11 NHS and 5 control consented to join PD COMM Pilot, most were female and spouses).

27. Page 10, lines 14-15: Regarding retention for the LSVT group. In the "real world" individuals interested in receiving treatment are educated about why LSVT LOUD needs to be delivered in this intensive manner, and are provided with information regarding the
research evidence so that they can make an informed decision about treatment. It is ethical to discuss the pros/cons of treatments and existing research evidence of why each treatment is done in a particular intensity so the individuals then have an informed decision. Decisions regarding which treatments to deliver should not be based on how many individuals enroll, but rather, what is the outcome.

This education is also important to provide to the caregivers so that they see why intensity of treatment is positive and the benefits, rather than having it presented as overwhelming and complex. It would be expected, when presented with an option to come 4 times a week for 4 weeks, versus once a week, that individuals would pick the "easier" route if they were not educated about the rationale/research for intensity of treatment.

If this is "real world" study, individuals and care givers randomized to the LSVT LOUD arm should be provided with education regarding the rationale for intensive treatment and the research evidence, as this is what would happen in a typical clinic. When clinicians are trained and certified in LSVT LOUD they are trained to provide patients and family members with this information.

We found in the LSVT arm that on average patients had 16 sessions, lasting 61 minutes over 4.7 weeks which is very close to what constitutes LSVT Loud. There will be fluctuations in this as the reports are of real world clinical contacts. We are accurately reporting the data returned to us.

As already stated ALL therapists delivering LSVT were LSVT certified.

28. Page 11 Lines 9-10: It should be noted, that even when correlations of vocal loudness are made specifically to measures of patient reported outcome for those who have participated in voice related therapies, one may see low correlations between vocal loudness and the PRO-

This has been previously reported by Spielman et al. (2010). Vocal dysfunction in PD is multifaceted, vocal loudness is one variable, but other factors influence these ratings as well. Thus, this speaks directly to the necessity to include a variety of objective and subjective measures when reporting outcomes, and not just report on a few.

It is also essential that when investigating changes in vocal loudness that the data is collected in a reliable and consistent way without cues.

We note the reviewer’s opinions.

29. Page 11 Line 14: When asking people these types of questions it is important to frame the question properly. Increased loudness or increased articulation or increased respiration on their own are not the end goal of a therapeutic intervention, rather the goal is increased
loudness or improved articulation, which provides increased intelligibility for better communication. It is important when evaluating any type of therapeutic intervention that both qualitative and quantitative outcome measures are included in order to determine the mechanism of change that was the catalyst for improved communication noted by the patient and family. Otherwise it is not possible to replicate the therapeutic intervention.

Again- we are are conducting a feasibility trial not a mechanistic study. The full trial contains a process evaluation.

As one of the hallmark symptoms of communication issues in PD is reduced loudness it would be important in future studies to continue to include measures of vocal loudness in all 3 treatment arms to have a quantitative measure to add to the battery of qualitative measures.

We are using appropriate quantitative measures.

Reduced vocal loudness can have a significant impact on intelligibility and communication success. As stated previously, vocal loudness is indeed one piece to the puzzle, and needs to be evaluated with the other parts, but of objective measures that could be collected it is easy to collect sound pressure level data (the acoustic correlate of vocal loudness) and this data would speak directly and objectively to the impact of treatment on one of the hallmark voice and speech issues in PD. If improvements in communication are noted, it is essential to understand the mechanism for improvement. Further, when investigating voice and speech changes in a population that has known issues with self-perception, such as individuals with PD, it is essential that quantitative variables are included, otherwise results cannot be considered valid or reliable. An individuals' subjective perception of vocal loudness (which in PD is many times distorted) may be quite different from an objective, quantitative measure of vocal loudness as obtained from sound pressure level readings.

Again, this is a randomised trial we are not examining mechanisms.

30. Page 11 Line 18: It is important to be aware of the limitations of perceptual scales such as the VHI, which focus on voice, in this population. Due to the sensory issues that are well-documented in PD (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), many individuals with PD may not be fully aware of the issues with their voice (especially reduced loudness) pre treatment, thus they may rate themselves pre better than they actually are. Then, with increased awareness post treatment, they may not rate themselves as "improved" as would be expected from the outcomes that they report on their improved quality of life. Thus, comparisons of pre to post perceptual ratings that just focus on voice may not accurately reflect the impact that a therapeutic intervention has had on quality of life and communication overall.
This has been further demonstrated by a study conducted by Ford et al., 2015 which concluded that gains reported on scales of functional communication (Communication Effectiveness Survey) might be a more sensitive outcome measure to reflect LSVT LOUD treatment results than those reported on the VHI.

Thus, because of questionable confounds when the VHI is used by individuals with PD to rate their voice, the VHI data alone will not address the research question and should not be used as the main dependent variable. It is important when evaluating the impact of therapy to obtain information from a variety of qualitative and quantitative measures to account for this potential confound. It would be essential in future trials to also include a measure such as the Communication Effectiveness Survey, or the Living with Dysarthria Assessment survey, as well as a quantitative measure such as SPL, to provide a full picture of therapeutic outcomes.

Much focus is made in the discussion of this paper about the inclusion of the VHI in a larger trial based on the higher correlations of it to the therapist reported outcomes. However, the correlation outcomes cannot be interpreted with confidence as they are currently reported (all subjects grouped together, unclear if reliability was collected, timeline for collection not stated, etc.)

Again this is a randomised feasibility trial. We have chosen VHI as it did appear to be sensitive to change, is an assessment of how participants perceive their voice affects their day to day life, it was well correlated with other important outcome measures, and again, reviewer needs to see my earlier explanation of correlations.

31. Page 11 Line 24: At this 3 month point were data also included in the LSVT LOUD and traditional SLT group for people who had not fully completed LSVT LOUD or traditional SLT? If yes, this data is not valid.

We disagree- this is not an efficacy study and ITT is an appropriate approach

32. Page 12 Lines 16-17: When considering cost effectiveness of treatments, it is important to look at the long term data and maintenance of treatment effects. If one treatment has more sessions initially, but results last longer, thus, resulting in fewer follow-up treatments long term, this is important to consider as a part of this equation. This is where it would be important to also report the 6 and 12 month data that were collected, and not just the 3 months data, for making decisions of outcome measures for future implementation.

This is not a cost effectiveness study, the main trial includes an economic evaluation.
33. Table 1: Please double check the data to confirm that the mean score was in fact the same for all 3 groups on the VHI. Even though there is variance in the SD, this seems odd that the mean score would be the same.

This has been checked and is correct.

34. Table 3: Data was collected out to 12 months, data included here only reports out to 3 months. What were the results of the other outcome data? This is very important information when considering effectiveness of treatment options, as if one therapeutic intervention does not last as long as another, individuals will have to consider the need for more/additional sessions later.

Again, we not examining effectiveness in this study.

35. Figure: Please clarify the numbers in the chart for the LSVT LOUD group. According to the second box it says that 22 individuals went all the way through the LSVT protocol, however in the last box it says that 27 completed the full trial. Please refer to previous comments regarding if someone did not complete 16 sessions of LSVT LOUD, they did not complete the LSVT LOUD protocol and their data should not be included.

27 completed the trial without exit. This is ITT analysis so all were included.

36. A comment about follow-up during the post intervention phase:

An important consideration for the future, if this is to be implemented into a larger scale study, is the inclusion of monitoring, reporting and controlling for continued self-practice and participation in community voice exercise groups that individuals in any of the 3 arms may participate in during the 3, 6 and 12 month follow-up period. For example, a part of the LSVT LOUD protocol involves training individuals to continue with daily home practice following the end of the 4 weeks of treatment. Also, LSVT LOUD therapists may encourage their patients to participate in post treatment, community based voice exercise groups as a motivator for continued practice. Some of these groups (such as LOUD for LIFETM) meet weekly to continue to practice LSVT LOUD exercises together. Likewise, assignments for continued practice, and communication groups exist following other types of traditional SLT interventions as well. Finally, some individuals who are in the no intervention arm, might participate in individual voice/speech practice based on videos or participate in some type of community speech/voice/communication group in place of formal therapy. These will be important activities
to track and control for, as various levels of participation during the follow-up period could influence results.

We will document this during the main trial.

Reviewer 2

1. In the Methods section, some clarification is needed of how dementia was defined and which, if any, formal cognitive tests were used?

   Clarified in text - none were used

2. Who devised and provided the training for the SL therapists, and was there a standardized training programme across all centres?

   Clarified in text - it relates to trial procedures training NOT intervention training.

3. An important point that warrants further discussion is the very large number screened vs recruited (89/2223). Given the prevalence of speech problems in PD (50-70%), what in the authors' opinion is driving this apparent discrepancy (>60% found ineligible due to no speech problems) and how will it impact recruitment for PD-COMM?

   We have highlighted the issue but have no explanation, will be interesting to see if replicated in main trial.

4. Page 8 line 9 - there appears to be an error in reporting number of control subjects completing trial (from fig 1 should be 29 not 30)

   Thank you - error changed in text

5. In the discussion section, could a reference be quoted for the MCIC of the communication domain of PDQ-39?

   Now in text