RE: BMJ.2016.036958

Dear Dr. Loder:

We have received, and reviewed, the Committees' comments on our manuscript: Low intensity pulsed ultrasound for bone healing: a systematic review of randomised controlled trials.

We thank the BMJ editorial board and the reviewers for their many valuable comments which helped us to improve the quality of the manuscript. We have addressed comments in the order that they appear.

Comments by the Committee:

1. We wondered about the decision to exclude trials published only as abstracts without making any attempt to contact authors and determine whether they had been published.

Reply: We did attempt to contact the authors when we found no full text through our search, but did not make this clear in our original submission. We have now included the following statement in our eligibility criteria:

"We excluded trials published only as protocol or abstract if attempts to acquire final results from investigators were unsuccessful."

and updated our results section, which now reads as follows:

"… Our registry search yielded four protocols of potentially eligible RCTs; one was discontinued due to slow recruitment (ISRCTN90844675, personal communication, outcome data not available yet), one manuscript is under peer-review (NCT00744861, personal communication: "no difference between the control group and the ultrasound group"), one is completed but unpublished (JPRN-UMIN000002005, no response from investigators), and the last is still ongoing (NCT02383160). Attempts to acquire the full text of another potentially eligible RCT,(51) reported in a recent systematic review,(11) were unsuccessful."

We carefully checked our search results again to make sure the numbers and labels reported in the flow chart are correct and clear. We did not identify any eligible conference abstracts without subsequent publication.

2. We are concerned about comments from several reviewers that conclusions may be too broad given that the high quality trials did not include all types of fracture. This seems important and we wonder if the conclusions should be less sweeping.

Reply: Although baseline healing time differs by size of bone and the site of fracture, and there are likely differences in aetiology as mentioned by one of the reviewers, the effect of low intensity pulsed ultrasound compared with control on the time to fracture healing is therefore likely to be similar. We reasoned that pooling trials with the same intervention directed towards fractures of different bones would increase the generalisability of our results, but did formally explore clinical presentation as a possible source of heterogeneity for our pooled estimates. We found no subgroup effect based on clinical presentation, which increases our confidence in pooling across different types of bone lesions.

In our discussion, we expand on why we think it is reasonable to generalize the results to underrepresented populations in the absence of trustworthy evidence. However, we accept the reviewers' concerns and have rephrased the conclusions:

Abstract: "Based on moderate to high quality evidence from studies conducted in patients with fresh fracture, LIPUS does not improve patient-important outcomes and probably has no effect on radiographic bone healing."

Main text: "In conclusion, moderate to high quality evidence demonstrates that LIPUS fails to […]The evidence applies directly to patients with fresh fractures and indirectly to children and other underrepresented populations, particularly those with non-union, for which no trustworthy direct evidence exists."

An important point, we think, is that given the likely applicability to other populations, the burden of proof should now be on those inclined to use LIPUS in populations in which it has not been extensively tested. This, it seems to us, strengthens the tentative (i.e. pending direct study) applicability to other populations.

3. Low risk of bias is defined as no limitation for any of the 5 areas given in table 2 plus less than 20% loss to follow up. Since the conclusions depend heavily on this dichotomy of studies into low and high risk of bias groups this is a crucial area. Some trials do poorly in all 5 areas but others in only a few. Have you really considered all of these factors on an equal footing? How was the 20% level for the dichotomy determined? Can you reassure us that it was not post hoc? Kristiansen, Lubbert and Heckman all have no limitations, but Lubbert is considered low risk with 16% loss to follow up, Heckman high risk with 31% loss and Kristiansen high risk since loss was unclear although assumed to be 0%. No distinction is made for which or how many of the limitations are not met. The TRUST trial is categorised as low risk as the % loss for the 3 main outcomes are 19%, 11%, 9%. Some sensitivity analyses should be presented to determine the extent to which choice of cut-point (20%) has influences the conclusions.

We note that the dichotomisation into low vs. high risk of bias was prespecified in the register (https://www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42016050965): "Specifically, we will categorize trials at high risk of bias if they fail to report concealment of allocation, blinding of patients, caregivers, data collectors or outcome assessors, or report >20% loss to follow-up."

Reply: We apologize for an error in Table 2. Kristiansen had in fact 28% loss to follow-up; this finding provided the basis for classifying the study at high risk of bias. This proportion of missing data is consistent with previous meta-analyses that included this study.(1, 2) We thank the editors for pointing out this error, which we have now corrected.

The issue at hand is whether the subgroup analysis based on risk of bias for radiographic healing is robust, as the largest trial (TRUST trial) is classified as low risk of bias with a loss to follow-up of 19% - a figure that just falls below our threshold of 20%.

We did pre-specify the threshold before we conducted our analyses, and considering >20% LTFU as posing a serious threat to

validity is a widely-used (although imperfect) rule of thumb.(3) We have used the same threshold for designating studies at high risk of bias for missing data in other reviews, most recently in Wang et al.(4).

As requested, we conducted a sensitivity analysis using an alternative threshold of >10% missing data for considering a study at high risk of attrition bias (text of revised methods section provided in the next commentary). For each outcome, we added a statement in the text as to what extent the sensitivity analysis results differed from the original approach, and for those that differed we included a forest plot in appendix 4.

For pain, the threshold influenced the subgroup analysis: with the Lubbert study (16% missing data) changing to high risk of bias, the subgroup effect was no longer significant, and one might therefore place higher credibility in the overall effect rather than the effect in the low risk of bias studies. However, the overall effect remained not significant either, and the result was disproportionally influenced by one very small trial at high risk of bias for a number of reasons. Thus, the inference of no benefit of LIPUS on pain remains whether one credits the primary or the sensitivity analysis.

The more conservative threshold of 10% did not affect interpretation of the effect of LIPUS on days to radiographic healing, despite the fact that it moved the TRUST trial to the group at high risk of bias. The subgroup effect was still significant (interaction p=0.004). The summary effect estimate based on the low risk of bias group suggested no acceleration in days to healing (6.5% later healing with LIPUS, 14.6% earlier to 32.7% later, moderate certainty due to imprecision).

We added the following to the results for pain:
"The subgroup effect was no longer significant when we used a threshold of ≥10% missing data to designate a trial at high risk of attrition bias (p=0.35, fig 4 in Appendix 4)."

We added the following to the results for radiographic healing:
"In addition, the subgroup effect was robust to our sensitivity analysis using a more rigorous threshold for defining risk of attrition bias (interaction p=0.004, fig 5 in appendix 4)."

We added the following to our discussion:
"A post-hoc sensitivity analysis exploring a more conservative threshold for attrition bias (≥10% loss to follow-up) yielded, for all outcomes, results essentially consistent with the primary analyses."

Our conclusion acknowledges that while some readers may still not accept the results of our subgroup analysis for time to radiographic healing, whether readers accept the subgroup effect or not, the (absence of an) effect of LIPUS on patient-important outcomes remains unchanged:

"In conclusion, moderate to high quality evidence demonstrates that LIPUS fails to accelerate return to work, return to full weight bearing, pain, or the need for subsequent operation. If one gives highest credibility to combined effects from all available RCTs, low quality evidence would suggest a large reduction in time to radiographic healing. If, however, one gives higher credence to low risk of bias trials, moderate to high quality evidence suggests that LIPUS not only has no effect on patient-important outcomes, but also fails to accelerate radiographic healing…"

4. We agree with the reviewer who wants more information about the use of the modified ROB tool.

Reply: The questions of reviewer 3 is: "A modified Cochrane risk of bias instrument was used to assess risk of bias. The major elements included in the assessment involves selection bias and performance bias. There are some important items in the Cochrane tool that were not evaluated, such as random sequence generation or selective reporting. It is not sure why a modified tool was used instead of the original tool and whether the left-out items were all included in 'other bias'."

We have now added sequence generation and reporting bias as additional items for our risk of bias assessment. These additions have further confirmed our previous categorization of trials at high or low risk of bias.

The alteration to the original Cochrane risk of bias tool is removal of the 'unsure' category (except for reporting bias), and explicit rules to reviewers to guide a judgement of the likely risk of bias on an item-by-item basis: instead of 'unclear', reviewers are asked to make a judgement of 'probably low risk of bias' or 'probably high risk of bias'. Dichotomization provides greater clarity to readers and allows for subgroup analysis based on risk of bias. Reference 18 provides additional details about this approach, which we have previously validated. The Cochrane Risk of Bias Methods Group has recognized the desirability of this modification – a draft revised instrument from this group circulated at the last Cochrane Colloquium has adopted our approach to the instrument response options. Our revised section regarding risk of bias in the methods reads as follows:

"Two reviewers independently assessed risk of bias using a modified Cochrane risk of bias instrument that includes response options of "definitely or probably yes" (assigned a low risk of bias) or "definitely or probably no" (assigned a high risk of bias), an approach we have previously validated.(18) On the study level, we assessed generation of randomization sequence, concealment of allocation, blinding of patients, caregivers, and outcome reporting (by comparing each publication with its' corresponding published protocol, when available). For each outcome within studies, we assessed blinding of outcome assessors, loss to follow-up, and additional limitations. We considered ≥20% loss-to follow-up to represent a high risk of bias unless the investigators performed appropriate sensitivity analyses demonstrating the robustness of the results. As a sensitivity analysis, we alternatively considered a more conservative threshold of ≥10% loss to follow-up. We categorised a trial as being at low risk of bias for a particular outcome if we identified no limitation for any risk of bias item."

* We were very pleased with the readability of the paper - a number of editors remarked positively on that. Similarly, our patient advisor was very pleased with the patient information and involvement in the paper. Thank you!

Reply: Thank you!

Comments from Reviewers

Reviewer: 1

1. There are more modalities to accelerate bone healing than just LIPUS. Especially PEMF (pulsed electric magnetic fields) is subject of scientific research in a significant number of publications. It should be mentioned at least.

Reply: We have now added the following material to acknowledge this competing modality:

To our Introduction section:
"In 2008, 45% of Canadian trauma surgeons prescribed bone stimulators to manage tibia fractures, equally split between LIPUS and electrical stimulation (21% each)."

To our Discussion section:
"Our findings are similar to a 2016 systematic review of 15 small trials that explored electrical stimulation vs. sham therapy for fracture healing; only 4 of which were at low risk of bias.(52) This review found moderate quality evidence for a 35% reduction (95% CI 19% to 47%; I2=46%) in the rate of radiographic non-union. The authors found no evidence of a subgroup effect based on clinical presentation (i.e. fresh fractures, delayed union or non-union, spinal fusion, or surgical osteotomy; interaction p = 0.41) – they did not explore whether risk of bias explained heterogeneity, but all 4 trials at low risk of bias showed no significant effect on radiographic union.(53-56) This review found a small reduction in pain (mean difference of −7.7 mm on a 100mm visual analogue scale for pain, 95% CI −13.92 to −1.43), and low quality evidence for no difference in functional outcome (mean difference of −0.88 points on the 100 point Short Form 36 Physical Component Summary score, 95% CI −6.63 to 4.87)."

2. Underlying processes contributing to nonunion are fundamentally different from acute fracture repair and often multimodal. Therefore, I do not agree on including nonunions alongside fresh fractures.

Reply: We acknowledge the different clinical conditions for which LIPUS is marketed. We are not aware, however, of any evidence suggesting that the effects of LIPUS on bone healing differ among conditions, nor do the manufactures recommend different dosing for different categories, not does our subgroup analysis suggest a difference in effect on radiographic healing based on clinical condition (p=0.13).

We also note that the authors of a 2016 systematic review that explored the effect of electrical stimulation for fracture healing also found no evidence for a different in treatment effect based on clinical presentation (i.e. fresh fractures, delayed union or nonunion, spinal fusion, or surgical osteotomy; interaction p = 0.41).(5) We have now reported this finding in our Discussion section.
The bottom line for this issue is that we consider it reasonable to include all fracture-related conditions for which LIPUS is marketed in the primary analysis, explore (as we have done) possible subgroup effect related to condition (not found), but then also to acknowledge the limitation for the subgroup of nonunion in the discussion as follows:
"The evidence applies directly to patients with fresh fractures and indirectly to children and other underrepresented populations, in particular patients with non-union, for which no trustworthy direct evidence exists."

3. Selective outcome reporting (reporting bias) or incomplete outcome data (attrition bias) are not addressed in this model. It diminishes the strength if the evidence.

Reply: We did not originally include reporting bias, as the only manner to confidently assess this issue is to compare published trials with their study protocol. Aside from the TRUST definitive and pilot trials, we did not have access to any of the study protocols. However, we have now included an assessment of the risk of selective reporting bias. Owing to our inability to confirm or refute selecting reporting bias in 24 of 26 trials, we did not decrease the quality ratings on this basis.

We did explicitly consider the risk of attrition bias.

4. In conclusion: What is the added value to already existing reviews? This review is a summation of the already existing pooled evidence, with addition of the TRUST trial, that underpins the conclusions already drawn from previous reviews. Moreover, reported outcomes do not differ from previously published reviews despite of consultation of the guideline panel.

Reply: The methods, results and conclusions differ from previous reviews in a number of important aspects, including new outcomes. None of the previous reviews considered the outcome subsequent operation, and only one previous review addressed pain in one study. Strength of inference is now considerably greater given the addition of the TRUST trial. We expanded the justification for our review in the introduction:

"… In addition, recent reviews used suboptimal strategies for outcome selection, data synthesis analysis, and interpretation, leading to potentially misleading conclusions. For instance, the most recent systematic review, published in the top speciality journal for orthopaedic surgeons, considered radiographic union a "critically important outcome" and did not assess the effect of LIPUS on the patient-important outcomes of pain relief or re-operation. Their conclusion that "LIPUS treatment effectively reduces the time to radiographic fracture union" is questionable because it is based on the pooled absolute difference in days to healing, which does not account for the large variation in healing time, showed high unexplained heterogeneity (I2= 94%), and was driven by studies at high risk of bias. This positive conclusion has the potential to expand the already considerable use of a potentially ineffective therapy."

, and the following in the section "what this study adds":

"A guideline panel including patients and clinical experts informed outcome selection, importance of outcomes, subgroup analyses, and interpretation of results.

Subgroup analyses suggested that beneficial effects of LIPUS are restricted to trials at high risk of bias.
…"

Reviewer: 2

1. Overall this is a good paper which systemically reviewed the efficacy of LIPUS on bone fracture or osteotomy. Although lots of papers showed that LIPUS is effective in muscular skeletal system both clinic or experiment. Surprisingly, this manuscript proved that LIPUS is not effective for bone healing. Based on your description, you set a reasonable standard for study selection and avoiding the risk of bias. In addition, the conclusion of this review was based on the evaluation of patient's bone healing time, pain, radiographic image. So I think this is a very nice article and it is suitable for publication in BMJ.

Reply: Thank you for the positive feedback.

Reviewer: 3

This meta-analysis synthesized results from current RCTs on the effects of LIPUS on bone healing. The meta-analysis was conducted by a pre-defined guideline with a structured methodology. Below are a few comments/suggestions for the authors.

1. A modified Cochrane risk of bias instrument was used to assess risk of bias. The major elements included in the assessment involves selection bias and performance bias. There are some important items in the Cochrane tool that were not evaluated, such as random sequence generation or selective reporting. It is not sure why a modified tool was used instead of the original tool and whether the left-out items were all included in 'other bias'.
Reply: Please see the discussion of this issue above, in response to comment #4 from the Committee.

2. The final judgement of high or low risk should also be listed in Table 2.
Reply: This would considerably complicate the table, which is by study while our final judgement for risk of bias was done on an outcome-by-outcome basis. We specified in our methods the following rule for overall risk of bias: "We categorised a trial as being at low risk of bias for a particular outcome if we identified no limitation for any risk of bias item."

3. It appears that three trials were not included in any of the meta-analyses, without an explicit explanation of why (except the study by Zacherl 2009, which was excluded from time to return to work analysis due to insufficient data). If the studies were excluded because they did not report the intended outcomes, eligibility criteria should reflect selection criteria on study outcomes. The definition of 'insufficient data' and whether efforts have been made to retrieve such data should be mentioned.
Reply: We thank the reviewer for pointing out that this was unclear. Two studies did not contribute any outcomes of interest, another one had serious problems with reporting and clustering. We have amended our explanation of why these trials did not contribute to statistical pooling of effects in table 2.

In our study selection flow chart, we specifically describe that we included 26 trials in our review and 23 in our meta-analysis.

We did not approach authors of eligible RCTs for potentially unpublished outcomes.

4. As mentioned in the Methods, the authors graded the quality of evidence using GRADE approach. This part should be put in the Results, instead of Discussion. The table does not reflect the basis of GRADE's judgement about evidence's quality, which involves several domains. As the authors mentioned in the conclusion, the quality of evidence depended on whether or not studies with high risk of bias was included. And this should also be reflected on the GRADE table.
Reply: The optimal place of the GRADE judgements is not explicitly addressed in the official GRADE guidelines. We followed the reviewer's suggestion and now provide the following information under results/outcomes:
"Table 3 summarises findings of all outcomes. Interactive summary of findings tables are available online at www.magicapp.org/app#/guideline/1432."

Table 4 is a standard format for reporting GRADE Summary of Findings. Column 2 includes the information on whether or not we based our conclusions on studies at low risk of bias. We rephrased "high quality" to "low risk of bias", which is the more precise term

5. Other points: Outcomes - Functional recovery: 'Appendix 2 presents results…': there is no such appendix.
Reply: We apologize for this oversight, and have corrected the issue in the resubmission.

6. Outcomes - Number of subsequent operations: 'fig 6': it should be fig 5.
Reply: Thank you, we corrected the link.

7. Methods -synthesis of results: 'only one outcome, days to radiographic healing, included enough studies to perform': definition of 'enough' is not clear.
Reply: There is no common criteria but we had pre-specified at least three studies per group. We rephrased the section as follows: "Only one outcome, days to radiographic healing, included enough studies to perform all planned subgroup analysis. We had pre-specified in our protocol at least three studies per group."

Reviewer: 4

1. The paper is well written. The statistics are also comprehensive. However, the greatest concern is that the data from the authors' published RCT accounted for the majority of data included in this meta. Especially when the high-biased studies were excluded, the results were very similar with that of RCT. the results of this meta would significantly biased by their RCT.
Reply: The sample size of TRUST is larger than other trials, but not overwhelming compared to all data provided by the other 14 trials that contributed to the meta-analysis for radiographic healing.

The TRUST trial is potentially practice-changing, which is why the BMJ Rapid Rec was triggered. Our comprehensive meta-analysis of all RCTs, including the TRUST trial, provides critical information supporting a change in practice.

2. Additionally as for the final conclusion, the authors included all kinds of fractures: fresh or old fractures, operated or non-operated. That would result in a huge heterogeneity, which made the results hard to explain. For instance of the operated fractures, ideal reduction, fixation and rehabilitation plays the biggest role. the effect of ultrasound becomes little or even negligible…
Reply: This interpretation is consistent with our results. We found no relevant effects of ultrasound for operated tibia fracture. Regarding the heterogeneity of clinical groups, please see the discussion following comment #2 from the Committee.

… while for the non-operated chronic fractures, it may be different.
Reply: Regarding the heterogeneity of clinical groups, please see the discussion following comment #2 from the Committee.

3. The present conclusion that ultrasound is not helpful to all the fractures is too big. it is hard to believe the orthopaedic doctors would change their decision making.
Reply: As we have argued previously, the overall results indicate no benefit for any patient-important outcome, and in what we believe is the most credible analysis, no benefit in fracture healing. It remains plausible that there is some yet unidentified subgroup in which LIPUS is of benefit. We would argue, however, that the burden of proof is now on those advocating this expensive and burdensome intervention to demonstrate such a benefit.

Reviewer: 5

Overall impression

Schandelmaier et al. present a new, up-to-date meta-analysis of randomised control trials investigating the efficacy of low intensity pulsed ultrasound (LIPUS) to improve bone healing. This is a well-written, well-presented and robustly constructed analysis, which is of value to the general audience. Evidence is presented for the widespread use of LIPUS in the orthopaedic community, particularly in North America, justifying the importance of this research question.

Methods are comprehensively described, search criteria explained, two reviewers independently analysed the trials, standardized data abstraction methods were used and are presented.

The authors justify the rationale for their new meta-analysis in two main ways:
1. They focus primarily on outcomes that patients have reported to them as important, such as pain and time to return to work, rather than radiological evidence of bone healing.
2. They include the new TRUST trial, the largest RCT yet completed investigating the efficacy of LIPUS to improve bone healing in tibia fractures.

1. Point one, the focus on outcomes deemed important by patients, is I believe a significant step-forward in interpreting this data and an improvement on previous meta-analyses. The authors should be congratulated for this advance.
Reply: Thank you for the positive feedback!

2. However, I am less convinced that point 2, the inclusion of new TRUST data, significantly advances our understanding in addition to the findings of the TRUST trial alone.
Reply: Considering the whole body of evidence facilitates a number of important insights which would not have been clear or possible by considering the TRUST trial alone. For instance, it would have been unclear whether one should believe the conclusions of the TRUST trial versus the contradictory conclusions of the most recent meta-analysis which included 24 other RTCs: "LIPUS treatment effectively reduces the time to radiographic fracture union".(6) TRUST was restricted to tibial fractures – other trials permit wider applicability. The consistency of evidence for the patient-important outcomes substantially strengthens inferences from TRUST. The exploration of subgroup effects provides substantial additional insights, in particular explaining the differences in results from TRUST versus prior studies and reviews.

3. Their main finding is that there is no evidence for efficacy of LIPUS in all types of bone healing. This has major implications for patients, healthcare professionals and healthcare funders. However, to justify this broad conclusion, appropriate data must be available and be included.
Reply: Please see our responses to comment #2 from the Committee.

This leads on to my three Major Concerns:

4. There is a high quality randomised controlled trial recently published (TRUST). What is the need for this meta-analysis? Justification for the present study is provided by the statement: 'With inclusion of the recently published TRUST trial, sufficient high quality data for patients with fresh fractures has accumulated to conclude that LIPUS fails to improve patient-important outcomes and radiographic healing.' Could the authors explain and justify this position? What does this meta-analysis add further to the high-quality evidence from the randomised controlled TRUST trial? It may be reasonable that previous meta-analyses could not make firm conclusions because of the small size and poor quality trials they include. With the addition of TRUST this may change the robustness and quality of the possible meta-analysis, but the authors must convince us that this RCT genuinely adds more information on top of TRUST. Doesn't TRUST itself answer this question (in terms of fresh fractures at least)?
Reply: We have responded to the concern in comment #2 of this reviewer (reviewer 5).

5. In principle meta-analyses are most useful where numerous small studies are insufficiently powered to answer a question, and pooling them improves that power. This is not the case here, as indeed is stated by the authors, where TRUST data accounts for >1/3 total patients included in the meta-analysis
Reply: Once again, we make the argument for the important contributions of this systematic review and meta-analysis in response to this reviewer's point 1.

6. I do not find the wide-reaching, comprehensive, conclusion that LIPUS does not work for any bone healing sufficiently compelling. The authors state there has not been a high quality trial outside of tibia and clavicle fractures, and base their firm conclusions about other injuries, including non-unions, on data from trials on tibia and clavicle fractures only.

Extrapolation to other clinical indications, particularly non-unions, in my view is not justified. Certainly there is not enough justification for this assertion provided by the authors in the present manuscript. What evidence is there that stress fractures and non-unions will behave the same under LIPUS treatment? The biology of the non-union is different to an acute fracture; for example, cell types, collagen types and changes in molecules (chemokines, cytokines etc) expressed are different, and chronic inflammation is present.

Can the authors defend the homogeneity of studies included? It seems that the strong conclusions are valid for tibia and clavicle fractures (fresh fractures) and not for underrepresented types (non-unions, osteotomies, stress fractures). The biology and aetiology of these injuries is, in my view, different, certainly in the case of non-union. Therefore it may not be reasonable to make definitive conclusions about these particular subsets given that the substantial data used (19-46) came only from tibia and clavicle fractures. E.g. (2 – 41) 'considering only low risk of bias trials…'.

I would be more comfortable if distinction were made between the fracture types in this strong conclusion; however, this does lessen the impact of the paper.

On page 57 line 45 (57-45) and 58-32 is the statement: 'Further research is unlikely to alter the evidence'. This is a strong statement and fundamentally justifies the importance of this meta-analysis, which concludes that LIPUS does not work for any bone healing. The strong evidence presented is clearly on a particular type of injury and without being able to analyse data on other conditions then I am not satisfied that this question has been answered. The authors must either provide strong, compelling justification for including under-represented fractures in their conclusion, or temper their conclusions in accordance

with their findings.
Reply: Please see our response to comment #2 from the Committee.

7. Data abstraction
The quality of the meta-analysis is completely dependent on a) the quality of the data input and b) rigorous study selection to identify homogenous studies. In addition to point 2 above (not being convinced that these studies are homogeneous enough) the researchers pooled treatment effects of LIPUS on similar outcomes across eligible trials (9-32). The highest quality meta-analysis use raw data - have the raw data been obtained and analysed? If not this should be stated, justified, and rationale behind not doing this explained.
Reply: We agree that individual patient data meta-analysis of randomized controlled trials is optimal. However, an advantage of aggregated data meta-analysis over individual patient data meta-analysis is that most published studies can be included. It would have been very difficult and time-consuming to get the raw data from more than 20 RCTs published over the last 22 years. Therefore, we did not attempt to obtain and re-analyse the raw data.

8. Similarly 13-30 – were attempts made to obtain completed but unpublished datasets? If not why not? This could cause publication bias in the authors' findings.
Reply: We contacted investigators if we could not locate a final publication (please see page 1). Unfortunately, our attempts to get unpublished data were not successful. Our attempts were not clearly described in the initial submission.

If any publication bias was present, it would likely be directed in favour of LIPUS, given that almost all trials including TRUST were industry sponsored. Accounting for the bias in the interpretation would therefore be very unlikely to alter our conclusions that LIPUS has no effect. One could even argue that the plausible direction of possible reporting bias strengthens our conclusion.

Minor points:

9. Can the authors justify their inclusion of both sham and no treatment comparison trials? Are these sufficiently similar?
Reply: Trials with no sham could yield higher treatment effects, and is one source of risk of bias (lack of blinding). We have dealt with this in the subgroup analyses related to risk of bias, and indeed the higher risk of bias trials show effects not present in the low risk of bias trials, in particular with regard to radiographic healing.

10. 16-26 - 'transforming of all results' should be 'transforming all results' or 'transformation of all results'
Reply: Thank you, we have modified.

11. Insufficient data for inclusion in meta-analysis is given as a reason for not analyzing a study that reported functional recovery (14-40), pain (16-40). Could further justification for excluding these studies be provided?
Reply: Thank you for pointing this out. Regarding the first study: the small trial by Zacherl et al. had serious reporting problems and we did not trust their data. For instance, they confused group A and B, did not report any measures of variance, and randomized 44 feet but analyzed 52 toes as if they had randomized toes. We provided a link to table 2 where we provide some details.

Regarding the two other studies, we rephrased as follows:
"Two other small studies assessed pain intensity at 5 months but could not be included in the meta-analysis. One reported pain outcomes only narratively (no effect),(41) another used a modified instrument with unclear scale and variance (not significant).(49)"

12. 19-44 should be 'patients'
Reply: Thank you, corrected.

13. 18-42 'with lived' should be 'with personal experience of fractures…'
Reply: Thank you, we have modified according to the reviewer's wording preference.

Additional corrections:

For the outcome subsequent operation, we show now the forest plot for risk ratio instead of risk difference because it is probably more generalizable. We also changed it in the abstract. The text still describes both approaches.

References

1. Busse JW, Kaur J, Mollon B, Bhandari M, Tornetta P, Schunemann HJ, et al. Low intensity pulsed ultrasonography for fractures: systematic review of randomised controlled trials. Bmj. 2009;338:b351.
2. Griffin XL, Parsons N, Costa ML, Metcalfe D. Ultrasound and shockwave therapy for acute fractures in adults. Cochrane Database Syst Rev. 2014(6):CD008579.
3. Sackett DL, Richardson WS, Rosenberg W, Hayned RB. Evidence-Based Medicine - How to Practice and Teach EBM. New York: Churchill Livingstone; 1997. 250 p.
4. Wang L, Guyatt GH, Kennedy SA, Romerosa B, Kwon HY, Kaushal A, et al. Predictors of persistent pain after breast cancer surgery: a systematic review and meta-analysis of observational studies. CMAJ. 2016;188(14):E352-E61.
5. Aleem IS, Aleem I, Evaniew N, Busse JW, Yaszemski M, Agarwal A, et al. Efficacy of Electrical Stimulators for Bone Healing: A Meta-Analysis of Randomized Sham-Controlled Trials. Sci Rep. 2016;6:31724.
6. Rutten S, van den Bekerom MP, Sierevelt IN, Nolte PA. Enhancement of Bone-Healing by Low-Intensity Pulsed Ultrasound: A Systematic Review. JBJS Rev. 2016;4(3).

1. LIPUS SR response to reviewers_FINAL.docx PDF HTML