PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Sing Your Lungs Out- A community singing group for chronic obstructive pulmonary disease: a one-year pilot study
AUTHORS	McNaughton, Amanda; Weatherall, Mark; Williams, Mathew; McNaughton, Harry; Aldington, Sarah; Williams, Gayle; Beasley, Richard

VERSION 1 - REVIEW

REVIEWER	Nick Hopkinson Imperial College, London UK
REVIEW RETURNED	16-Sep-2016

GENERAL COMMENTS	The authors describe a feasibility study recruiting patients following PR to a weekly singing program. Lung function, exercise capacity and hospitalisation are recorded at 6 and 12 months. The data are encouraging but the absence of a control group limits interpretation of the findings which should focus on the feasibility result (it is feasible) and discuss this in the context of other post-pr interventions to sustain benefit. The claims to originality are a little overstated - they have missed citing Stephen Clift's large long term feasibility study. Major The statistical analysis is inappropriate for repeated measures - they should perform ANOVA with post hoc test. There is anyway no hypothesis advanced as to why singing should improve lung function tests - the RV result is almost certainly a chance finding. The size of effect is anyway well below the mcid and certainly cannot be described as 'important'. Of note ref 19 measured lung volumes immediately after the class ie an acute effect of the singing. It would be much more informative if participants' response to PR was included as well. A PR nurse attended every session- it is likely they reinforced exercise advice. This should be discussed. The CONSORT diagram should include how many people were approached to take part. Is there any relationship between level of participation and benefit - a dose response? Minor Disinterested is the wrong word.

	The term "real world" should be avoided Similarly, the fact that studies are hospital based simply reflects a context for intervention. It isn't a "limitation" in the sense that a small sample size or short follow up is. Page 3 line 51 - there is no evidence that this is the case and this sentence should be deleted. Were patients on the same medication? The data presented here are not "mixed methods"

REVIEWER	Renae McNamara Prince of Wales Hospital Australia
REVIEW RETURNED	26-Sep-2016

GENERAL COMMENTS	This prospective observational study examined the effect of group singing sessions over the course of a year, on lung function and clinical outcomes in people with COPD. Weekly singing sessions were found to be feasible, with high attendance. General feedback: - Greater attention needs to be paid to the accuracy of the methods and reporting of results. - The discussion is limited due to repeated reference to the qualitative component of the study, and little discussion of the findings in relation to other interventions for COPD. Specific comments and questions: - This is referred to as a 'mixed methods' study throughout the manuscript however, only the quantitative results are presented (with the qualitative results previously published). 'Mixed methods' is misleading for this paper, and should be removed. - As qualitative results are not reported in this manuscript, the reference in the introduction (3rd paragraph) to whether singing 'is acceptable', and in the first paragraph of the discussion ('supports the acceptability') should also be removed as this was not an aim. - In the abstract it is stated that hospital admission days for AECOPD were collected 'for one year before and after joining the singing group', and in the methods it is stated that this data were recorded 'for the 12 months before and 12 months after enrolment'. Please clarify whether the hospital admission data was collected during the 12 months of the singing group, or in the 12 months after completion of the 12 months of singing. Please also consider the significant effect of attending PR and the time period following completion of PR before commencing the singing group, as PR has been shown to have a significant effect on hospital admissions. - Please review the number of decimal points reported throughout the manuscript. Generally, percentages should be reported without decimal places in sample sizes less than 100. Also, do not imply greater precision than the measurement instrument and unit, e.g. 6-minute walk distance is not measured in meters with decimal points. - The study concludes that there was an important reduction in residual volume and total lung capacity after 4 months. What evidence do you have of these reductions being important? Hartman et al (ERJ, 2014) report an MCID for RV ranging between -0.31 and -0.43 L (-6.1 to -8.6% for percentage change). - The participants were recruited after completing a pulmonary rehabilitation (PR) program. How long after completion of the PR were they recruited? What were the changes in clinical outcomes

	following participation in PR? - How were this participant group 'unselected' (referred to in strengths and limitation dot points, introduction (3rd paragraph) and discussion (3rd paragraph))? The methods state that participants were invited to join the group, thus it appears participants self-selected to join the group. - How many participants attending the maintenance PR group were offered the singing group and how many accepted? - In the methods, it states the trial was prospectively registered on www.anzctr.org.au, however the date of registration on this website is July 2015 and the enrolment of participants occurred between Oct 2014 and Feb 2015. - Data collection – please provide a reference to the ATS/ERS standards used for pulmonary function tests. - Analysis – very brief description provided and no justification for lack of adjustment. Please provide further details. - Results – the first paragraph is very hard to follow in regards to the participant numbers at different stages of the recruitment and testing, and it appears different numbers have been reported in the abstract and Figure 1. Please carefully review. - Reporting of results – there is a lack of consistency in reporting of data, with a combination of mean (SD), medians and ranges reported for the same outcomes. Please review each outcome measure for normal distribution, and report consistently as parametric or non-parametric statistics. Hospital admission data were reported as being skewed – results in Table 2 should be reported as median (IQR). - The range of FEV1 % predicted is reported as up to 110%. Did this participant/s really have COPD? - How were comorbidities assessed? Self-reported or chart-based? - The group of participants had mild to moderate COPD. This should be highlighted in the discussion/limitations as it limits the generalisability of the findings. - The group of participants had very low anxiety and depression scores on the HADS at baseline, thus a reduction in scores would not be considered clinically significant. Please comment on the clinical relevance of this and add to discussion. - Table 1: o The BODE score is not an indicator of COPD severity, but was developed to predict mortality. o All outcomes reported in Table 2 should also be reported in Table 1 (and would be helpful for the reader) (e.g. RV and TLC % predicted). o The HADS should only be reported as an anxiety and depression score (not a total score). - This group of participants were recruited from a PR maintenance exercise class. Did the participants continue to attend this weekly exercise group as well as the singing group? If so, this needs to be highlighted, as their participation in this group may have contributed significantly to the reported change in outcomes. - The methods state the 6MWT was performed according to international guidelines, but the discussion reports only one 6MWT was performed, thus the 6MWT was not performed according to international guidelines. - Figure 2 – please reconsider the inclusion of Figure 2 as it does not appear to add to the presentation of results. In fact, the huge variability leads to the question of whether the data was normally distributed and whether mean (SD) is an appropriate way to report these results. - Discussion:

	o 2nd paragraph – please clarify what 'music practice' was. o Additional limitation to consider – the reduction in RV at 4 months was discussed as a possible effect of singing on the respiratory muscles, but it appears inspiratory and expiratory muscle strength was not measured. o The methods state a CD was provided to participants for home practice, however the discussion states there was no practice at home. Was home practice recorded or reported in the qualitative feedback? - The 2nd sentence of last paragraph of discussion and the last 2 sentences of the conclusion should be removed and added as an appendix/online supplement.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 (NH)
1. The claims to originality are a little overstated - they have missed citing Stephen Clift's large long term feasibility study.

We regret the omission. Morrison and Clift's paper escaped our attention due to its non-traditional source (UNESCO Observatory Multi-Disciplinary Journal in the Arts). We have included the reference (#20) and added sentences in the introduction (line 82-7) and discussion (line 234-7)
2a. The statistical analysis is inappropriate for repeated measures - they should perform ANOVA with post hoc test.

This is a small data set and we think that sophisticated analyses may obscure rather than describe the findings, which was the reason for selecting the paired t-tests; easy for clinicians to understand and see what was happening. Paired t-tests also already account for the correlation between adjacent measurements but don't force the correlation to be the same between more distant measurements (as does a repeated measures ANOVA). The method suggested by the reviewer has probably been superseded by mixed linear models which allow more flexible modelling of the repeated measures covariance (other than the compound symmetric covariance implied by the repeated measures ANOVA) and also controls for Type I error rate within each analysis for comparison of different measurement times (rather than having to do a post hoc adjustment). However as we have done a lot of statistical tests even within variable adjustments have to be regarded with a lot of caution. At the reviewers request we have repeated the analyses with a mixed linear model with the individual participants as a random effect and a fully flexible, so-called unstructured, variance-covariance matrix. In general it made very little or no difference to the paired t-test estimates and confidence intervals and P values – see Table 2 in the 'tracked changes' document to appreciate the size of the changes – no change at all in the four month comparisons, minor changes only in the 12 month comparisons. The Analysis section has been modified (line 150-4)

2b. There is anyway no hypothesis advanced as to why singing should improve lung function tests.

An hypothesis is suggested in the Introduction (line 73-4) ie active exhalation may theoretically reduce residual volume (RV)

3. the RV result is almost certainly a chance finding. The size of effect is anyway well below the mcid and certainly cannot be described as 'important'

We agree that chance is a possible cause. The reference in the Abstract/Results section to 'important' (line 41) was an error and has been removed. The Abstract/Conclusion (line 48-9) and Conclusion (line 283-4) sections have been rewritten, removing any reference to improved lung function. The discussion has been rewritten with references to this finding more circumspect (line 216-8). Specifically (line 254-5), we have included the statement that this result may be due to the play of chance.

4. Of note ref 19 measured lung volumes immediately after the class ie an acute effect of the singing.

Yes. We have added a statement to this effect in the Discussion (line 222-3)

5. It would be much more informative if participants' response to PR was included as well.

The data for preceding PR response are limited (14/21 participants). We have included a summary statement in the Results section (line 158-60)

6. A PR nurse attended every session- it is likely they reinforced exercise advice. This should be discussed.

The PR nurse did not actively participate in the sessions themselves but had a role in 'meeting and greeting' participants, general encouragement to attend, supervising refreshments, organising transport for performances. There was no formal or informal exercise advice provided. We think her 'caring role' was more important than in encouraging exercise and, at the risk of referring again to the qualitative data we have added a few lines to the Discussion (lines 252-7)

7. The CONSORT diagram should include how many people were approached to take part

Figure 1 has been modified

8. Is there any relationship between level of participation and benefit - a dose response?

We agree that this is an interesting question. An analysis of dose-response was not attempted because of the high rate of attendance – a statement has been added to the Discussion (line 260-3). We also note the concerns of the reviewer above that it is already difficult to tell which of the statistical tests might be spuriously positive (Type I error rate inflation) and we did not wish to add another.

9. Disinterested is the wrong word.

Agreed and amended (Figure 1 and line 161)

10. The term "real world" should be avoided

Agreed and amended (multiple places)

11. Similarly, the fact that studies are hospital based simply reflects a context for intervention. It isn't a "limitation" in the sense that a small sample size or short follow up is.

Agreed, sentence amended (line 81-2)

12. Page 3 line 51 - there is no evidence that this is the case and this sentence should be deleted.

This sentence has been deleted from the Introduction

13. Were patients on the same medication?

Medication use was not specifically recorded. Participants were managed by their usual doctors.

14. The data presented here are not "mixed methods"

Agreed, see Reviewer 2 comment below.

Reviewer 2 (RM)
1. This is referred to as a ‘mixed methods’ study throughout the manuscript however, only the quantitative results are presented (with the qualitative results previously published). ‘Mixed methods’ is misleading for this paper, and should be removed.

References to ‘mixed methods’ have been removed from Abstract/Method section, Strengths statement and Discussion.

2. As qualitative results are not reported in this manuscript, the reference in the introduction (3rd paragraph) to whether singing ‘is acceptable’, and in the first paragraph of the discussion (‘supports the acceptability’) should also be removed as this was not an aim.

We accept the reviewer’s viewpoint but feel that the high attendance rate in itself is evidence of ‘acceptability’ by the participants. We have removed the reference to ‘enjoyment’ of the intervention (previously in the Discussion) which comes from the qualitative data. We have only used findings from the qualitative component of the study to help explain high attendance and lower anxiety in the Discussion and not as outcomes for this report.

3. - In the abstract it is stated that hospital admission days for AECOPD were collected ‘for one year before and after joining the singing group’, and in the methods it is stated that this data were recorded ‘for the 12 months before and 12 months after enrolment’. Please clarify whether the hospital admission data was collected during the 12 months of the singing group, or in the 12 months after completion of the 12 months of singing. Please also consider the significant effect of attending PR and the time period following completion of PR before commencing the singing group, as PR has been shown to have a significant effect on hospital admissions.

Hospital admission days were collected for the 12 months before and the 12 months starting from singing group entry. The Methods section has been clarified (line 129-30). We accept the point about the effect of PR on admissions although mean time from PR end to singing group enrolment was 1.2 years so we may have expected any benefit on admissions from PR to reduce the admissions in the 12 months prior to singing group entry.

4. - Please review the number of decimal points reported throughout the manuscript. Generally, percentages should be reported without decimal places in sample sizes less than 100. Also, do not imply greater precision than the measurement instrument and unit, e.g. 6-minute walk distance is not measured in meters with decimal points.

Results and abstract amended for 6MWT and attendance. We wished to preserve precision as much as possible so that if readers wished to do their own statistical tests or use this data for their own studies (e.g. sample size estimates) that there is sufficient precision to do this.

5. The study concludes that there was an important reduction in residual volume and total lung capacity after 4 months. What evidence do you have of these reductions being important? Hartman et al (ERJ, 2014) report an MCID for RV ranging between -0.31 and -0.43 L (-6.1 to -8.6% for percentage change)

See Reviewer 1 point 3

6. The participants were recruited after completing a pulmonary rehabilitation (PR) program. How long after completion of the PR were they recruited? What were the changes in clinical outcomes following participation in PR?

See also Reviewer 1 comment 5

7. How were this participant group 'unselected' (referred to in strengths and limitation dot points, introduction (3rd paragraph) and discussion (3rd paragraph))? The methods state that participants were invited to join the group, thus it appears participants self-selected to join the group.

Agreed. We have changed the references to 'unselected' to 'broad inclusion criteria'

8. How many participants attending the maintenance PR group were offered the singing group and how many accepted?

See Reviewer 1 comment 7

9. In the methods, it states the trial was prospectively registered on www.anzctr.org.au, however the date of registration on this website is July 2015 and the enrolment of participants occurred between Oct 2014 and Feb 2015

Yes, there was a delay in registration although it was initiated early in the study. The word 'prospectively' has been deleted from Methods section (line 112).

10. Data collection – please provide a reference to the ATS/ERS standards used for pulmonary function tests
Ref #25 added

11. Analysis – very brief description provided and no justification for lack of adjustment. Please provide further details.

See Reviewer 1, point 2 also. This study has a relatively small number of participants and a large number of variables. In our view it is unusual for studies of this sort to carry out statistical testing with, for example, a Bonferroni adjustment (divide the nominal P value for significance of 0.05 by the number of P values assessed). This is why we have left it as a weakness for readers to judge and also with reporting of P values to sufficient precision readers could carry out their own informal Bonferroni adjustment

12. Results – the first paragraph is very hard to follow in regards to the participant numbers at different stages of the recruitment and testing, and it appears different numbers have been reported in the abstract and Figure 1. Please carefully review.

We apologise for the lack of clarity. The Results paragraph (line 157-68) and the Abstract/Findings section (line 36-38) have been rewritten. Figure 1 has also been modified and checked for consistency.

13. Reporting of results – there is a lack of consistency in reporting of data, with a combination of mean (SD), medians and ranges reported for the same outcomes. Please review each outcome measure for normal distribution, and report consistently as parametric or non-parametric statistics. Hospital admission data were reported as being skewed – results in Table 2 should be reported as

median (IQR).

We respectfully disagree. It was not our intention and nor do we imply that the reporting of summary descriptors implies our view on whether normality assumptions are strongly enough violated for normal distribution tests to be invalid. Rather our approach is to report data summaries in sufficient detail that readers can understand the data distributions. In a data set of this size formal tests of violation of normality assumptions themselves lack statistical power to detect departures from this assumption and in any case the methods we have chosen are robust to mild non-normality. We have modified Table 1 to present summary data as mean (sd), median (IQR) and min/max.

14. The range of FEV1 % predicted is reported as up to 110%. Did this participant/s really have COPD?
Yes – FEV1/FVC =56%, TLC 6.1L

15. How were comorbidities assessed? Self-reported or chart-based?
Chart-based

16. The group of participants had mild to moderate COPD. This should be highlighted in the discussion/limitations as it limits the generalisability of the findings.

Although the majority had moderate COPD, 20% had severe or very severe COPD. This is a reasonable reflection of a community living population with COPD in New Zealand eg Shirtcliffe et al in a community study of COPD in New Zealand (Eur Respir J 2007; 30: 232–239) 46% mild, 45% moderate, 7% severe, 3% very severe. We have added a statement in the Results section (line 171-3)

17. The group of participants had very low anxiety and depression scores on the HADS at baseline, thus a reduction in scores would not be considered clinically significant. Please comment on the clinical relevance of this and add to discussion.

We respectfully disagree that our study population had 'very low anxiety and depression' as suggested. For example, in Coulton et al's RCT (BJPsych 2015, 207:250-255) of singing group intervention in older people baseline mean HADS anxiety was 6.4 and depression 4.3 compared to our values anxiety = 5.8, depression = 4.1. In the 2 RCTs of singing in COPD of Lord et al, baseline HADS anxiety was 5.5-5.8 although depression scores were higher than ours (5.5-6). We have not modified the Discussion section, having pointed out already that the change in HADS anxiety score (0.9) is below the MCID (1.32) [line 218-9] and thus may not be considered clinically significant.

18. Table 1:
The BODE score is not an indicator of COPD severity, but was developed to predict mortality.

Agreed. We have included both BODE scores and severity grades using FEV1% predicted (GOLD criteria) to Table 1

19. All outcomes reported in Table 2 should also be reported in Table 1 (and would be helpful for the reader) (e.g. RV and TLC % predicted).

Our intention was to enable readers to follow across the changes in the variables with time in one table and to also provide separation of the analysis from the data description trying to keep Table 1 as simple as possible. We would prefer to keep the tables as they are for this reason rather than duplicate more data across the 2 tables.

20. The HADS should only be reported as an anxiety and depression score (not a total score).

We were unable to find support for the reviewer's statement. Most papers we could find using the HADS in COPD patients report HADS Total as well as HADS-D and HADS-A separately, for example: Nowak C, et al Accuracy of the Hospital Anxiety and Depression Scale for Identifying Depression in Chronic Obstructive Pulmonary Disease Patients. Pulmonary Medicine (2014), http://dx.doi.org/10.1155/2014/973858
Ian McDowell in his book 'Measuring Health: A guide to rating scales and questionnaires' (3rd ed 2006, OUP) pages 294-300 supports the use of the total HADS score as having better discriminant ability than the HADS-D score alone for depression

21. This group of participants were recruited from a PR maintenance exercise class. Did the participants continue to attend this weekly exercise group as well as the singing group? If so, this needs to be highlighted, as their participation in this group may have contributed significantly to the reported change in outcomes.

This has been clarified in Methods (line 147-8) and discussed in the Discussion section (line 248-251)

22. The methods state the 6MWT was performed according to international guidelines, but the discussion reports only one 6MWT was performed, thus the 6MWT was not performed according to international guidelines.

Yes. The Methods section has been modified (line 122-4)

23. - Figure 2 – please reconsider the inclusion of Figure 2 as it does not appear to add to the presentation of results. In fact, the huge variability leads to the question of whether the data was normally distributed and whether mean (SD) is an appropriate way to report these results.

We have removed Figure 2. See the point above (Reviewer 2, point 13) regarding normality assumptions

24. Discussion: 2nd paragraph – please clarify what 'music practice' was.

We were trying to express the idea that participants didn't see singing group as primarily something that they did to improve their singing but saw it first of all as a fun social event each week. That it was also enjoyable to sing was an added bonus. We have changed the phrase 'music practice' to 'musical endeavour' (line 216-7) to try and clarify this point.

25. Additional limitation to consider – the reduction in RV at 4 months was discussed as a possible effect of singing on the respiratory muscles, but it appears inspiratory and expiratory muscle strength was not measured.

Agreed. We were simply proposing an hypothesis

26. The methods state a CD was provided to participants for home practice, however the discussion states there was no practice at home. Was home practice recorded or reported in the qualitative feedback?

Practice at home was optional. Some participants certainly did practice at home, especially before performances. Participants took the performances (sometimes in front of audiences of 100 people or more) seriously and the qualitative feedback emphasises the idea that they had achieved something that they thought unimaginable prior to singing group (having COPD and yet being able to sing in public). The Methods section has been amended to clarify that practice at home was optional (line 144)

27. The 2nd sentence of last paragraph of discussion and the last 2 sentences of the conclusion should be removed and added as an appendix/online supplement.

We have removed these from the Discussion and Conclusions and included them as a supplemental file 'Setting up a singing group'

VERSION 2 – REVIEW

REVIEWER	Nick Hopkinson Imperial College, London
REVIEW RETURNED	28-Oct-2016

GENERAL COMMENTS	The paper has been improved but there are still some significant issues. This is a feasibility study and the temptation to over-claim about the outcomes still needs to be avoided. There is undue emphasis on small lung function changes and the discussion makes claims about generalizability which don't altogether stand up. In particular the abstract: 1) needs to report the outcome measures from the study fairly rather than selectively – from the protocol the primary (power calculation) outcome was the HADS score – either restrict to that only or report all the outcome measures from protocol at 6 and 12 months (i.e. FEV1, RV, 6MWD, CCQ.) or at 12 months only. 2) It needs to reflect how this group was selected. People, who had done PR isn't enough – it's people who have done PR and are in a maintenance program (I hadn't quite appreciated this when I read it previously) – also the protocol says recruitment within 12 months of PR but the mean time to enrolment was 1.2 years 3) Need to specify that hospital admission days was per year. Also In COPD raised RV is caused largely by premature airway closure rather than expiratory muscle weakness. Considering RV in isolation is anyway inappropriate. The fall in RV (120mls) is actually numerically smaller than the fall in TLC (160mls) – presumably the authors are not going to attribute the latter fall to the development of inspiratory muscle weakness? No other measure of hyperinflation improves (IC/TLC, RV/TLC, IC) nor vital capacity so this is not likely to be meaningful even if it isn't due to chance or to a change in patient technique doing the tests - which do rely to an extent on cooperation and breathing control. The fact that they are in an ongoing maintenance program makes the walk data hard to interpret – performance might continue to improve due to that without the singing. The methods need to make it clear that this is a study adding singing to maintenance exercise Discussion First para – reiterate that patients "who have completed PR and are in an ongoing exercise program" and take out references to generalizability. One year of lessons did not improve lung function – no parameter was better at 12 months. The improvement in lung volumes is not clinically significant (bottom page 14) Remove "pragmatic community context" – they're in a maintenance exercise program which sadly is not the case for the vast majority of COPD patients.

	The recruitment criteria (people with COPD) are also standard and not a particular strength of the study. The economic analysis is highly speculative and should be removed.

REVIEWER	Renae McNamara Prince of Wales Hospital Australia
REVIEW RETURNED	21-Oct-2016

GENERAL COMMENTS	Thank you for the opportunity to re-review this paper and to the authors for their revision and response to reviewers comments. Regarding previous reviewers comments and responses: Reviewer 1 Comment 11. It appears the sentence has not been amended in the revised manuscript. Reviewer 2 Comment 3. Although the mean time from PR completion to singing group enrolment was 1.2 years, the SD was 1.2, therefore some patients were attending the singing group in the 12 months following PR, so the effect of a reduction in hospital admissions cannot be totally attributed to the singing intervention and needs to be highlighted. Comment 15. Collection of comorbidities from the medical chart was not mentioned in the Methods – please include. Comment 21. The fact that patients also attended an exercise group for the duration of the trial is a major confounder, and changes the whole context of the paper and results. This is actually a trial of a community singing group in addition to a weekly maintenance exercise group following PR and should be referred to as such in the abstract, introduction, methods and conclusion (this was the study design, therefore it is not sufficient to only add this to the discussion). It is not possible to ascertain whether the results were due to the singing group, the exercise group or a combination of the two, and this should be discussed. Additional comments: - Figure 1 is very useful, however it is important to report the reasons why 112 patients declined to participate. It could be assumed that singing is not appealing to the large majority of people with COPD. - Table 2 footnotes: it is not necessary to repeat the information about the spirometric reference values, bronchodilator and 6MWT as this information was provided in the methods. - Discussion, lines 209-211: Please review “Factors favouring….”, as this information has been repeated 3 times in the discussion (lines 209—211, 232-233 and 283-284). This information is best placed towards the end of the discussion where limitations and strengths are discussed.

	- Discussion, line 239: reference number missing. - Discussion, line 275: it should be noted that previous RCT's have used singing classes of longer duration (1 hour) in people with COPD and they were well tolerated (Bonihla 2009, Lord 2010, Lord 2012).

VERSION 2 – AUTHOR RESPONSE

1 In the abstract need to report the outcome measures from the study fairly rather than selectively – from the protocol the primary (power calculation) outcome was the HADS score – either restrict to that only or report all the outcome measures from protocol at 6 and 12 months (i.e. FEV1, RV, 6MWD, CCQ.) or at 12 months only.

There is insufficient room to report all the 4 and 12 month outcomes in the abstract. We have reported the 'primary outcome' (acknowledging that this is a non-experimental study) first and feel it is reasonable to include 2 of the 3 statistically significant secondary outcome measures (avoiding mentioning the RV result at the request of Reviewer 1) also. Lines 39-43

2 In the abstract need to reflect how this group was selected. People, who had done PR isn't enough – it's people who have done PR and are in a maintenance program (I hadn't quite appreciated this when I read it previously) – also the protocol says recruitment within 12 months of PR but the mean time to enrolment was 1.2 years

We have added the phrase 'enrolled in a weekly community exercise group' in multiple places (Abstract, already in Methods, Discussion and Conclusions), at the risk of a reader finding this repetition rather tedious. Although the original plan was to recruit people within 1 year of PR in practice we couldn't refuse to accept people attending exercise classes who wanted to join the singing group but had attended PR between 1 and 3 years previously so they were also included.

3 Need to specify that hospital admission days was per year.

This result has been removed from the abstract and 'per year' added line 228

4 In COPD raised RV is caused largely by premature airway closure rather than expiratory muscle weakness. Considering RV in isolation is anyway inappropriate. The fall in RV (120mls) is actually numerically smaller than the fall in TLC (160mls) – presumably the authors are not going to attribute the latter fall to the development of inspiratory muscle weakness? No other measure of hyperinflation improves (IC/TLC, RV/TLC, IC) nor vital capacity so this is not likely to be meaningful even if it isn't due to chance or to a change in patient technique doing the tests - which do rely to an extent on cooperation and breathing control.

Although we believe that there is some chance that the statistically significant 4 month RV result is real, we accept that Type I error is the likeliest cause and have said so in the paper already (lines 221-223) and removed all other discussion about possible other causes. We have added detail from 2 other studies which largely support the contention that RV is not significantly affected by singing group intervention (lines 224-8).

5 The fact that they are in an ongoing maintenance program makes the walk data hard to interpret – performance might continue to improve due to that without the singing.

There is no evidence of which we are aware to support the contention that 6MWT distance improves with maintenance exercise following PR, although some conflicting data as to whether PR gains are maintained. We have addressed this issue in the Discussion already but have tried to clarify it further for this revision (Lines 251-259)

6 The methods need to make it clear that this is a study adding singing to maintenance exercise Discussion First para – reiterate that patients "who have completed PR and are in an ongoing exercise program" and take out references to generalizability. One year of lessons did not improve lung function – no parameter was better at 12 months.

Changes made – see point 2

7 The improvement in lung volumes is not clinically significant (bottom page 14)

Agreed. Removed.

8 Remove "pragmatic community context" – they're in a maintenance exercise program which sadly is not the case for the vast majority of COPD patients.

We respectfully disagree – all patients in our region who have completed a PR program are encouraged to enrol in a community exercise class. This doesn't alter the nature of the 'pragmatic community context' which we are using here to describe the nature of the singing group intervention – community hall, amateur singing group facilitator, low cost. Nevertheless, we have removed the specific words 'pragmatic community context' at the reviewer's request.

9 The recruitment criteria (people with COPD) are also standard and not a particular strength of the study.

Agreed, removed.

10 The economic analysis is highly speculative and should be removed.

Possibly. One of the questions that an interested reader may ask, thinking of setting up a singing group, is 'How will I pay for it'? The answer may be to either do as we did and fundraise or by convincing their hospital trust that there may be a pay off in money saved through fewer admissions – this was the basis for the simple cost equation. We have substantially simplified this but feel there is still a place for it in the Discussion (lines 231-233).

Reviewer 2
1 Comment 11. It appears the sentence has not been amended in the revised manuscript.

Now revised (line 83)

2 Comment 3. Although the mean time from PR completion to singing group enrolment was 1.2 years, the SD was 1.2, therefore some patients were attending the singing group in the 12 months following PR, so the effect of a reduction in hospital admissions cannot be totally attributed to the singing intervention and needs to be highlighted.

Agreed – point added to discussion (lines 244-5)

3 Comment 15. Collection of comorbidities from the medical chart was not mentioned in the Methods – please include.

This has been added to the Methods line 131

4 Comment 21. The fact that patients also attended an exercise group for the duration of the trial is a major confounder, and changes the whole context of the paper and results. This is actually a trial of a community singing group in addition to a weekly maintenance exercise group following PR and should be referred to as such in the abstract, introduction, methods and conclusion (this was the study design, therefore it is not sufficient to only add this to the discussion). It is not possible to ascertain whether the results were due to the singing group, the exercise group or a combination of the two, and this should be discussed.

See Reviewer 1, points 2 and 5 above plus our preamble to these detailed comments.

5 Figure 1 is very useful, however it is important to report the reasons why 112 patients declined to participate. It could be assumed that singing is not appealing to the large majority of people with COPD.

We did not collect this information and did not have ethics approval to do so.

6 Table 2 footnotes: it is not necessary to repeat the information about the spirometric reference values, bronchodilator and 6MWT as this information was provided in the methods.

Done

7 Discussion, lines 209-211: Please review "Factors favouring….", as this information has been repeated 3 times in the discussion (lines 209—211, 232-233 and 283-284). This information is best placed towards the end of the discussion where limitations and strengths are discussed.

We have simplified the Discussion to limit this repetition

8 Discussion, line 239: reference number missing.

Now added

9 Discussion, line 275: it should be noted that previous RCT's have used singing classes of longer duration (1 hour) in people with COPD and they were well tolerated (Bonihla 2009, Lord 2010, Lord 2012).

We have removed this sentence.

VERSION 3 – REVIEW

REVIEWER	Nick Hopkinson Imperial College, London
REVIEW RETURNED	10-Nov-2016

GENERAL COMMENTS	The authors have improved the description of the study so the context in which it took place. Is now clear. A few small things: Should make it clear that the Portuguese study includes RV data on only 4 out of 8 patients who completed and was not statistically significant. As a feasibility study the discussion should make some mention of recruitment. Is there no information at all on why only 1 in 7 took

	part? Even if there isn't the discussion should have a sentence to say so. P 14 The health economic speculation has been reworded rather than removed. This is an uncontrolled finding which is not statistically significant and should be removed. P 14 The high retention rate is not a strength of the study surely it's an outcome of the study?

REVIEWER	Renae McNamara Prince of Wales Hospital Australia
REVIEW RETURNED	24-Nov-2016

GENERAL COMMENTS	Thank you for the opportunity to re-review this paper and to the authors for their revision. Regarding authors responses: It is important to note that one of the main issues raised by both reviewers was not that patients were recruited from a maintenance exercise class following PR, but that participants were attending a maintenance exercise class at the same time as the singing group. There is no way to identify whether the results were due to the singing group, or the maintenance group, or a combination of the two groups together. The changes to the manuscript now more clearly describe this feature of the study. Specific comments (line numbers refer to the clean copy of the manuscript): 1. The conclusion of the abstract and manuscript refer to 'breathless adults with COPD' however a specific measurement of breathlessness has not been reported to justify this statement. Although the mMRC score is a component of the BODE index and the BODE index scores have been reported, the mMRC score cannot be determined from the presentation of the BODE index score. 2. Introduction, line 93: it is stated that the researchers were interested in the potential of singing to 'sustain the benefits of PR', however this question has not been adequately addressed in the results or discussion. Post-PR data was not available for the whole group of participants and a comparison of post-PR and 12 month follow-up data was not presented. 3. Discussion, line 256: there is the introduction of additional data (reference is made to 6MWT data) which was not presented in the results section. This data should be presented in the results section if a discussion is to occur. 4. Discussion, lines 259-262: the paragraph commencing on line 234 is regarding the strengths and limitations (line 248) of the study. The information about the PR nurse (lines 259-262) is not a limitation of the study – this information should sit within the methods. Additionally, the information in lines 262-264 is not a limitation, and if the results of the qualitative paper are included here, then they should be discussed in the context of the quantitative findings of this study. 5. Discussion, line 264-265: 'all participants received usual medical care so therapeutic changes may have affected some, but not all, of the participants' - I don't understand this sentence and why it has been placed here. Furthermore, why 'not all'? 6. Discussion, lines 269-272: Why has a discussion of physical

	activity related to ‘sustaining the benefits’ been introduced here? This information is not relevant to the quantitative results presented in this manuscript (no measure of physical activity was reported), and the reference to the qualitative findings have not been related to any of the quantitative findings. 7. Conclusion, lines 299-300: the final sentence ‘these results….’ are not a conclusion of the study. Minor comments: ABSTRACT - Line 33: ‘tests’ should be ‘test’ MANUSCRIPT - Results tables (1 & 2): please review the number of decimal places reported for each outcome as there is inconsistency within some outcomes (eg. sometimes 1, 2 and/or 3 decimal places reported), and consider the appropriateness of the number of decimal places reported for each outcome (eg. HADS results reported with 2 decimal places). - Results, line 191: suggest remove the words ‘strong evidence for’ – the strength or otherwise of results should sit within the discussion and not in the results. - Table 2, footnotes: please review the footnotes - FRC is not reported in this table; 6MWT, CCQ and HADS should be added. - Discussion, line 209: ‘is’ should be replaced with ‘was’ - Discussion, line 210: ‘are’ should be replaced with ‘were’ - Discussion, line 246: please review whether ‘not been offered’ should actually be ‘not attended’ Notes for editor: - Methods, line 114: should the ethics approval number be reported? - References: should the dates that electronic data was accessed be provided?

VERSION 3 – AUTHOR RESPONSE

Reviewer 1 (NH)
1 Should make it clear that the Portuguese study includes RV data on only 4 out of 8 patients who completed and was not statistically significant.

L 230 further clarification

2 As a feasibility study the discussion should make some mention of recruitment. Is there no information at all on why only 1 in 7 took part? Even if there isn’t the discussion should have a sentence to say so.

Further clarification in Methods L108-109 and Discussion L209-212

3 P 14 The health economic speculation has been reworded rather than removed. This is an uncontrolled finding which is not statistically significant and should be removed.

Speculation removed L234-239

4 P 14 The high retention rate is not a strength of the study surely it’s an outcome of the study?

Removed L240.

Reviewer 2

1 The conclusion of the abstract and manuscript refer to 'breathless adults with COPD' however a specific measurement of breathlessness has not been reported to justify this statement. Although the mMRC score is a component of the BODE index and the BODE index scores have been reported, the mMRC score cannot be determined from the presentation of the BODE index score.

Although we disagree with this point – in the Methods (lines 109-111), all patients enrolled in PR have mMRC dyspnoea scale of 2 or more and, by definition, have at least symptoms of breathlessness on exertion – we have removed the reference to 'breathless' in the abstract and conclusions

2 Introduction, line 93: it is stated that the researchers were interested in the potential of singing to 'sustain the benefits of PR', however this question has not been adequately addressed in the results or discussion. Post-PR data was not available for the whole group of participants and a comparison of post-PR and 12 month follow-up data was not presented.

Yes, we are interested in sustaining the benefits of PR and the statement in the Introduction is appropriate in representing that interest as a general focus of our research. We accept that this manuscript is far from able to answer the question definitively. As pointed out throughout the manuscript, this is a feasibility study

3 Discussion, line 256: there is the introduction of additional data (reference is made to 6MWT data) which was not presented in the results section. This data should be presented in the results section if a discussion is to occur.

It is already in the results section (line 164 -166 and Table 1). The information in line 164 and the discussion was specifically requested by both the reviewers.

4 Discussion, lines 259-262: the paragraph commencing on line 234 is regarding the strengths and limitations (line 248) of the study. The information about the PR nurse (lines 259-262) is not a limitation of the study – this information should sit within the methods. Additionally, the information in lines 262-264 is not a limitation, and if the results of the qualitative paper are included here, then they should be discussed in the context of the quantitative findings of this study.

Agree. We have rearranged this information with 'limitations' following 'strengths' (now lines 244-248) followed by 'factors that may have influenced the results' (lines 251-273). Regarding the PR nurse discussion, this was another expanded statement requested by one of the reviewers. The decision about putting the information about the specific role of the PR nurse in the Methods is another 'style' issue – we have chosen to include it here where it is directly relevant to the question of the PR nurse potentially influencing the results. We would prefer to avoid repeating the information in the Methods section as well.

5 Discussion, line 264-265: 'all participants received usual medical care so therapeutic changes may have affected some, but not all, of the participants' - I don't understand this sentence and why it has been placed here. Furthermore, why 'not all'?

This statement was inserted at revision 1 at the request of one of the reviewers because it was felt that therapeutic changes, which we were unable to quantify (not collected) could have affected the results and was therefore a potential limitation of the study. We have removed the 'but not all'. L272-273

6 Discussion, lines 269-272: Why has a discussion of physical activity related to 'sustaining the benefits' been introduced here? This information is not relevant to the quantitative results presented in

this manuscript (no measure of physical activity was reported), and the reference to the qualitative findings have not been related to any of the quantitative findings.

We are simply trying to tie together some of the potential strands as is the purpose of the Discussion section – if singing group is effective, why? PR is thought to be effective because it promotes physical activity. Singing group participants did more physical activity just getting to the singing group, according to their reports in the qualitative study. Perhaps these are connected?

7 Conclusion, lines 299-300: the final sentence 'these results….' are not a conclusion of the study.

This is a 'style' issue. Some journals are happy to have this statement in the Conclusions rather than the Discussion. This is a feasibility study, looking forward to a definitive RCT, therefore it could be argued that this statement is an appropriate 'conclusion' of the study. We looked at moving the statement to the end of the Discussion but feel it makes more sense to be in the Conclusion and would welcome advice from the Editor about the placement of the statement.

8 ABSTRACT
- Line 33: 'tests' should be 'test'

Changed

9 Results tables (1 & 2): please review the number of decimal places reported for each outcome as there is inconsistency within some outcomes (eg. sometimes 1, 2 and/or 3 decimal places reported), and consider the appropriateness of the number of decimal places reported for each outcome (eg. HADS results reported with 2 decimal places).

Reporting of one p value to 3 decimal places (HADS anxiety in Table 2) changed to 2 decimal places. Our statistician is comfortable with the reporting of the other results.

10. Results, line 191: suggest remove the words 'strong evidence for' – the strength or otherwise of results should sit within the discussion and not in the results.

Our statistician disagrees. This is a standard statistical statement reflecting the statistical test result and is appropriate in the Results section.

11. Table 2, footnotes: please review the footnotes - FRC is not reported in this table; 6MWT, CCQ and HADS should be added.

Done

12. Discussion, line 209: 'is' should be replaced with 'was'

Done

13. Discussion, line 210: 'are' should be replaced with 'were'

Done

VERSION 4 – REVIEW

REVIEWER	Nick Hopkinson NHLI Imperial College London UK
REVIEW RETURNED	01-Dec-2016

GENERAL COMMENTS	The revised version is fine.