PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Electroacupuncture as a complement to usual care for patients with non-acute low back pain after back surgery: A pilot randomised controlled trial
AUTHORS	Heo, In; HWANG, MAN-SUK; Hwang, Eui-Hyoung; CHO, JAE HEUNG; Ha, In-Hyuk; Shin, Kyung-Min; Lee, Jun-Hwan; Kim, Nam-Kwen; Son, Dong-Wuk; Shin, Byung-Cheul

VERSION 1 – REVIEW

REVIEWER	María Villarreal Santiago Physio Villarreal México
REVIEW RETURNED	12-Aug-2017

GENERAL COMMENTS	Does all the participants had the same back surgery procedure? For how long they have been in pain before having the surgery? Is it possible for future study to take away the drug therapy from the patients that will received EA. Regarding the back pain education, what it is the information that you provide to the patients? It is important to highlight that pain is multifactorial and the way you explain pain to the patients it is really important for their fully recovery. Finally I recommend to check the literature about of the neuroscience of pain and include it on your future research.

REVIEWER	Emma Jonsson Quantify Research, Sweden
REVIEW RETURNED	26-Sep-2017

GENERAL COMMENTS	1. Page 6 line 34: The wording should be that acupuncture is cost-effective compared with usual care. 2. Page 11 line 20: Suggest to add which types of visits or week number instead of assessment number. 3. Page 12 line 57: Please add more details on the calculations and assumptions for the calculation of sample size for future trial, rather than just referring to the statistical program. 4. Page 15 line 27: Remove "And". 5. Page 16 line 34: The conclusion is made that EA+UC is more effective than UC alone based on e.g. that there was a numerical difference in VAS score between the groups. The CIs of the differences are clearly overlapping and the authors also state there was no statistical difference between the groups. Therefore, it is not very convincing that these results support the conclusion of EA+UC being more effective. Please explain.

REVIEWER	Dr Simon Skene University College London, UK
REVIEW RETURNED	22-Oct-2017

GENERAL COMMENTS	The authors have described a trial completed according to a protocol previously published in BMJ Open. The trial is described in the introduction as a “pilot feasibility to compare the effectiveness of EA in combination with UC to UC alone”, where “the primary purpose of the study was to explore whether EA in combination with UC is beneficial in patients with non-acute pain and dysfunction after back surgery”. It is not exactly clear whether the trial is (a) a feasibility; to determine whether the intervention can be effectively delivered before an RCT could be undertaken, (b) a pilot; to check the elements of the protocol before rolling out into a larger study, or to gather sufficient data to power a definitive study, or (c) a ‘proof-of-concept’; to check additionally whether there is sufficient signal of efficacy to justify further investigation of the intervention The primary outcome measure (and purpose) suggests the latter ‘proof of concept’ model, whereas the discussion states that “the purpose of the pilot study was to confirm feasibility of such a study rather than determine the effectiveness of EA”. The investigators could make clearer the purposes of the study at the outset, and whether there were any a priori metrics in each these areas which would warrant a recommendation to proceed to a definitive trial. The formal sample size suggests a clinically significant change (>=20mm on VAS) was not achieved, although the direction of effect is positive in favour of EA. The paper refers to the protocol submission for the sample size calculation, but it should be possible to check the details without reference, as a way of judging the success of the study. The details of all assumed parameters should be in this paper. Notwithstanding comments in the discussion about cultural experience and ‘biases’ concerning acupuncture in Korea, consideration of acupuncture (without electric stimulation) as the control of interest rather than usual care versus EA would be useful up front. ie Did UC preclude the use of traditional acupuncture? This is not clear. Much is made about the blinded assessments, but it should be noted that the three main outcome measures, VAS, ODI and EQ5D are all patient reported. Did the protocol allow for replacement of patients who consented but withdrew before treatment? It is not clear whether these were randomised. If so, it would be useful to know how these 8 were distributed between arms. Likewise, it would be important to know at what stage the patients who withdrew after treatment did so, since this may affect the judgement of LOCF as an appropriate method for dealing with missing data. The dropout/replacement statistics should be discussed in assessing the feasibility of a future trial, as should the delivery of the

	intervention. The analysis does not conform fully to the details given in the protocol paper, which describe both ITT and per-protocol analysis and subgroup analyses based on surgery type etc, and responder analysis. Was a statistical analysis plan prepared in advance of the analysis of unblinded data. It would be useful to lodge this as supplemental material alongside a submission of results. P-values should not be given for differences on baseline demographics, which by difference would occur only to chance. Should the primary differences in scores not be assessed (between groups) from week 0 (baseline/randomisation) and week 4, rather than between weeks 1 and 4? A more powerful analysis would consider the differences between groups at week 4 adjusted for baseline. The data for latter assessments at 4 and 8 weeks following treatment are not summarised here. If not in the paper these should certainly be included as supplementary material to give a complete picture of the study. It would be natural in planning a larger trial to consider the timing of assessments and outcomes. Ie is any benefit of EA retained? Whilst it is usual to base the sizing of a future trial on the observed standard deviation from 'pilot' data, it would be sensible to account for uncertainty in this estimate (particularly given the small sample size). Additionally, the MCID should be informed by clinical judgement, not an observed difference in a pilot. Consideration by the authors of these points would make for a more credible paper in recommending whether a future trial is indeed feasible, and justified, allowing readers to consider more fully for themselves the true value of this work.

VERSION 1 – AUTHOR RESPONSE

Editorial Comments and Requests:

Please clarify why there is such a large gap between study completion (2015) and submission of this paper.

Response: First, patient recruitment was slower than expected, so the research period was prolonged. In addition to our studies on the effectiveness of the electroacupuncture for low back pain after surgery, the qualitative research and the economic evaluation studies conducted by other researchers were conducted concurrently. Because the statistical analysis was performed after all data collection was completed, our findings have been postponed unfortunately. We added these contents to the "Study design".

Abstract: the study does not appear to be testing feasibility. What are the feasibility outcomes? It's described as a pilot trial in your protocol so it is not clear why you are talking about feasibility here.

Response: Thank you for your evaluation. This study is a pilot study for estimating appropriate powered sample size for future large pragmatic RCT of the comparative effectiveness of electroacupuncture for low back pain after back surgery. Our main aim was to test the appropriateness of our study design to know the feasibility of study design and adequate sample size

in study condition, therefore we used the word of ‘feasibility’. We revised the manuscript based on these points of view.

Along with your revised manuscript, please provide a completed copy of the CONSORT extension for pilot studies. This can be found in Table 2 of the following paper:
http://www.bmj.com/content/355/bmj.i5239.

Response: Thank you for your suggestion. We will provide a completed copy of the CONSORT extension for pilot studies.

Thanks to all reviewers for their valuable evaluation.

Reviewers' comments to Author:

Reviewer: 1
Reviewer Name: María Villarreal Santiago
Institution and Country: Physio Villarreal, México

Does all the participants had the same back surgery procedure?
For how long they have been in pain before having the surgery?

Response: Thank you for your comment. Our Case report form included this checklist to know which surgery patients had, but when most of patients asked about the types of surgery did not remember correctly, so we had difficulty obtaining information about it. And the duration of pain before having the surgery is found to be insufficient, therefore, we failed to conduct subgroup analysis by the type of surgery or duration of pain. We will reflect this point in the main trial following this pilot trial. We added these contents to the "Disscussion".

Is it possible for future study to take away the drug therapy from the patients that will received EA. Regarding the back pain education, what it is the information that you provide to the patients? It is important to highlight that pain is multifactorial and the way you explain pain to the patients it is really important for their fully recovery.

Response: Thank you for your comment. In many professional conferences, it was difficult to completely rule out medication when considering the realistic aspects of pain management. As also this pilot trial was pragmatic comparative effectiveness RCT, therefore, we reflected real world condition in clinical current status. Therefore we included drug therapy for reflecting current use of medication. We added these contents to the "Disscussion". And We provided physiology, pathology, and epidemiology of pain after back surgery in the form of a brochure, besides Korean medical doctors educated posture and exercise appropriate for back pain management in a 15-min face-to-face education session. In order to supplement the education conducted in the pilot study, we plan to use additional video materials in the main trial following the pilot trial.

Finally I recommend to check the literature about of the neuroscience of pain and include it on your future research.

Response: Thank you for your suggestion.We have revised manuscript for your suggestion by reflecting related references in "Discussion"

Thanks you for your valuable evaluation.

Reviewer: 2
Reviewer Name: Emma Jonsson
Institution and Country: Quantify Research, Sweden

1. Page 6 line 34: The wording should be that acupuncture is cost-effective compared with usual care.

Response: Thank you for your comment.We have revised manuscript for your suggestion.

2. Page 11 line 20: Suggest to add which types of visits or week number instead of assessment number.

Response: Thank you for your suggestion.We have revised manuscript for your suggestion.

3. Page 12 line 57: Please add more details on the calculations and assumptions for the calculation of sample size for future trial, rather than just referring to the statistical program.

Response: Thank you for your comment. We revised the manuscript based on the point of view. Briefly, the mean difference in the pain VAS for LBP between the experimental and control groups was 14.02 mm. And the SD between the two groups was 22.12, based on results. When a two-tailed test with a test power of 80% (β error) and a significance level of 5% (α error) was applied to the formula for calculating sample size, the number of participants required for each group was found to be 40. Considering a dropout rate of 25% and a 1:1 allocation ratio, the total sample size was calculated to be 54 per group. This calculated sample size is the same as the result from G*Power program. We added these contents to the "Statistical analysis".

4. Page 15 line 27: Remove "And".

Response: We have revised manuscript for your suggestion.

5. Page 16 line 34: The conclusion is made that EA+UC is more effective than UC alone based on e.g. that there was a numerical difference in VAS score between the groups. The CIs of the differences are clearly overlapping and the authors also state there was no statistical difference between the groups. Therefore, it is not very convincing that these results support the conclusion of EA+UC being more effective. Please explain.

Response: Thank you for your comment on this point. This study is a pilot study for estimating appropriate sample size for future large pragmatic RCT of the comparative effectiveness of electroacupuncture for low back pain after back surgery. The results of pilot study in VAS score should only be used for sample size estimation. We revised the manuscript based on this point of view and reflected this in discussion.

Thanks you for your valuable evaluation.

Reviewer: 3
Reviewer Name: Dr Simon Skene
Institution and Country: University College London, UK

The authors have described a trial completed according to a protocol previously published in BMJ Open.

The trial is described in the introduction as a “pilot feasibility to compare the effectiveness of EA in combination with UC to UC alone”, where “the primary purpose of the study was to explore whether EA in combination with UC is beneficial in patients with non-acute pain and dysfunction after back surgery”.

It is not exactly clear whether the trial is

(a) a feasibility; to determine whether the intervention can be effectively delivered before an RCT could be undertaken,
(b) a pilot; to check the elements of the protocol before rolling out into a larger study, or to gather sufficient data to power a definitive study, or
(c) a ‘proof-of-concept’; to check additionally whether there is sufficient signal of efficacy to justify further investigation of the intervention

The primary outcome measure (and purpose) suggests the latter ‘proof of concept’ model, whereas the discussion states that “the purpose of the pilot study was to confirm feasibility of such a study rather than determine the effectiveness of EA”.
The investigators could make clearer the purposes of the study at the outset, and whether there were any a priori metrics in each these areas which would warrant a recommendation to proceed to a definitive trial. The formal sample size suggests a clinically significant change (>=20mm on VAS) was not achieved, although the direction of effect is positive in favour of EA.

Response: Thank you for your comment. Like your comment, this study is a pilot study for estimating appropriate sample size for future large pragmatic study of the comparative effectiveness of electroacupuncture for low back pain after back surgery. Although the results of VAS score, primary outcome, were not clinically significant change, it is necessary to conduct additional larger scale trial using estimated powered sample size. So we revised the manuscript based on this point of view.

The paper refers to the protocol submission for the sample size calculation, but it should be possible to check the details without reference, as a way of judging the success of the study. The details of all assumed parameters should be in this paper.

Response: Response: Thank you for your comment. This was reflected in " Statistical analysis", and this point was also answered in the above comment. Briefly, the mean difference in the pain VAS for LBP between the experimental and control groups was 14.02 mm. And the SD between the two groups was 22.12, based on results. When a two-tailed test with a test power of 80% (β error) and a significance level of 5% (α error) was applied to the formula for calculating sample size, the number of participants required for each group was found to be 40. Considering a dropout rate of 25% and a 1:1 allocation ratio, the total sample size was calculated to be 54 per group. This estimated sample size is the same as the result from G*Power program. We revised the manuscript based on the point of view.

Notwithstanding comments in the discussion about cultural experience and ‘biases’ concerning acupuncture in Korea, consideration of acupuncture (without electric stimulation) as the control of interest rather than usual care versus EA would be useful up front. ie Did UC preclude the use of traditional acupuncture? This is not clear.

Response: Thank you for your comment. Before the trial process, we discussed many this point with experts for designing the trial. Also, this study is a pilot study for estimating appropriate sample size for future large pragmatic study of the comparative effectiveness of electroacupuncture for low back pain after back surgery. The basic purpose of this design is to confirm the effectiveness of another treatment, which is the electroacupuncture, for the treatment of low back pain after surgery using conventional western medicine so called usual care. From the viewpoint of Korean medicine doctors, it is considered that electroacupuncture is a part of the acupuncture treatment in a broad sense, and the electroacupuncture treatment is more suitable for the pain treatment than the acupuncture without electric stimulation. Therefore, we conducted clinical trial of this design despite some of the mentioned limitations. We added these contents to the "Discussion".

Much is made about the blinded assessments, but it should be noted that the three main outcome measures, VAS, ODI and EQ5D are all patient reported.

Response:Thank you for your comment. As you point out, our three main outcome measures are all patients reported. This can serve as a limitation from the outcomes used, although using assessor blinding as possible as we could. We added this point in the "Discussion".

Did the protocol allow for replacement of patients who consented but withdrew before treatment? It is not clear whether these were randomised. If so, it would be useful to know how these 8 were distributed between arms. Likewise, it would be important to know at what stage the patients who withdrew after treatment did so, since this may affect the judgement of LOCF as an appropriate method for dealing with missing data.

Response: Thank you for your point. The current analysis did not include data from 8 patients (5 in the EA plus UC group and 3 in the UC alone group), who did not receive any treatment after randomization and who withdrew their consent. They were randomly assigned at the same time as the study participation agreement, but were excluded from the ITT analysis because they withdrew their consent for the next visit without receiving any treatment for their family reasons. In addition, 6 patients whose participation in continuous research was limited due to the reason of moving or working etc., withdrew their consent while the treatment was in progress, and 3 persons who missed important visit schedules were dismissed for violating the protocol.These nine data were included in the LOCF analysis according to pre-published protocol as planned.

The dropout/replacement statistics should be discussed in assessing the feasibility of a future trial, as should the delivery of the intervention.

Response: In order to overcome the possible problems related to withdrawal of consent before the treatment progress, we plan to adjust the timing of randomization and the timing of initiation of treatment to reflect the findings in the subsequent confirmative trial protocol. We added this point in the "Discussion"

The analysis does not conform fully to the details given in the protocol paper, which describe both ITT and per-protocol analysis and subgroup analyses based on surgery type etc, and responder analysis. Was a statistical analysis plan prepared in advance of the analysis of unblinded data. It would be useful to lodge this as supplemental material alongside a submission of results.

Response: Thank you for your comment. In our study, not only researchers but also statisticians participated in the planning stage and developed a statistical analysis plan and published it as a protocol. Because of the nature of the treatment method used in the study, the evaluator and the analyst were blinded and the results were analyzed according to the protocol.
There were many dropouts in this pilot trial, per-protocol analysis could not be conducted due to less than the estimated sample size, the number of subjects required for each group was 16. And the subgroup analyses based on surgery type etc were also could not be conducted due to lack of collected information because of the patients could not remember about the past condition. For this reason, the results were reported based on the ITT analysis, and the number of samples of the follow-up study was estimated. Problems such as dropout and subgroup analysis will be reflected in the protocol of the follow-up study so that they can be overcome. We added this point in the "Discussion"

P-values should not be given for differences on baseline demographics, which by difference would occur only to chance.

Response: We have revised manuscript for your comment. Thank you.

Should the primary differences in scores not be assessed (between groups) from week 0 (baseline/randomisation) and week 4, rather than between weeks 1 and 4? A more powerful analysis would consider the differences between groups at week 4 adjusted for baseline. The data for latter assessments at 4 and 8 weeks following treatment are not summarised here. If not in the paper these should certainly be included as supplementary material to give a complete picture of the study. It would be natural in planning a larger trial to consider the timing of assessments and outcomes. Ie is any benefit of EA retained?

Response: Thank you very much for the keen comment of the reviewer. Based on your comments, we reviewed Table 2 and found some incorrect entries. In rewriting Table 2, we have included the baseline and the results of the 4th, 8th, and 12th weeks in the light of your opinion. Based on the results of the baseline and primary endpoint (assessment 10) as protocol, we estimated the number of samples in the follow-up study. This is the same as previously reported.

Whilst it is usual to base the sizing of a future trial on the observed standard deviation from ‘pilot’ data, it would be sensible to account for uncertainty in this estimate (particularly given the small sample size). Additionally, the MCID should be informed by clinical judgement, not an observed difference in a pilot.

Response: Thank you for your comment.We have confirmed that in the estimation of the sample sizes of the follow-up trial estimated from the pilot study, it is possible to have uncertainties due to a small sample size and this is reflected in the manuscript. In addition, the MCID from the other study was confirmed and compared with the results of pilot study. We added this point in the "Discussion"

Consideration by the authors of these points would make for a more credible paper in recommending whether a future trial is indeed feasible, and justified, allowing readers to consider more fully for themselves the true value of this work.

Response: Thank you for your constructive criticism and comment. Although we have some limitations, we think that it is necessary to confirm the effectiveness of electroacupuncture conducting a large-scale trial using the estimated powered sample size through the results of pilot study.

Thanks you for your valuable evaluation.

VERSION 2 – REVIEW

REVIEWER	Emma Jonsson Quantify Research
REVIEW RETURNED	16-Dec-2017

GENERAL COMMENTS	The authors responded sufficiently to my comments. I have no further comments.

REVIEWER	Dr Simon Skene University College London, UK
REVIEW RETURNED	02-Jan-2018

GENERAL COMMENTS	The authors have adressed most of the substantive points raised in my previous review, and the manuscript is much clearer now about the study's aims as a pilot. There are however, one or two remaining statistical issues which if resolved would provide further clarity. Sample size of pilot. This is written as an experiment to determine superiority of EA+UC, whereas the purpose is to estimate the sample size for a future RCT. It would be better to demonstrate the precision of the estimated standard deviation from a sample of 32 (given 20% loss-to-follow-up). Otherwise, the sample size description should include the effect size and assumed standard deviation rather than give the formula. Sample size is 40 in total (20 per arm). Future RCT The sample size of the future RCT should be based on the MCID of lower back pain VAS which is the primary outcome. Elsewhere in the article this is quoted as 22.5 (Discussion), so 14 from the pilot study would not seem to be a clinically relevant target.s suggested previously, there will be a considerable increase in power if a future comparison adjusts for baseline score, and this could be reflected in the calculation. Statistical reporting p-values should be removed from the baseline characteristics, Table 1, as any differences noted here are due to chance (because of randomisation).Similarly in Table 2. In Table 2 it would be better to simply give the means (standard deviations) of the scores at the various timepoints, and the mean difference between groups with confidence interval for the between-group comparisons. The p-values for the latter may be useful, but elsewhere the reliance on p-values is not helpful given the focus of the pilot on estimation.

VERSION 2 – AUTHOR RESPONSE

Editorial Comments and Requests:
- Please remove the following on page 7: “Because the statistical analysis was performed after all data collection was completed, our reporting have been postponed unfortunately” One would expect

the statistical analysis to be carried out after the data is collected so it is not clear why this would delay the publication of the study's results by two years. We just needed an explanation in the rebuttal letter. The reason(s) do not need to be incorporated into the manuscript.

Response: Thank you for your comment. We have removed the sentence for your suggestion. As noted earlier, additional time was required to collect economic evaluation data and qualitative research data, and after all of these data were collected, analysis was delayed and submission was delayed.

- The quality of English still needs improving in places. Please thoroughly proofread the paper, paying particular attention to the revised sections e.g. page 18: "This can serve as a limitation from the outcomes used, although using assessor blinding as possible as we could." (what is meant by "..using assessor blinding as possible as we could?") Please proofread the paper one more time. We recommend consulting a native English speaker if possible.

Response: Thank you for your comment. We have revised manuscript for your suggestion.

Reviewers' comments to Author:

Reviewer: 3
Reviewer Name: Dr Simon Skene
Institution and Country: University College London, UK
Competing Interests: None declared

The authors have adressed most of the substantive points raised in my previous review, and the manuscript is much clearer now about the study's aims as a pilot.

There are however, one or two remaining statistical issues which if resolved would provide further clarity.

Sample size of pilot.
This is written as an experiment to determine superiority of EA+UC, whereas the purpose is to estimate the sample size for a future RCT. It would be better to demonstrate the precision of the estimated standard deviation from a sample of 32 (given 20% loss-to-follow-up).Otherwise, the sample size description should include the effect size and assumed standard deviation rather than give the formula. Sample size is 40 in total (20 per arm).

Response: As mentioned in the manuscript, the aim of this pilot study was to estimate sample size of a large pragmatic study using the mean difference and standard deviation. As there was no same trial with our design of RCT, we calculated the sample size of the pilot study using the mean difference (20) and standard deviation (19) derived from othersimilar studies. Because those studies were not exactly the same our trial in terms of the patient, method of treatment and design of trial, they differed from the mean difference (14.02) and standard deviation (22.12) of pilot study. These differences were caused the underpowered results of pilot study that is a reason the large-scale follow-up study is needed. We revised the manuscript based on these points of view.

Future RCT
The sample size of the future RCT should be based on the MCID of lower back pain VAS which is the primary outcome. Elsewhere in the article this is quoted as 22.5 (Discussion), so 14 from the pilot

study would not seem to be a clinically relevant target.s suggested previously, there will be a considerable increase in power if a future comparison adjusts for baseline score, and this could be reflected in the calculation.

Response: The MCID is the smallest change in an outcome that a patient would identify as important. The change of VAS in EA+UC was over the MCID from VAS change of FBSS patients and the effect size by the mean difference (14.02) and standard deviation (22.12) of the two groups means medium-sized effect. The aim of follow-up study is to compare the effectiveness of electroacupuncture (EA) with usual care (UC) versus UC alone on pain control and functional improvement after back surgery. Therefore, we think that it is reasonable to use the mean difference and the standard deviation confirmed in the pilot study results, not the MCID, for the sample size estimation of the follow-up study .So we calculated the sample size of the follow-up study using the G * power program using the mean difference between the two groups (14.02) and the standard deviation (22.12) according to the pre-planned protocol. And baseline calibration for power increase is reflected in the follow-up study, and I am grateful for the advice.

Statistical reporting
p-values should be removed from the baseline characteristics, Table 1, as any differences noted here are due to chance (because of randomisation).Similarly in Table 2.

In Table 2 it would be better to simply give the means (standard deviations) of the scores at the various timepoints, and the mean difference between groups with confidence interval for the between-group comparisons. The p-values for the latter may be useful, but elsewhere the reliance on p-values is not helpful given the focus of the pilot on estimation.

Response: Thank you for your comment. We have revised manuscript for your suggestion.

VERSION 3 – REVIEW

REVIEWER	Professor Simon Skene University of Surrey, UK
REVIEW RETURNED	25-Mar-2018

GENERAL COMMENTS	The authors have addressed all my previously stated points, and the manuscript is much improved. There are two areas in the revised response that I think could be simplified further in a minor way to ensure total clarity. 1. Estimating sample size of a future trial. Again, there is no need to give the formula, since the text allows for checking the calculation by anyone with knowledge of statistics. I'd suggest simply (from line 55 page 17) "..., based on ITT analysis. On this basis, using the G*Power program, 40 participants per group would be required. Allowing for a dropout rate of 25%, a total of 108 participants (54 per group) would need to be recruited." 2. The MCID point is well addressed, but I feel the following text (or similar) would be more reasoned, replacing the final two paragraphs (beginning on lines 30 and 43) of page 18. "The observed change in VAS scores in the EA plus UC group

	(23.11) is greater than the minimum clinically important difference (MCID) value (22.50 in low back pain) reported in a previous study, [38] and the mean difference (14.02) and standard deviation (22.12) of the two groups indicates a medium-sized effect, justifying the need for a larger scale follow-up study. The pilot study was underpowered, the sample size being based on the mean difference (20) and standard deviation (19) derived from other similar studies. However, those studies differed from our trial in terms of patients, methods of treatment and study design. It follows that the sample size for our future RCT based on a similar protocol to the pilot study should be calculated using our observed parameters so that a future study would be conservatively powered for a meaningful effect."

VERSION 3 – AUTHOR RESPONSE

Response to reviewers' comments

Manuscript ID bmjopen-2017-018464 entitled "Electroacupuncture as a complement to usual care for patients with non-acute low back pain after back surgery: A pilot randomised controlled trial"

We would like to appreciate the reviewers for his/her constructive comments, which have helped us to improve the manuscript. We make an earnest effort to revise the manuscript in accordance with the reviewer's suggestions.

Editorial Comments and Requests:

Please carefully proofread the paper one more time before resubmitting. In the abstract, "groupata1:1 ratio" should be "group at a 1:1 ratio".

Response: Thank you for your comment. We have revised manuscript for your suggestion.

Reviewers' comments to Author:

Reviewer: 3

Reviewer Name: Professor Simon Skene

Institution and Country: University of Surrey, UK

Competing Interests: None declared

The authors have addressed all my previously stated points, and the manuscript is much improved.

There are two areas in the revised response that I think could be simplified further in a minor way to ensure total clarity.

1. Estimating sample size of a future trial. Again, there is no need to give the formula, since the text allows for checking the calculation by anyone with knowledge of statistics. I'd suggest simply (from line 55 page 17)

"..., based on ITT analysis. On this basis, using the G*Power program, 40 participants per group would be required. Allowing for a dropout rate of 25%, a total of 108 participants (54 per group) would need to be recruited."

Response: Thank you for your comment. We have revised manuscript for your suggestion.

2. The MCID point is well addressed, but I feel the following text (or similar) would be more reasoned, replacing the final two paragraphs (beginning on lines 30 and 43) of page 18.

"The observed change in VAS scores in the EA plus UC group (23.11) is greater than the minimum clinically important difference (MCID) value (22.50 in low back pain) reported in a previous study, [38] and the mean difference (14.02) and standard deviation (22.12) of the two groups indicates a medium-sized effect, justifying the need for a larger scale follow-up study.

The pilot study was underpowered, the sample size being based on the mean difference (20) and standard deviation (19) derived from other similar studies. However, those studies differed from our trial in terms of patients, methods of treatment and study design. It follows that the sample size for our future RCT based on a similar protocol to the pilot study should be calculated using our observed parameters so that a future study would be conservatively powered for a meaningful effect."

Response: Thank you for your comment. We have revised manuscript for your suggestion.