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)
John 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: 2 Date: 29 Aug 2017

Author’s response to reviews:

Reviewer reports:

Reviewer #1: Thank you for providing these responses to my initial review. An over-arching comment after reviewing the author responses is that the initial review included recommendations for extensive revision based on the many issues raised. While the authors
responded to concerns listed below in the comments it is noted that the manuscript was only minimally revised, and did not include updates to address many of the issues and concerns that were identified. Thus, the recommendation still remains that this article should not be published unless the paper can be rewritten and significant revisions are made.

We would like to make it clear that although we have responded directly to the reviewer’s comments and fully appreciate that an alternative viewpoint is expressed, we do not agree that the manuscript requires major changes. In particular, our group consists of world leading statisticians, trialists and a health economics expert. All of the group has published multiple trials and pilot trials in PD and rehabilitation. It is therefore very difficult to continue to argue about trial methodology, such as the use of intention to treat analysis. Looking at the reviewer’s publications it is clear that their speciality is not statistics and so it feels rather a circular argument to keep disagreeing with our approach. Perhaps the journal could seek advice from methodological experts with a track record of design and conducting trials of complex interventions?

I will start with a summary of the most critical issues.

Number One:

Without consideration of the points listed below, it is not possible to draw the conclusion that the VHI should be used as a primary outcome measure for the larger RCT based on the results obtained in this study. The limitations and possible confounds need to be added to the discussion.

It is imperative to note that the VHI has not been validated for use with people with PD. The 1997 Jacobsen study did not include a study of validation for the PD population. Further, when determining which outcome measures might be appropriate for the larger trial, it is critical to assess reliability (e.g. stability of the outcome measures). This is a standard and important component to include. "Reliability is a critical measurement property for health-related quality of life (QOL) instruments. Reliability refers to the consistency of scores obtained by the same persons when re-examined with the same test on different occasions or with different sets of equivalent sets of items" (Anastasi A. Psychological testing. New York: Macmillan; 1988, as cited by Marx et al., 2003).

To clarify previous confusion regarding comments about this point from my first review. In this instance, reliability is in relation to the intra-subject reliability of the ratings of the patients, not reliability of the VHI (or the other measures) themselves.

We disagree, the intra rater reliability has been published, and is referenced in the paper. ‘Test-retest stability for subscale and total scores was found to be strong for the functional (r = 0.84),
emotional (r = 0.92), physical (r = 0.86) subscales, and total score (r = 0.92).’ These values were
determined from a group of patients with a wide range of impairments.

If the VHI ratings (and other measures) were not collected more than once at each time point, the
results from the patient ratings cannot be shown to be stable, conclusions cannot be drawn about
it as a primary outcome measure and it should not be used in the larger trial without first
establishing this.

Additionally, as described at length in my first review, it is well established that individuals with
PD have problems with self-perception of their speech (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).

For example, individuals with PD may state that they don't have problems with their voice being
too soft, but rather that others need a hearing aid, when in fact their voice is very quiet. It is clear
how this issue with self- perception could skew pre/post treatment results of self-perceived
ratings of voice and speech (pre, they don't think their voice is too soft, thus ratings are higher
than they would be in a population without this self-perception issue) and how it could affect
correlations of self-perception and actual acoustic measurements of loudness.

Given this situation of a questionable primary outcome variable, and the fact that the VHI has
not been previously validated with the PD population, it is essential that the reliability of
patients' ratings on this variable be assessed in the feasibility study before recommending this as
a primary outcome variable for a larger study. Without this information the primary outcome
variable maybe invalid (untrue) and unstable.

The Voice Handicap Index is a commonly used measure and as with any HRQL measure it is
used and reported without repeat / multiple testing at each time point. Heath services research
frequently uses self reported measures and participants do not complete them multiple times
(unlike mechanistic measures of impairment, such as peak flow spirometry), for example
participants are not asked to report the SF36 three times at every assessment?

The VHI is a well used measure, included in many thousands of publications (5,500 on scoping
search), with over 350 publications in PD research, indeed including work from the reviewers
own research group-


‘Werthiemer J…, C Walton, A Duboille, M Tuchman, L Ramig - The impact of STN deep brain
stimulation on speech in individuals with Parkinson's disease: the patient's perspective
Indeed, it is widely used and investigated— even including validations of the translations of the Voice Handicap Index (VHI) and its impact on disease severity. Majdinasab et al. 


Furthermore, the introduction of the VHI as a routine measure has been suggested to improve clinical practice.


While we are interested in the reviewer’s view of the design of our ongoing large trial, we have submitted the pilot study report for consideration for publication and we feel we have demonstrated that the measure is feasible and acceptable. The patients and carers involved in our research supported the measurement of handicap.

In light of this, it is important to include these limitations in the discussion and also to highlight the need to include additional outcome measures in a larger RCT, to offset potential issue with the VHI and provide a full picture of outcomes. These should include other more global communication measures of communication effectiveness as well as objective measures such as SPL. The positive relationship between increased loudness and increased intelligibility in PD (Neel 2009; Tjadän, K., & Wilding, G. E. (2004), has been established. Thus when considering outcomes to demonstrate improved communication, this would be an important outcome measure to include.

Information needs to be added to the discussion regarding this issue of decreased sensory awareness in people with PD, the effect this could have on their VHI ratings, the need for repeated measures for intra-subject reliability, and the need to include other types of outcome measures in a larger RCT.

As explained in the previous response, the purpose of a randomised design is to accommodate such issues, variability is addressed by randomisation, there is no reason to expect systematic bias. We have included several other outcome measures in our trial, including the PDQ39 and the Living with Dysarthria scale.
Also, if no objective measures (i.e. SPL), are reported in a larger RCT, then it is likely that patient reported outcomes PRO and clinician reported outcomes (CRO) are more variable, and thus such trials may require more subjects.

Finally, when determining the most appropriate measures to include in a larger study, the determination should be based on those that measure the construct that you want to measure, and should not be based on feasibility for deriving a particular sample size.

We do not understand the above statements. The purpose of a feasibility trial with appropriate references was included in our previous response. The sample size for our main train was calculated allowing for variability.

‘To detect a 10 point difference in VHI total score between arms at 3 months (using a 2-sided t-test and the upper standard deviation of 26.27 obtained from the VHI baseline data from the pilot trial; effect size 0.38), with 80% power and α=0.01, we need 163 participants per arm. Allowing for 10% drop-out will require 182 participants per arm, so 546 participants in total.’

This will be (by far) the largest trial addressing the question of effectiveness of SLT for PD.

Number Two:

The paper should be re-organized to focus on the outcomes of the feasibility of recruitment, retention, return rate of forms, etc., and actual results of VHI and other data collected should not be included.

The purpose of a feasibility study is to determine, "can the study be done", and may or may not include which outcome measures to include in a larger study (Eldridge et al., 2016). The main focus, when following CONSORT guidelines for feasibility studies should be to include outcomes and information on recruitment, retention, and completion rates of forms, with confidence interval information reported for these measures. According to the CONSORT guidelines, the CI for these variables need to be included in the abstract. Thus, it is recommended that these be added to the abstract.

The Consort guidelines recommend reporting the results (which we have done). Looking at the papers published in the Journal, this would fit the house style.
These CONSORT guidelines further caution against reporting results of outcome measures of interventions, and if results are reported, then the limitations of these results need to be clearly stated (see previous points above in number 1).

The authors of this paper state that "As this was a feasibility study, definitive comparisons of the interventions were not appropriate." However, differences among the interventions were in fact reported. This implies that statistical analyses were conducted in spite of the above statement and the fact that the study was not powered adequately. Thus, no conclusions can be drawn from this study regarding treatment outcomes, and this information should not be included.

We made it explicit that definitive analysis was not performed and that an exploratory analysis was performed ‘An exploratory analysis of differences between the arms’. P values were not reported.


"The purpose of most feasibility and pilot studies should be to describe information and evidence related to the successful implementation and validity of a planned main trial. Null hypothesis significance testing is not appropriate for these studies unless the sample size is properly powered. The primary tests of the intervention effectiveness hypotheses should occur in the main study, not in the studies that are serving as feasibility or pilot studies”.

Also, "The primary purposes of a feasibility study are to ensure that study implementation is practical and to re-duce threats to the validity of the study's outcomes". This relates back to the previous point of questionable validity of the VHI as a primary outcome measure.

We have not tested the null hypothesis.

Also found in Arain et al. BMC Medical Research Methodology 2010, 10:67 http://www.biomedcentral.com/1471-2288/10/67, 

"Crucially, feasibility studies do not evaluate the outcome of interest; that is left to the main study".

We have not reported efficacy.
Finally, as stated in the first review, regarding results that are currently being reported. 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 mis-leading and should not be included. Additionally, all references to directional comparisons (e.g. this mean is greater than that) imply statistical significance, thus these should be removed as well. In conclusion, the authors should remove confidence intervals and directional language regarding the differences between means for treatment outcome measures in the abstract and in the paper.

We disagree, the reporting of the clinical outcome measures is to provide data to inform the sample size of the main trial.

Number Three:

An important component of a feasibility study is to identify potential issues with methods and solidify procedures so that data collected for a larger trial will be valid. Training of data collectors and ensuring that post treatment data is collected in the same way, that SPL and other objective data are collected in a consistent manner across sites, that data is not collected by a treating clinician and is not cued for performance is essential. The feasibility study is the time to identify the process for this, what worked and did not work and what will be done prior to a RCT to ensure that clinicians and data collectors understand their role in impacting the data. Information should be included in the feasibility article that describes issues that arose regarding implementation and solutions that were derived to address this in a larger RCT.

Please see below for a few other additional points that were not mentioned at the time of the first review.

1. A point to take into consideration for the larger RCT, and that should be mentioned in the discussion of this paper is the design flaw of the study. For example, standard treatment was given over a longer period of time than LSVT LOUD. If it is determined in a larger study that SLT had more optimal estimates of outcomes, it would be difficult to draw a straightforward conclusion. It would be impossible to determine if SLT outcomes were better because of the SLT treatment itself, or if outcomes were better because of giving treatment over a longer period of time that coincides with treatment assessment points.
Complex interventions frequently differ in terms of dose, intensity and duration, this does not prevent comparisons of effectiveness.

2. The following statement should include data, so that the statement does not appear to be conjecture.

* A number of patients decided not to enter the trial because of the intensity of LSVT LOUD® which did affect recruitment rates.

20 patients decided not to enter the trial because of the intensity of LSVT LOUD® which did affect recruitment rates.

Please see below for additional specific comments to some of the author comments. For these items, I have left in my original comments from February 21st, followed by a notation of the author comment, followed by a notation of my latest comments as indicated by Reviewer 1, June 2017.

Reviewer 1: Feb., 2017

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.
Author Comment: Thank you for your comments but we disagree with this statement, the study was multi entre 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.

Reviewer 1, June 2017: Because this is a feasibility study and not an efficacy study, please see points in the initial summary of what should be removed from the article and how the article should be reorganized.

Response; We disagree that this does not follow the guidelines or structure of many other articles published in the journal.

Reviewer 1, Feb. 2017: 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.

Author Comment: 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.
Reviewer 1, June 2017: The language in the article with regard to the outcomes implies relative efficacy. Thus, please see previous points about re-writing these sections. Also, randomization with such a small sample size, does not guarantee that the groups are balanced.

Additionally, as stated above, reliability in this instance was related to intra-subject reliability and the need for repeated measures. Randomization for equal sensory awareness problems in all groups, does not justify the use of the VHI. Rather it suggests that the ratings will not be valid or reliable in all groups and thus not an appropriate way to assess the impact of either treatment when used on its own. While the VHI can provide important information, due to the issues with decreased sensory awareness it should be paired with more global communication scales and quantitative measures as well.

There is no data to support this viewpoint.

It is important to alert the reader to the limitations of the VHI with this population, by including an explanation of this for the reader in the paper. And to also note in the paper the importance of including additional communication measures and quantitative measures (e.g. measurements of SPL) in the larger trial.

We have justified the inclusion of the VHI and explained the purpose of randomisation already.

Reviewer 1, Feb. 2017: 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.

Author Comment: 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.

Reviewer 1, June 2017: Thank you for this information, a description of this should be included in the paper to address feasibility of the methods. When you say the "therapist will have been blinded when doing the…" was this for pre only? Who collected this information at 3, 6 and 12 months? If this was collected at any of the post time points by the treating therapists, the data would be cued (treating therapists by nature of being present serve as an external cue) and therefore biased and not valid. An explanation of this needs to be included in the paper. In future trials it will be essential that the therapists who provided treatment do not collect the post treatment data.

They are not
I would refer the authors to Neel, 2003, JSHLR. This article provides a nice description of the role of SPL in intelligibility and controlling for SPL when evaluating the impact of SPL on intelligibility.

This is not the purpose of the trial, which compares the effectiveness and cost effectiveness of interventions delivered in the NHS vs control.

The assessments were blinded, which is what we say in the paper, we do not understand the

Reviewer 1, Feb. 2017: Reliability of the outcome measures is not reported, if reliability data was not collected there is no measurement stability – we disagree – addressed above and the data cannot be considered valid. See item 18 for more details regarding this issue.

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


Reviewer 1, June 2017: As mentioned previously, the VHI has not been validated with the PD population. It is mentioned that the Jacobson and Rosen were referenced. I do not see these references included in the reference list? A reference to this VHI scale needs to be added to the paper. Also, it appears that the regular VHI, not the VHI-10 was used in the current paper, thus reference to the Rosen article and the VHI-10 should not be included in the paper or this review discussion. Without repeated measures data cannot be interpreted with confidence. As mentioned above, if this is a feasibility study, not an efficacy study, it is not appropriate to reports results or make directional comparisons.

Reviewer 1, Feb., 2017: 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.
Author Comment: 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.

Reviewer 1, June 2017: Stating that the correlations are just at the 3 month interval would be important to add to the text and to the correlation chart in Table 2, as this was not clear. Please see previous point in the initial summary statements about only reporting correlations for the entire group lumped together if this is baseline data. It is not statistically appropriate to report it on the whole sample post treatment, without accounting for treatment groups.

As explained previously, the statisticians (authors) are career statisticians.

Reviewer 1, Feb. 2017: 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.

Author Comment: 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.

Reviewer 1, June 2017: Therefore, all references to later time points beyond 3 months should be removed from the paper.

If one of the aims of the feasibility study was to determine the best outcome measures to include for a trial that will last out to 12 months, it is essential to also look at the correlations of suggested outcome measures at 6 months and 12 months, as the results might be quite different than what is seen at 3 months.

For the larger study, it will be essential to have data at 6 and 12 months, as when investigating the impact of treatments on PD, and which treatments are the most effective, it is essential that the studies follow subjects long term in order to determine which treatments have lasting effects and are not just a result of short term cues or performance behavior.

Again, we feel the reviewer is criticising the design of the full study, yet the protocol is unpublished? The full trial will examine follow up outcomes.

Reviewer 1, Feb., 2017: 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.

Author Reply: 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.

Reviewer 1, June 2017: ITT should not be included in the manuscript because no efficacy analyses were done.

Including data from people who did not actually complete a particular protocol could result in diluted results and falsely make two group look more similar than they are. In the larger trial, when comparing the effectiveness of LSVT LOUD versus traditional SLT versus no-treatment, it would be important to be aware of these confounds. The potential confounds of continuing to include data from someone in the LSVT LOUD arm who did not actually receive LSVT LOUD (if someone did not receive 1 hour sessions, 4 times a week for 4 weeks, they did not receive LSVT LOUD) should be included in the discussion.

Intention to treat is the analysis of choice, patient do not all receive a full course of intervention within any health system (due to illness, death, drop out etc etc).

Reviewer 1, Feb., 2017: 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; Sapi et al., 2011).

Author Comment: 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.

Reviewer 1, June 2017: In summary to my previous description of the data, there have been three RCTS investigating LSVT LOUD:

1) Two PD treatment groups comparing LSVT LOUD vs RESPIRATORY treatment, studied out to 24 months. Published

2) LSVT LOUD versus untreated PD and healthy controls out to 6 months. Published

3) LSVT LOUD and LSVT ARTIC out to 6 months (preparing for publication, and currently many posters and presentations on this study) as well as data published from other groups. In order to more accurately reflect the literature available (comparison of LSVT LOUD to alternative treatment, inclusion of untreated PD, and healthy control, long term data, randomized controlled trials, studies from other groups), it would be recommended to change the "little evidence" statement in the abstract.

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

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

Reviewer 1, June 2017: A publication is currently in preparation for the 3rd RCT.

In addition to the Halpern, 2016 that you have already cited, please see below for citations for presentations/abstracts that have been done thus far.


The evidence behind LSVT LOUD has been documented by additional research groups, separate from the work done by Ramig et al. Please see below for an example of some of these.


Please also refer to Atkinson-Clement et al., 2015 for a meta-analysis of behavioral speech treatments in PD.

We cannot find between group data with confidence intervals of this 3rd trial. We are reporting high quality randomised studies in our Cochrane review, and there is little high quality data.

Reviewer 1, Feb., 2017: 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?

Author Comment: 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.
Reviewer 1, June 2017: Thank you for this additional information. However, it is still not clear who collected the 3, 6 and 12 month information. When you say this information was collected by blinded therapists, was this just for pre. Or at post and 3 month did a different therapist collect that information? This is important information that should be included in the paper to speak to the feasibility of the proposed method for a larger study.

Can you please clarify what is meant by "arbitrary 4th test" that was used for the correlation of vocal loudness and patient reported measures? This would also be an important clarification to include in the text and in the footnote on Table 2.

We are not sure what is unclear, the therapists were blinded to group allocation? The correlation was taken against the 4th measure of loudness and this was an arbitrary decision.

Reviewer 1, Feb., 2017: 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.

Author Comment: 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.

Reviewer 1, June 2017: Then it might be more accurate to change the text to say, "as LSVT LOUD may not be appropriate for some individuals with these previous medical conditions".

The trial inculdes other forms of therapy

Reviewer 1, Feb., 2017: 7. Page 5 line 20: LSVT LOUD Treatment sessions should be 60 minutes in length, not 50-60 minutes.

Author Comment: 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.
Reviewer 1, June 2017: Therapists are trained globally to deliver the treatment for 60 minute sessions. If the treatment is not delivered in 60 minute sessions, it cannot be considered LSVT LOUD. Participants who do not receive 1 full hour of individualized treatment for 16 sessions in 4 weeks would not have actually received LSVT LOUD treatment. Dosage and intensity are an important part of the treatment protocol. A note about this should be included in the paper that some individuals may not have received the protocol, and thus did not receive LSVT LOUD as described. It would be essential to control for this in a larger study.

Discussed previously

Reviewer 1, Feb., 2017: 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.

Author Reply: Corrected

Reviewer 1, June 2017: Thank you for the printed correction. Can it be confirmed that the therapists did this exercise correctly and not as originally written in the paper? If not a piece of the protocol was not correctly delivered and should be noted as such.

The correction has already been made.

Reviewer 1, Feb., 2017: 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.

Author Comment: Noted we have expanded the description.

Reviewer 1, June 2017: Thank you for the printed correction. Can it be confirmed that the therapists did incorporate sensory calibration training into the treatment sessions? If not a piece of the protocol was not correctly delivered and should be noted as such.

We have already corrected this

Reviewer 1, Feb., 2017: How was it ensured that LSVT LOUD was delivered according to protocol?
Author Comment: 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?

Reviewer 1, June 2017: Since information was obtained regarding duration and sessions attended for this study, at a minimum, this information could be reviewed to ensure that anyone in the LSVT LOUD arm received 60 minutes of treatment 4 times a week for 4 weeks (which it appears you did). After this analysis, then it should be noted which individuals did not receive LSVT LOUD. If this is a study of how therapies are delivered within the NHS, and people assigned to LSVT LOUD, did not actually receive LSVT LOUD (e.g. treatment was not 1 hour, 4 times a week for 4 weeks), then this is very important to note in the discussion, as then general conclusions cannot be made of outcomes from LSVT LOUD on the whole if people are not actually receiving the treatment.

Already discussed purpose of intention to treat analysis

There are also skills checklists that LSVT LOUD therapists are provided with during their training that should be used by the therapists to ensure that they are adhering to the components of the protocol. This is important to establish fidelity of treatment. If this was not utilized it should be listed as a limitation to the study in the discussion. This checklist should be utilized for a larger RCT.

Reviewer 1, Feb., 2017: 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.

Author response: 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.

Reviewer 1, June 2017: See previous comments above.

See previous comment

Reviewer 1, Feb., 2017: 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?

Author comment: See '4' 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.

Reviewer 1, June 2017: It is not clear what is meant above by "Then the 3, 6 and 12 months were done by post." I think some information is missing from your response above following the word post? Can you please clarify?

The questionnaires were returned by post- what would the USA equivalent be? By mail?

Reviewer 1, Feb., 2017: 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.

Author Reply: 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.

Reviewer 1, June 2017: It would be important to add this caveat to the paper regarding "if treatment was not done correctly" and that LSVT LOUD is intended to decrease vocal strain.

Reviewer 1, Feb., 2017: 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, and variability that can occur day to day when collecting data from any individual, 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.

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.

Author Reply: 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.

We do not routinely conduct further validation of measures.
Reviewer 1, June 2017: Please see previous point regarding the need for this to establish intra-subject reliability in terms of participant performance, versus validation of the measurement tool itself. Also, the previous point that the VHI has not been validated for use with PD.

Please see above

Reviewer 1, Feb., 2017: 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 treatment 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.

Author Response: 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;


Pilot and Feasibility Studies, 2 (25)


Reviewer 1, June 2017: Please see previous comment about not including actual results of outcomes measures or directional comparisons, but rather just focusing on descriptive statistics, and an explanation of the limitations of study design for the reader.

Discussed above-
Reviewer 1, Feb., 2017: 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.

Author Reply: 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.

Reviewer 1, June 2017: Please see previous comment about use of ITT in this instance. Further information should be included in the paper about the individuals who did not complete LSVT LOUD according to protocol and thus did not actually receive LSVT LOUD therapy.

Discussed above

Does the reviewer have implementation data on the number of patients who clinically fail to complete 16 sessions in Europe, USA etc?

Reviewer 1, Feb., 2017: 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.

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.

Author Response: Again, this is a true reflection of how care is delivered within a health care system and we were assessing feasibility for a phase 111 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.

We were not assessing efficacy with this study. It is a feasibility for a phase iii study of clinical and cost effectiveness.

Reviewer 1, June 2017: Feasibility of comparing traditional SLT to LSVT LOUD cannot be determined if individuals are not actually receiving LSVT LOUD. This is an important topic that needs to be discussed in this paper, and would also need to be controlled for in a larger trial.

The question of "how is health care delivered" and what typically happens in the NHS system is a different question than the investigation of comparison of LSVT LOUD and standard SLT and no treatment.

The full trial has a work stream investigating and recording treatment fidelity.

Reviewer 1, Feb. 2017: 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.

Author Response: 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.

Reviewer 1, June 2017: Please see previous comments about correlations.

We have to present the results we recorded, all our results are recorded and analysed independently from myself as chief investigator.

Reviewer 1, Feb., 2017: 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.

Author Reply: We note the reviewer's opinions.

Reviewer 1, June 2017: It would be important to include the rationale behind the need for inclusion of quantitative (such as SPL) and a variety of qualitative measures for a larger trial in the discussion.

We disagree and will use a primary measure not compound end points.

Reviewer 1, Feb., 2017: 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.

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

Reviewer 1, June 2017: It would be important to include the rationale behind the need for inclusion of quantitative and a variety of qualitative measures (beyond just the VHI) for a larger trial in the discussion.

See above

Reviewer 1, Feb., 2017: 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.

Author Reply: We are using appropriate quantitative measures.

Reviewer 1, June 2017: If this is the case, please add to the discussion the plan to use quantitative measures such as SPL in the larger trial, as this is currently not clear. I refer the
authors to a few studies that nicely demonstrate the relationship between loudness (SPL) and intelligibility, and the importance of including this type of an objective measure when evaluating outcomes for the larger RCT.


This is a feasibility study for a large rct of over 500 participants.

Reviewer 1, Feb., 2017: 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.

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

Reviewer 1, June 2017: It would be important to include the rationale behind the need for inclusion of quantitative and a variety of qualitative measures for a larger trial in the discussion.

As above

Reviewer 1, Feb., 2017: 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) pretreatment, 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 (Halpern et al., 2012).

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.)

Author Reply: 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.

Reviewer 1, June 2017: Please see previous comments about the invalidity of the correlations and confounds for use of the VHI as a primary outcome variable, and the need to discuss these confounds in the paper.

We do not think that our data is invalid-conflicting results from small studies from other investigators does not mean ours are invalid just because the results are not what people would wish.

Reviewer 1, Feb., 2017: 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.

Author Reply: We disagree- this is not an efficacy study and ITT is an appropriate approach
Reviewer 1, June 2017: Please see previous explanation of why ITT is not an appropriate approach in this instance.

Disagree, ITT is appropriate

Reviewer 1, Feb., 2017: 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.

Author Reply: This is not a cost effectiveness study, the main trial includes an economic evaluation.

Reviewer 1, June 2017: This comment was made due to the statements by the authors in the paper on page 3 line 22, page 10, line 6 and page 12, line 22 about investigating cost effectiveness. This will be an important consideration to include in the discussion of this paper, and as consideration for a larger trial, as fewer sessions initially may appear most cost effective, but if a treatment does not last, more sessions will be required later.

Dr Sue Jowett is a reader in health economics, she has designed and published many health economic evaluations from RCTS.

Reviewer 1, Feb., 2017: 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.

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

Reviewer 1, June 2017: Please see previous comments about use of ITT in this instance.