PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Sham treatment effects in manual therapy trials on back pain patients: a systematic review and pair-wise meta-analysis
AUTHORS	Lavazza, Carolina; Galli, Margherita; Abenavoli, Alessandra; Maggiani, Alberto

VERSION 1 – REVIEW

REVIEWER	van Tulder, Maurits University of Amsterdam, Health Sciences
REVIEW RETURNED	01-Nov-2020

GENERAL COMMENTS	The article is clearly written, but needs some language editing; there are several typos. Page 9, lines 31-32: “Placebo effect, also called placebo response, is the reported improvement in symptoms among patients in randomized controlled trials (RCT).” This is not an adequate definition of the placebo effect. The effect is not only in patients in RCTs. Please provide an adequate definition. Page 9, lines 47-48 “placebo might be influenced by the same therapist beliefs which are actively providing the inactive treatment.” The authors should elaborate a bit more on the challenges of developing or using a placebo for manual treatments. In pharmaceutical trials, a placebo ensures that both patient and clinician are blinded. Is it possible to have a doubel-blind placebo for manual treatments? If not, it is not a true placebo. Page 9, lines 50-53: physical placebo treatments might have a greater effect on patient reported outcomes, but the original intervention (in this case manual therapy) is also a physical treatment. Doesn’t that cancel each other out? Page 10, lines 10-11: “to analyse the effects, possible harm and the reliability of different kinds of sham treatments provided in RCTs involving MT”. Here the authors suggest that the focus is on sham treatments and not on placebo. Although it is a subtle difference, it is important. In physical treatments like manual therapy, it would be impossible to develop a true placebo. The best alternative is a sham treatment that mimics the active treatment (manual therapy) best. However, the therapists will not be blinded and cannot be blinded, which increases the risk of bias of the results, especially when subjective, patient-reported outcomes are being used. I would strongly advise to consistently use ‘sham treatment’ instead of ‘placebo’ throughout the text and title. Page 10, lines 55-57: physiotherapy and kinesiology are not types of

	manual therapy. I assume the authors mean manual therapy provided by a physiotherapist or kinesiologist? Page 12, risk of bias assessment: “Bias risk was assessed by CL and agreed by MG using the Cochrane Risk of bias (CRB) tool….Each possible risk was evaluated as “high”, “medium” or “low” by CL and a revision of the judgments was performed by MG.”. It is highly recommended two have two authors independently assessing the risk of bias and have a consensus meeting to discuss discrepancies. In this study, MG had a final say. The method used here is sub-optimal. There was only one assessor and the other was asked to agree. That is less valid than having two independent assessors. The authors should describe this as a limitation in the discussion section. Page 16, line 6: “All 24 studies included in this review were RCT.” Only RCTs were included, so this information is redundant. Just report that 24 studies (or 24 RCTs) were included. Page 16, lines 7-8: “Only arms involving the interested treatments were included in analyses.” This can be deleted, because this information has already been reported in the methods section. Please avoid duplications. Page 16, lines 10-11: “Three trials did not report where their were conducted.” Change into “Three trials did not report in which clinical setting they were conducted. ‘Where’ may also refer to the country, and that information is reported and provided in the next sentence. Page 16, lines 26-27: “and a percentage of women that ranged from 20% to 82%.” This can be deleted. If you know the percentage of men, the percentage of women is known as well. This information is redundant. Page 16, lines 33-35: “included participants with mechanical pain (described as pain exacerbated by movement”. Why is this reported separately? Mechanical pain is part of ‘unspecified cause of back pain’. Mechanical pain, pain execerbated by movement, doesn’t refer to the cause of the pain, just to the symptoms. Page 17, lines 24-25: “Different active treatments were provided”. I would suggest to change ‘active’ into ‘manual’. Manual therapy is by many considered an inactive/passive treatment. This is often debated and I would suggest to avoid that discussion here. Page 17, lines 40-41: “Just one trial using reflexology provided both active and inactive manipulation in a different zone. (37)”. Change ‘active manipulation’ into ‘manual therapy’ or ‘manual treatment’ and ‘inactive’ into ‘sham’. The comparison here is ‘sham’ treatment, so use this term consistently throughout the text. Page 17, lines 44-45: “generally defined physical therapists”. What are ‘generally defined physical therapists’? Probably cleared if you change this into ‘physical therapists’. Page 20, line 7: “Figure 2 shows risks of bias judged by two authors.” Delete ‘judged by two authors’. This information has

	already been provided in the methods section. Page 20, lines 14-22: This is the most important finding of this study. This shows that it is very difficult and maybe impossible to develop a true placebo or sham treatment for manual therapy. The main reason for having a placebo/sham treatment is to reduce the risk of bias by blinding participants. The authors should elaborate on this in the discussion session. On page 27, lines 3-6, the authors write "Future researches should also evaluate the real effects of placebo comparing it both with active treatment and to control groups. Only with this kind of design the real placebo effect in MT could be defined." What I miss in the discussion is a broader perspective. There have been many trials on manual therapy for low back pain (see, for example, Rubinstein et al. Benefits and harms of spinal manipulative therapy for the treatment of chronic low back pain: systematic review and meta-analysis of randomised controlled trials. BMJ 2019;364:I689). The effectiveness of manual therapy is still uncertain because there are no truly placebo controlled trials. Rubinstein et al suggested that it is debatable whether studies that examine the effect of SMT compared with non-recommended therapies or sham (placebo) therapies will add further to our understanding. The effect of manual therapy seems small; will development of a better placebo a better and larger studies add anything to our knowledge? Page 23, placebo vs. control. Please explain why this comparison is relevant in a study evaluating placebo effect in manual therapy. Please describe the types of control. The authors should also explain how this comparison helps answering that question. Page 23, line 39: "Of the six studies reporting AE, only two compared placebo and control." Change this into "of the five studies comparing placebo and control, only two reported AE." Page 24, lines 7-9: "In the treatment of back pain very low quality of evidence suggests a slight improvement of pain, not clinically meaningful, in favour of MT at short-term." I would suggest to change this statement as it doesn't reflect the results optimally. I would suggest changing this into "Results show a small, not clinically meaningful effect in favour of MT for short-term pain relief compared with sham treatment. However, the quality of evidence is very low, indicating that the true effect is probably markedly different from the estimated effect."

REVIEWER	Paolucci, Teresa Università G D'Annunzio di Chieti
REVIEW RETURNED	08-Nov-2020

GENERAL COMMENTS	Dear Authors, the topic is intriguing and very interesting. The paper develops well. In my opinion, remains a basic problem that should be introduced in more depth in the introduction and then taken up and discussed adequately (in the discussion section). From this premise the title should be reshaped. 1. The question is: how could one define the "placebo effect" in rehabilitation? even a light touch (can a light touch massage be assimilated to a placebo?). For instrumental physical therapy it is simpler (medical device off / on) but for example, we are sure that a "Neck massage" or "Soft-tissue manipulation of the foot" or "Intermittent

	traction of the spine "is PLACEBO? for example, Soft-tissue manipulation of the foot could have an analgesic reflex effect on pain, why not. The above is not a problem with your review, but it is a problem with studies that perhaps define placebo as something that is not. I strongly believe that this debate should be included in the introduction and addressed in discussion 2. How do you define "Manual therapy"? are you referring to manipulative therapy or to all those techniques within the rehabilitation approaches that use contact techniques regardless? I ask you this because different rehabilitation approaches that are not part of manipulative therapy such as "Kinesiology" or "Physiotherapy" are put on the same level. So I suggest to better address the definition of "Manual Therapy" in the introduction to more coherently introduce the purpose of the review, or what you intend to include in "Manual Therapy" and why, with the appropriate bibliographic references.

REVIEWER	Brox, Jens Oslo University Hospital, Phys med & rehab
REVIEW RETURNED	08-Nov-2020

GENERAL COMMENTS	This is a systematic review conducted according to the PRISMA check list. It is supported by a PRISMA flow diagram, include independent reviewers, Cochrane Risk of bias tool, funnel plot and assessment of heterogenety. 24 trials with 2019 participants are included. A clinically meaningless difference favoring MT was found for pain, no difference for other outcomes. Blinding was not successful, suggesting that there was a bias favoring MT. Nevertheless, the authors conclude that there is very low quality evidence for these findings. This is a typical technical Cochrane language. With > 2000 participants, no effectiveness of MT versus sham MT or control and bias likely to favor MT, a more pragmatic conclusion should omit the Cochrane language and conclude that according to the current state of knowledge MT is not effective compared with sham MT or control. This is slightly different from the conclusion in the European Guideline for Low back pain published 14 years ago but in agreement with more recent studies published in Lancet (Hancock) and this year in JAMA Open comparing MT to other types of placebo (detuned ultrasound) or control. To me it seems meaningless to make such a conservative conclusion in the current study when both results and bias point to no difference. The study has included both neck and low back pain and this should be clearly stated in the abstract and in the introduction. Studies are merged which may be criticized but acceptable in my opinion. In the abstract correct detecting range from 47 to 84%, suggesting that studies may be biased in favor of MT. I presume that expectations for MT is higher than for sham MT. The authors label assessment of blinding as reliability which is not incorrect because it is a part of the reliability and validity assessment of a study. In their SR, blinding is the only aspect of reliability assessed and it is more precise to label this success of blinding. Pain is the primary outcome and is reported on visual analogue scales. The metric scale is just an analogue of pain intensity and meaningless itself. I would prefer to label it pain intensity and delete the metrics (mm). Additionally, the authors state that the minimal

	important difference is 30, which is probably too high, but the observed mean differences are much lower. They do not report the number of patients with MID > 30 which is probably not available from the individual studies. The MID of 30 is therefore not really applied in the current study. The focus of successful blinding is an advantage of the study. The authors state that future studies should develop reliable kinds of placebo. With the nature of MT I think that this is very difficult to obtain but it might not be impossible. The meta analyses merge the small individual studies and larger individual studies in my opinion are most likely to report changes within the confidence intervals reported in this SR. Are placebo effect and placebo response identical (see Kaptchuk et al, BMJ 2020)? The introduction about placebo might be somewhat revised. On page 12, lines 42-44, which subgroup and sensitivity analyses? Page 14, line 20, please replace the majority with the exact number of studies. Page 16, line 49, N=11, most of the studies? Separate analyses for neck and low back pain as a sensitivity analysis? Does MCID or MID, which may be the preferred short form from 2020, indicate patients values and preferences? Is it not just the difference between unchanged and slightly improved when by example a balanced cut-off of sensitivity and specificity is derived from ROC analysis of two subjective scores, VAS pain and global change. The value of MID is to translate the mean outcomes of a study into clinical practice, by example would my patient pay for manual therapy if the likely benefit is 4 or 6 on a scale from 0 to 100 and how can you explain this difference to the patient when the individual measurement error is about 15? The latter do not represent a to the point critic of the current study but on the other hand - is it likely that future studies would find a difference > 15 favoring MT? This is three times the observed difference and besides higher quality future studies are likely to be less biased and actually reduce the observed difference.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

The article is clearly written, but needs some language editing; there are several typos.

Dear reviewer, thank you for your feedback and to have spent your time reading our work. We really appreciated your precise review, we have gone through your comments and we hope that our changes meet your expectations.

You have highlighted different interesting points of discussion which we have tried to assess in our revised version, especially in introduction and discussion sections.

Page 9,lines 31-32: “Placebo effect, also called placebo response, is the reported improvement in symptoms among patients in randomized controlled trials (RCT).” This is not an adequate definition of the placebo effect. The effect is not only in patients in RCTs. Please provide an adequate definition.

We agree with you comment, we have update this information.
Page 9, lines 47-48 “placebo might be influenced by the same therapist beliefs which are actively providing the inactive treatment.” The authors should elaborate a bit more on the challenges of developing or using a placebo for manual treatments. In pharmaceutical trials, a placebo ensures that both patient and clinician are blinded. Is it possible to have a double-blind placebo for manual treatments? If not, it is not a true placebo.

This is a very interesting point, we agree with your opinion and we have tried to develop this concept better in the introduction. We hope we have adequately addressed your concerns.
Page 9, lines 50-53: physical placebo treatments might have a greater effect on patient reported outcomes, but the original intervention (in this case manual therapy) is also a physical treatment. Doesn’t that cancel each other out?
We might have misunderstand your comment, so, please, our apologies if we are not replying appropriately. What we meant is that if physical placebo is suggested to influence PROs, manual therapy that uses a physical contact could potentially have the same influence on PROs.
Page 10, lines 10-11: “to analyse the effects, possible harm and the reliability of different kinds of sham treatments provided in RCTs involving MT”. Here the authors suggest that the focus is on sham treatments and not on placebo. Although it is a subtle difference, it is important. In physical treatments like manual therapy, it would be impossible to develop a true placebo. The best alternative is a sham treatment that mimics the active treatment (manual therapy) best. However, the therapists will not be blinded and cannot be blinded, which increases the risk of bias of the results, especially when subjective, patient-reported outcomes are being used. I would strongly advise to consistently use ‘sham treatment’ instead of ‘placebo’ throughout the text and title.
Thank you for your comment, this is a very interesting point. We agree with your opinion, placebo can not be achieved in manual therapy trials due to the lack of blinding of therapists. We apologize for this mistake but we have decided to use this terminology because it was the one most used in MT trials. As a matter of fact only very few trials used the definition “sham treatment”. We have replaced the word placebo with sham therapy in all the paper.
Page 10, lines 55-57: physiotherapy and kinesiology are not types of manual therapy. I assume the authors mean manual therapy provided by a physiotherapist or kinesiologist?
Yes, thank you for pointing it out.

Page 12, risk of bias assessment: “Bias risk was assessed by CL and agreed by MG using the Cochrane Risk of bias (CRB) tool….Each possible risk was evaluated as “high”, “medium” or “low” by CL and a revision of the judgments was performed by MG.”. It is highly recommended two have two authors independently assessing the risk of bias and have a consensus meeting to discuss discrepancies. In this study, MG had a final say. The method used here is sub-optimal. There was only one assessor and the other was asked to agree. That is less valid than having two independent assessors. The authors should describe this as a limitation in the discussion section.

We agree with your opinion, MG was the one that had the last say. We understand your point of view and we will report it in the discussion section.

Page 16, line 6: “All 24 studies included in this review were RCT.” Only RCTs were included, so this information is redundant. Just report that 24 studies (or 24 RCTs) were included.

Page 16, lines 7-8: “Only arms involving the interested treatments were included in analyses.” This can be deleted, because this information has already been reported in the methods section. Please avoid duplications.

Page 16, lines 10-11: “Three trials did not report where their were conducted.” Change into “Three trials did not report in which clinical setting they were conducted. ‘Where’ may also refer to the

country, and that information is reported and provided in the next sentence.

Page 16, lines 26-27: “and a percentage of women that ranged from 20% to 82%.” This can be deleted. If you know the percentage of men, the percentage of women is known as well. This information is redundant.

Page 16, lines 33-35: “included participants with mechanical pain (described as pain exacerbated by movement”. Why is this reported separately? Mechanical pain is part of ‘unspecified cause of back pain’. Mechanical pain, pain exacerbated by movement, doesn’t refer to the cause of the pain, just to the symptoms.

Page 17, lines 24-25: “Different active treatments were provided”. I would suggest to change ‘active’ into ‘manual’. Manual therapy is by many considered an inactive/passive treatment. This is often debated and I would suggest to avoid that discussion here.

Page 17, lines 40-41: “Just one trial using reflexology provided both active and inactive manipulation in a different zone. (37)”. Change ‘active manipulation’ into ‘manual therapy’ or ‘manual treatment’ and ‘inactive’ into ‘sham’. The comparison here is ‘sham’ treatment, so use this term consistently throughout the text.

Page 17, lines 44-45: “generally defined physical therapists”. What are ‘generally defined physical therapists’? Probably cleared if you change this into ‘physical therapists’.

Page 20, line 7: “Figure 2 shows risks of bias judged by two authors.” Delete ‘judged by two authors’. This information has already been provided in the methods section.

Thank you for your suggestions. We have amended the text as you requested.
Page 20, lines 14-22: This is the most important finding of this study. This shows that it is very difficult and maybe impossible to develop a true placebo or sham treatment for manual therapy. The main reason for having a placebo/sham treatment is to reduce the risk of bias by blinding participants. The authors should elaborate on this in the discussion session. On page 27, lines 3-6, the authors write “Future researches should also evaluate the real effects of placebo comparing it both with active treatment and to control groups. Only with this kind of design the real placebo effect in MT could be defined.” What I miss in the discussion is a broader perspective. There have been many trials on manual therapy for low back pain (see, for example, Rubinstein et al. Benefits and harms of spinal manipulative therapy for the treatment of chronic low back pain: systematic review and meta-analysis of randomised controlled trials. BMJ 2019;364:I689). The effectiveness of manual therapy is still uncertain because there are no truly placebo controlled trials. Rubinstein et al suggested that it is debatable whether studies that examine the effect of SMT compared with non-recommended therapies or sham (placebo) therapies will add further to our understanding. The effect of manual therapy seems small; will development of a better placebo a better and larger studies add anything to our knowledge?

This is very interesting point. We have tried to develop this concept in the discussion section.
The trials included in this review used an inappropriate design, with improper sample size. With these assumptions, even if an important effect for patients was present, those types of designs would have not been able to see it. Moreover the research questions were often not clinically relevant, which also raises some ethical concern.
This is why we strongly suggest to future investigators to find clinically meaningful research questions and power a trial that takes into account an effect size which is significant for participants. This could help to evaluate the real clinical effect, both if an effect is present or not. With the information collected until now, we are not able to reply with confidence to your question.

Page 23, placebo vs. control. Please explain why this comparison is relevant in a study evaluating placebo effect in manual therapy. Please describe the types of control. The authors should also explain how this comparison helps answering that question.
We apologize for this missing information, we have reported it in the methods section.

Page 23, line 39: “Of the six studies reporting AE, only two compared placebo and control.” Change this into “of the five studies comparing placebo and control, only two reported AE.”
We have correct it, thank you.

Page 24, lines 7-9: “In the treatment of back pain very low quality of evidence suggests a slight improvement of pain, not clinically meaningful, in favour of MT at short-term.” I would suggest to change this statement as it doesn’t reflect the results optimally. I would suggest changing this into “Results show a small, not clinically meaningful effect in favour of MT for short-term pain relief compared with sham treatment. However, the quality of evidence is very low, indicating that the true effect is probably markedly different from the estimated effect.”
We think that your sentence is more readable and clear than ours. We have used it in the results section. Thank you for your help.

Reviewer 2
Dear Authors, the topic is intriguing and very interesting. The paper develops well.
In my opinion, remains a basic problem that should be introduced in more depth in the introduction and then taken up and discussed adequately (in the discussion section). From this premise the title should be reshaped.
Dear reviewer, thank you for your revision and time spent on our project. We really appreciated your feedback and we think you have underlined very interesting points.
We agree with your opinion and have reshaped our project following your suggestions and those of other reviewers. We have used a more correct terminology,(implemented????) especially in the introduction and the discussion section. We apologize for the wrong use of the term “placebo” but we used the terminology most used in manual therapy trials.
1. The question is: how could one define the "placebo effect" in rehabilitation? even a light touch (can a light touch massage be assimilated to a placebo?). For instrumental physical therapy it is simpler (medical device off / on) but for example, we are sure that a "Neck massage" or "Soft-tissue manipulation of the foot" or "Intermittent
traction of the spine "is PLACEBO? for example, Soft-tissue manipulation of the foot could have an analgesic reflex effect on pain, why not.
The above is not a problem with your review, but it is a problem with studies that perhaps define placebo as something that is not.
I strongly believe that this debate should be included in the introduction and addressed in discussion
Thank you for your comment, this is a very difficult and interesting point to address and it is definitely the key point of our research. We have tried to clarify this concept in the introduction and discussion sections. We hope that those changes address your requests.
2. How do you define "Manual therapy"? are you referring to manipulative therapy or to all those techniques within the rehabilitation approaches that use contact techniques regardless?
I ask you this because different rehabilitation approaches that are not part of manipulative therapy such as "Kinesiology" or "Physiotherapy" are put on the same level.
So I suggest to better address the definition of "Manual Therapy" in the introduction to more coherently introduce the purpose of the review, or what you intend to include in "Manual Therapy" and why, with the appropriate bibliographic references.
We agree with you that we did take for granted this definition. Although physiotherapy and kinesiology not involve hand contact, some physical therapists and kinesiologist use hand contact techniques. We have included all trials where manual therapy, expressed as hand contact therapy, provided by a physical therapist. We have tried to clarify this information in the introduction and in methods sections.

Reviewer 3

1. This is a systematic review conducted according to the PRISMA check list. It is supported by a PRISMA flow diagram, include independent reviewers, Cochrane Risk of bias tool, funnel plot and assessment of heterogenety. 24 trials with 2019 participants are included. A clinically meaningless difference favoring MT was found for pain, no difference for other outcomes. Blinding was not successful, suggesting that there was a bias favoring MT. Nevertheless, the authors conclude that there is very low quality evidence for these findings. This is a typical technical Cochrane language. With > 2000 participants, no effectiveness of MT versus sham MT or control and bias likely to favor MT, a more pragmatic conclusion should omit the Cochrane language and conclude that according to the current state of knowledge MT is not effective compared with sham MT or control. This is slightly different from the conclusion in the European Guideline for Low back pain published 14 years ago but in agreement with more recent studies published in Lancet (Hancock) and this year in JAMA Open comparing MT to other types of placebo (detuned ultrasound) or control. To me it seems meaningless to make such a conservative conclusion in the current study when both results and bias point to no difference.

We would like to thank you for taking the time and effort to read our work and give us your feedback. We really appreciate your feedback and we hope that our replies address your concerns.We have also tried to reshape the results in order to highlight the meaningless results found. We hope our findings are clearer now.

2. The study has included both neck and low back pain and this should be clearly stated in the abstract and in the introduction. Studies are merged which may be criticized but acceptable in my opinion.

Thank you for your comment. We agree with you, this could be criticized. Nevertheless this clinical difference did not have an impact on studies' heterogeneity so we really think that this difference did not have an impact on the overall meta-analysis and on review results.

3. In the abstract correct detecting range from 47 to 84%, suggesting that studies may be biased in favor of MT. I presume that expectations for MT is higher than for sham MT. The authors label assessment of blinding as reliability which is not incorrect because it is a part of the reliability and validity assessment of a study. In their SR, blinding is the only aspect of reliability assessed and it is more precise to label this success of blinding.

Thank you for your suggestion - we have renamed this outcome variable.

4. Pain is the primary outcome and is reported on visual analogue scales. The metric scale is just an analogue of pain intensity and meaningless itself. I would prefer to label it pain intensity and delete the metrics (mm). Additionally, the authors state that the minimal important difference is 30, which is probably too high, but the observed mean differences are much lower. They do not report the number of patients with MID > 30 which is probably not available from the individual studies. The MID of 30 is therefore not really applied in the current study.

We apologize but in this case we have a different opinion. We have used the metrics because we think it is a more precise description of our outcome so we would like to leave it as it is. As you supposed, the values for individual studies were not available. Nevertheless MID was used to interpret review findings and to assess whenever a clinically important effect was found.

5. The focus of successful blinding is an advantage of the study. The authors state that future studies should develop reliable kinds of placebo. With the nature of MT I think that this is very difficult to obtain but it might not be impossible. The meta analyses merge the small individual studies and larger individual studies in my opinion are most likely to report changes within the confidence intervals reported in this SR.

We agree with your point that success of blinding in manual therapies trials is very difficult to achieve. Nevertheless we think that different strategies could be implemented to increase success of blinding.

For instance, different studies included in this review have analyzed the effect of a single technique on pain. This is clinically and ethically questionable - what effect do you want to achieve with a single technique performed? This does not represent the real clinical context. A manual treatment usually includes different techniques (that could be of the same nature (as HVLA) or different (such as soft-tissue, articular techniques etc). So what clinical significance has a trial to assess the effect of a single technique compared to sham on patient pain immediately after its performance? It follows that providing a sham therapy that guarantees blinding is definitely more difficult. Instead if researchers investigate a treatment closer to clinical context, such as different techniques in a single session, this could help in increasing success of blinding. To verify this hypothesis, investigators should investigate the reliability of blinding as study outcome. Additionally, although blinding of therapists can not be achieved, investigators should implement strategies to guarantee the reliability of data collection and data analysis such as blinding of outcome assessors and statistician.

6. Are placebo effect and placebo response identical (see Kaptchuk et al, BMJ 2020)? The introduction about placebo might be somewhat revised.
We have modified the introduction. We hope that this concept is clearer now.
7. On page 12, lines 42-44, which subgroup and sensitivity analyses?
We apologize for this missing information. Our chapter highlights only the additional analyses actually performed in order to avoid any impact on word count. Nevertheless, it is possible to find this information in the registered protocol available online.
Here is our pre-planned analysis in detail:
“Analysis of subgroups or subsets
If data are available, the statistically significant presence of heterogeneity will be explored further with a subgroup analysis. The analysis will be performed on the following:
• Patient characteristics (type of back pain – cervical, dorsal or lumbar)
• Type of intervention (singular or multiple techniques, type of manual therapy applied, number of sessions)
• Type of placebo (similar or dissimilar to the comparator)
• Follow-up duration (short, medium or long-term)”

8. Page 14, line 20, please replace the majority with the exact number of studies.
We apologize for this missing information. We have not found this on page 14, we think you mean page 16 line 20. We have reported the information you requested.
9. Page 16, line 49, N=11, most of the studies?
We apologize for this error and we have corrected this information.
10. Separate analyses for neck and low back pain as a sensitivity analysis?
Thank you for highlighting this - it is a very interesting point to debate. Per protocol, we decided to perform this analyses only if statistically significant heterogeneity was found.
In our project we found high heterogeneity levels related to methodological diversity of the trials included, not related to their clinical differences. Therefore, to increase the precision of the estimate made, we decided to include all types of back pain together increasing the sample size of the meta-analysis performed.
11.. Does MCID or MID, which may be the preferred short form from 2020, indicate patients values and preferences? Is it not just the difference between unchanged and slightly improved when by example a balanced cut-off of sensitivity and specificity is derived from ROC analysis of two subjective scores, VAS pain and global change. The value of MID is to translate the mean outcomes of a study into clinical practice, by example would my patient pay for manual therapy if the likely benefit is 4 or 6 on a scale from 0 to 100 and how can you explain this difference to the patient when the individual measurement error is about 15? The latter do not represent a to the point critic of the current study but on the other hand - is it likely that future studies would find a difference > 15 favoring MT?
This is three times the observed difference and besides higher quality future studies are likely to be

less biased and actually reduce the observed difference.
We definitely agree with your point of view. We have decided to include MID to highlight any significant change that could be clinically meaningful for patients. However, the small effect found was not clinically significant. We think that a good starting point to detect if there is any difference between these two interventions is to implement a proper sample size calculation using MID. If investigators use a superiority design, the sample size calculation should be implemented using a clinically meaningful effect size. This could increase the chances of seeing any clinical effect, if present. In our review we found only 2 studies that performed this type of sample size calculation.

VERSION 2 – REVIEW

REVIEWER	van Tulder, Maurits University of Amsterdam, Health Sciences
REVIEW RETURNED	12-Feb-2021

GENERAL COMMENTS	The manuscript is clear, but needs some language editing. Abstract, methods: First time RCT is used, provide full description "randomized controlled trials (RCTs)". Abstract, conclusion: If MT is not clinically more effective than sham treatment, shouldn't the conclusion be that MT doesn't seem to have a clinically relevant effect? The authors research question is "To assess the effects and reliability of sham procedures in manual therapy (MT) trials…" However, sham treatments are methods used in clinical trials to help researchers determine the effectiveness of a treatment. Sham treatments are 'fake' treatments in which the care provider goes through the motions without actually performing the treatment. So the effectiveness of sham treatments can only be established by comparing it to 'no treatment'. Limitations: "This study did not include a comparison with machine provided placebo, its aim focused on hand contact sham treatment." I don't consider this an adequate limitation. What is a 'machine provided placebo'? How could you develop a machine provided placebo for spinal manipulation? Limitations: "Insufficient number of studies were included to conduct a network meta-analysis." Why would a network meta-analysis be better? The effectiveness of sham treatment will not be more clear if you compare it to other comparators. Introduction, page 7 first paragraph: the authors provide a definition of placebo: "an inert substance or sham procedure that is provided to research participants with the aim of making it impossible for them, and usually the researchers themselves, to know who is receiving an active or inactive intervention." I would suggest adding "Placebo interventions are methodological tools used to treat participants in the study arm and the control arm in exactly the same way, except that the study group receives an active substance and the control group does not. A true placebo doesn't exist for manual therapy, and testing the effectiveness of manual therapy requires a sham intervention." Page 7, second paragraph: the authors state "Since a placebo has

	no inherent therapeutic power, it rarely cures the disease". I would suggest changing this into "Since a placebo has no inherent therapeutic power, it cannot cure a disease". Page 7: the authors suggest that "where other important characteristics should be considered as part of this therapeutic context such as the tactile interaction between patient and practitioner and clinician beliefs.". However, the sentence above they suggest that also in pharmacological studies "individual patient and clinician factors (e.g. beliefs,.." also play a role. These characteristics are the same, so 'other important characteristics' doesn't seem adequate. Page 7: the authors suggest that "in MT trials, the blinding of clinicians is almost impossible to achieve". I would suggest deleting 'almost'. I cannot think of a way of blinding clinicians for MT. Page 7: the authors state that a sham treatment "aims to ensure at least the blinding of participants". I would suggest changing this into "aims at blinding of participants". Page 9, second sentence: "This review involved all types of placebo that include hand contact". I would suggest using 'sham treatment' consistently throughout the manuscript instead of 'placebo'. Page 9, third sentence: "Studies where placebo was provided by machines (such as inactive ultrasound) were excluded". The authors should make a stronger statement that many trials of MT that suggest having a placebo control are misleading. They may use an inactive ultrasound machine, which would be a placebo in a trial on the effectiveness of ultrasound, but is not an adequate placebo for MT. Therefore, the studies were excluded. Page 9: the authors state that "ST was compared to other manual therapies such as: physiotherapy, chiropractic, osteopathy, massage, kinesiology and reflexology and to control." Physiotherapy is not 'another manual therapy'. I would suggest changing this into " ST was compared to other manual therapies provided by any type of health care provider such as: physiotherapist, chiropractor, osteopath, massage therapist, kinesiologist and reflexologist." Also, delete 'and to control'. I don't think that comparing ST to any non-MT control would add anything to answering the research question. This also refers to the next sentence "To assess if touch itself could have a positive health effect, ST was also compared to control." Delete this sentence or change 'to control' into 'to no treatment', although I would be surprised if there are any studies comparing sham MT to no treatment. Page 13, population: "16 trials enrolled participants with low back pain (LBP), nine included participants with cervical pain (CP)." Please add references of the studies on CP. Page 13, interventions: "Generally the trials used a single therapy session (N=11) with a single technique performed (N=8)." Eleven out of 24 trials is not really 'generally'. I would suggest stating "Eleven trials used a singly therapy session with a single technique performed in eight of those trials." Page 14: "Different manual treatments were provided: Physiotherapy (2 trials, 288 participants)". Physiotherapy is not a manual treatment.

	Manual treatment can be part of a physiotherapy treatment, but a physiotherapy treatment typically includes exercise therapy and other treatments as well. Page 14: “Five trials with multiple arms compared ST to control group”. Describe what type of control groups were used in these studies. Page 14: “Most of the trials involved physiotherapists (N=8), physical therapists (N=4), osteopaths (N=3) and students (N=1).” Delete “Most” – 8 out of 24 trials is not the majority – and just state that 8 trials involved physios etc.. Also, what type of students, physiotherapy, chiropractic, osteopathy? Table 2 and page 20: table 2 reports the results of ST vs MT on pain improvement as an MD of 3.86 higher (3.29 higher to 4.43 lower). I assume that this should be 4.43 higher? Discussion: here the authors use “sham treatment” while in the methods and results section “ST” is used. Please be consistent in using ST or sham treatment. Discussion, implications for research and practice: the authors state that “Future studies should address meaningful research question…”. What do the authors mean. What would be a meaningful research question? Discussion: the discussion is still a bit weak. The objective of this review was to evaluate the effect, reliability and safety of sham manual therapy. The review shows that there is no good sham procedure that has proven to be effective, reliable and safe. Having a good sham procedure could help understanding the true effect of the intervention. If the intervention has a ‘true’ effect, the total effect should be larger that the effect of a sham procedure. However, there are many randomized controlled trials that have evaluated the effectiveness of manual therapy to sham procedures, no treatment, ‘inactive’ treatments that have not shown a clinically relevant effect in favour of manual therapy. How does this review relate to all the other reviews on effectiveness of manual therapy? What can we learn from that and what would be the best step to take to improve treatment for back pain patients? I still miss a broader perspective in the discussion.

REVIEWER	Brox, Jens Oslo University Hospital, Phys med & rehab
REVIEW RETURNED	10-Feb-2021

GENERAL COMMENTS	The manuscript is considerably improved and the answers from reviewers properly answered in my view. I have responded yes to all questions on the check list but I am not fully satisfied. I accept some disagreement and have hopefully made my critics clear in the first review, still I suggest that the authors conduct a minor revision. In agreement with reviewer 1 it is my opinion that the suggestions about future studies on the sham effect have got too much focus. Given the methodological problems, the likely bias favoring manipulative therapy, the very small and clinically meaningless difference observed, the number of studies already conducted, and the effort to conduct future studies that are likely to be biased, the authors should focus less on this part of their review.They should

	mainly focus on their findings, not on their recommendation for future studies. The discussion about the bias that are likely to favor manipulative therapy is missing and not mentioned in any of the conclusions. Lack of blinding in many patients is likely to have negative impact on outcome in the control and sham groups. This should be clearly outlined. I do disagree in the first and new part of the first paragraph in the discussion which has been suggested by reviewer 1. How do we know that this study indicates that the true effect is probably markedly different from the estimated effect? I would change the wording, giving a more conservative and balanced interpretation: " suggesting that the true effect may be different from the estimated effect." Probably reviewer 1 would agree. I disagree about the metrics used in the assessment of pain but I accept that the authors use it. Pain intensity as measured on an analogue scale do not improve precision. Pain can by example be measured on a 0 to 100 mm scale or on a scale simply with numbers from 1 to 10 and the measurement error is about the same, 15-20 or 1,5 to 2. There are many questions concerning pain assessment, by example worst pain, average pain, pain last week, pain today etc. It is not mm precise estimate, and at least not 3,87 - on a 0 to 100 scale. I do agree that the main advantage of MCID or MID is the sample size calculation in future studies but we still do not know if the difference from unchanged to slightly improved (that is the definition) is important and we have to consider the individual measurement error of the questionnaire or VAS because results are to be used in individual patients for shared decision making. There are still some minor typos, by example the double commas on line 15 and the comma space on line 17, page 11; lack of space before (on line 56 page 12 and line 12 page 21 and line 32 and 37 and 42; page 23 line 30, replace reliability by lack of blinding; line 3 on page 24, space after not, page 25 line 3, questions. Please check the manuscript carefully.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3
Dr. Jens Brox, Oslo University Hospital
Comments to the Author:
The manuscript is considerably improved and the answers from reviewers properly answered in my view. I have responded yes to all questions on the check list but I am not fully satisfied. I accept some disagreement and have hopefully made my critics clear in the first review, still I suggest that the authors conduct a minor revision.

Dear Dr. Brox, thank you very much for spending time in revising our work. We really apologize but we misunderstood some of your comments in the first place. We hope that this latest revised copy addresses better your concerns.

In agreement with reviewer 1 it is my opinion that the suggestions about future studies on the sham effect have got too much focus. Given the methodological problems, the likely bias favoring manipulative therapy, the very small and clinically meaningless difference observed, the number of studies already conducted, and the effort to conduct future studies that are likely to be biased, the authors should focus less on this part of their review. They should mainly focus on their findings, not on their recommendation for future studies.

We definitely agree with your point of view. Our aim was to give methodological guidance for the development of RCT in MT therefore we decided to put more attention on the “implication for researches” section. Nevertheless we understood that this chapter was too mono-focused and we have tried to develop further other discussion chapters. We hope that this version is closer to your suggestions.

The discussion about the bias that are likely to favor manipulative therapy is missing and not mentioned in any of the conclusions. Lack of blinding in many patients is likely to have negative impact on outcome in the control and sham groups. This should be clearly outlined.

We definitely agree with your opinion. We did not analyse adequately this part. We have updated both the discussion and conclusion chapters. We feel our revision is enough to address this issue.

I do disagree in the first and new part of the first paragraph in the discussion which has been suggested by reviewer 1. How do we know that this study indicates that the true effect is probably markedly different from the estimated effect? I would change the wording, giving a more conservative and balanced interpretation: " suggesting that the true effect may be different from the estimated effect." Probably reviewer 1 would agree.

Thank you for pointing this out we agree with you and we have corrected the sentence accordingly.

I disagree about the metrics used in the assessment of pain, but I accept that the authors use it. Pain intensity as measured on an analogue scale do not improve precision. Pain can by example be measured on a 0 to 100 mm scale or on a scale simply with numbers from 1 to 10 and the measurement error is about the same, 15-20 or 1,5 to 2. There are many questions concerning pain assessment, by example worst pain, average pain, pain last week, pain today etc. It is not mm precise estimate, and at least not 3,87 - on a 0 to 100 scale.

Your point of view here is compelling., But, respectfully, we prefer to retain the text as is since the metric does not influence our findings.

I do agree that the main advantage of MCID or MID is the sample size calculation in future studies but we still do not know if the difference from unchanged to slightly improved (that is the definition) is important and we have to consider the individual measurement error of the questionnaire or VAS because results are to be used in individual patients for shared decision making.

Thank you for this insight too. We certainly agree with you that studies are needed to assess which size of improvement is important especially for acute back pain. We have used a MID established by other two trials, one on neck pain and one on chronic lower back pain. We agree with you that this is not sufficient. However, our area of interest was precisely to assess the clinical importance of findings based on already-existing evidence.

There are still some minor typos, by example the double commas on line 15 and the comma space on

line 17, page 11; lack of space before (on line 56 page 12 and line 12 page 21 and line 32 and 37 and 42; page 23 line 30, replace reliability by lack of blinding;
line 3 on page 24, space after not, page 25 line 3, questions. Please check the manuscript carefully.

Thank you, we have combed through the document many times since. We hope it is OK now.

Reviewer: 1
Prof. Maurits van Tulder, University of Amsterdam
Comments to the Author:
The manuscript is clear but needs some language editing.

Dear Prof van Tulder,
Thank you for spending time to give us this second feedback. We have gone through the document very carefully and revised completely the discussion chapter. We apologise if our previous revision did not meet your requirements. We think we misunderstood some of your observations. We hope that this version will address better your concerns.

Abstract, methods: First time RCT is used, provide full description “randomized controlled trials (RCTs)”.

We apologise, we missed this error and have corrected it.

Abstract, conclusion: If MT is not clinically more effective than sham treatment, shouldn’t the conclusion be that MT doesn’t seem to have a clinically relevant effect?

Yes, we have updated the conclusions.

The authors research question is “To assess the effects and reliability of sham procedures in manual therapy (MT) trials…” However, sham treatments are methods used in clinical trials to help researchers determine the effectiveness of a treatment. Sham treatments are ‘fake’ treatments in which the care provider goes through the motions without actually performing the treatment. So, the effectiveness of sham treatments can only be established by comparing it to ‘no treatment’.

We agree with you, this is why our review included the comparison with control. Thank you for pointing it out. We think that you underlined a very big difference between effect and effectiveness. To assess the ST effectiveness, we have made a comparison between ST and control (now ‘no treatment’). We understand your point of view, nevertheless we think that “effect” is a correct word to use in this context because we are speaking generically about MT trials and not about any specific comparison. This is why, respectfully, we think the sentence is important to our meaning.

Limitations: “This study did not include a comparison with machine provided placebo, its aim focused on hand contact sham treatment.” I don’t consider this an adequate limitation. What is a ‘machine provided placebo’? How could you develop a machine provided placebo for spinal manipulation?

We have modified the limitations and we hope they are better now.
Some studies used a placebo provided, for example, with detuned ultrasound. Even when SM was provided. It would have been interesting to include those studies in our review and assess the effect of a “machine provided placebo”. Nevertheless, we decided to not include those studies for two main reasons: we wanted to study the effects of manual contact on participants and we strongly believe that the inclusion of these studies could have potentially increased performance bias presence and consequently heterogeneity levels. This is why we thought it was a limit on our study.

Limitations: “Insufficient number of studies were included to conduct a network meta-analysis.” Why would a network meta-analysis be better? The effectiveness of sham treatment will not be clearer if you compare it to other comparators.

We apologize but, in this case, we do not agree with your point of view. A NMA would provide more precise estimates. A NMA could allow us add results from direct and indirect comparisons and that is not possible with pair-wise meta-analysis (especially for the comparison between placebo and control). Maybe results would not have changed but we still consider this as a limit of our study. This is why we reported it as a limit in the limitations chapter, but we have updated the “Strengths and limitations of this study” chapter as you requested.

Introduction, page 7 first paragraph: the authors provide a definition of placebo: “an inert substance or sham procedure that is provided to research participants with the aim of making it impossible for them, and usually the researchers themselves, to know who is receiving an active or inactive intervention.” I would suggest adding “Placebo interventions are methodological tools used to treat participants in the study arm and the control arm in exactly the same way, except that the study group receives an active substance and the control group does not. A true placebo doesn’t exist for manual therapy and testing the effectiveness of manual therapy requires a sham intervention.”

Thank you for your suggestion. We have added this part to the introduction chapter.

Page 7, second paragraph: the authors state “Since a placebo has no inherent therapeutic power, it rarely cures the disease”. I would suggest changing this into “Since a placebo has no inherent therapeutic power, it cannot cure a disease”.

We apologize for having missed this mistake.

Page 7: the authors suggest that “where other important characteristics should be considered as part of this therapeutic context such as the tactile interaction between patient and practitioner and clinician beliefs.”. However, the sentence above they suggest that also in pharmacological studies “individual patient and clinician factors (e.g. beliefs,..” also play a role. These characteristics are the same, so ‘other important characteristics’ doesn’t seem adequate.

Thank you for highlighting this. We have corrected the text accordingly.

Page 7: the authors suggest that “in MT trials, the blinding of clinicians is almost impossible to achieve”. I would suggest deleting ‘almost’. I cannot think of a way of blinding clinicians for MT.

We agree and we have deleted it.

Page 7: the authors state that a sham treatment “aims to ensure at least the blinding of participants”. I would suggest changing this into “aims at blinding of participants”.

Thank you, we have adjusted this sentence.

Page 9, second sentence: “This review involved all types of placebo that include hand contact”. I would suggest using ‘sham treatment’ consistently throughout the manuscript instead of ‘placebo’.

Thank you - we missed it in the first revision but have now made use consistent.

Page 9, third sentence: “Studies where placebo was provided by machines (such as inactive

ultrasound) were excluded". The authors should make a stronger statement that many trials of MT that suggest having a placebo control are misleading. They may use an inactive ultrasound machine, which would be a placebo in a trial on the effectiveness of ultrasound, but is not an adequate placebo for MT. Therefore, the studies were excluded.

We had thought of developing this in the limitations section, but we do agree that it is better to explain this decision in the methods section. We have adjusted the text accordingly.

Page 9: the authors state that "ST was compared to other manual therapies such as: physiotherapy, chiropractic, osteopathy, massage, kinesiology and reflexology and to control." Physiotherapy is not 'another manual therapy'. I would suggest changing this into "ST was compared to other manual therapies provided by any type of health care provider such as: physiotherapist, chiropractor, osteopath, massage therapist, kinesiologist and reflexologist." Also, delete 'and to control'. I don't think that comparing ST to any non-MT control would add anything to answering the research question. This also refers to the next sentence "To assess if touch itself could have a positive health effect, ST was also compared to control." Delete this sentence or change 'to control' into 'to no treatment', although I would be surprised if there are any studies comparing sham MT to no treatment.

Thank you for identifying this problem. Yes, four studies had a no treatment control group. Only one study used medical treatment as control group. To avoid any misleading results, we have corrected the meta-analysis including only the four studies with no treatment as control group. This correction did not affect our findings but we have decided to maintain this comparison because is in our view fundamental to assess any effect of sham therapy.

Page 13, population: "16 trials enrolled participants with low back pain (LBP), nine included participants with cervical pain (CP)." Please add references of the studies on CP.

We have included the references.

Page 13, interventions: "Generally the trials used a single therapy session (N=11) with a single technique performed (N=8)." Eleven out of 24 trials is not really 'generally'. I would suggest stating "Eleven trials used a singly therapy session with a single technique performed in eight of those trials."

We have corrected this part.

Page 14: "Different manual treatments were provided: Physiotherapy (2 trials, 288 participants)". Physiotherapy is not a manual treatment. Manual treatment can be part of a physiotherapy treatment, but a physiotherapy treatment typically includes exercise therapy and other treatments as well.

Thank you for highlighting this error. We missed it in the first review. The papers where physiotherapy was used were classified based on the types of passive manual treatment provided. One paper used soft-tissue technique (Hansen F 1993) another used articular mobilization technique (Erdogmus S. 2007).

Page 14: "Five trials with multiple arms compared ST to control group". Describe what type of control groups were used in these studies.

We have reported this information.

Page 14: "Most of the trials involved physiotherapists (N=8), physical therapists (N=4), osteopaths

(N=3) and students (N=1).” Delete “Most” – 8 out of 24 trials is not the majority – and just state that 8 trials involved physios etc.. Also, what type of students, physiotherapy, chiropractic, osteopathy?

We apologize for this error and we have corrected it.

Table 2 and page 20: table 2 reports the results of ST vs MT on pain improvement as an MD of 3.86 higher (3.29 higher to 4.43 lower). I assume that this should be 4.43 higher?

Thank you, we missed this error. You will also find “MD 5.84 lower (20.46 lower to 8.78 higher)”. It is not an error, GRADEpro software reported the negative number 20.46 with “lower” label to identify the number -20.46. The same happened with MD.

Discussion: here the authors use “sham treatment” while in the methods and results section “ST” is used. Please be consistent in using ST or sham treatment.

Thank you, we missed it. We have replaced every “sham treatment” with “ST” but sometimes we left “sham therapy” in order to have a synonym that helps the reader by not repeating the same expression too often and too close together.

Discussion, implications for research and practice: the authors state that “Future studies should address meaningful research question…”. What do the authors mean. What would be a meaningful research question?

We have updated the discussion chapter and this sentence has been excised.

Discussion: the discussion is still a bit weak. The objective of this review was to evaluate the effect, reliability and safety of sham manual therapy. The review shows that there is no good sham procedure that has proven to be effective, reliable and safe. Having a good sham procedure could help understanding the true effect of the intervention. If the intervention has a ‘true’ effect, the total effect should be larger that the effect of a sham procedure. However, there are many randomized controlled trials that have evaluated the effectiveness of manual therapy to sham procedures, no treatment, ‘inactive’ treatments that have not shown a clinically relevant effect in favour of manual therapy. How does this review relate to all the other reviews on effectiveness of manual therapy? What can we learn from that and what would be the best step to take to improve treatment for back pain patients? I still miss a broader perspective in the discussion.

We apologise but we did not understand this comment in the first review. We think that we have comprehended what you meant with “broader perspective” and in accordance with the comments of reviewer 3, we have updated all the discussion section. We hope it now addresses this request.