PEER REVIEW HISTORY

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

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Scoping Review of Systematic Reviews of Complementary Medicine for Musculoskeletal and Mental Health Conditions.
AUTHORS	Lorenc, Ava; Feder, Gene; MacPherson, Hugh; Little, Paul; Mercer, Stewart; Sharp, Deborah

VERSION 1 – REVIEW

REVIEWER	Lesley Ward University of Oxford, UK.
REVIEW RETURNED	29-Nov-2017

GENERAL COMMENTS	Thank you to the authors for a well-written scoping review of complementary medicine for musculoskeletal and mental health. This is a large volume of work, which the authors are to be commended for. I have three areas I would like to address regarding this article, which I feel require clarification to make the article ready for publication. In particular, I have listed 27 points which require further work: A. Overall area of focus of the review. I am unclear whether this article is about multi-morbidity or co-morbidity. The term 'co-morbidity' is used in the Abstract and Discussion; however, the Introduction refers to 'multi-morbidity.' These are two quite different concepts, and there is inconsistently in the article as to what you are focused on. As such, when, on pages 14 and 15 you list 'five areas which we feel have potential' for future research, I am unclear why you have grouped those particular conditions and symptoms together. For example, in the first of the 5 areas, why group yoga, LBP, anxiety, and sleep together? Are they identified co-morbidities? Are you more likely to have sleep issues if you have low back pain? Why exclude older adults? As such, two specific points need addressing: 1. The articles needs better clarification of its focus (multi- or co-morbidity), and consistency of focus throughout. 2. If the article is to focus on co-morbidities, then more reasoning is needed for grouping certain conditions and symptoms together. For example, there is good data out there on the increased rate of depression in people with rheumatoid arthritis. B. Methodology questions. You give a lot of detail on the Methods on how you identified systematic reviews for inclusion in this paper. However, as a researcher in the field of CAM, I am aware of multiple systematic reviews of yoga for various musculoskeletal conditions that have not been included in this review. As such, would you

please address the following points:

3. Why was 'yoga' not included as a search term?
4. Page 4, lines 53-56 you state you excluded articles pre-2005 because they 'were more likely to be of poorer quality.' Provide justification of your reasoning for this.
5. There are 206 articles in the final review, of which 111 were 'high quality.' However, as you only reference a handful of these included articles, the reader is left unsure of what articles were actually included, and, as such, from what articles your results and conclusions are based on. Please include a list of all included articles, preferably in the reference list, and reference the articles more within the text when reporting results.

C. Results section. I find the results section is lacking in a lot of detail. There are many broad overview statements, but little in-depth explanation, leaving the reader with little information on which to form their own opinions as to the robustness of the evidence. In particular, safety and harm are not described clearly, and these will be important issues in multi-morbid and/or co-morbid populations. Please clarify the following:

6. Page 10, lines 54-57. This sentence is unclear: 'The MSK disorders with the most evidence were low back pain …'. Please clarify what this 'most evidence' is referring to.
7. Pg 11, line 7: 'There is some evidence of safety,[37-39]' What evidence? What does safety mean? What types of adverse events were reported?
8. Pg 11, line 11-12: 'osteopathy for low back pain compared to spinal manipulation/heat/physical therapy, although there is some evidence of harm.[40]' As above, What is the evidence? What does harm mean? What types of adverse events were reported?
9. Pg 11, line 18: 'as well as as some evidence of safety[43 ,44]'. What evidence? What does safety mean? What types of adverse events were reported?
10. Pg 11, line 20-21: 'and some evidence of safety.[40 ,41, 45-47].' What evidence? What does safety mean? What types of adverse events were reported?
11. Pg 11, line 39: 'There was good quality evidence: in a large population for acupuncture for knee or peripheral joint OA compared to placebo,[52 ,53]' evidence for what outcomes??
12. Pg 11, line 40: 'although there is some evidence of harm [53,54]. What is the evidence? What does harm mean? What types of adverse events were reported?
13. Pg 11, line 43: 'and some evidence of safety.[55-57]'. What evidence? What does safety mean? What types of adverse events were reported?
14. Pg 11, line 46-50: 'For manual therapy plus exercise compared to exercise alone there was good quality evidence in a medium sized population,[58] and moderate quality evidence in a large population ….' Evidence for what outcomes???
15. Pg 11, line 54-55: 'Acupuncture for neck pain compared to placebo or usual care had moderate quality evidence in a medium sized population for,[62] …' Evidence for what outcomes???
16. Pg 12, line 9: 'There is some evidence of safety[63-67]…' As above, please explain.
17. Pg 12, line 13-14: 'with some evidence of safety[69]…' As above, please explain.
18. Pg 12, line 17-18: 'Moderate quality evidence was documented ….' For what outcomes???

	19. Pg 12, lines 27-28: 'as well as some evidence of effectiveness in an MSK population.[71 ,72]'. Evidence for MBSR or meditation, as you have mentioned them both in this sentence? 20. Pg 12, lines 27-28: 'as well as some evidence of effectiveness in an MSK population.[71 ,72]'. Unclear – do you mean evidence for improving depression in an MSK population? And if so, which MSK conditions was this in? 21. Pg 12, line 31: 'MBSR/mindfulness/meditation all have some evidence of safety.[72]' In which populations, MSK or depression, as you mentioned both in this section. What type of evidence? This sentence needs much more explanation. 22. Pg 12, lines 37-41: 'no evidence of safety …. some evidence of safety.' What evidence??? 23. Pg 12, lines 46-50: 'meditation for anxiety, compared to usual care, placebo or other active interventions,[78] and moderate quality evidence in a medium sized population for MBSR compared to usual care,[71] and for moving meditation compared to static meditation,[78] as well as some evidence of safety[72 ,78 ,79] and effectiveness in an MSK population.[71 ,72]'. Far too much in one sentence, leading to confusion. Please explain and expand on each point, in separate sentences. 24. Pg 12, lines 54-57. Evidence for what outcomes of sleep disorders? And what was the safe evidence? 25. Pg 13, line 7. What was the MSK population? 26. Pg 13, line 10. What was the MSK population? 27. Pg 13, lines14-34. I find this section on 'Evidence for effectiveness in multimorbidity' a bit confusing. You state: 'Very few of the SRs considered multimorbid populations. However, there are some CAM which appear to be effective for both MSK and MH conditions.' So is the rest of this section about multimorbid conditions or not. i.e. in the first example, lines 22-26, are the low back pain and anxiety multi-morbid conditions, or separate conditions? If the three examples of acupuncture, yoga, and tai chi relate to multi-morbidity, please clarify. If you are talking about them as separate conditions, then the title of this section needs to be changed to reflect this.

REVIEWER	Katri Laimi Department of Physical and Rehabilitation Medicine University of Turku and Turku University Hospital, Finland
REVIEW RETURNED	19-Dec-2017

GENERAL COMMENTS	**Review "A Scoping Review of Systematic Reviews of Complementary** **Medicine for Musculoskeletal and Mental Health."** **Journal: BMJ Open** **Manuscript ID bmjopen-2017-020222** **General:**

	Thank You for Your interesting manuscript on complementary and alternative medicine (CAM) for musculoskeletal (MSK) and mental health (MH). To clarify the research question, which is now slightly differently stated in abstract, introduction and methods, I would reconsider formulating the title of the review, focusing on CAM on concomitant MSK and MH. Also shortening the manuscript would help readers to find the essential message. Now it seems based on the title, that the review focuses thoroughly on both musculoskeletal symptoms and on mental health, but the amount of information is too large for one comprehensive review of all these subjects. From my point of view, this review was a preliminary search to find gaps of previous research. If I understood right, the original aim was only to find all systematic reviews (SRs) of possibly usable forms of CAM for primary health care in patients with concomitant musculoskeletal and mental symptoms, not in them with only MSK or with only MH. All results, however, concentrate to report the conclusions (not numerical results) of reviews on musculoskeletal pain, and separately those of reviews on mental symptoms. It seems that the large number of included studies have prevented a thorough analysis of previous results, and when only relying on previously reported conclusions, this scoping review is not as reliable as meant to be. In SRs and meta-analysis of treatment options for one symptom only (low back pain, depression etc), it is more obvious than in Your review, that if a certain treatment is compared to no treatment or waiting list, especially not blinded, the studied treatments have a good treatment effect for days, weeks or even for a few months. But when comparing blindly to a credible active control treatment, the effectiveness decreases or even disappears. When studying CAM or

	any practitioner-depending treatments, treatment effect is also depending on researchers' belief in their own treatment option. Unfortunately this bias is seen in many original studies and even in SRs. This type of discussion would add the scientific value of the present review. **Abstract** - Background: Many of MSK symptoms are benign and do not require any special treatment. To change the activity of a pain patient to being a passive receiver of CAM treatment is not reasonable without a strong evidence supporting CAM. - A clinician without scientific background reads Your results seeing that CAM is highly effective in all studied symptoms. If original articles of reviewed narrative reviews had been checked, the results –section would probably show either slightly or remarkably different results. When I reflected part of results in Your abstract to recent meta-analysis on same subjects, the effectiveness in meta-analysis were not as favourable to CAM as in Your review relying only on qualitative data. - In myofascial pain the therapy (myofascial trigger point needling) claimed to be acupuncture in myofascial pain, was not acupuncture. In results, the sentence: "very few SRs considered comorbid populations, however, acupuncture,… have evidence for both MSK and MH conditions." is misleading, if there were no SRs evaluating CAM on patients having both MSK and MH. If there were such SRs, the number of these studies should be mentioned, as this was the main reason for Your review.

	Strengths and limitations - section - Even if a wide range of topics has been regarded as a strength, it can also be a limitation, because a wide scope of view makes profound discussion inaccurate. - The strength of this study is including only high-quality systematic reviews and only practitioner-delivered treatments, because otherwise the field would have been too difficult to be analysed. **Introduction** Introduction is appropriate. However, as authors are opinion-leaders in public health, I suggest that if using epidemiological terms, the used terms should be exact. In defining prevalences, prevalence numbers are now reported without knowing, if they are life-time, point prevalence, or something else, or if these prevalences are in general population, primary health or tertiary care, and if they are prevalences of symptoms or diseases. If of symptoms only, prevalence is different for chronic, and for episodic symptoms, and for intensive vs mild pain. The review does not necessarily need all these numbers, but the chosen numbers are important only if reported in the right way. In the last paragraph, it is straightly assumed that CAM has a potential positive role in MSK and MH. Always when considering a change in present treatment strategies, we have to remember that presently used strategies hopefully rely on evidence. It is important to discuss, if recommending more passive treatment options instead of active role of patient as shown to be important in most of pain research, has also negative influences on the outcome of symptoms and functioning.

	Methods In methods –section, despite of the wide search, the physiotherapeutic database (Pedro) has been forgotten. The methods –section claim that only one author (AL) read the full texts of SRs. As assessing the results of previous reviews is challenging, all numbers and conclusions would be important to be seen and interpreted by two independent authors. Also in data extraction, quality assessments by two independent authors are recommended. On the other hand, if a review of reviews aims at pooling previous research results, conclusions are more commonly based on previously reported numbers instead of conclusions only. The quality of previous reviews was estimated by AMSTAR checklist, but even if this checklist is a valuable tool, it does not tell, if the conclusions of a "high-quality SR" are right or if the evidence from a high-quality SR is strong. Regarding side effects of CAM treatments or manual treatments, straight conclusions of safety is not possible, if original studies concentrate mostly on effectiveness, not accurate reports of side effects. In effectiveness adjustment, You prioritized one SR on another. It is not clear from methods –section, if this decision changed results. When reading SRs, the quality of studies is relatively easy to estimate when the treatment can be exactly defined and exact credible placebo treatments have been used. In CAM studies, quality assessments vary a lot between assessors. As You had a huge number of SRs to be estimated, I understand believing the

	conclusions of previous reviews regarding the quality of included studies was the easiest solution. This choice, however, has to be remembered when interpreting final results. In the methods, it seemed that conclusions of SRs were copied without checking the tables of SRs with exact information of included original articles. In methods, it is not defined, what is meant by "high-quality evidence" regarding effectiveness. It cannot mean straightly reported conclusions of SRs without pooling results. The minimal clinically important difference is also important to remember, when interpreting pooled results. The effectiveness in multimorbidity (MH + MSK) cannot be shown, if all studies are only in patients with MH or in patients with MSK. It cannot be proven either, if the effectiveness is shown in a population with many symptoms (such as elderly), but not analysed for MH + MSK. Stage two: The size of the pooled study population (small, moderate or large) does not influence the reliability of results, if results rely on narrative conclusions, not on pooled quantitative results. Appendix 1: It should be clearly stated, if cancer was included or excluded (different information in PRISMA flowchart, and in appendix 1). **Results -section** Search results are thoroughly explained in PRISMA flowchart. The results –section with a huge number of included studies is impossible to be evaluated without reading also cited SRs. In results

	–section, the reader has to rely on authors without any pooled numbers of results. Also moderate/good quality evidence of effectiveness from prioritised SRs has been reported without any quantitative information. The chapter of "Evidence for MSK disorders": Does this review really give such profound answers as given here? If a previous review has been regarded to be of high quality in AMSTAR-scale, this fact does not justify decision, that also the quality of evidence based on this review is of high quality. A high-quality review based on RCTs with a high risk of bias does not always add high-quality evidence. Authors claim that there is good quality evidence for yoga for low back pain. This statement relies on one previous review (37) of eight included RCTs, and in four of these eight studies, yoga was compared with waiting list. In pooled effect sizes, on the other hand, minimal clinical differences were not achieved. Regarding myofascial trigger point pain, relying on one previous review (49), conclusions have been made of acupuncture even if the treatment used was dry needling vs wet needling of trigger points. These treatments are not based on acupuncture. As other trigger point treatments are left outside the scope of this review, it is impossible to raise this one treatment over others. In neck pain, reference 59 was checked. Authors concluded that based on this reference, there is moderate quality evidence in a large population for manipulation compared to other mobilization or medication. In this cited SR, there were, however, no differences between effectiveness of manipulation and mobilization (pSMD -0.07 (-0.72- 0.59)).

	As in all these three checked SRs, conclusions of the authors were different than I would have done, I concluded, that the manuscript does not add our knowledge of CAM without a thorough revision of results. **6. Discussion** I am afraid that if results of this review are based on subjective conclusions of evaluated SRs instead of pooling quantitative results, also discussion may give a wrong picture of previous research. The study group members probably continue their valuable and huge work also after this preliminary report of possible future study questions. Before publishing this narrative preliminary review, I would recommend that authors would check result tables of included SRs and change both results, discussion and abstract of this manuscript based on more quantitative analysis taking also MCID (minimal clinically important difference) into account if possible.

REVIEWER	L. Susan Wieland, MPH PhD, Assistant Professor Center for Integrative Medicine, University of Maryland School of Medicine Baltimore MD, USA
REVIEW RETURNED	26-Dec-2017

GENERAL COMMENTS	This is a reasonable approach to identifying the best available current evidence on CAM interventions for MH conditions and MSK conditions, and any available current evidence on CAM interventions for comorbid MH and MSK conditions, as preparation for development of an RCT on CAM for comorbid MH and MSK. The authors are to be commended for their transparency, as well as their explicit inclusion of information on cost-effectiveness and safety. The list of exclusion criteria in Appendix 1 is helpful, as some of the excluded CAM therapies and excluded medical conditions are not intuitively obvious from the descriptions in the text. Given the vast number of SRs published, the exclusion of pre-2005 reviews also seems very reasonable. However, I could not find in the manuscript that the authors explicitly excluded SRs that were available in abstract form only. It seems that they did this so if it is not in the text, please add it. I could not find in the manuscript where there was an operationalization of good quality vs moderate quality vs poor quality

	evidence from the reviews. I could also not find where large vs medium vs small size populations were pre-defined. These need to be added if they are not already there. A practical step which is however a limitation of the overview is the restriction of included reviews to those with an AMSTAR score above 5. The AMSTAR checklist is a mix of methodological and reporting items, some of which are difficult to interpret, and which furthermore are not likely of equal importance. It might have been preferable for the authors to identify key AMSTAR items which they required reviews to meet for inclusion in the overview. Could the authors briefly acknowledge and justify this limitation in the discussion? Minor issues: Page 11 line 55 'medium sized population for [what?]' The breakdown of numbers of SRs for MSK and MH conditions are a bit confusing. Nowhere in Table 1 or Table 2 is the total corresponding to the totals in the text given, although Table 2 has a note on the Total that reviews may appear in more than one row or column. Could the totals be added to the bottommost right cells of the tables? In the text, when the percentage of reviews on particular conditions or interventions is given, could numbers also be provided? For example, instead of simply 40% of SRs were on low back pain, would this be 34/84?

REVIEWER	Sydne Jennifer Newberry Southern CA Evidence-based Practice Center, The RAND Corporation, US
REVIEW RETURNED	31-Dec-2017

GENERAL COMMENTS	As the lead on a systematic review of interventions to treat OA of the knee, I deeply appreciate the goals the authors set for themselves and the reasons, but I'm concerned that in trying to cover such a broad area, they have had to sacrifice really important detail. I will try to address my concerns about this manuscript in the order of the questions provided in the review template at least to the extent possible). 1. The title of the article, specifically inclusion of the phrase, "MSK-MH comorbidity," implied that the focus was on interventions aimed at treating individuals with co-occuring musculoskeletal and mental health disorders. The authors speak to the relevance of treatments aimed at these co-occuring disorders. And they are quite relevant, if difficult to disentangle: E.g., who would not be depressed by constant unremitting lower back pain, but under what conditions is this considered co-occuring depression and LBP? Yet it becomes clear further along that the review includes reviews of both MSK and MH separately. I believe a review could have been conducted of MSK interventions that assessed the impact of treatment on MH outcomes. 2. For the reasons I just elaborated, the objective stated in the abstract is not clear or accurate. I would also have appreciated some reference in the abstract to the challenges presented by the choices of control groups in these kinds of studies. Finally, the conclusions are really confusing as to whether they pertain to individuals with both conditions, MH aspects of MSK, or both

	conditions separately. 3. I appreciate the ambitious undertaking of trying to assess the impact of alternative treatments on these prevalent conditions, but this was far too broad a scope even for a scoping review, even to be useful to identify topics for future RCTs. Far better, in my opinion, would have been to identify one or a small number of MSK conditions with evident MH implications and to assess SRs or even original RCTs that include MH outcome measures.Also, I have concerns about the choice of conditions and alternative treatments, as I will outline further below. 6. As I've alluded to, I believe the review would have been stronger had you included only studies of MSK conditions with outcome measures that assessed MH outcomes. Although the standard measures of improvement for osteoarthritis only indirectly assess MH, many studies also assess MH outcomes, although SRs might focus only on functional and/or pain outcomes. The decision to include such a broad range of conditions meant that you had to sacrifice any consideration of whether the outcome measures were really appropriate to assess treatment effectiveness, as you've (implicitly) defined it. 9. My previous response also pertains to the Results. An additional concern is that the category of manual therapy includes evidence-based treatments that are ordinarily performed by a physical therapist or osteopathic physician (at least in the US) and would not be considered alternative; in contrast, chiropractic has a much weaker evidence base, especially for the wide range of conditions it purports to treat. A further, probably more important, concern is that with the large number of interventions included, it is not possible to address the problem of appropriate comparators. When assessing these kinds of interventions for these kinds of conditions, the devil is in the details. For example, acupuncture studies are relevant only if they employ a sham control. Yoga and Tai Chi studies that use wait list or other passive controls can't be compared with those that use another physical activity intervention (such as strength training, physical therapy, and even treatment as usual) as the control. Another concern, alluded to in the discussion, is that the kinds of treatments covered by the NHS would affect the population of patients who would seek non-covered vs. covered treatments, which would, in turn, affect the applicability of findings. 11 and 12. The points I just raised limit the usefulness of the Discussion section. 15. The mansucript would need a major proofreading and edit, all else being acceptable.For example, the text includes numerous incomplete sentences and missing words. Page 5, line 45: I'm unclear what is meant by "we searched for minor mental health issues and symptoms." Some additional concerns include the following: Homeopathy lacks an evidence base and plausibility, not to mention that the NHS recently abolished coverage for it; hence I would exclude it on principle. I agree with the decision to exclude dietary supplements and herbal remedies. AMSTAR, in spite of its intended purpose, is really not a great tool for assessing SR quality. Most importantly, I would look for an accepted tool for risk of bias assessment (e.g., Cochrane), use of GRADE to assess strength of evidence, listing of reasons for study exclusion, and dual inclusion screening. Assessment of publication bias may or may not be meaningful. My overall recommendation would be to focus on a small number of MSK conditions, possibly with MH sequelae, and a small number of

	commonly used alternative treatments, preferably those administered by a licensed health care provider and covered by health insurance.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Thank you to the authors for a well-written scoping review of complementary medicine for musculoskeletal and mental health. This is a large volume of work, which the authors are to be commended for.

I have three areas I would like to address regarding this article, which I feel require clarification to make the article ready for publication. In particular, I have listed 27 points which require further work:

A. Overall area of focus of the review. I am unclear whether this article is about multi-morbidity or co-morbidity. The term 'co-morbidity' is used in the Abstract and Discussion; however, the Introduction refers to 'multi-morbidity.' These are two quite different concepts, and there is inconsistently in the article as to what you are focused on. As such, when, on pages 14 and 15 you list 'five areas which we feel have potential' for future research, I am unclear why you have grouped those particular conditions and symptoms together. For example, in the first of the 5 areas, why group yoga, LBP, anxiety, and sleep together? Are they identified co-morbidities? Are you more likely to have sleep issues if you have low back pain? Why exclude older adults?

As such, two specific points need addressing:
1. The articles needs better clarification of its focus (multi- or co-morbidity), and consistency of focus throughout.
Response: We agree we were inconsistent in the use of the terms multi-morbid or comorbid. We have changed to comorbid throughout (except where referencing a paper which refers to multi-morbidity).

2. If the article is to focus on co-morbidities, then more reasoning is needed for grouping certain conditions and symptoms together. For example, there is good data out there on the increased rate of depression in people with rheumatoid arthritis.
Response: We accept that it was unclear as to how we chose the 'five areas with potential'. We have added to and retitled this section to make it clearer that these conditions/CAM were chosen as having good evidence for at least one MSK and at least one MH condition, and some of them have new evidence on comorbidity (since our review was conducted). We did not take into account the prevalence of comorbidity in the scoping review. We have also moved the 'new evidence' from the Discussion to the Results to improve consistency.

B. Methodology questions. You give a lot of detail on the Methods on how you identified systematic reviews for inclusion in this paper. However, as a researcher in the field of CAM, I am aware of multiple systematic reviews of yoga for various musculoskeletal conditions that have not been included in this review. As such, would you please address the following points:

3. Why was 'yoga' not included as a search term?
Response: Yoga is included in the MeSH term 'complementary therapies.'

4. Page 4, lines 53-56 you state you excluded articles pre-2005 because they 'were more likely to be of poorer quality.' Provide justification of your reasoning for this.
Response: we have deleted this reference to quality and replaced it with a statement on pre-2005 exclusions being a pragmatic choice which we feel minimally impacted the results given that the majority of SRs were published in or after 2010 (122/158 SRs).

5. There are 206 articles in the final review, of which 111 were 'high quality.' However, as you only reference a handful of these included articles, the reader is left unsure of what articles were actually included, and, as such, from what articles your results and conclusions are based on. Please

include a list of all included articles, preferably in the reference list, and reference the articles more within the text when reporting results.
Response: we have added the references for all the high-quality reviews, both in the first section of Results and under their relevant subheading.

C. Results section. I find the results section is lacking in a lot of detail. There are many broad overview statements, but little in-depth explanation, leaving the reader with little information on which to form their own opinions as to the robustness of the evidence. In particular, safety and harm are not described clearly, and these will be important issues in multi-morbid and/or co-morbid populations. Please clarify the following:

6. Page 10, lines 54-57. This sentence is unclear: 'The MSK disorders with the most evidence were low back pain …'. Please clarify what this 'most evidence' is referring to.
7. Pg 11, line 7: 'There is some evidence of safety[37-39]' What evidence? What does safety mean? What types of adverse events were reported?
Response (addressing points 7 -13 and 16-17): we have added details of safety/harm to all results
8. Pg 11, line 11-12: 'osteopathy for low back pain compared to spinal manipulation/heat/physical therapy, although there is some evidence of harm.[40]' As above, What is the evidence? What does harm mean? What types of adverse events were reported?
Please see response to point 7 above
9. Pg 11, line 18: 'as well as as some evidence of safety[43 ,44]'. What evidence? What does safety mean? What types of adverse events were reported?
Please see response to point 7 above
10. Pg 11, line 20-21: 'and some evidence of safety.[40 ,41, 45-47].' What evidence? What does safety mean? What types of adverse events were reported?
Please see response to point 7 above
11. Pg 11, line 39: 'There was good quality evidence: in a large population for acupuncture for knee or peripheral joint OA compared to placebo,[52 ,53]' evidence for what outcomes??
Please see response to point 7 above
12. Pg 11, line 40: 'although there is some evidence of harm [53,54]. What is the evidence? What does harm mean? What types of adverse events were reported?
Please see response to point 7 above
13. Pg 11, line 43: 'and some evidence of safety.[55-57]'. What evidence? What does safety mean? What types of adverse events were reported?
Please see response to point 7 above
14. Pg 11, line 46-50: 'For manual therapy plus exercise compared to exercise alone there was good quality evidence in a medium sized population,[58] and moderate quality evidence in a large population ….' Evidence for what outcomes???
Response (addressing points 14 and 15): thank you for this. We have added outcomes to the manuscript in all cases, and checked throughout the paper for any other similar omissions
15. Pg 11, line 54-55: 'Acupuncture for neck pain compared to placebo or usual care had moderate quality evidence in a medium sized population for,[62] …' Evidence for what outcomes???
Please see response to point 14 above
16. Pg 12, line 9: 'There is some evidence of safety[63-67]…' As above, please explain.
Please see response to point 7 above
17. Pg 12, line 13-14: 'with some evidence of safety[69]…' As above, please explain.
Please see response to point 7 above
18. Pg 12, line 17-18: 'Moderate quality evidence was documented ….' For what outcomes???
Please see response to point 14 above
19. Pg 12, lines 27-28: 'as well as some evidence of effectiveness in an MSK population.[71 ,72]'. Evidence for MBSR or meditation, as you have mentioned them both in this sentence?
Response: We have added text to clarify this
20. Pg 12, lines 27-28: 'as well as some evidence of effectiveness in an MSK population.[71 ,72]'. Unclear – do you mean evidence for improving depression in an MSK population? And if so, which MSK conditions was this in?
Response: the SR authors did not state which MSK conditions the studies were on
21. Pg 12, line 31: 'MBSR/mindfulness/meditation all have some evidence of safety.[72]' In which populations, MSK or depression, as you mentioned both in this section. What type of evidence? This sentence needs much more explanation.
Response: thank you; we have added more detail

22. Pg 12, lines 37-41: ‘no evidence of safety …. some evidence of safety.’ What evidence???
Response: we have added details of safety/harm to all results
23. Pg 12, lines 46-50: ‘meditation for anxiety, compared to usual care, placebo or other active interventions,[78] and moderate quality evidence in a medium sized population for MBSR compared to usual care,[71] and for moving meditation compared to static meditation,[78] as well as some evidence of safety[72 ,78 ,79] and effectiveness in an MSK population.[71 ,72]’. Far too much in one sentence, leading to confusion. Please explain and expand on each point, in separate sentences.
Response: thank you we have separated this sentence into bullet points and two other sentences on safety and MSKs.
24. Pg 12, lines 54-57. Evidence for what outcomes of sleep disorders? And what was the safe evidence?
Please see responses to point 7 and 14 above
25. Pg 13, line 7. What was the MSK population?
Response: the SR authors did not state which MSK conditions the studies were on
26. Pg 13, line 10. What was the MSK population?
Response: the SR authors did not state which MSK conditions the studies were on
27. Pg 13, lines14-34. I find this section on ‘Evidence for effectiveness in multimorbidity’ a bit confusing. You state: ‘Very few of the SRs considered multimorbid populations. However, there are some CAM which appear to be effective for both MSK and MH conditions.’ So is the rest of this section about multimorbid conditions or not. i.e. in the first example, lines 22-26, are the low back pain and anxiety multi-morbid conditions, or separate conditions? If the three examples of acupuncture, yoga, and tai chi relate to multi-morbidity, please clarify. If you are talking about them as separate conditions, then the title of this section needs to be changed to reflect this.
Response: Reviewers 1, 2, and 4 made a number of comments regarding whether we focussed on comorbid MSK and MH or not, as our aim was to find evidence for treating comorbidity, but we included reviews of MSK or MH, and our results are primarily from reviews of one or the other. We had not made it clear enough that we wanted to review evidence on comorbidity, but there was none/very little, so we decided to identify which CAM appear effective for MSKs and for MH. We have amended the title, abstract, and aims to clarify this, and, most importantly, added a sentence to the end of the Introduction: "Preliminary searches showed that evidence for comorbid populations was very limited, so we chose to include independent evidence on MSK and MH conditions". We have also tried to make it clearer that our aim was to identify which CAM were effective for both MSKs and MH conditions (and thus potentially a comorbid population)

Reviewer: 2
Thank You for Your interesting manuscript on complementary and alternative medicine (CAM) for musculoskeletal (MSK) and mental health (MH). To clarify the research question, which is now slightly differently stated in abstract, introduction and methods, I would reconsider formulating the title of the review, focusing on CAM on concomitant MSK and MH.
Response: Please see response to Reviewer 1's comment number 27

Also shortening the manuscript would help readers to find the essential message. Now it seems based on the title, that the review focuses thoroughly on both musculoskeletal symptoms and on mental health, but the amount of information is too large for one comprehensive review of all these subjects.
Response: as a scoping review the purpose of this work was to identify which MSK-MH comorbidity may benefit from a CAM approach. We hope the information we have added to the Introduction about the purpose of scoping reviews helps to address this point

From my point of view, this review was a preliminary search to find gaps of previous research. If I understood right, the original aim was only to find all systematic reviews (SRs) of possibly usable forms of CAM for primary health care in patients with concomitant musculoskeletal and mental symptoms, not in them with only MSK or with only MH. All results, however, concentrate to report the conclusions (not numerical results) of reviews on musculoskeletal pain, and separately those of reviews on mental symptoms. It seems that the large number of included studies have prevented a thorough analysis of previous results, and when only relying on previously reported conclusions, this scoping review is not as reliable as meant to be.
Response: Please see response to Reviewer 1's comment number 27. We have added a section to the end of the Introduction clarifying the purpose of scoping reviews and that they often rely on the conclusions of previous reviews

In SRs and meta-analysis of treatment options for one symptom only (low back pain, depression etc), it is more obvious than in Your review, that if a certain treatment is compared to no treatment or waiting list, especially not blinded, the studied treatments have a good treatment effect for days, weeks or even for a few months. But when comparing blindly to a credible active control treatment, the effectiveness decreases or even disappears. When studying CAM or any practitioner-depending treatments, treatment effect is also depending on researchers' belief in their own treatment option. Unfortunately this bias is seen in many original studies and even in SRs. This type of discussion would add the scientific value of the present review.
Response: we have added a reference to the challenges of blinding CAM in the Limitations section.

Abstract
- Background: Many of MSK symptoms are benign and do not require any special treatment. To change the activity of a pain patient to being a passive receiver of CAM treatment is not reasonable without a strong evidence supporting CAM.
Response: thank you for this comment; we agree that a strong evidence base is needed to justify changes to practice, which is why we have conducted this review as a precursor to conducting an RCT in this area. We would not view all CAM as 'passive' as many involve significant commitment and action from the patient, especially yoga, tai chi and mindfulness which were prioritised in our review.
-
A clinician without scientific background reads Your results seeing that CAM is highly effective in all studied symptoms. If original articles of reviewed narrative reviews had been checked, the results – section would probably show either slightly or remarkably different results. When I reflected part of results in Your abstract to recent meta-analysis on same subjects, the effectiveness in meta-analysis were not as favourable to CAM as in Your review relying only on qualitative data.
Response: We feel that this comment does not take into account the purpose of scoping reviews, particularly when compared with systematic reviews/meta-analyses. We have added to the Methods section and also the Introduction paragraphs about scoping studies, including that "the scoping study does not seek to 'synthesize' evidence or to aggregate findings from different studies"

- In myofascial pain the therapy (myofascial trigger point needling) claimed to be acupuncture in myofascial pain, was not acupuncture.

Response: myofascial trigger point acupuncture is considered a style of acupuncture, which in turn is commonly synthesised with other acupuncture trials – as in Vickers et al 2012 (https://www.ncbi.nlm.nih.gov/pubmed/22965186), and therefore can be considered part of CAM.

- In results, the sentence: "very few SRs considered comorbid populations, however, acupuncture,2 have evidence for both MSK and MH conditions." is misleading, if there were no SRs evaluating CAM on patients having both MSK and MH. If there were such SRs, the number of these studies should be mentioned, as this was the main reason for Your review.
Response: This is an important point and made us rethink how we have written this section. We have restructured and retitled this section and have added that only one of the reviews gave results for an MSK-MH comorbid population (also in Abstract)

Strengths and limitations - section
- Even if a wide range of topics has been regarded as a strength, it can also be a limitation, because a wide scope of view makes profound discussion inaccurate.
Response: yes, we have added a new sentence to the Discussion stating that the breadth of the review was both a strength and a limitation.

- The strength of this study is including only high-quality systematic reviews and only practitioner-delivered treatments, because otherwise the field would have been too difficult to be analysed.
Response: Thank you, we have added more on these two points in the Discussion

Introduction
- Introduction is appropriate. However, as authors are opinion-leaders in public health, I suggest that if using epidemiological terms, the used terms should be exact. In defining prevalences, prevalence numbers are now reported without knowing, if they are lifetime, point prevalence, or something else, or if these prevalences are in general population, primary health or tertiary care, and if they are prevalences of symptoms or diseases. If of symptoms only, prevalence is different for chronic, and for episodic symptoms, and for intensive vs mild pain. The review does not necessarily need all these numbers, but the chosen numbers are important only if reported in the right way.
Response: we have added more detail on the prevalence figures for MSK and MH conditions in the introduction
- In the last paragraph, it is straightly assumed that CAM has a potential positive role in MSK and MH. Always when considering a change in present treatment strategies, we have to remember that presently used strategies hopefully rely on evidence. It is important to discuss, if recommending more passive treatment options instead of active role of patient as shown to be important in most of pain research, has also negative influences on the outcome of symptoms and functioning.
Response: we are not entirely sure we understand this comment, but if the reviewer is suggesting that CAM is passive we would respond that this is a misconception, as many CAM involve and indeed heavily emphasise self-management, exercise, patient activation.

Methods
- In methods –section, despite of the wide search, the physiotherapeutic database (Pedro) has been forgotten.
Response: We apologise for this omission. However, we believe that the majority of the references in PEDro are also indexed in EMBASE, MEDLINE/PubMed, PsycINFO, and CINAHL databases, as described in this article https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1629414/. We also found PEDro difficult to perform our searches in due to being unable to combine 'AND' and 'OR' Boolean operators.

- The methods –section claim that only one author (AL) read the full texts of SRs. As assessing the results of previous reviews is challenging, all numbers and conclusions would be important to be seen and interpreted by two independent authors. Also in data extraction, quality assessments by two independent authors are recommended. On the other hand, if a review of reviews aims at pooling previous research results, conclusions are more commonly based on previously reported numbers instead of conclusions only.

Response: We have added more detail to the Methods section to describe how the team members were involved in the data extraction and analysis. However, the majority of the data extraction was done by one author - we have added this to the limitations section of the Discussion.

The quality of previous reviews was estimated by AMSTAR checklist, but even if this checklist is a valuable tool, it does not tell, if the conclusions of a "high-quality SR" are right or if the evidence from a high-quality SR is strong.
Response: We are slightly unclear about this comment, but if we are right in understanding that the reviewer is concerned that high-quality (based on AMSTAR) reviews do not necessarily present high-quality evidence, as in their comment below under Results, then we agree. We did take into account the quality of the RCTs in our matrix (as assessed by the review authors). Scoping review guidance (Arksey & O'Malley) does not usually recommend assessing quality of individual studies.
Also see response to reviewer 3's comment on AMSTAR

Regarding side effects of CAM treatments or manual treatments, straight conclusions of safety is not possible, if original studies concentrate mostly on effectiveness, not accurate reports of side effects.
Response: yes we agree that safety is rarely adequately reported in SRs of effectiveness. We did aim to include any reviews which were only on safety, but there were none. We have added some more detail regarding safety/AEs as per reviewer 1's minor comments 21-30. We have added this as a limitation in the Discussion

In effectiveness adjustment, You prioritized one SR on another. It is not clear from methods –section, if this decision changed results.
Response: on re-reading our manuscript we agree that it is not clear how the prioritised SRs were used. We have revised the Analysis section in Methods to emphasise the importance of the prioritised SR/s

When reading SRs, the quality of studies is relatively easy to estimate when the treatment can be exactly defined and exact credible placebo treatments have been used. In CAM studies, quality assessments vary a lot between assessors. As You had a huge number of SRs to be estimated, I understand believing the conclusions of previous reviews regarding the quality of included studies was the easiest solution. This choice, however, has to be remembered when interpreting final results.
Response: Thank you for your understanding of the limitations of our methodology. We have added sentences in the Results and the Discussion about the potential variability in quality assessment by SR authors.

In the methods, it seemed that conclusions of SRs were copied without checking the tables of SRs with exact information of included original articles.
Response: We were unable to check the results of individual studies included in reviews. We did rely on the conclusions of the SRs, which we feel is appropriate for a scoping review.

In methods, it is not defined, what is meant by "high-quality evidence" regarding effectiveness. It cannot mean straightly reported conclusions of SRs without pooling results.
Response: We did not pool results, and, as above, we relied on the SR author's conclusions regarding quality, via qualitative analysis of the review author's conclusions regarding quality. This is in line with current guidance on conducting scoping reviews (Arksey & O'Malley). We have added to the Analysis section of Methods

The minimal clinically important difference is also important to remember, when interpreting pooled results.
Response: as we did not and do not plan to pool data (in this scoping review), we feel that the reference to MCID is not relevant. In addition, the MCID is rarely reported in SRs, and the MCID in pain can be misleading (https://www.ncbi.nlm.nih.gov/pubmed/28215182)

The effectiveness in multimorbidity (MH + MSK) cannot be shown, if all studies are only in patients with MH or in patients with MSK. It cannot be proven either, if the effectiveness is shown in a population with many symptoms (such as elderly), but not analysed for MH + MSK.
Response: Please see response to Reviewer 1's comment number 27. If we understand your comment correctly, we agree that studies are needed in comorbid populations, including the elderly, who may have more than two (index) conditions.

Stage two: The size of the pooled study population (small, moderate or large) does not influence the reliability of results, if results rely on narrative conclusions, not on pooled quantitative results.
Response: We would disagree and argue that, although a large population does not necessarily mean a high-quality trial, it is one factor in the quality of the evidence (along with the review's conclusions on study quality)

Appendix 1: It should be clearly stated, if cancer was included or excluded (different information in PRISMA flowchart, and in appendix 1).
Response: Cancer was indeed an exclusion not an inclusion criteria, and we have corrected the inconsistency in the manuscript.

Results –section Search results are thoroughly explained in PRISMA flowchart.
The results –section with a huge number of included studies is impossible to be evaluated without reading also cited SRs. In results –section, the reader has to rely on authors without any pooled numbers of results. Also moderate/good quality evidence of effectiveness from prioritised SRs has been reported without any quantitative information.
We have taken the SR's authors word for it regarding quality of evidence. This is a scoping review so yes, one purpose could be for the reader to use this as a starting point to then read the SRs themselves for more information. We have added this to the section in the Introduction on what scoping studies can and can't do - as a scoping review the purpose of this work was to identify which MSK-MH comorbidity may benefit from a CAM approach. We hope the information we have added to the Introduction about the purpose of scoping reviews helps to address this point.

The chapter of “Evidence for MSK disorders”: Does this review really give such profound answers as given here? If a previous review has been regarded to be of high quality in AMSTAR-scale, this fact does not justify decision, that also the quality of evidence based on this review is of high quality. A high-quality review based on RCTs with a high risk of bias does not always add high-quality evidence.
Response: Please see response to this reviewer’s comments above (under Methods) regarding AMSTAR

Authors claim that there is good quality evidence for yoga for low back pain. This statement relies on one previous review (37) of eight included RCTs, and in four of these eight studies, yoga was compared with waiting list. In pooled effect sizes, on the other hand, minimal clinical differences were not achieved.
Response: We re-read reference 37 but were unable to identify the conclusion referred to regarding minimal clinical differences

Regarding myofascial trigger point pain, relying on one previous review (49), conclusions have been made of acupuncture even if the treatment used was dry needling vs wet needling of trigger points. These treatments are not based on acupuncture. As other trigger point treatments are left outside the scope of this review, it is impossible to raise this one treatment over others.
Response: myofascial trigger point acupuncture is considered a style of acupuncture, which in turn is commonly synthesised with other acupuncture trials – as in Vickers et al 2012 (https://www.ncbi.nlm.nih.gov/pubmed/22965186), and therefore can be considered part of CAM.

In neck pain, reference 59 was checked. Authors concluded that based on this reference, there is moderate quality evidence in a large population for manipulation compared to other mobilization or medication. In this cited SR, there were, however, no differences between effectiveness of manipulation and mobilization (pSMD -0.07 (-0.72- 0.59)).
Response: we thank the reviewer for their eagle eye - this was incorrectly reported and has been amended.

As in all these three checked SRs, conclusions of the authors were different than I would have done, I concluded, that the manuscript does not add our knowledge of CAM without a thorough revision of results.
Response: We hope that the revisions we have made will address your concerns

6. Discussion

I am afraid that if results of this review are based on subjective conclusions of evaluated SRs instead of pooling quantitative results, also discussion may give a wrong picture of previous research. The study group members probably continue their valuable and huge work also after this preliminary report of possible future study questions. Before publishing this narrative preliminary review, I would recommend that authors would check result tables of included SRs and change both results, discussion and abstract of this manuscript based on more quantitative analysis taking also MCID (minimal clinically important difference) into account if possible.
Response: As discussed above, we feel that the reviewer's comments do not take into account the purpose of scoping reviews, particularly when compared with systematic reviews/meta-analyses. We have added to the Methods section and also the Introduction paragraphs further clarification on the scope and design of scoping studies, including that "the scoping study does not seek to 'synthesize' evidence or to aggregate findings from different studies" (Arksey and O'Malley). We have also obtained statistical advice which suggests that meta-analysis of SRs is risky (due to overlap in trials and thus double weighting, and narrow CIs)

Reviewer: 3

This is a reasonable approach to identifying the best available current evidence on CAM interventions for MH conditions and MSK conditions, and any available current evidence on CAM interventions for comorbid MH and MSK conditions, as preparation for development of an RCT on CAM for comorbid MH and MSK. The authors are to be commended for their transparency, as well as their explicit inclusion of information on cost-effectiveness and safety. The list of exclusion criteria in Appendix 1 is helpful, as some of the excluded CAM therapies and excluded medical conditions are not intuitively obvious from the descriptions in the text. Given the vast number of SRs published, the exclusion of pre-2005 reviews also seems very reasonable. However, I could not find in the manuscript that the authors explicitly excluded SRs that were available in abstract form only. It seems that they did this so if it is not in the text, please add it.
Response: yes we had to exclude abstract-only reviews as we could not get the information we needed from them. We have added this information to Appendix 1

I could not find in the manuscript where there was an operationalization of good quality vs moderate quality vs poor quality evidence from the reviews. I could also not find where large vs medium vs small size populations were pre-defined. These need to be added if they are not already there.
Response: We did not operationalise the quality of the studies in the reviews, but performed a qualitative analysis of the review author's conclusions regarding quality. This is in line with current guidance on conducting scoping reviews (Arksey & O'Malley). We have added to the Analysis section of Methods. The final sentence of the Methods section states the cut-offs for small/medium/large populations (small (<500 participants); moderate (501 to 3000); or large (>3000).)

A practical step which is however a limitation of the overview is the restriction of included reviews to those with an AMSTAR score above 5. The AMSTAR checklist is a mix of methodological and reporting items, some of which are difficult to interpret, and which furthermore are not likely of equal importance. It might have been preferable for the authors to identify key AMSTAR items which they required reviews to meet for inclusion in the overview. Could the authors briefly acknowledge and justify this limitation in the discussion?
Response: We used AMSTAR as a pragmatic choice, as it is useful for rapid and reproducible assessments of the quality of SRs. We notice that the AMSTAR tool has recently been revised and published in BMJ (http://www.bmj.com/content/358/bmj.j4008.full.print). We have added a sentence in the discussion highlighting the limitations of AMSTAR and suggesting that future reviews use a different checklist.

Minor issues:

Page 11 line 55 'medium sized population for [what?]'
Response: as per the comments from reviewer 1 we have clarified this, added outcomes to the manuscript in all cases, and reviewed the manuscript for any other similar omissions

The breakdown of numbers of SRs for MSK and MH conditions are a bit confusing. Nowhere in Table 1 or Table 2 is the total corresponding to the totals in the text given, although Table 2 has a note on

the Total that reviews may appear in more than one row or column. Could the totals be added to the bottommost right cells of the tables?
Response: We have now added the same footnote to Table 1 as Table 2 and have added totals to the bottom right cell.

In the text, when the percentage of reviews on particular conditions or interventions is given, could numbers also be provided? For example, instead of simply 40% of SRs were on low back pain, would this be 34/84?
Response: This has been done

Reviewer: 4
As the lead on a systematic review of interventions to treat OA of the knee, I deeply appreciate the goals the authors set for themselves and the reasons, but I'm concerned that in trying to cover such a broad area, they have had to sacrifice really important detail. I will try to address my concerns about this manuscript in the order of the questions provided in the review template at least to the extent possible).

1. The title of the article, specifically inclusion of the phrase, "MSK-MH comorbidity," implied that the focus was on interventions aimed at treating individuals with co-occurring musculoskeletal and mental health disorders. The authors speak to the relevance of treatments aimed at these co-occurring disorders. And they are quite relevant, if difficult to disentangle: E.g., who would not be depressed by constant unremitting lower back pain, but under what conditions is this considered co-occurring depression and LBP? Yet it becomes clear further along that the review includes reviews of both MSK and MH separately. I believe a review could have been conducted of MSK interventions that assessed the impact of treatment on MH outcomes.
Response: Please see response to Reviewer 1's comment number 27. We have changed the title of the article. It now states "Musculoskeletal and mental health conditions". Regarding the last sentence in this paragraph, this would be quite a different review and not in line with our aim to review comorbidity (see earlier comments). To clarify this, we have added to Appendix 1 and in the Methods that for MH we only included reviews which used MH symptoms/diagnoses not just MH outcomes. For MSK SRs which mention mental health we have added further clarification regarding whether they mention MH outcomes or diagnoses.

2. For the reasons I just elaborated, the objective stated in the abstract is not clear or accurate. I would also have appreciated some reference in the abstract to the challenges presented by the choices of control groups in these kinds of studies. Finally, the conclusions are really confusing as to whether they pertain to individuals with both conditions, MH aspects of MSK, or both conditions separately.
Response: Please see response to Reviewer 1's comment number 27

3. I appreciate the ambitious undertaking of trying to assess the impact of alternative treatments on these prevalent conditions, but this was far too broad a scope even for a scoping review, even to be useful to identify topics for future RCTs. Far better, in my opinion, would have been to identify one or a small number of MSK conditions with evident MH implications and to assess SRs or even original RCTs that include MH outcome measures. Also, I have concerns about the choice of conditions and alternative treatments, as I will outline further below.
Response: We did not make it clear enough in the manuscript that the starting point for this review is the common presence of MSK-MH comorbidity in UK NHS primary care, and our intention to develop an intervention to help this group of patients - therefore the need to prioritise which MSK-MH comorbidity we should investigate in a trial. We have now added this information to the Introduction. Please also see our response to reviewer 2's second comment.

6. As I've alluded to, I believe the review would have been stronger had you included only studies of MSK conditions with outcome measures that assessed MH outcomes. Although the standard measures of improvement for osteoarthritis only indirectly assess MH, many studies also assess MH outcomes, although SRs might focus only on functional and/or pain outcomes. The decision to include such a broad range of conditions meant that you had to sacrifice any consideration of whether the outcome measures were really appropriate to assess treatment effectiveness, as you've (implicitly) defined it.

Response: please see our response to your comment 1 above. Also we have added a section to the end of the Introduction clarifying the purpose of scoping reviews and highlighting that they often rely on the conclusions of previous reviews

9. My previous response also pertains to the Results. An additional concern is that the category of manual therapy includes evidence-based treatments that are ordinarily performed by a physical therapist or osteopathic physician (at least in the US) and would not be considered alternative; in contrast, chiropractic has a much weaker evidence base, especially for the wide range of conditions it purports to treat.
Response: We agree. We had difficulty deciding whether to include manipulation/manual therapy etc. We have added a sentence to the Methods section explaining our decision, which was to include manipulation, manual therapy and mobilisation as techniques commonly practised by some CAM practitioners (as well as conventional practitioners). In our full report we say "Manipulation includes chiropractic, osteopathy, manual therapy, spinal manipulation and mobilisation. The latter two are techniques rather than professions, and are commonly practiced by both chiropractors and osteopaths, as well as physiotherapists and other manual therapists. We used the SR authors' term for the CAM, but acknowledge that there is likely to be significant overlap between these topics"

A further, probably more important, concern is that with the large number of interventions included, it is not possible to address the problem of appropriate comparators. When assessing these kinds of interventions for these kinds of conditions, the devil is in the details. For example, acupuncture studies are relevant only if they employ a sham control. Yoga and Tai Chi studies that use wait list or other passive controls can't be compared with those that use another physical activity intervention (such as strength training, physical therapy, and even treatment as usual) as the control.
Response: as a team we disagree regarding "appropriate comparators" as we have a pragmatic rather than explanatory perspective. Our manuscript contains the information on the comparators for all the data reported, so that the reader can draw their own conclusions if they are interested in a particular comparison.

Another concern, alluded to in the discussion, is that the kinds of treatments covered by the NHS would affect the population of patients who would seek non-covered vs. covered treatments, which would, in turn, affect the applicability of findings.
Response: This is an interesting point but not one that can be addressed in our review. We have added a sentence to the Introduction and the Discussion explaining that the other phases of the study of which this review was a part included evaluation of other aspects, such as NHS availability of CAM, patients' views, GPs' views etc.

11 and 12. The points I just raised limit the usefulness of the Discussion section.
Response: we hope that the revisions we have made address your concerns about the usefulness of the Discussion

15. The manuscript would need a major proofreading and edit, all else being acceptable. For example, the text includes numerous incomplete sentences and missing words.
Response: Done.

Page 5, line 45: I'm unclear what is meant by "we searched for minor mental health issues and symptoms."
Response: we have changed this sentence in the manuscript to clarify that we excluded major MH disorders

Some additional concerns include the following:
Homeopathy lacks an evidence base and plausibility, not to mention that the NHS recently abolished coverage for it; hence I would exclude it on principle. I agree with the decision to exclude dietary supplements and herbal remedies.
Response: We disagree. - we did not exclude any CAM 'on principle'. Practitioner-led CAM was our main criteria for selection of CAM, as we are interested in CAM which a GP could refer to safely and easily. However, in the end homeopathy did not have any appropriate evidence so was not reported.

AMSTAR, in spite of its intended purpose, is really not a great tool for assessing SR quality. Most importantly, I would look for an accepted tool for risk of bias assessment (e.g., Cochrane), use of

GRADE to assess strength of evidence, listing of reasons for study exclusion, and dual inclusion screening. Assessment of publication bias may or may not be meaningful.
Response: This is similar to the comment above from reviewer 3. We have added a sentence in the discussion highlighting the limitations of AMSTAR and suggesting that future reviews use a different checklist.

My overall recommendation would be to focus on a small number of MSK conditions, possibly with MH sequelae, and a small number of commonly used alternative treatments, preferably those administered by a licensed health care provider and covered by health insurance.
Response: please see our response to your comment 3 above. As a UK-based team, we do not feel the comment on health insurance is relevant

VERSION 2 – REVIEW

REVIEWER	Sydne Newberry RAND Corporation, US
REVIEW RETURNED	18-Mar-2018

GENERAL COMMENTS	I hope it's acceptable that I inserted my comments in the report text using the Comments function. I just have a few points I want to address for emphasis. 1. The inclusion of manual therapy among the CAM therapies is problematic without further characterization. Physical therapy can be considered among the manual therapies, yet PT is an evidence-based therapy for numerous musculoskeletal complaints (it would be akin to including medical nutrition therapy in a review of CAM). 2. Although I realize scoping reviews have their place, sometimes the devil is in the details: when appraising studies of the efficacy of these kinds of therapies, it's important to know what the comparator is. E.g., if yoga is being compared to strength and agility training, and it doesn't have greater benefit than that control, it might mean yoga is pretty darned effective, but if yoga is being compared to a minutes' worth of doctor advice on moving, you would hope the yoga does lots better than the control. Unfortunately, this distinction sometimes gets lost. 3. I would have much preferred the authors to select some elements of AMSTAR, e.g., assessing study quality, using GRADE to assess strength of evidence (more are included in my in-text comments). 4. The manuscript is in need of a heavy edit and reformatting. 5. The results would be much clearer and easier to follow if they were presented graphically as a matrix.

VERSION 2 – AUTHOR RESPONSE

Reviewer 4 (second review)

Thank you for this further helpful set of comments which has helped us clarify our thoughts. We have addressed all the comments – see responses below. Major changes are:

• We have replaced the Results text with tables as suggested – we hope this makes it easier to read. We welcome the Editor's advice as to whether new tables (table 3 to table 7) should be combined into one/two tables.

• We have added further detail on type of comparator

• We have deleted the RCT update information

• Although the reviewer did not specifically request this change, we realised that we had not been clear on what level of quality of evidence we were including. We have made sure this paper only reports those areas where there was moderate/good quality evidence (quality as rated by SR authors) and have thus removed any information which was poor quality.

• We have removed acupuncture from our conclusions about the best CAM for MSK-MH comorbidity as the MH evidence was poor quality.

Reviewer's individual comments:

1. The inclusion of manual therapy among the CAM therapies is problematic without further characterization. Physical therapy can be considered among the manual therapies, yet PT is an evidence-based therapy for numerous musculoskeletal complaints (it would be akin to including medical nutrition therapy in a review of CAM). [Also in pdf re Table 1 "This [inclusion of manual therapy] worries me: many studies of manual therapy are not CAM by any stretch of the imagination…]

Response: we recognise that there is a lack of agreement around the inclusion of manual therapy as CAM, and we had lengthy discussions as a team as to whether to include it. We also examined the manual therapy SRs to see if they included treatments provided by a CAM practitioner, which most did – see table below. We would like to keep the sentence which we added in response to the reviewer's comments on our original submission, which rationalises our decision ("We included manipulation, manual therapy and mobilisation as techniques commonly practised by some CAM practitioners (as well as conventional practitioners)")

Spinal manipulation/mobilisation/manual therapy for low back pain

Rubinstein 2011 Most treatments were delivered either by physiotherapists or chiropractors, in other cases either an osteopathic physician, combination physiotherapist or medical manipulator, medical manipulator or osteopath.

Rubinstein 2012 Most treatments were delivered either by physiotherapists or chiropractors.

Furlan 2010 In 32 trials, spinal manipulation or mobilization was provided by experienced and licensed chiropractors. In the remaining studies manipulation or mobilization was provided by physical therapists (17 trials), general practitioners (five trials), licensed or qualified manual therapy practitioners (six trials), physical therapists with manual therapy training (three trials), clinicians or experienced clinicians (four trials), neurologists or rheumatologists with chiropractic training (three trials), folk healers (one trial), and osteopaths (one trial). The information regarding treatment provider was not reported for the remaining 29 trials.

Manual therapy for Osteoarthritis

Corbett et al No details on what this was

Massage/manual therapy for Fibromyalgia

Yuan et al 2015 Swedish massage, connective tissue massage, manual lymphatic drainage, myofascial release, shiatsu and a combination of different massage styles.

Manual therapy/manipulation for Neck pain/disorders

Lin et al 2012 Chinese manipulation

Gross et al 2015 Not stated

Miller et al 2010 some of the trials they included were chiropractic but some were physiotherapy

D'Sylva et al 2010 Not stated

Carlesso et al 2010 9 trial were chiropractors, 6 physiotherapists, 2 osteopaths

Manual therapy for Shoulder pain/disorders

Page et al 2014 In Background only: Manual therapy and exercises are delivered by various clinicians, including physiotherapists, physical therapists, chiropractors and osteopaths.

2. Although I realize scoping reviews have their place, sometimes the devil is in the details: when appraising studies of the efficacy of these kinds of therapies, it's important to know what the comparator is. E.g., if yoga is being compared to strength and agility training, and it doesn't have greater benefit than that control, it might mean yoga is pretty darned effective, but if yoga is being compared to a minutes' worth of doctor advice on moving, you would hope the yoga does lots better than the control. Unfortunately, this distinction sometimes gets lost.

Also comment from PDF: Did the SR specifically compare one intervention to another active intervention? This issue needs to be considered: if the intervention of interest is compared to another active intervention with known efficacy, then it’s kind of like a superiority trial, and a lack of difference may be a good thing, whereas comparison to a placebo or passive intervention should show a positive effect if the intervention is really effective.

Response: We had not used the type of control in determining relative effectiveness as although active comparators provide really useful explanatory evidence their key limitation is that they do not provide the best evidence for overall effectiveness. We have added the comparator to our Results, and mention it in the Discussion.

3. I would have much preferred the authors to select some elements of AMSTAR, e.g., assessing study quality, using GRADE to assess strength of evidence (more are included in my in-text comments).

Response: see point 8 below

4. The manuscript is in need of a heavy edit and reformatting.

Response: We have edited the text but welcome any further suggestions

5. The results would be much clearer and easier to follow if they were presented graphically as a matrix.

Response: we have added tables of Results for each condition and reduced the accompanying text. We realise this results in a large number of tables and would appreciate the Editor's advice on this – perhaps we could combine them into one MSK and one MH? We have moved Box 1 to a new Appendix 2

Comments from pdf of manuscript:

6. How are you defining 'high-quality' (in Aims)

Response: We relied on SR authors' assessment of the quality of the evidence from their included RCTs, and prioritised areas where the authors reported high quality.

7. I'm confused. I'm assuming you considered studies of patients with symptoms of depression or anxiety secondary to a physical condition (e.g., cancer), but then why not include studies of

patients experiencing these symptoms secondary to any of these conditions, since they all are seen in primary care? (in Methods Searches)

Response: This is a valid point, but we had to make a pragmatic decision to make the review manageable. As a team we decided to exclude certain conditions which we felt are mainly dealt with in secondary care and not commonly treated/dealt with in primary care. We have deleted much of this text and instead refer the reader to the full list in Appendix 1. An additional issue to clarify is that we only included SRs which specified that patients had to meet a specified threshold for anxiety/depression.

8. I believe that you could have saved time by focusing on particular elements of AMSTAR, e.g., listing

inclusion/exclusion criteria, describing quality control on the review process, assessing study quality, using GRADE, listing studies excluded at full-text and reasons for exclusion, including the PRISMA flow.

Response: this may have been a more efficient approach but unfortunately we did not do this

9. Not sure what you mean by "ranking:" ranking implies some prioritization. What was the basis? (in Methods > Analysis)

Response: The criteria for ranking are given below the reviewer's comment, under 'Step 2'. We have changed the wording of this paragraph to make the 3-step process clearer. We have also changed 'analysis' to 'evidence synthesis'

10. Analysis? Do you mean you ranked reviews by total sample size? Weighted them somehow? Ok, based on the description below, I think you mean computation of sample size for a particular intervention and condition

Response: Further detail is provided in ‘Step 3’. We have also changed ‘analysis of sample size’ to ‘computation of sample size’

11. I’m confused. The prioritization was based on AMSTAR and the quality of the studies was based on the review prioritization? If not, what was study quality based on? The appraisal of the review authors?

Response: The prioritisation of SRs was based on recency, breadth, quality (AMSTAR Score) and level of analysis (e.g. meta-analysis). Study (individual trial) quality was as reported by the SR authors

12. If areas were ranked highly for safety independently of effectiveness, this is not useful…(In Methods > Analysis > Step2)

Response: We maintain that any indication of safety regardless of effectiveness is worth taking into consideration.

13. I think you mean interventions that showed this indication were ranked highly? …(In Methods > Analysis > Step2)

Response: We have changed the wording

14. Not sure why homeopathy included

Response: We are not sure why we should exclude homeopathy SRs as long as they fitted our inclusion criteria – we did identify some SRs of homeopathy but we were unable to determine if a practitioner was involved in the interventions so we could not include them.

15. 49 or 48? (top of page 15)

Response: We have corrected this typo

16. This paragraph is very confusing. The first sentence suggested the ranking was based on number of studies that included the condition. The second sentence suggests the ranking was based on a combination of condition, intervention, and comparator.

Response: The second interpretation is correct – this is based on our 3 step analysis process. We have added a sentence at the start of this paragraph to clarify this.

17. Please check the journal’s formatting guidelines and make sure formatting is uniform throughout this section. It’s really a bit disorganized and difficult to read. A matrix would be ideal. (In Results)

Response: See point 5 above

18. Page 19 change “There was only poor quality evidence for all other combinations of MH condition/CAM/comparator”

Response: done

19. Page 19 “than non-comorbid”?

Response: We have changed to ‘depression alone’

20. I thought individual original studies were not included.(in Results > Update)

Response: Thank you - we have moved the individual study data to discussion

21. Cochrane and GRADE are not interchangeable with AMSTAR (Discussion)

Response: We have looked into this further and agree that AMSTAR aims to assess methodological quality rather than risk of bias and appropriateness . We have therefore suggested instead the Oxford Centre for Evidence-based Medicine quality appraisal tool as an alternative, and the risk of bias tools (ROBIS) as additional.

VERSION 3 – REVIEW

REVIEWER	Sydne Jennifer Newberry Southern California Evidence Based Practice Center, The RAND Corporation, Santa Monica CA US
REVIEW RETURNED	03-Jul-2018

GENERAL COMMENTS	I believe this review merits publication, if for no other reason than to illustrate the lack of quality research in this area. However, I would strongly advise that the publication be accompanied by a commentary on the limitations to this literature in general and the challenges to doing these kinds of studies.