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

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Acupuncture for Low Back and/or Pelvic Pain during Pregnancy: A Systematic Review and Meta-Analysis of Randomized Controlled Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Yang, Jiaman; Wang, Yi; Xu, Jianmei; Ou, Zhiwen; Yue, Ting; Mao, Zhitao; Lin, Ying; Wang, Tao; Dong, Wei; Shen, Zhen</td>
</tr>
</tbody>
</table>

VERSION 1 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Levett, Kate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>University of Western Sydney, National Institute of Complementary Medicine Research</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>01-Oct-2021</td>
</tr>
</tbody>
</table>

GENERAL COMMENTS

This is an important paper, and reasonably presented. However, there are some major flaws that require addressing prior to being considered for publication.

ABSTRACT,

1. safety outcomes are not addressed/listed
2. The STRICTA reporting guidelines are recommended for acupuncture systematic reviews (not PRISMA)
3. It is not clear what period the analgesic consumption refers to? Intrapartum or during the study period?
4. Why are Apgar scores considered as safety outcomes, and no maternal outcomes listed? Important maternal safety concerns such as pre-term birth, pre-term contractions, premature rupture of membranes etc. as well as adverse effects from acupuncture, such as headache, fatigue, nausea, fainting etc.

The conclusion states that acupuncture had no observable adverse effects on the newborn, but with no reference to maternal outcomes. Newborn effects are more likely to be correlated with intrapartum events.

INTRODUCTION

1. Many of the references are more than 10-15 years old. Need updating (e.g. ref 2, 6, 7, 9, 10, 11, 13, 14)
2. Line 4 - what does 'chronic situation' refer to?
4. Paragraph 2, line 2 - instead of saying 'specific parts', please refer to acupoints along the meridians - and explain a bit of TCM theory here - what are the meridians and acupoints.
6. You comment that the reviews have not assessed newborn safety, but maternal safety outcomes would seem to be more relevant here.
7. You report that you found only high quality evidence, but do not report the method that you used to assess this. Later you mention the PEDRO tool, but I would suggest that the Cochrane RoB tool is more relevant here (Refer to following articles: a) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0028130 b) https://journals.sagepub.com/doi/10.1177/0964528420946043)

METHODS
8. PRISMA is not the correct reporting guideline - see STRICTA
9. Acupuncture terms are insufficient for the review - need to increase terms - (see Cochrane review of acup for prenancy, labour and birth. Smith et al 2021)
9. Does 'no treatment' mean 'usual care'?
10. Clarify period of time that you're measuring analgesic consumption
11. Clarify the safety outcomes, beyond newborn Apgars, that you report on - in particular maternal outcomes

QUALITY ASSESSMENT
12. Cochrane RoB tool should be used
13. Line 3. Do you mean similarity (not similarly)?
14. What does 'completeness of follow up mean'? Is this different to differential attrition rates? If so, be specific about attrition being measured, as this is a significant factor in acupuncture trials
15. Statistical analysis. Usual trials with >30% I2 score will use a random effects model, as per Cochrane Reviews. The cut off of 50% seems very high

RESULTS,
STUDY SELECTION
16. Did you hand search the reference lists?

STUDY CHARACTERISTICS
17. What does the 'therapists were trained' mean. Trained as what? Were they trained acupuncturists? If not, then they are not trained for this intervention. This needs to be clear, and should be reported via STRICTA
18. Line 9 - it is 'midwives' not 'midwifes'
19. Refering to auricular press needles - how did you find this study with the search terms that you used? None of these terms were included. Search terms need updating to include ‘auricular’, ‘acup’*, ‘needl*’ etc.

RISK OF BIAS
20. You state that 2 studies used subject blinding. However, only 1 study is listed as using placebo in the table. How was blinding achieved where no placebo was used?
21. The drop out rates are not specifically reported fully, although you state that the drop out rates are adequately described in the studies. 22. Additionally, you report that all studies had a <15% drop out rate. This is not correct - The studies by Wedenberg (2009); Ekdahl (2010) and Lund (2006) all have drop out rates >15%, in fact they are >20%, which is considered a critical flaw in the risk of bias. please refer to the paper: Levett et al 2019. Using Forbidden Points in Pregnancy: Adverse Outcomes and Quality of Evidence in Randomized Controlled Trials—A Systematic Narrative Review: https://www.liebertpub.com/doi/abs/10.1089/acu.2019.1391

TABLE 1
23. You report the acupoints used in Table 1, and note that so called 'forbidden points' are used, yet you do not report on risks from this, including preterm labour. Nor make mention of 'forbidden points' which is a controversial area

ADVERSE EVENTS
24. You state in this section "No preterm labor or other adverse pregnancy outcomes were reported during study period". This is not correct and is misleading for the review to state. Please refer to the
Levett paper previously cited. This is repeated in the conclusion, but remains incorrect.

QUALITY OF EVIDENCE
25. Inadequate reporting of the GRADE constituent parts

This paper requires serious attention prior to publication, but it would be good to see a revised attempt, as this is an important area for pain relief for pregnant women, and one of significant disability.

REVIEWER
Kern, Alexandra
Institute for Complementary and Integrative Medicine, University Hospital Zurich and University Zurich, Zurich Switzerland

REVIEW RETURNED
29-Oct-2021

GENERAL COMMENTS
With their meta-analysis Yang et al. aimed to investigate the effects of acupuncture on primary outcomes (such as pain) as well as secondary outcomes (e.g., Apgar scores for newborns) for women with low back and/or pelvic pain during pregnancy. Overall, the manuscript is well written and the process that lead to the results is sufficiently described. I appreciated the choice of the primary and secondary outcomes. However, I would suggest some revisions in order to improve the quality of the manuscript.

1. In certain parts of the manuscript, the use of words is somewhat confusing. Examples are:
   Line 66/67: What do the authors mean by “specific parts”?
   Line 69: What are “other situations”?
   Additionally, I would suggest an overall revision of language.
2. Table 1: By looking at the average age of the females, it appears that the participants were rather young (mostly < 35 years of age). I did not find any mention of this in the discussion section. Would you like to comment on this?
3. Line 283/284: “These results revealed that painkillers might be unhelpful for pain during pregnancy …”. Either this sentence needs clarification or I do not fully agree with the conclusion. Why do you conclude that the absence of an effect means that painkillers are “unhelpful” during pregnancy? Wouldn’t it be possible that pregnant women simply do not rely on painkillers because of safety concerns?

REVIEWER
Symanowski, James
Levine Cancer Institute, Department of Cancer Biostatistics

REVIEW RETURNED
30-Nov-2021

GENERAL COMMENTS
This research study was generally well constructed, and the results of the meta-analysis are clearly described, and the conclusions appropriately reflect the reported results.

More explanation of the statistical methods is warranted. Specifically, more description regarding "heterogeneity" using the Higgins I^2 statistic should be provided. This description should include heterogeneity of the treatment effect sizes across the sited publications included in the meta-analysis. The information in Table S2 in the supplementary material document needs further clarification. While the "leave-one-out" approach is referenced in the main manuscript, a brief description should be provided explaining what the study-specific I^2 statistic means (i.e., the heterogeneity between studies after leaving that study out). Also,
an interpretation of the p-values in Table S2 should be described explaining what test these p-values are associated with.

The Higgins I^2 statistic is classified into four categories: very low, low, moderate, and high heterogeneity. More explanation should be provided regarding how these categories impacted the modeling. A more complete description of the models should be provided and a rationale for why fixed effects models were used for I^2 < 50% while random effects models were used for I^2 > 50%.

A more complete explanation of the Funnel plots is needed explaining how they are constructed how to interpret the results. Additionally, a description of what the statistical test for the Funnel plot is testing needs to be described.

The sentence beginning on line 239 ("The heterogeneity was significantly reduced of pain outcome after removing the study by Elden H. et al (32), which suggested that this study could be the primary source of heterogeneity.") needs to be re-worded. It does not make sense as current worded. What does "The heterogeneity was significantly reduced of pain outcome after removing..." mean? I believe this intent was to say the heterogeneity in the pain treatment effect size was significantly reduced after removing....
Response: Thank you very much for your valuable advice. We have added maternal outcomes and adverse effects from acupuncture in the paper and re-written this part according to your suggestion. Revised portion are marked in red in the paper (page 11 line 268-286 and Table 2)

INTRODUCTION
1. Response to comment: Many of the references are more than 10-15 years old. Need updating (e.g. ref 2, 6, 7, 9, 10, 11, 13, 14)
Response: Thank you very much for your constructive suggestions. We updated the references. Revised portion are marked in red in the reference part (page 20)
2. Response to comment: Line 4 - what does ‘chronic situation’ refer to?
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper (page 3 line 84)
3. Response to comment: An important and related reference that you have not mentioned is: Levett et al 2019. Using Forbidden Points in Pregnancy: Adverse Outcomes and Quality of Evidence in Randomized Controlled Trials—A Systematic Narrative Review:
Response: Thank you very much for your constructive suggestions. We have inserted the references in the paper. Revised portion are marked in red in the paper (page 3 line 94-99)
4. Response to comment: Paragraph 2, line 2 - instead of saying ‘specific parts’, please refer to acupoints along the meridians - and explain a bit of TCM theory here - what are the meridians and acupoints.
Response: Thank you very much for your valuable advice, and we have made correction and explained a bit of TCM theory. The meridian theory was proposed based on ancient physiological system (not western scientific empiricism). This theory suggested that the body’s vital energy (qi) flows along meridians, which are associated with the internal organs. It is believed that internal disharmony of organs is reflected at specific points, known as acupoints. Revised portion are marked in red in the paper (page 3 line 78-82)
5. Response to comment: Paragraph 2 - evidence for pain relief: Evidence for chronic pain
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper (page 4 line 105)
6. Response to comment: You comment that the reviews have not assessed newborn safety, but maternal safety outcomes would seem to be more relevant here.
Response: Thank you very much for your valuable advice. We have added safety outcomes in the paper according to your suggestion. Revised portion are marked in red in the paper (page 4 line 102, page 2 line 39-41, page 11 line 268-286 and Table 2)
7. Response to comment: You report that you found only high quality evidence, but do not report the method that you used to assess this. Later you mention the PEDRO tool, but I would suggest that the Cochrane RoB tool is more relevant here
Response: Thank you very much for your valuable advice. We have made correction about the method for assessing methodological quality of included studies. Revised portion are marked in red in the paper (page 7 line 216-221 and Figure 2 and 3)

METHODS
8. Response to comment: PRISMA is not the correct reporting guideline - see STRICTA
Response: Thank you very much for your valuable advice. We have added STRICTA reporting guidelines in the paper. Revised portion are marked in red in the paper (page 13 line 325-339 Table 3)
9. Response to comment: Acupuncture terms are insufficient for the review - need to increase terms - (see Cochrane review of acup for pregnacy, labour and birth. Smith et al 2021)
Response: Thank you very much for your suggestion. We increased terms in the paper and researched them. The search strategy is detailed in Supplementary Table 1. Revised portion are marked in red in the paper (page 4 line 128-130)
10. Response to comment: Does ‘no treatment’ mean ‘usual care’?
Response: We are very sorry for our negligence. 'No treatment' means 'no intervention'.

11. Response to comment: Clarify period of time that you're measuring analgesic consumption
Response: We are very sorry for our negligence of explaining what period the analgesic consumption refers to. The period of analgesic consumption refers to during the study period. Revised portion are marked in red in the paper (page 2 line 37, page 5 line 138-139 and page 8 line 253)

12. Response to comment: Clarify the safety outcomes, beyond newborn Apgars, that you report on - in particular maternal outcomes
Response: Thank you very much for your suggestion. We have clarified the safety outcomes Revised portion are marked in red in the paper (page 4 line 102, page 2 line 39-41, page 11 line 268-286 and Table 2)

QUALITY ASSESSMENT
13. Response to comment: Cochrane RoB tool should be used
Response: Thank you very much for your valuable advice. We have made correction about the method for assessing methodological quality of included studies. Revised portion are marked in red in the paper (page 7 line 216-221 and Figure 2 and 3)

14. Response to comment: Line 3. Do you mean similarity (not similarly)?
Response: Thank you very much for your careful review of this manuscript. We re-write this part and delete this word

15. Response to comment: What does 'completeness of follow up mean'? Is this different to differential attrition rates? If so, be specific about attrition being measured, as this is a significant factor in acupuncture trials
Response: Thank you very much for your careful review of this manuscript. 'Completeness of follow up' means dropout, which is one of items of the PEDro scale. We have replaced PEDro with Cochrane RoB tool to assess methodological quality of included studies. Revised portion are marked in red in the paper (page 7 line 216-221 and Figure 2 and 3)

16. Response to comment: Statistical analysis. Usual trials with >30% I2 score will use a random effects model, as per Cochrane Reviews. The cut off of 50% seems very high
Response: Thank you very much for your valuable advice, and we searched the relevant literature in Cochrane Database of Systematic Reviews. Revised portion are marked in red in the paper (page 6 line 170-174)

RESULTS, STUDY SELECTION
17. Response to comment: Did you hand search the reference lists?
Response: Thank you for your essential advice. The search strategy is detailed in Supplementary Table 1, and the electronic databases were re-searched from the inception dates to 15 January 2022. Revised portion are marked in red in the paper (page 4 line 116 and Supplementary Table 1)

STUDY CHARACTERISTICS
18. Response to comment: What does the 'therapists were trained' mean. Trained as what? Were they trained acupuncturists? If not, then they are not trained for this intervention. This needs to be clear, and should be reported via STRICTA
Response: We are very sorry for our improper writing. The therapists involved were trained for acupuncture intervention. Additionally, thank you for your essential advice.
We have added STRICTA reporting guidelines in the paper. Revised portion are marked in red in the paper (page 13 line 325-339 Table 3)

19. Response to comment: Line 9 - it is 'midwives' not 'midwifes'
Response: Thank you very much for your careful review of this manuscript. We have made correction. Revised portion are marked in red in the paper (page 7 line 207).

20. Response to comment: Refering to auricular press needles - how did you find this study with the search terms that you used? None of these terms were included. Search terms need updating to include 'auricular', 'acup*', 'needl*' etc.
Response: Thank you very much for your valuable advice. We have corrected the Intervention of Selection Criteria, acupuncture, including traditional acupuncture (using thin and solid metallic needles to insert into specific parts), needle, and auricular acupuncture. Revised portion are marked in red in the paper (page 4 line 128-130) and the electronic databases were re-searched from the inception dates to 15 January 2022.

RISK OF BIAS
21. Response to comment: You state that 2 studies used subject blinding. However, only 1 study is listed as using placebo in the table. How was blinding achieved where no placebo was used?
Response: Thank you very much for your careful review of this manuscript. We are very sorry for our negligence of listing placebo acupuncture, and we have made correction. Revised portion are marked in red in the paper (page 9 Table 1)
22. Response to comment: The drop out rates are not specifically reported fully, although you state that the drop out rates are adequately described in the studies. 22. Additionally, you report that all studies had a <15% drop out rate. This is not correct - The studies by Wedenberg (2009); Ekdahl (2010) and Lund (2006) all have drop out rates >15%, in fact they are >20%, which is considered a critical flaw in the risk of bias.
Response: Thank you very much for your careful review of this manuscript. We have corrected the drop-outs rate and specifically reported fully in the study. Revised portion are marked in red in the paper (page 7 line 210-211, page 9 Table 1)

TABLE 1
23. Response to comment: You report the acupoints used in Table 1, and note that so called 'forbidden points' are used, yet you do not report on risks from this, including preterm labour. Nor make mention of 'forbidden points' which is a controversial area
Response: Thank you very much for your valuable advice. We described the use of “forbidden points” during pregnancy in the study. Revised portion are marked in red in the paper (page 7 line 210-211). We also reported on risks from this in the discussion. Revised portion are marked in red in the paper (page 16 line 381-390)

ADVERSE EVENTS
24. Response to comment: You state in this section "No preterm labor or other adverse pregnancy outcomes were reported during study period". This is not correct and is misleading for the review to state. Please refer to the Levett paper previously cited. This is repeated in the conclusion, but remains incorrect.
Response: Thank you very much for your careful review of this manuscript. We have re-written these two parts according to your suggestion. Revised portion are marked in red in the paper (page 2 line 39-48, page 11 line 268-286 and Table 2, page 19 line 477-483)

QUALITY OF EVIDENCE
25. Response to comment: Inadequate reporting of the GRADE constituent parts
Response: Thank you for your essential advice. We have re-written this part about the Quality of evidence Revised portion are marked in red in the paper (page 14 line 341-363)
Reviewer 2:
1. Response to comment: In certain parts of the manuscript, the use of words is somewhat confusing. Examples are:
   Line 66/67: What do the authors mean by “specific parts”?
   Line 69: What are “other situations”?
   Additionally, I would suggest an overall revision of language.
Response: Thank you very much for your careful review of this manuscript. We have made correction according to your suggestion. Revised portion are marked in red in the paper (page 3 line 78-82 and line 84) We regret there were problems with the English. The paper has been carefully revised by a professional language editing service to improve the grammar and readability.

2. Response to comment: Table 1: By looking at the average age of the females, it appears that the participants were rather young (mostly < 35 years of age). I did not find any mention of this in the discussion section. Would you like to comment on this?
Response: Thank you very much for your constructive suggestions. We have added some comments on this. Revised portion are marked in red in the paper (page 18 line 465-469)

3. Response to comment: Line 283/284: “These results revealed that painkillers might be unhelpful for pain during pregnancy …”. Either this sentence needs clarification or I do not fully agree with the conclusion. Why do you conclude that the absence of an effect means that painkillers are “unhelpful” during pregnancy? Wouldn’t it be possible that pregnant women simply do not rely on painkillers because of safety concerns?
Response: Thank you very much for your valuable advice. It is really true as you suggested that we could not conclude that painkillers might be unhelpful for pain during pregnancy. We have re-written this part according to your suggestion. Revised portion are marked in red in the paper (page 16 line 375-380)

Reviewer 3:
1. Response to comment: More explanation of the statistical methods is warranted. Specifically, more description regarding “heterogeneity” using the Higgins I**2 statistic should be provided. This description should include heterogeneity of the treatment effect sizes across the sited publications included in the meta-analysis. The information in Table S2 in the supplementary material document needs further clarification. While the "leave-one-out" approach is referenced in the main manuscript, a brief description should be provided explaining what the study-specific I**2 statistic means (i.e., the heterogeneity between studies after leaving that study out). Also, an interpretation of the p-values in Table S2 should be described explaining what test these p-values are associated with.
Response: Thank you very much for your careful review of this manuscript. We have made explanation of the statistical methods according to your suggestion. Revised portion are marked in red in the paper (page 6 line 166-185) In addition, we provided a brief description to explain the heterogeneity between studies after leaving that study out and the p-values in Table S2. Revised portion are marked in red in the paper (page 11 line 289-302)

2. Response to comment: The Higgins I**2 statistic is classified into four categories: very low, low, moderate, and high heterogeneity. More explanation should be provided regarding how these categories impacted the modeling. A more complete description of the models should be provided and a rationale for why fixed effects models were used for I**2 < 50% while random effects models were used for I**2 > 50%.
Response: Thank you very much for your constructive suggestions. We have added some comments on this. Revised portion are marked in red in the paper (page 6 line 166-185).

3. Response to comment: A more complete explanation of the Funnel plots is needed explaining how they are constructed how to interpret the results. Additionally, a description of what the statistical test for the Funnel plot is testing needs to be described.
Response: Thank you very much for your valuable advice. We have re-written this part according to your suggestion. Revised portion are marked in red in the paper (page 13 line 310-320)

4. Response to comment: The sentence beginning on line 239 (“The heterogeneity was significantly reduced of pain outcome after removing the study by Elden H. et al (32), which suggested that this study could be the primary source of heterogeneity.”) needs to be re-worded. It does not make sense as current worded. What does “The heterogeneity was significantly reduced of pain outcome after removing...” mean? I believe this intent was to say the heterogeneity in the pain treatment effect size was significantly reduced after removing....
Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing, and we have re-written this part according to your suggestion. Revised portion are marked in red in the paper (page 11 line 289-302)

VERSION 2 – REVIEW

REVIEWER
Levett, Kate
University of Western Sydney, National Institute of Complementary Medicine Research

REVIEW RETURNED
05-Mar-2022

GENERAL COMMENTS

Abstract.
- The secondary outcomes: overall effects (include in brackets what overall effects included)
- Other outcomes should be identified as important for inclusion, even if the studies didn’t include them. This will highlight the need for future studies to collect outcome data on this. This includes: Gestational age at birth (weeks and days), induction of labour (spontaneous/induced) Mode of birth (normal vaginal birth / instrumental vaginal birth / caesarean section).
- Attrition bias needs to be addressed – some studies had greater than 20% dropout rate. This is considered a critical flaw.

Introduction:
- Line 88-90 where it says: …but the evidence is equivocal. Besides, some studies have reported that there is no significant difference for relieving low back pain compared acupuncture with sham acupuncture.
  Should say: difference for relieving low back pain when acupuncture is compared with sham

- Line 94…… You discuss reference 24: “in view of concerns about using “forbidden points” during pregnancy and the risk of preterm contractions, one systematic narrative review examined randomized controlled trials of acupuncture using forbidden points prior to 37 weeks of pregnancy to treat pain conditions, with the primary outcome of preterm contractions. It suggested that high-quality trials are required to confirm the use of “forbidden points” during pregnancy.24”
  You have not taken into consideration that this review discusses the high attrition rate of some trials included, and the critical flaw that it introduces. It highlights that high quality trials are required, as the evidence has significant bias due to the dropout rates.

- Line 105… where is says: acupuncture for chronic pain and improvement in function status and quality of life for pregnant
  It should say functional status
- Line 393: improvement in function status
  It should say functional status

- This should be revised throughout

- Line 106: Where it says: women with LBPP. Secondary objectives included the assessment of Apgar scores and adverse 107 events

  However, secondary objectives in abstract state ‘overall effects’ – what does this refer to? And why is it not stated at the introduction section in this statement?

- Use reviewer initials not names in the text (search strategy) – both methods are used variably in the text.

- Line 182: ‘that cannot be explained. An extended discussion of this option appears in Section. Appears to be missing the end of the sentence

Adverse Events:

- Lund 2006 33% drop out rate, Wedenberg 23% drop out rate – anything more than 20% is considered critical – needs to be stated here.

- Line 345: There is low evidence in pain regarding the GRADE confidence in this estimate. 348-9 And this line: There is moderate evidence in pain regarding the 349 GRADE confidence in this estimate.

  What does low /moderate evidence in pain mean? Do you mean evidence for pain reduction?

- Line 375: where it says…….‘effects. However, there were no significant differences about analgesic consumption during 376 treatment and Apgar scores of newborns after treatment. On the one hand, these results revealed 377 that the reduction of pain was not sustained for every participant. Because the intervention in both 378 studies was auricular acupuncture.

  Should say in the sentence preceding that this outcome was measured in two studies, before saying ‘in both...’ This is not clear what you’re referring to.

- Outcome measures – I’m aware that some outcome measures are generally not captured in these studies, and you’re limited with what you can do in an SR, as some outcomes are not within your control.

  However, there are some important outcome measures that should be considered, which I’ve mentioned earlier. These are important because if a woman has ongoing pelvic pain, there
may be changes to the management of her pregnancy including induction of labour, pain relief required during labour, which may impact mode of birth, with the consequence of increased rates of caesarean section, then this is a vital component that needs to be understood. Part of the discussion and recommendations should suggest consideration of additional outcome measures as routine for primary studies, including: mode of birth (vaginal birth, instrumental vaginal birth, caesarean section), pain relief used during labour (opioid, epidural, sterile water injections etc), and onset of labour (induced, spontaneous) should be included.

If the researchers are able to capture Apgar scores, by presumably accessing patient records following the birth, then these outcomes are part of routinely collected data and could feasibly be included.

- Discussion of comparison treatments is lacking. In the discussion, there is not sufficient elucidation of what the comparison/control includes and its appropriateness
- Line 394…… Where is says: ‘Two systematic reviews reported that there was no clear evidence showing acupuncture to be more effective for low back pain than other therapies.41 42’.

Please review the references used: Reference 41 is for migraine, not low back pain, this should be removed as evidence for this statement.

Reference 42 was published in 1999, and is nearly 25 years old, so would not be considered an adequate reference for evidence for this statement. In the CSR reported in reference 42, low quality trials are used, and the comparison groups included: trigger-point injection or transcutaneous electrical nerve stimulation, which is now known to be erroneous, as it’s an active treatment and not an appropriate ‘sham’ comparison. Additionally, this CSR did not include any outcomes for acute lower back pain, only chronic lower back pain. Acute would probably be the outcome of interest for a common presentation in pregnancy.

- Line 401…… Reference 44 is cited as evidence for typical care on low back pain during pregnancy. However, from all the studies included in this meta-analysis only one study used acupuncture intervention in the intervention group.

You have neglected to mention that this study, while including only the one study of auricular acupuncture, showed significant improvement in pain compared with standard care. Active placebo treatments are also somewhat effective, which is why verum acupuncture is not significantly greater. The right comparison is standard care, not other active sham treatments.
We are all well aware that they demonstrate effectiveness of their own.

- Line 418 Where is says ...... ‘It is challenging for pregnant......419 women to receive drug therapy during pregnancy because they feared that the drugs might be....420 harmful to the fetus.46’

Please amend this to include this to state that drug therapies can actually be harmful to pregnant women, not just that pregnant women feared this. It is important to recognise that pregnancy is a time where interventions are necessarily minimal, and should not include pharmaceuticals unless absolutely required. This is due to concerns of teratogenicity for the fetus and harm to the mother.

- Line 425. Where is says..... ‘acupuncture significantly increase blood flow of local skin and muscle to relieve various pain’ (add an ‘s’ to increase – should be increases)

- Line 432..... Where is says …Moreover, exercise and yoga as potentially non.......433 pharmacological therapies are difficult to keep on for pregnant women due to the inconvenience.10’

There is no need to introduce this here. This is not a review of other therapies, which may or may not be effective for LBPP. You have not made any comparisons with other therapies so far in the review, and you can promote the potential of acupuncture effectiveness without it.

- Line 434......where it says: ‘Importantly, No preterm labor or other adverse pregnancy outcomes were reported during.....435 acupuncture. Although 3 participants in all studies experienced transient pain and/or redness during’

In this section, you should mention that some studies had high dropout rates, so we are not certain of some outcomes, or potential bias due to attrition. (also the ‘N’, where it appears in No preterm labor, does not need capitalisation).

- Line 449. Where it says...... ‘In addition, only one study....450 of all included studies referred the costs of acupuncture. We found that few pregnant women with....451 LBPP receive active treatment might, which might associate with high costs that patients cannot ....452 afford.34’

Please revise sentence – it does not make sense (underlined section).

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Kern, Alexandra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institute for Complementary and Integrative Medicine, University Hospital Zurich and University Zurich, Zurich Switzerland</td>
<td></td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>20-Feb-2022</td>
</tr>
</tbody>
</table>
### GENERAL COMMENTS

All of my comments and suggestions have been appropriately addressed.

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Symanowski, James</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levine Cancer Institute, Department of Cancer Biostatistics</td>
<td></td>
</tr>
</tbody>
</table>

| REVIEW RETURNED | 02-Mar-2022 |

<table>
<thead>
<tr>
<th>GENERAL COMMENTS</th>
<th>Responses to Reviewer # 3 Comments:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comment 1:</td>
</tr>
<tr>
<td></td>
<td>Response Document</td>
</tr>
<tr>
<td></td>
<td>Response is acceptable.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revised Manuscript</td>
</tr>
<tr>
<td></td>
<td>Editorial comment on line 173: &quot;P &lt; 0.1 of chi-square test…&quot; should read “A chi-square test p-value &lt;0.1…&quot;.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comment 2:</td>
</tr>
<tr>
<td></td>
<td>Response Document</td>
</tr>
<tr>
<td></td>
<td>Response is acceptable.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revised Manuscript</td>
</tr>
<tr>
<td></td>
<td>The explanation for the four categories of heterogeneity adds some clarity to the manuscript. However, the specific I<strong>2 intervals described are overlapping. For example: 1. I</strong>2 = 35% could be either “might not important” or “moderate heterogeneity”; 2. I<strong>2 = 55% could be either “moderate heterogeneity” or “substantial heterogeneity”; 3. I</strong>2 = 80% could be either “substantial heterogeneity” or “considerable heterogeneity”. The source of this rough guide should be provided and an explanation for the overlapping I**2 intervals should be provided.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Editorial comment on line 182: “An extended discussion of this option appears in Section.&quot; Is missing something. What does “Section&quot; mean?</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comment 3:</td>
</tr>
<tr>
<td></td>
<td>Response Document</td>
</tr>
<tr>
<td></td>
<td>Response is acceptable.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revised Manuscript</td>
</tr>
<tr>
<td></td>
<td>Revision to the manuscript is acceptable.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comment 4:</td>
</tr>
<tr>
<td></td>
<td>Response Document</td>
</tr>
<tr>
<td></td>
<td>Response is acceptable.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revised Manuscript</td>
</tr>
<tr>
<td></td>
<td>Revision to the manuscript is acceptable.</td>
</tr>
</tbody>
</table>

### VERSION 2 – AUTHOR RESPONSE

Responds to the reviewer’s comments:

Reviewer 1:

**ABSTRACT**
1. Response to comment: The secondary outcomes: overall effects (include in brackets what overall effects included)

Response: Thank you very much for your valuable advice. We have added what overall effects included in brackets. Revised portion are marked in red in the paper. (page 1 line 29-31)

2. Response to comment: Other outcomes should be identified as important for inclusion, even if the studies didn’t include them. This will highlight the need for future studies to collect outcome data on this. This includes: Gestational age at birth (weeks and days), induction of labour (spontaneous/induced) Mode of birth (normal vaginal birth / instrumental vaginal birth / caesarean section).

Response: Thank you very much for your valuable advice. We have added these outcomes in the manuscript. Revised portion are marked in red in the paper. (page 1 line 31-32, page 2 line 42-45, page 4 line 111-115 and page 9 line 274-277)

3. Response to comment: Attrition bias needs to be addressed – some studies had greater than 20% dropout rate. This is considered a critical flaw.

Response: We are very sorry for our negligence of addressing the attrition bias. Revised portion are marked in red in the paper. (page 7 line 222-225)

INTRODUCTION

1. Response to comment: Line 88-90 where it says: …but the evidence is equivocal. Besides, some studies have reported that there is no significant difference for relieving low back pain compared acupuncture with sham acupuncture.

Should say: difference for relieving low back pain when acupuncture is compared with sham

Response: Thank you very much for your constructive suggestions. We have made correction according to your comments. Revised portion are marked in red in the paper. (page 3 line 94-95)

2. Response to comment: Line 94……. You discuss reference 24: “in view of concerns about using “forbidden points” during pregnancy and the risk of preterm contractions, one systematic narrative review examined randomized controlled trials of acupuncture using forbidden points prior to 37 weeks of pregnancy to treat pain conditions, with the primary outcome of preterm contractions. It suggested that high-quality trials are required to confirm the use of “forbidden points” during pregnancy.24”

You have not taken into consideration that this review discusses the high attrition rate of some trials included, and the critical flaw that it introduces. It highlights that high quality trials are required, as the evidence has significant bias due to the dropout rates.

Response: Thank you very much for your valuable advice. We are very sorry for our negligence of the dropout rates and we have made correction. Revised portion are marked in red in the paper. (page 4 line 104)

3. Response to comment: Line 105… where is says: acupuncture for chronic pain and improvement in function status and quality of life for pregnant

It should say functional status

Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 4 line 111)

4. Response to comment: Line 393…. improvement in function status. It should say functional status.
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 11 line 358 and page 17 line 425)

5. Response to comment: Line 106...... Where it says: women with LBPP. Secondary objectives included the assessment of Apgar scores and adverse 107 events

However, secondary objectives in abstract state ‘overall effects’ – what does this refer to? And why is it not stated at the introduction section in this statement?

Response: Thank you very much for your valuable advice. We are very sorry for our negligence of stating at the introduction section. We have added what overall effects refer to and re-written this part according to your comments. Revised portion are marked in red in the paper. (page 4 line 111-115)

6. Response to comment: Use reviewer initials not names in the text (search strategy) –both methods are used variably in the text.

Response: Thank you very much for your careful review of this manuscript. Revised portion are marked in red in the paper. (page 4 line 128, page 5 line 155, line 161 and 171 and page 7 line 230)

7. Response to comment: Line 182 ‘that cannot be explained. An extended discussion of this option appears in Section. Appears to be missing the end of the sentence

Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing. This sentence should be deleted and we have deleted in the new manuscript.

ADVERSE EVENTS

1. Response to comment: Lund 2006 33% drop out rate, Wedenberg 23% drop out rate – anything more than 20% is considered critical – needs to be stated here.

Response: Thank you for your essential advice. We have stated the dropout rate according to your suggestion. Revised portion are marked in red in the paper. (page 9 line 298-301)

2. Response to comment: Line 345 There is low evidence in pain regarding the GRADE confidence in this estimate. 348-9 And this line: There is moderate evidence in pain regarding the 349 GRADE confidence in this estimate.

What does low /moderate evidence in pain mean? Do you mean evidence for pain reduction?

Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing. We have made correction. Revised portion are marked in red in the paper. (page 11 line 358, 362, 366, 370 and 377) Additionally, low /moderate evidence means the quality of the included studies in the outcomes of pain, functional status, quality of life, overall effects, analgesic consumption and Apgar scores.

3. Response to comment: Line 375 where it says…….’effects. However, there were no significant differences about analgesic consumption during treatment and Apgar scores of newborns after treatment. On the one hand, these results revealed that the reduction of pain was not sustained for every participant. Because the intervention in both studies was auricular acupuncture.

Should say in the sentence preceding that this outcome was measured in two studies, before saying ‘in both...’ This is not clear what you’re referring to.

Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing. We have re-written this part according to your suggestion. Revised portion are marked in red in the paper. (page 16 line 398-399)
DISCUSSION

1. Response to comment: I’m aware that some outcome measures are generally not captured in these studies, and you’re limited with what you can do in an SR, as some outcomes are not within your control.

However, there are some important outcome measures that should be considered, which I’ve mentioned earlier. These are important because if a woman has ongoing pelvic pain, there may be changes to the management of her pregnancy including induction of labour, pain relief required during labour, which may impact mode of birth, with the consequence of increased rates of caesarean section, then this is a vital component that needs to be understood.

Part of the discussion and recommendations should suggest consideration of additional outcome measures as routine for primary studies, including: mode of birth (vaginal birth, instrumental vaginal birth, caesarean section), pain relief used during labour (opioid, epidural, sterile water injections), and onset of labour (induced, spontaneous) should be included.

If the researchers are able to capture Apgar scores, by presumably accessing patient records following the birth, then these outcomes are part of routinely collected data and could feasibly be included.

Response: Thank you very much for your essential advice. We have added this part according to your suggestion. Revised portion are marked in red in the paper. (page 16 line 410-416)

2. Response to comment: Discussion of comparison treatments is lacking. In the discussion, there is not sufficient elucidation of what the comparison/control includes and its appropriateness.

Response: Thank you for your essential advice. We have discussed the differences of treatment in the control group among included studies according to your suggestion. Revised portion are marked in red in the paper. (page 16 line 416-421)

REFERENCES

1. Response to comment: Line 394……Where is says: ‘Two systematic reviews reported that there was no clear evidence showing acupuncture to be more effective for low back pain than other therapies.41 42’.

Please review the references used: Reference 41 is for migraine, not low back pain, this should be removed as evidence for this statement.

Reference 42 was published in 1999, and is nearly 25 years old, so would not be considered an adequate reference for evidence for this statement. In the CSR reported in reference 42, low quality trials are used, and the comparison groups included: trigger-point injection or transcutaneous electrical nerve stimulation, which is now known to be erroneous, as it’s an active treatment and not an appropriate ‘sham’ comparison. Additionally, this CSR did not include any outcomes for acute lower back pain, only chronic lower back pain. Acute would probably be the outcome of interest for a common presentation in pregnancy.

Response: Thank you very much for your careful review of this manuscript. We have re-examined the references. We have deleted this sentence and the references 41 and 42 according to your suggestion.

2. Response to comment: Line 401……Reference 44 is cited as evidence for typical care on low back pain during pregnancy. However, from all the studies included in this meta-analysis only one study used acupuncture intervention in the intervention group, You have neglected to mention that this study, while including only the one study of auricular acupuncture, showed significant
improvement in pain compared with standard care. Active placebo treatments are also somewhat effective, which is why verum acupuncture is not significantly greater. The right comparison is standard care, not other active sham treatments. We are all well aware that they demonstrate effectiveness of their own.

Response: Thank you very much for your careful review of this manuscript. We have re-examined the reference and re-written this part according to your suggestion. Revised portion are marked in red in the paper. (page 17 line 435-437)

3. Response to comment: Line 418 Where is says ...... ‘It is challenging for pregnant.....419 women to receive drug therapy during pregnancy because they feared that the drugs might be....420 harmful to the fetus.46’ Please amend this to include this to state that drug therapies can actually be harmful to pregnant women, not just that pregnant women feared this. It is important to recognise that pregnancy is a time where interventions are necessarily minimal, and should not include pharmaceuticals unless absolutely required. This is due to concerns of teratogenicity for the fetus and harm to the mother.

Response: Thank you for your essential advice. We have re-written this part according to your suggestion. Revised portion are marked in red in the paper. (page 17 line 450-452)

4. Response to comment: Line 425. Where is says….. ‘acupuncture significantly increase blood flow of local skin and muscle to relieve various pain’ (add an ‘s’ to increase – should be increases).

Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing. Revised portion are marked in red in the paper. (page 18 line 457)

5. Response to comment: Line 432…. Where is says …’Moreover, exercise and yoga as potentially non.......433 pharmacological therapies are difficult to keep on for pregnant women due to the inconvenience.10’ There is no need to introduce this here. This is not a review of other therapies, which may or may not be effective for LBPP. You have not made any comparisons with other therapies so far in the review, and you can promote the potential of acupuncture effectiveness without it.

Response: Thank you very much for your essential advice. There is no need to introduce this here. We have deleted this sentence according to your suggestion.

6. Response to comment: Line 434……where it says: ‘Importantly, No preterm labor or other adverse pregnancy outcomes were reported during..... 435 acupuncture. Although 3 participants in all studies experienced transient pain and/or redness during’

In this section, you should mention that some studies had high dropout rates, so we are not certain of some outcomes, or potential bias due to attrition. (also the ‘N’, where it appears in No preterm labor, does not need capitalisation).

Response: Thank you very much for your essential advice. We have stated the dropout rate according to your suggestion. Revised portion are marked in red in the paper. (page 18 line 470-474)

7. Response to comment: Line 449. Where it says….. ‘In addition, only one study....450 of all included studies referred the costs of acupuncture. We found that few pregnant women with....451 LBPP receive active treatment might, which might associate with high costs that patients cannot afford.34’ Please revise sentence – it does not make sense.

Response: Thank you very much for your essential advice. It does not make sense to introduce this here. We have deleted this sentence according to your suggestion.
1. Response to comment: line 173: “P < 0.1 of chi-square test...” should read “A chi-square test p-value <0.1...”.

Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing. We have made correction according to your suggestion. Revised portion are marked in red in the paper. (page 6 line 184-185)

2. Response to comment: The explanation for the four categories of heterogeneity adds some clarity to the manuscript. However, the specific I2 intervals described are overlapping. For example: 1. I2 = 35% could be either “might not important” or “moderate heterogeneity”; 2. I2 = 55% could be either “moderate heterogeneity” or “substantial heterogeneity”; 3. I2 = 80% could be either “substantial heterogeneity” or “considerable heterogeneity”. The source of this rough guide should be provided and an explanation for the overlapping I2 intervals should be provided.

Response: Thank you very much for your constructive suggestions. We are very sorry for our negligence of providing the guide. Considering the Reviewer’s suggestion, we have provided the guide in the manuscript. In terms of the overlapping I2, we did note the overlapping ranges and the equivocation (“may”). After all, this was a rough guide according to Cochrane Handbook for Systematic Reviews of Interventions, and all statistical approaches have limitations. However, we believed that a new edition of the Cochrane database of systematic reviews will update the guidance. Revised portion are marked in red in the paper. (page 6 line178 and 181-183) Additionally, a chi-square test p-value < 0.1 or I2 > 50% indicates statistically significant heterogeneity. Moreover, fixed effect models would be applied to combine results if I2 < 50%; If P < 0.1 or I2 > 50% we combined results using a random-effect model. These were demonstrated in the paper.

3. Response to comment: line 182: “An extended discussion of this option appears in Section.” Is missing something. What does “Section” mean?

Response: Thank you very much for your careful review of this manuscript. We are very sorry for our incorrect writing. This sentence should be deleted and we have deleted this sentence in the new manuscript.

VERSION 3 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Levett, Kate</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Western Sydney, National Institute of Complementary Medicine Research</td>
<td></td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>24-Apr-2022</td>
</tr>
</tbody>
</table>

GENERAL COMMENTS

Thank you for the opportunity to review this paper. It is thorough and well conceived. There are some minor issues with English expression and some wording. These are outlined as follows:

- English grammar/expressions for paper
- Throughout paper – use the word ‘birth’ instead of ‘delivery’ – e.g. preterm birth not preterm delivery
- Throughout paper – use the word ‘woman’ instead of ‘patient’
- Throughout paper – use the term ‘preterm’ contractions, not ‘premature’ contractions
- Throughout paper – Apgar scores are usually defined as >7 at 1 minute and 5 minutes, or <7 at 1 minute or 5 minutes (rather than 7-10). No time frames are provided here. Please report if the outcomes of >7 are at 1 minute or 5 minutes (5 mins being more relevant).
For whole paper – usually the following is expressed as: "this meta-analysis included 10 studies, reporting on a total of 1040 women.
Usually the expression is ‘evidence for….’ – rather than evidence in.

Specific line feedback.
Line 80 – wording – try: However, these therapies are not recommended according to the evaluation of results of clinical trials due to their unclear effectiveness
Line 93-95 – wording and concept – try: Additionally, some studies that report a significant improvement when acupuncture is compared with usual care, but no significant difference for relieving low back pain when acupuncture is compared with sham acupuncture. This highlight the importance of standardising control groups for acupuncture, as many sham treatments use an active placebo, distorting the validity of results (18, 19).
Line 95 – wording – try: Importantly, there are serious concerns about drug use for pain relief in pregnancy, and the side effects of these treatments. Acupuncture is a safe, non-pharmacological treatment option that is being increasingly used for relieving discomfort during pregnancy (1, 20).
Line 102-3 – the study you are referring to (24) did NOT suggest that high-quality trials are required to confirm the use of ‘forbidden points’ during pregnancy. It was recommending that due to the concern of high drop out rates from studies that use forbidden points, that alternative points should be considered in the treatment of LBPP, which have demonstrated effectiveness. It also suggested that studies collect and report on safety outcomes for trials involving pregnancy women. Please revise comment.
Line 108-9 – wording – high quality evidence ‘for’ acupuncture, not ‘on’ acupuncture
Line 151: grammar and punctuation (underlined): Study design: randomized controlled trials (plural) (RCT). Studies that met any of the following criteria would be excluded: (colon not full stop) 1. Pain etc etc.
Line 156: grammar – These included (not including) participant characteristics.....
Line 205: wording – 228 potentially eligible studies were identified (not acquired) by searching electronic databases....... Line 206-8 – wording/grammar – All studies were imported into EndNote X9 (ref) for detection of duplicates. Following that, 188 were removed through title and abstract screening, and the remaining 20 studies were screened through full text review, and 10 further studies were removed. Finally, 10 RCTs (22, 31-39) were included in the final analysis (figure 1).
Line 225 – grammar – excluded from analysis due to noncompliance (not noncompliant) in submitting pain diaries not due to the adverse events rate. (occurs again in Line 301)
Line 231 – line starts with ‘the’ – use capitalisation of first word.
Line 234-6 – grammar/wording – Two of the studies used subject (not plural) blinding, three studies blinded assessors, and none of the studies blinded the therapists.
Line 225 – do not need to capitalise ‘meta’ in meta-analysis
Line 263-4 – grammar – overall effects were recorded in four studies..............significant difference in overall effects when acupuncture was compared with other interventions or no intervention....
Line 268–9 – grammar/word ordering – difference in analgesic consumption during the study period when acupuncture was compared with no intervention……
Line 271–2 – grammar – Apgar scores >7 were recorded in four studies……………… no significant difference in Apgar scores of newborns when acupuncture was compared with other interventions or no intervention
THIS APPLIES TO WHOLE OF PAPER – It is difference IN outcomes, acupuncture compared to an outcome (ordering of words), and interventions (plural), or no intervention (singular)
ADVERSE EVENTS – Line 281 – grammar - All studies examined the adverse event rates, or all studies examined adverse events (no ‘the’)
Line 285 – word missing and punctuation – the most common being pain and slight bleeding at the needle sites, slight bruising……
Line314 – grammar – did not change when other studies were removed.
Line317 – grammar – difference in pregnant women’s age
Line 335 – grammar – related to the fact that some trials showing adverse effects from acupuncture may not be published and are not included……
Quality of evidence
Line 355 – grammar – None studies… contributed data for the outcome of pain (not 'with' data) – applies to rest of paper
Line 357 Low evidence for pain outcomes regarding the GRADE Applies to rest of paper – evidence 'for'; contributed data (not with data),
Apgar scores should be expressed as >7 (rather than 7-10) and were they measured at 1 minute or 5 minutes? Please report this

Conclusion
Line 394 – grammar – this meta-analysis included 10 studies, reporting on a total of 1040 women (conventional way to express meta-analysis inclusions).
No significant difference in analgesic consumption
Difference in Apgar scores should be reported as a safety measure. That acupuncture is a safe treatment and there was no difference in Apgar scores – rather than demonstrating a lack of effect.
Line 416 – grammar – With regard to Apgar scores, if researchers were able to capture this data, presumably by accessing patient records following birth as part of routinely collected data, it could feasibly be included in the outcome data.
Line 466 – Information included here is repeated from above. Pregnancy was a time where interventions were necessarily minimal………… harm to the mother’ (appears in paragraph above)
Discussion could be tidied up – is a bit inconsistent with reporting and flow of argument – requires review of flow by native English speaker
Is blinding of subjects and therapist required for treatment effectiveness – this methodological argument is old and the use of sham acupuncture is problematic.
Non-blinded trials (similar to those included in surgical trials or trials of psychological interventions) are not usually blinded. Why is acupuncture held to a different standard? See work from Claudia Witt and Hugh McPherson

THIS APPLIES TO WHOLE OF PAPER – It is difference IN outcomes, acupuncture compared to an outcome (ordering of words), and interventions (plural), or no intervention (singular)
Responds to the reviewer's comments:

Reviewer 1:

1. Response to comment: Throughout paper – use the word 'birth' instead of 'delivery' – e.g. preterm birth not preterm delivery
   Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 2 line 40, page 8 line 267, Table 2, page 16 line 386 and line 388, page 17 line 440)

2. Response to comment: Throughout paper – use the word 'woman' instead of 'patient'
   Response: We are very sorry for our improper writing, and we have made correction.
   Revised portion are marked in red in the paper. (page 9 line 277 and line 303)

3. Response to comment: Throughout paper – use the term 'preterm' contractions, not 'premature' contractions
   Response: We are very sorry for our improper writing, and we have made correction.
   Revised portion are marked in red in the paper. (page 2 line 41, page 8 line 268, Table 2, page 16 line 385 and 388)

4. Response to comment: Throughout paper – Apgar scores are usually defined as >7 at 1 minute and 5 minutes, or <7 at 1 minute or 5 minutes (rather than 7-10). No time frames are provided here. Please report if the outcomes of >7 are at 1 minute or 5 minutes (5 mins being more relevant).
   Response: We are very sorry for our improper writing, and we have made correction.
   Revised portion are marked in red in the paper. (page 1 line 31, page 5 line 138, page 8 line 257, and Table 1)

5. Response to comment: For whole paper – usually the following is expressed as: “this meta-analysis included 10 studies, reporting on a total of 1040 women.
   Response: We are very sorry for our improper writing, and we have made correction.
   Revised portion are marked in red in the paper. (page 1 line 33, and page 16 line 380)

6. Response to comment: Usually the expression is ‘evidence for….’ – rather than evidence in.
   Response: We are very sorry for our improper writing, and we have made correction.
   Revised portion are marked in red in the paper. (page 3 line 93, page 4 line 105, page 11 line 344, line 351, line 355, line 359, and line 363)

Specific line feedback.

1. Response to comment: Line 80 – wording – try: However, these therapies are not recommended according to the evaluation of results of clinical trials due to their unclear effectiveness.
   Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 3 line 76-77)

2. Response to comment: Line 93-95 – wording and concept – try: Additionally, some studies that report a significant improvement when acupuncture is compared with usual care, but no significant difference for relieving low back pain when acupuncture is compared with sham acupuncture. This highlight the importance of standardising control groups for acupuncture, as many sham treatments use an active placebo, distorting the validity of results (18, 19).
   Response: We are very sorry for our improper writing. There is no need to highlight this and we have deleted this sentence.

3. Response to comment: Line 95 – wording – try: Importantly, there are serious concerns about drug use for pain relief in pregnancy, and the side effects of these treatments. Acupuncture is a safe, non-pharmacological treatment option that is being increasingly used for relieving discomfort during pregnancy (1, 20).
   Response: We are very sorry for our improper writing, and we have made correction. Revised portion marked in red in the paper. (page 3 line 89-92)

Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 3 line 93)
5. Response to comment: Line 102-3 – the study you are referring to (24) did NOT suggest that high-quality trials are required to confirm the use of ‘forbidden points’ during pregnancy. It was recommending that due to the concern of high drop out rates from studies that use forbidden points, that alternative points should be considered in the treatment of LBPP, which have demonstrated effectiveness. It also suggested that studies collect and report on safety outcomes for trials involving pregnancy women. Please revise comment.
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 3 line 98-101)
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 4 line 105)
7. Response to comment: Line 151: grammar and punctuation (underlined): Study design: randomized controlled trials (plural) (RCT). Studies that met any of the following criteria would be excluded: (colon not full stop) 1. Pain etc etc.
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 5 line 141, 142)
8. Response to comment: Line 156: wording – These included (not including) participant characteristics…
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 5 line 146)
9. Response to comment: Line 205: wording – 228 potentially eligible studies were identified (not acquired) by searching electronic databases.
Line 206-8 – wording/grammar – All studies were imported into EndNote X9 (ref) for detection of duplicates. Following that, 188 were removed through title and abstract screening, and the remaining 20 studies were screened through full text review, and 10 further studies were removed. Finally, 10 RCTs (22, 31-39) were included in the final analysis (figure 1).
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 6 line 194, page 6 line 196 and 198)
10. Response to comment: Line 225 – grammar – excluded from analysis due to noncompliance (not noncompliant) in submitting pain diaries not due to the adverse events rate. (occurs again in Line 301)
Response: We are very sorry for our improper writing, and we have made correction. This sentence was deleted from Study Characteristics because it is repeated. Revised portion are marked in red in the paper. (page 9 line 286-287)
11. Response to comment:Line 231 – line starts with ‘the’ – use capitalisation of first word.
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 9 line 217)
12. Response to comment: Line 225 – do not need to capitalise ‘meta’ in meta-analysis
Line 263-4 – grammar – overall effects were recorded in four studies………….significant difference in overall effects when acupuncture was compared with other interventions or no intervention.…
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 8 line 241, line 249-251)
13. Response to comment: Line 268-9 – grammar/word ordering – difference in analgesic consumption during the study period when acupuncture was compared with no intervention……
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 8 line 254-255)
14. Response to comment: Line 271-2 – grammar – Apgar scores >7 were recorded in four studies………….no significant difference in Apgar scores of newborns when acupuncture was compared with other interventions or no intervention
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 8 line 258-259)
15. Response to comment: THIS APPLIES TO WHOLE OF PAPER – It is difference IN outcomes, acupuncture compared to an outcome (ordering of words), and interventions (plural), or no intervention (singular)
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 8 line 250-251, line 255, line 259, page 9 line 271 and 275, page 11 line 354, line 358 and 362)
16. Response to comment: Line 281 – grammar - All studies examined the adverse event rates, or all studies examined adverse events (no ‘the’) 
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 8 line 267)
17. Response to comment: Line 285 – word missing and punctuation – the most common being pain and slight bleeded at the needle sites, slight bruising……
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 9 line 271)
18. Response to comment: Line314 – grammar – did not change when other studies were removed. Line317 – grammar – difference in pregnant women’s age
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 9 line 300 and line 303)
19. Response to comment: Line 335 – grammar – related to the fact that some trials showing adverse effects from acupuncture may not be published and are not included……
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 10 line 321-322)
20. Response to comment: Line 355 – grammar – None studies… contributed data for the outcome of pain (not ‘with’ data) – applies to rest of paper
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 10 line 342, line 345, line 349, line 352, line 356, and line 356)
21. Response to comment: Line 357 Low evidence for pain outcomes regarding the GRADE
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 11 line 344, line 347, line 351, line 355, line 359, and line 363)
22. Response to comment: Line 394 – grammar – this meta-analysis included 10 studies, reporting on a total of 1040 women (conventional way to express meta-analysis inclusions).
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 16 line 380, page 8 line 304-305)
23. Response to comment: Difference in Apgar scores should be reported as a safety measure. That acupuncture is a safe treatment and there was no difference in Apgar scores – rather than demonstrating a lack of effect.
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 8 line 258)
24. Response to comment: Line 416 – grammar – With regard to Apgar scores, if researchers were able to capture this data, presumably by accessing patient records following birth as part of routinely collected data, it could feasibly be included in the outcome data.
Response: We are very sorry for our improper writing, and we have made correction. Revised portion are marked in red in the paper. (page 17 line 439-442)
25. Response to comment: Line 466 – Information included here is repeated from above. Pregnancy was a time where interventions were necessarily minimal………… harm to the mother’ (appears in paragraph above)
Response: Thank you very much for your careful review. We have deleted this sentence.
26. Response to comment: Discussion could be tidied up – is a bit inconsistent with reporting and flow of argument – requires review of flow by native English speaker
Response: Thank you very much for Reviewers’ careful review. Discussion has been tidied up by native English speaker. Her name is Pearl Tin and email address is pearltin@gmail.com. She has a background in Traditional Chinese Medicine.

27. Response to comment: Line 466 – Is blinding of subjects and therapist required for treatment effectiveness – this methodological argument is old and the use of sham acupuncture is problematic. Non-blinded trials (similar to those included in surgical trials or trials of psychological interventions) are not usually blinded. Why is acupuncture held to a different standard? See work from Claudia Witt and Hugh McPherson.

Response: Thank you very much for your valuable advice. We searched and studied Claudia Witt and Hugh McPherson’s relevant literatures. It’s very helpful for revising our paper, as well as the important guiding significance to our future researches. We have deleted them in the discussion part.

---

**VERSION 4 – REVIEW**

| REVIEWER             | Levett, Kate  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>University of Western Sydney, National Institute of Complementary Medicine Research</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>25-Jun-2022</td>
</tr>
</tbody>
</table>

**GENERAL COMMENTS**

BMJOpen – 2021-056878.R3_Review

Thank you for the opportunity to review the revision of this manuscript. The authors have done a very thoughtful and comprehensive systematic review, and attended to all issues raised.

I have some very minor typographical/grammatical suggestions, which do not require further review if the authors can attend to these minor suggestions.

1. Forest plots – no headings – what are the outcomes being presented here? This may be just the presentation of the review, but ensure that the headings for outcomes are included in the publication

Typos/grammar

1. SELECTION CRITERIA

- Outcomes: “……….Secondary outcomes were overall effects, analgesic consumption (using analgesic drugs as an adjunct for pain treatment during the study period), Apgar scores > 7 at 5 minutes (apgar scores were used to assess the physical condition of the newborns……”
- Include the measures of overall effect in brackets after the statement
- Apgar – has a capital ‘A’ as it’s a person’s name

ADVERSE EVENTS

4. Where a paper is cited, include the reference number

COMPARISON WITH OTHER REVIEWS

5. Line 410 – “….. the safety of acupuncture for newborns was not assessed…. “

It makes it sound like acupuncture was performed on the newborn. Perhaps you could say that the
“safety of maternal acupuncture on the outcomes for the newborn was not assessed....” Or something similar.